Computational Studies of Signalling at the Cell Membrane

Craig Nicholas Lumb

Department of Biochemistry and Merton College, Oxford

Trinity 2012

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy at the University of Oxford

A copy of this thesis has been deposited in the Bodleian Library, University of Oxford.

The *viva voce* examination was held on Wednesday 3rd October 2012. The examiners were Professor Iain D. Campbell FRS (Department of Biochemistry, University of Oxford) and Dr Alan J. Robinson (Medical Research Council Mitochondrial Biology Unit, University of Cambridge).

Computational Studies of Signalling at the Cell Membrane

Craig Nicholas Lumb

Department of Biochemistry and Merton College, Oxford

Submitted for the degree of Doctor of Philosophy, Trinity 2012

Abstract

In order to associate with the cytoplasmic leaflet of the plasma membrane, many cytosolic signalling proteins possess a distinct lipid binding domain as part of their overall fold. Here, a multiscale simulation approach has been used to investigate three membrane-binding proteins involved in cellular processes such as growth and proliferation. The pleckstrin homology (PH) domain from the general receptor for phosphoinositides 1 (GRP1-PH) binds phosphatidylinositol (3,4,5)-trisphosphate (PI(3,4,5)P_3) with high affinity and specificity. To investigate how this peripheral protein is able to locate its target lipid in the complex membrane environment, Brownian dynamics (BD) simulations were employed to explore association pathways for GRP1-PH binding to PI(3,4,5)P_3 embedded in membranes with different surface charge densities and distributions. The results indicated that non-PI(3,4,5)P_3 lipids can act as decoys to disrupt PI(3,4,5)P_3 binding, but that at approximately physiological anionic lipid concentrations steering towards PI(3,4,5)P_3 is actually enhanced. Atomistic molecular dynamics (MD) simulations revealed substantial membrane penetration of membrane-bound GRP1-PH, evident when non-equilibrium, steered MD simulations were used to forcibly dissociate the protein from the membrane surface. Atomistic and coarse grained (CG) MD simulations of the phosphatase and tensin homologue deleted on chromosome ten (PTEN) tumour suppressor, which also binds PI(3,4,5)P_3, detected numerous non-specific protein-lipid contacts and anionic lipid clustering around PTEN that can be modulated by selective *in silico* mutagenesis. These results suggested a dual recognition model of membrane binding, with non-specific membrane interactions complementing the protein-ligand interaction. Molecular docking and MD simulations were used to characterise the lipid binding properties of kindlin-1 PH. Simulations demonstrated that a dynamic salt bridge was responsible for controlling the accessibility of the binding site. Electrostatics calculations applied to a variety of PH domains suggested that their molecular dipole moments are typically aligned with their ligand binding sites, which has implications for steering and ligand electrostatic funnelling.

Acknowledgements

First and foremost, I am enormously grateful to Mark Sansom for being such an excellent supervisor and for all of his help, support and encouragement over the past four years. I would also like to thank Mark for his seemingly inexhaustible supply of patience, for having a sense of humour, and for showing admirable restraint when I told him I was going to go and work in the City. Aside from Mark, special thanks must also go to Phill Stansfeld for showing me the simulation ropes when I first arrived. Phill has since become my go-to guy in the lab for advice on practically everything, and has patiently endured all of my questions, anxieties and existential crises for nearly four years. Thanks Mark and Phill!

I am indebted to Oliver Beckstein, Benjamin Hall and Philip Fowler for always taking an interest in my work and for all of their helpful comments and suggestions over the years. Special thanks go to Phil Biggin for always being available and happy to chat, and also to Elspeth Garman for giving me the opportunity to teach undergraduate students. I would also like to acknowledge our collaborators: Rob Gilbert and Luke Yates (Oxford); David Calderwood (Yale); Rob Stahelin (Notre Dame and Indiana); and Tanya Kutateladze (Colorado) for their interest in and enthusiasm for our work. I am particularly grateful to Tanya for hosting me in Denver for a lab visit, and also for pointing me in the direction of some good skiing in the Rockies.

I have been incredibly fortunate to have been surrounded by a fantastic group of people who have made Oxford such a great place to live and work during my DPhil. From the lab, there are many people I would like to thank, but first among equals are Alan Chetwynd, Joe Goose and Sarah Rouse. For the good times inside and outside SBCB thanks also to: Sonya Hanson; Jean Hélie; Heidi Koldsø; Jerome Ma; Maria Musgaard; Greg Ross; Matthias Schmidt; David Shorthouse; Lukas Stelzl; Jemma Trick and of course Shabana Vohra. For making me feel exceptionally welcome when I first arrived back in September 2008 thanks go to some previous members: Jussi Aittoniemi; Tim Carpenter; Rob D'Rozario; Jochen Klingelhoefer; Michelle Sahai; Jayne Wallace; and Chze Ling Wee. Beyond the lab, huge thanks for assorted advice, support and good times go to a variety of ex-Cambridge and Merton folk: Sonia Bracegirdle; Steve Buckley; Genevieve Clutton; Derek Hollman; Mike Lee; Gus Logan; Alex Matthews; Robin Neatherway; Graeme Spence and, last but by no means least, Will Upcher.

The Medical Research Council and Merton College, Oxford are gratefully acknowledged for generous funding through the provision of a studentship and travel grants respectively. Finally I would like to thank my parents for all of their support, both financial and emotional, over what has turned out to be quite an eventful four years. In particular, I owe a huge debt of gratitude to my mother, Julie, for everything she has done and all she continues to do. Without her, none of this would have been possible.

C.N.L.
Oxford
November 2012

List of publications originating from this thesis

Lumb C.N., He J., Xue Y., Stansfeld P.J., Stahelin R.V., Kutateladze T.G., and Sansom M.S.P.
Biophysical and computational studies of membrane penetration by the GRP1 pleckstrin homology domain.
Structure 19, 1338–1346, 2011.

Lumb C.N. and Sansom M.S.P.
Finding a needle in a haystack: The role of electrostatics in target lipid recognition by PH domains.
PLoS Computational Biology 8, e1002617, 2012.

Yates L.A., Lumb C.N., Brahme N.N., Zalyte R., Bird L.E., De Colibus L., Owens R.J., Calderwood D.A., Sansom M.S.P., and Gilbert R.J.C.
Structural and functional characterisation of the kindlin-1 pleckstrin homology domain.
Journal of Biological Chemistry (in press)

Lumb C.N. and Sansom M.S.P.
Defining the membrane associated state of the PTEN tumour suppressor protein.
Biophysical Journal (under revision)

Other publications as a result of work completed during the DPhil

Brown M.T., Steel B.C., Silvestrin C., Wilkinson D.A., Delalez N.J., Lumb C.N., Obara B., Armitage J.P., and Berry R.M.
Flagellar hook flexibility is essential for bundle formation in swimming *Escherichia coli* cells
Journal of Bacteriology 194, 3495–3501, 2012.

Contents

List of Figures

List of Tables

Chapter 1

Background

EUKARYOTIC cells are surrounded by a cell membrane, a selectively permeable layer that separates the contents of the cell from its exterior. The cell membrane is approximately 40 Å thick, and its composition and morphology is described by the fluid mosaic model of biological membranes first outlined by Singer and Nicolson [1]. Although the details of the model have been refined over the years [2], the underlying concept has endured. In the fluid mosaic model the lipids arrange themselves into a bilayer, with the hydrophobic lipid tails forming the core of the bilayer and the polar headgroups exposed to the solvent. The lipid bilayer is predominantly fluid, composed of phospholipids, sphingolipids and sterols. Embedded within this lipid matrix are proteins, which can be wholly intrinsic or may have regions that protrude from the surface of the membrane. Other, extrinsic, proteins may bind to the surface of the membrane from the cytosol or the extracellular medium (Figure 1.1).

The precise lipid composition of the cell membrane depends upon the cell type, but all mammalian cells possess phosphatidylcholine (PC), phosphatidylethanolamine (PE),

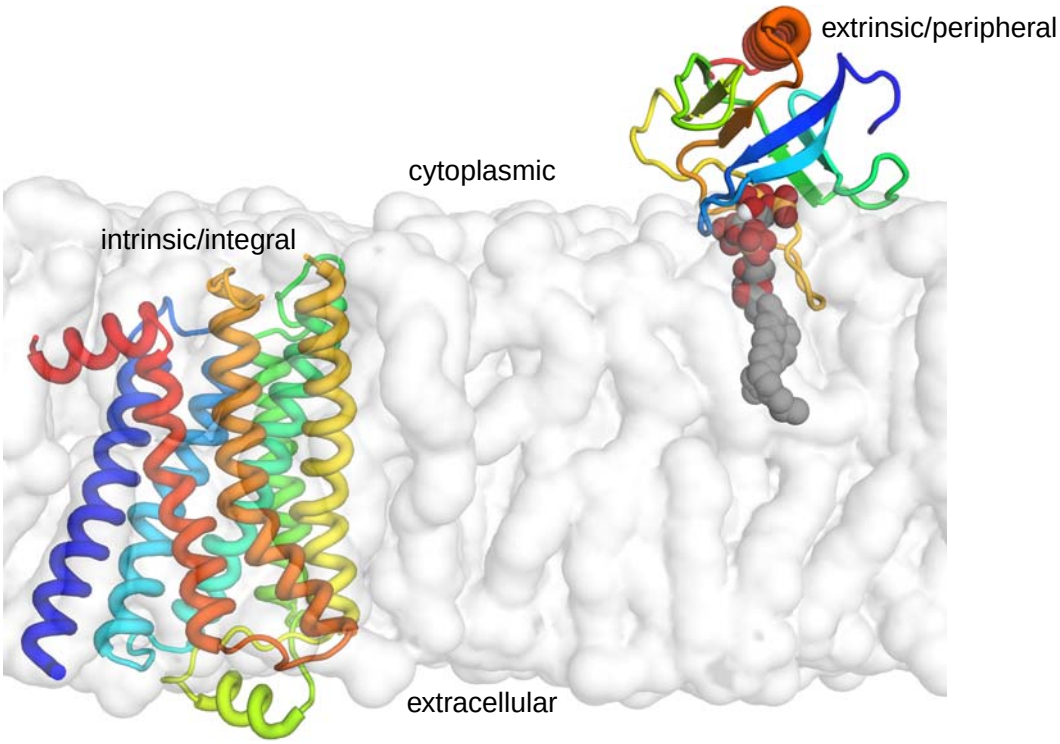

Figure 1.1: Intrinsic/integral and extrinsic/peripheral membrane proteins. On the left is shown the X-ray crystal structure of the human β_2-adrenergic receptor (PDB 2RH1) [3], which has a typical G protein coupled receptor (GPCR) fold of seven transmembrane helices that form a helical bundle. The protein is shown in its approximate position in a lipid bilayer, depicted as a white translucent surface, and it is thus an intrinsic or integral membrane protein. On the right is shown the X-ray crystal structure of a lipid binding domain, GRP1-PH (PDB 1FGY) [4], bound to its cognate lipid PI(3,4,5)P_3. GRP1-PH is ordinarily cytosolic but binds to the surface of the membrane on production of PI(3,4,5)P_3, shown as a collection of van der Waals spheres. GRP1-PH is therefore an example of an extrinsic or peripheral membrane protein. Protein sequences are coloured across the colour spectrum from blue (N-terminus) to red (C-terminus).

sphingolipids and cholesterol in varying amounts. However, the composition of cell membranes can differ substantially between cell types, and many membranes also contain more exotic lipid species in low concentrations. The full lipid profile of a cell has been termed the 'lipidome' and the science of 'lipidomics', describing the relationship between lipids and cellular signalling and metabolism, has recently gained prominence [5]. Aside from the composition, the dynamic organisation of these lipids within the membrane is also thought to be important, perhaps best illustrated by the concept of lipid 'rafts'. These are microdomains in the plasma membrane that are enriched in cholesterol and sphingolipids. Lipid rafts are thought to be important in cell signalling [6], among other things, though their existence is still rather controversial [7].

Lipid	Relative level / %	Fold increase on stimulation
phosphatidylserine (PS)	8.5	1
phosphatidic acid (PA)	1.5	1
phosphatidylinositol (PI)	1.0	1
PI(3)P	2×10^{-3}	1
PI(4)P	0.05	0.7
PI(5)P	2×10^{-3}	3-20
$PI(4,5)P_2$	0.05	0.7
$PI(3,4)P_2$	1×10^{-4}	10
$PI(3,5)P_2$	1×10^{-4}	2-30

Table 1.1: Relative abundance of anionic lipids in mammalian cells. Values are approximate and represent *total* lipid by weight in erythrocytes, but the asymmetric distribution of anionic lipids across the two bilayer leaflets means that the relative abundance in the cytoplasmic leaflet is likely to be higher, between 10-20%. Values are taken from the review by Lemmon [12].

PC, PE and sphingolipids are zwitterionic with no net charge, and cholesterol is uncharged. However, cell membranes also contain lipids with negatively charged headgroups. These lipids are predominantly found in the inner leaflet of the plasma membrane and are also found in the cytoplasmic leaflets of membrane-bound cell organelles. While estimates vary, anionic lipids are thought to constitute between 10-20% of the total lipids in the cytoplasmic leaflet of the plasma membrane, depending on the cell type [8, 9, 10, 11]. A breakdown of the relative abundances of anionic lipids by species is shown in Table 1.1.

The principal anionic lipid present in mammalian cells is phosphatidylserine (PS). PS is monovalent and participates in cell signalling by helping to recruit signalling proteins to the plasma membrane through electrostatic interactions [13], as the prevalence of anionic lipids in the cytoplasmic leaflet of the eukaryotic cell membrane produces a net negative surface charge [14, 15]. Aside from PS, other anionic lipids also exist, albeit in much lower concentrations. Some of the most important are the various phosphorylated derivatives of phosphatidylinositol (PI), known as the phosphatidylinositol phosphates (PIPs) or phosphoinositides.

1.1 Phosphoinositides

PI has an inositol (or cyclohexanehexol) headgroup, which is a six-membered ring that can be phosphorylated at the sites of the hydroxyl groups to produce various PI derivatives. The most common isomeric form of inositol encountered in biology is D-*myo*-inositol [16].

PI is the dominant inositol lipid in unstimulated mammalian cells, and various enzymes act to generate its phosphorylated derivatives. These PIPs are present in relatively low concentrations (Table 1.1) but nonetheless they have emerged as key regulators of cell signalling. PI can be phosphorylated at the 3′, 4′ and/or 5′ positions of the inositol ring, and the extent of phosphorylation at these three locations determines the net negative charge on the lipid [17]. The variable number of phosphate group substituents, coupled with structural isomerism arising from several possible phosphorylation sites, generates a total of seven distinct physiological PIPs (Figure 1.2).

Of the phosphorylated PI derivatives, the singly and doubly phosphorylated PI(4)P and PI(4,5)P_2 are the most abundant, though their concentrations are still typically an order of magnitude smaller than that of PI itself [18]. The other phosphoinositides are synthesised from these precursors in response to stimuli, and therefore their presence is much more transient.

Polyphosphatidylinositol lipids were discovered in 1949 [19], but the link between phosphoinositides and cell signalling was first proposed by Michell in 1975 [20]. At this point, work on phosphoinositide signalling revolved around the two most abundant inositol phospholipids, PI(4)P and PI(4,5)P_2, and their connection to Ca^{2+} signalling. PI(4,5)P_2 is hydrolysed by phospholipase C (PLC) to produce soluble I(1,4,5)P_3 and membrane bound diacylglycerol (DAG) [21]. The soluble I(1,4,5)P_3 is a second messenger that binds to I(1,4,5)P_3 receptors in the endoplasmic reticulum, triggering release of stored Ca^{2+} that goes on to regulate Ca^{2+}-dependent enzymes. DAG binds to protein

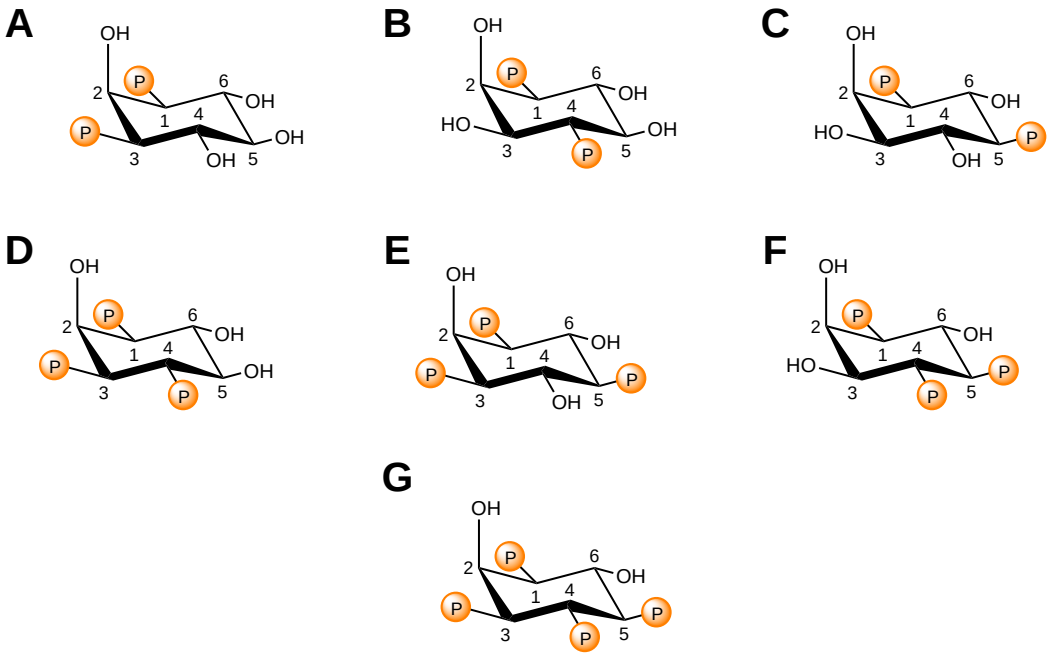

Figure 1.2: Chemical structures of phosphorylated derivatives of D-*myo*-inositol. Seven phosphatidylinositol phosphates occur naturally, and these are generated by either changing the number of phosphate groups on the ring or permuting the positions of these phosphate groups. Headgroups are shown for **(A)** I(1,3)P_2, **(B)** I(1,4)P_2, **(C)** I(1,5)P_2, **(D)** I(1,3,4)P_3, **(E)** I(1,3,5)P_3, **(F)** I(1,4,5)P_3 and **(G)** I(1,3,4,5)$_4$.

kinase C (PKC) and activates the kinase, which then goes on to promote downstream signalling.

This picture of phosphoinositide signalling was substantially modified by the discovery of phosphatidylinositol 3-kinase (PI3-K) in 1988 by Cantley and co-workers [22]. The PIPs phosphorylated at the 3′ position of the inositol ring (PI(3)P, PI(3,4)P_2 and PI(3,4,5)P_3) are not substrates for hydrolysis by phospholipases, and therefore they cannot lead to the production of the soluble inositol phosphate second messengers involved in Ca^{2+} signalling [23]. Instead, these PI3-K lipid products function as second messengers by directly recruiting proteins to the surface of cell membranes.

Specific phosphoinositide recognition by a lipid binding domain was first demonstrated for the interaction between PI(4,5)P_2 and a protein domain from pleckstrin [24], which was christened the pleckstrin homology (PH) domain when it was discovered in several other lipid binding proteins. It quickly became clear that many proteins con-

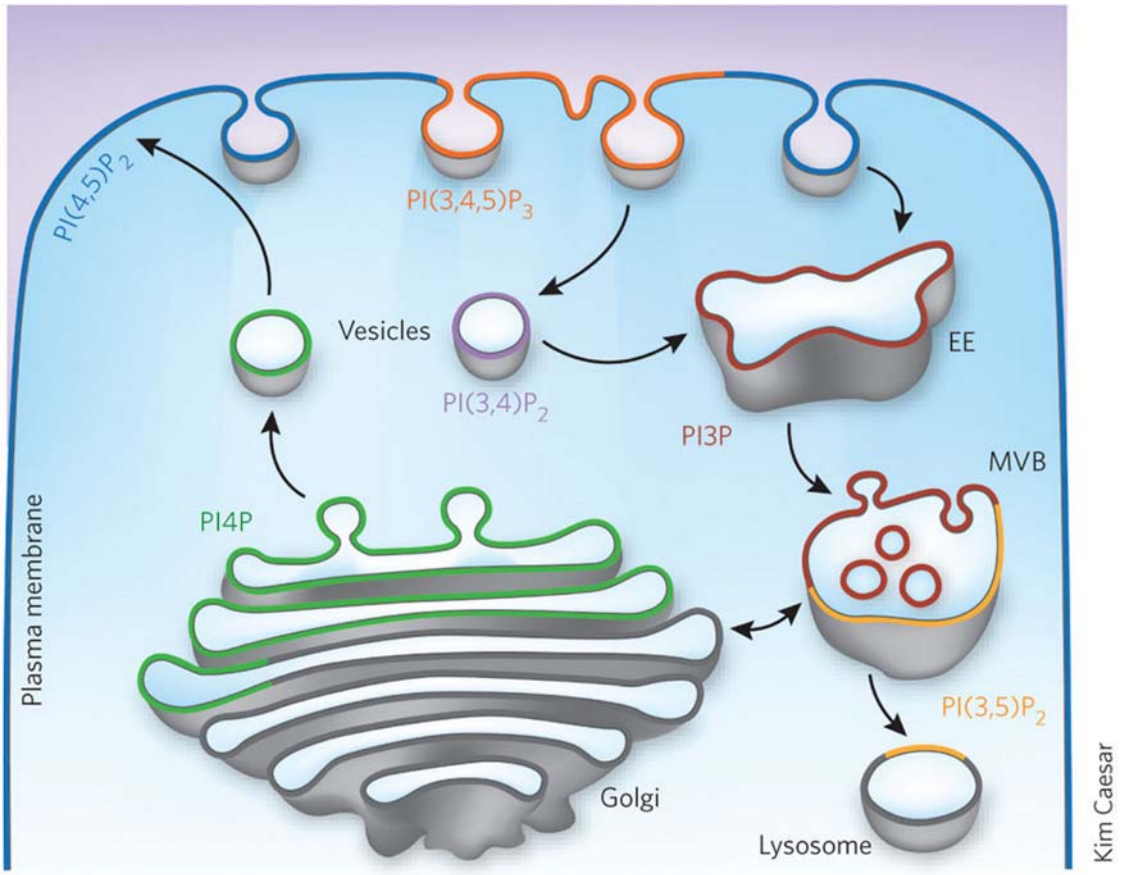

Figure 1.3: Subcellular localisation of PIPs. The distribution of PIPs across the intracellular membranes allows for spatial and temporal control over recruitment of PIP binding domains to subcellular compartments. EE: early endosome. MVB: multivesicular bodies. Reprinted by permission from Macmillan Publishers Ltd: Nature Chemical Biology, Kutateladze T.G. **6**, 507–513 (2010), © 2010.

tained a well defined lipid binding module as part of their fold, and that this conferred lipid binding specificity upon the host protein. At least 11 of these lipid binding modules have now been identified [12, 25], and many of them are adapted to recognise one or more of the seven species of PIPs [26, 27].

As the different PIP species are non-uniformly distributed across the cell organelles [28] they can be used as a marker to distinguish between different cellular compartments (Figure 1.3). For example, a large pool of $PI(4,5)P_2$ is constitutively present in the plasma membrane, whereas PI(4)P is mainly restricted to the Golgi. $PI(3,4,5)P_3$ is only produced at the plasma membrane in response to PI3-K activation, and so this heterogeneous and dynamic distribution of PIPs across the subcellular membranes allows

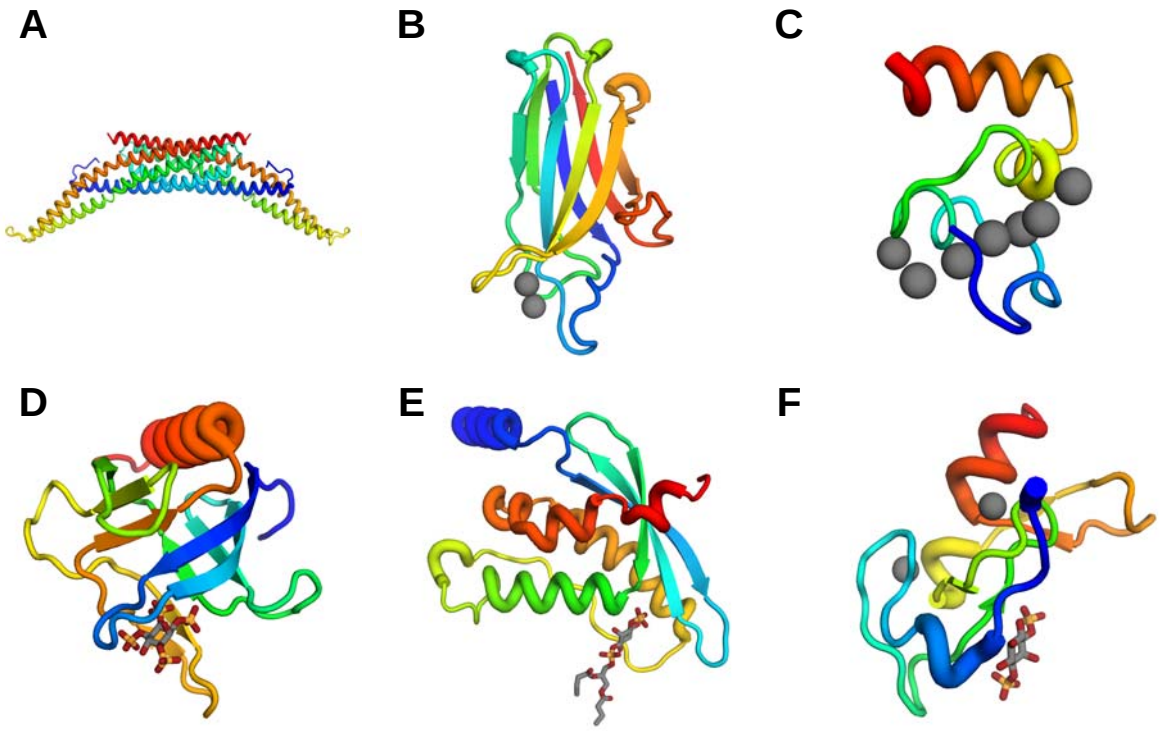

Figure 1.4: A selection of membrane-binding domains. **(A)** N-BAR domain dimer from amphiphysin (PDB 1URU) [29]. **(B)** C2 domain from PLCδ1 (PDB 1DJI) [30] with Ca^{2+} ions shown as grey van der Waals spheres. **(C)** GLA domain from blood coagulation factor VIIa (PDB 1DAN) [31], again with Ca^{2+} ions shown as grey van der Waals spheres. **(D)** PH domain from GRP1 (PDB 1FGY) [4] with the bound I(1,3,4,5)P_4 headgroup shown as a stick representation. **(E)** PX domain from p40^{phox} (PDB 1H6H) [32] with the bound I(1,3)P_2 headgroup shown as a stick representation. **(F)** FYVE domain monomer from EEA1 (PDB 1JOC) [33], with Zn^{2+} ions shown as grey van der Waals spheres and the bound PI(3)P ligand shown as a stick representation. Protein sequences are coloured across the colour spectrum from blue (N-terminus) to red (C-terminus). Proteins are shown in their approximate membrane-binding orientation, with the position of the membrane lying at the base of the figure in each case.

for spatial and temporal regulation of signalling. The relative specificity or promiscuity of a given lipid binding domain therefore provides the host protein with the means to preferentially target the correct cellular compartment(s).

1.2 Phosphoinositide binding domains

Membrane binding domains occur frequently in the human genome, and four of the most common are the PH domain (667 unique sequences in the Pfam database[1] [34]), the C2 domain (440 unique sequences), the Phox homology (PX) domain (177 unique

[1] `http://pfam.sanger.ac.uk`

sequences) and the Fab1, YOTB, Vac1, EEA1 (FYVE) domain (75 unique sequences). Lipid binding domains are usually soluble and associate only transiently with the membrane, and structure determination is therefore less difficult than would be the case for integral membrane proteins. As a result many structures of lipid binding domains solved using NMR and X-ray crystallography techniques are available in the Protein Data Bank. According to the Pfam database, structures of some of the major phospholipid binding domains are relatively abundant, and there are currently 109 PH domain structures, 41 C2 domain structures, 26 PX domain structures and 7 FYVE domain structures. A brief summary of the structural characteristics and lipid binding properties of each of these domains follows.

1.2.1 PH domains

The PH domain was first identified in the N-terminal domain of pleckstrin, a platelet protein, and was subsequently found to be present in a variety of proteins involved in signal transduction or cytoskeletal function [35, 36]. PH domains are unusual in that they share relatively low levels of sequence similarity but are nonetheless highly structurally conserved [37]. PH domains are known to bind to PI(3,4)P_2, PI(3,5)P_2 and PI(3,4,5)P_3, though the ligand specificity and promiscuity of different PH domains varies greatly [38].

A detailed discussion of the structural and functional characteristics of PH domains can be found in Chapter 3 for the case of GRP1-PH and in Chapter 5 in the context of kindlin-1 PH. Briefly, the PH domain consists of around 100 amino acid residues and its structure comprises a seven stranded β barrel, closed at one end by an amphipathic α helix and open at the other (Figure 1.4). The open end is the location of the PIP binding site, which is surrounded by three flexible, variable loops. These loops typically incorporate arginine and lysine residues to coordinate the phosphate groups of the PIP headgroup, and the length and precise amino acid composition of these loops determines the PIP binding specificity and affinity. Interestingly, although PH domains are known

primarily for their lipid binding properties, in fact only around 10% of them bind strongly to phosphoinositides [39] and the functions of the remainder are still unclear.

1.2.2 C2 domains

The C2 domain is typically around 130 amino acid residues in length, and was first discovered during investigations of the Ca^{2+}-dependent binding exhibited by protein kinase C [40]. The C2 domain fold is an eight stranded antiparallel β barrel (Figure 1.4), with the end of the barrel featuring variable loops much like the PH domain. C2 domains do not typically have a well defined lipid binding pocket and thus they tend to bind non-specifically to a variety of anionic [30, 41] and zwitterionic [42] lipid species through electrostatic interactions. This is often, but not always, Ca^{2+}-dependent. Membrane binding is therefore thought to be controlled by an electrostatic switch mechanism or possibly a Ca^{2+}-mediated bridging interaction. Structural differences between C2 domains that exhibit Ca^{2+}-dependent and Ca^{2+}-independent binding are described in Chapter 6. A more detailed discussion of electrostatic switches and bridging in C2 domains is provided later in this chapter (Section 1.6).

1.2.3 PX domains

The existence of the PX domain was first established by sequence analysis of the subunits of the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex, which identified high sequence similarity between the $p47^{phox}$ and the $p40^{phox}$ subunits. These results suggested that the two subunits both contained a homologous domain [43] and that this domain was also present in other proteins, for example, PI3-K. The first structure of a PX domain was solved by Bravo *et al.* [32], and this structure of $p40^{phox}$ bound to PI(3)P detailed the PIP recognition mechanism of PX domains. Interestingly, as is also the case for PH domains, PX domains show a broad spectrum of behaviour. While $p40^{phox}$ specifically binds PI(3)P, the $p47^{phox}$ domain binds PI(3,4)P_2 [44] and the PX domain

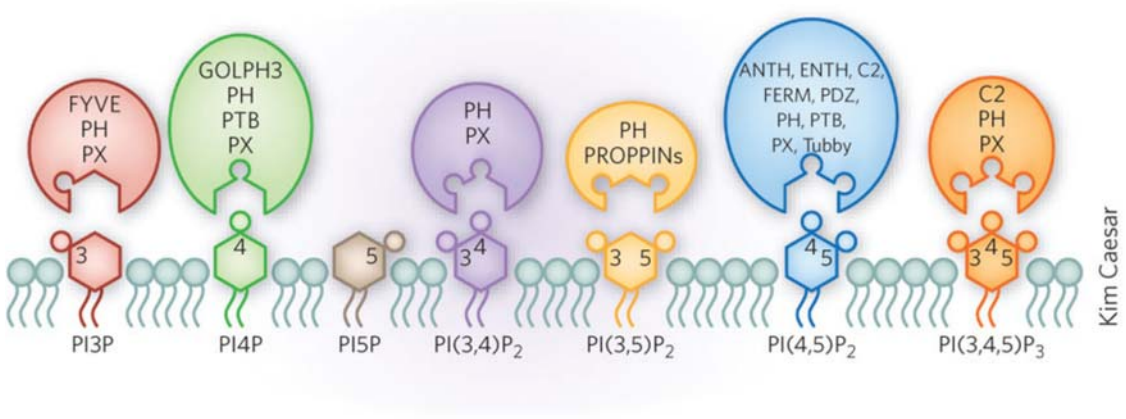

Figure 1.5: Lipid binding domains and their target PIPs. Colouring is the same as in Figure 1.3, reflecting the predominant distribution of PIPs across the cellular membranes. Reprinted by permission from Macmillan Publishers Ltd: Nature Chemical Biology, Kutateladze T.G. **6**, 507–513 (2010), © 2010.

from PI3-K C2α preferentially interacts with PI(4,5)P_2 [45].

1.2.4 FYVE domains

FYVE domains were first identified around 15 years ago [46], and are small, cysteine-rich domains comprising ~70-80 amino acid residues. FYVE domains are zinc fingers and bind two Zn^{2+} ions that are coordinated by the network of cysteine residues within the domain. FYVE domains are extremely specific for PI(3)P [47, 48, 49], and do not bind to other phosphoinositides. The first structure of a FYVE domain was reported by Hurley and Misra [50], and later a *holo* structure reported by Dumas *et al.* was able to shed light on the PI(3)P binding mode [33] (Figure 1.4).

1.2.5 Other phosphoinositide binding domains

Of the protein domains described above, FYVE and PX domains typically bind to PI(3)P and PH domains normally bind specifically to one or more of the following phosphoinositides: PI(3,4)P_3; PI(4,5)P_2 and PI(3,4,5)P_3 (Figure 1.5). However, comparatively little is known about possible effectors of the other physiological PIPs: PI(4)P; PI(5)P and PI(3,5)P_2 [12]. Recently, evidence has emerged that these PIPs may interact with a variety of other domains. For example, the PH domain from phosphatidylinositol four

phosphate adaptor protein 1 (FAPP1) has been shown to bind PI(4)P [51], and a recent crystal structure of the *apo* state of the FAPP1 PH domain suggested that binding was likely to occur in a similar fashion to the PIP recognition mode observed in other PH domains such as GRP1-PH [52].

The most recently discovered phosphoinositides are PI(3,5)P_3 [53, 54] and PI(5)P [55]. Dove *et al.* [56] identified a protein with a β propeller fold called ATG18, which interacts with PI(3,5)P_2. ATG18 is thought to be the prototype for a new family of PI(3,5)P_2 effectors, known as the β propeller binding phosphoinositides (PROPPINs), though no structure of a PROPPIN bound to PI(3,5)P_2 has yet been reported. The role and the subcellular distribution of PI(5)P have yet to be established (Figure 1.3 and Figure 1.5), though the plant homeodomain (PHD) finger of ING2 is believed to interact with PI(5)P [57]. Gozani *et al.* proposed that the recognition mechanism for PI(5)P was similar to that of the EEA1 FYVE domain for PI(3)P on the basis of homology modelling [57].

1.2.6 Probing PIP spatiotemporal distribution using lipid binding domains

The existence of such a broad variety of protein domains that selectively bind to different PIP species raised the prospect that these proteins could be used as biophysical probes to measure the *in vivo* spatiotemporal localisation and concentration of these lipids *in vivo*. Stauffer *et al.* [58] pioneered the use of confocal microscopy in combination with fluorescent tagging of PH domains to examine *in vivo* the subcellular distribution of PI(4,5)P_2 in real time. They produced fusion constructs of green fluorescent protein (GFP) and the PH domain of phospholipase C-δ1 (PLCδ1) and expressed the resulting GFP-PH by transfection into rat basophilic leukaemia cells. As PLCδ1 binds to PI(4,5)P_2 with high affinity, the GFP-PH probe first localised to the plasma membrane causing it to fluoresce. When Stauffer *et al.* added an agonist to induce PI(4,5)P_2 hydrolysis, they showed that GFP-PH dissociated from the plasma membrane and became uniformly distributed in the cytosol. In a related study, Venkateswarlu *et al.* [59] were able to use this technique in reverse to

observe localisation of the Arf nucleotide binding site opener (ARNO) PH domain to the plasma membrane in response to the addition of insulin. Insulin is able to stimulate PI3-K signalling, leading to elevated levels of PI(3,4,5)P_3 in the plasma membrane. ARNO, like the general receptor for phosphoinositides 1 (GRP1) (see Chapter 3), is a member of the cytohesin family of proteins that bind to PI(3,4,5)P_3 with high affinity.

This approach has the advantage of imaging PIPs in real time, but its utility can be limited by the comparatively low resolution of the confocal microscope and the fact that the large pool of PIPs in the plasma membrane can make it difficult to detect smaller intracellular pools of PIPs in, for example, the Golgi. An alternative method reported by Watt *et al.* [60] used immunoelectron microscopy on fixed tissue samples, again employing PLCδ1 to detect PI(4,5)P_2 but using gold labelled antibodies as a reporter. This technique provides improved quantification of PI(4,5)P_2 concentration in all parts of the cell, but does not provide any information on the real time dynamics. The strengths and weaknesses of these two approaches were briefly reviewed by Irvine [61].

A recent study by Cho and co-workers [62] described an innovative new method using an epsin N-terminal homology (ENTH) domain tagged with a polarity sensitive fluorophore as a biophysical probe for PI(4,5)P_2. This technique exploits the different polarities of the plasma membrane and the aqueous environment of the cytosol, as the spectral properties of the tagged ENTH sensor change depending upon the local dielectric constant. ENTH domains bind to PI(4,5)P_2, and the incorporation of a polarity sensitive tag allowed Cho and co-workers to quantify changes in PI(4,5)P_2 concentration using the observed blue shift in the emission spectrum on membrane binding of ENTH.

An approach that could perhaps be adapted to study PIP localisation is one used recently to detect syntaxin-1A clustering [63]. Super resolution stimulated emission depletion (STED) microscopy was used to show that PI(4,5)P_2 accumulates in large PI(4,5)P_2-rich microdomains approximately 750 Å in diameter, and that clusters of syntaxin-1A co-localise at the surface of these microdomains in the inner leaflet of PC12 cell plasma

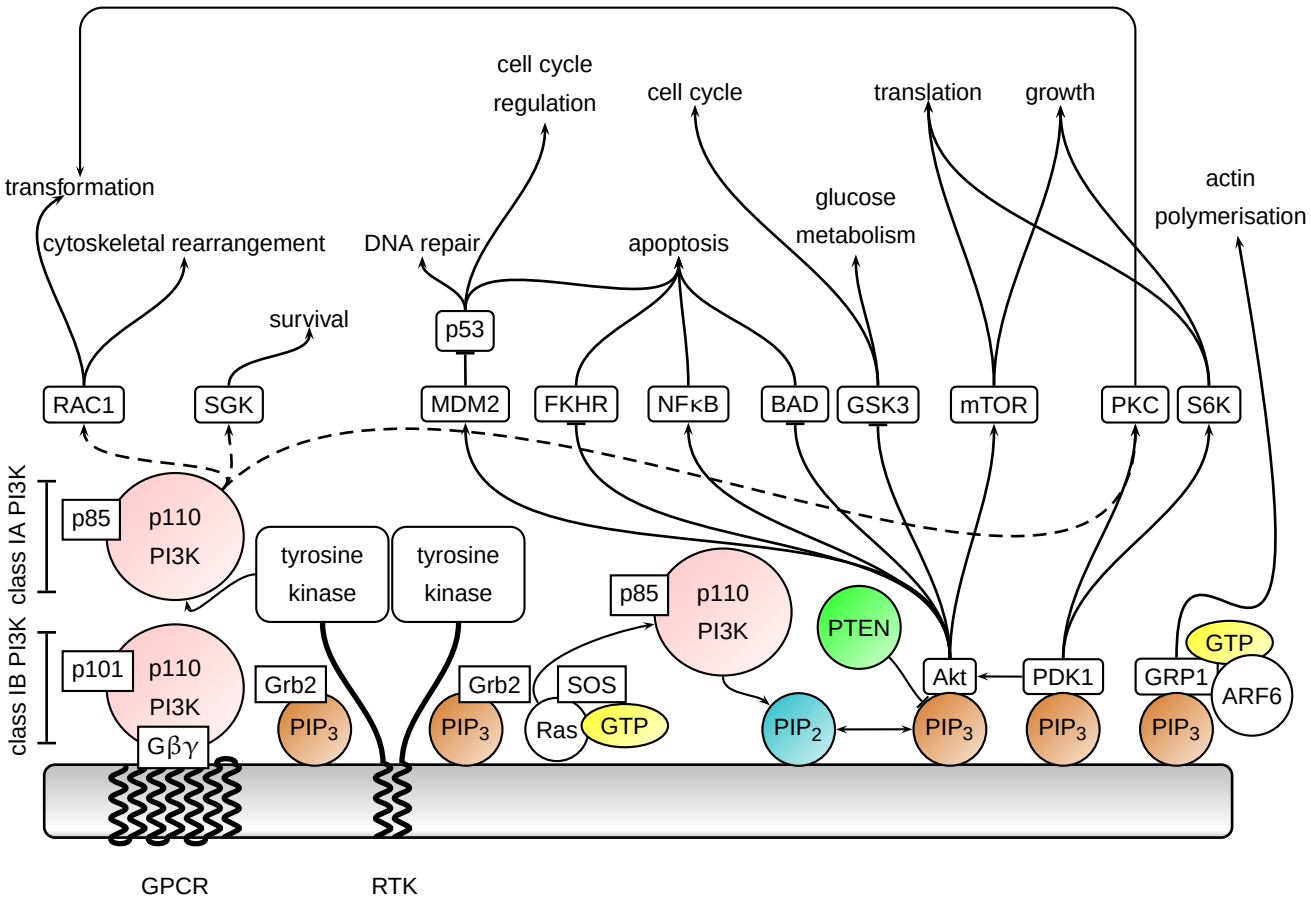

Figure 1.6: Schematic of PI3-K signalling. Class IA PI3-K is recruited to the membrane and activated through receptor tyrosine kinases. Ras [64] can also contribute to activation of class IA PI3-Ks [65]. Biosynthesis of $PI(3,4,5)P_3$ from $PI(4,5)P_2$ by PI3-K leads to recruitment of PH domain containing proteins such as PKB/Akt and GRP1 to the surface of the membrane, where they are activated and initiate cellular processes involved in growth and survival. Figure adapted from Refs [66], [67] and [68]. In general, node connections with arrows (→) represent activation while capped node connections (⊣) represent inhibition.

membranes.

1.3 Phosphoinositides in cell signalling

Membrane binding proteins typically need to bind to the cytoplasmic leaflet of the membrane before being able to carry out their function, and PIP lipids are often responsible for successfully recruiting these proteins to the membrane surface. Since the discovery of PI3-K by Cantley and co-workers [22], PIP lipids have been shown to play a greatly

expanded role in cell signalling, beyond the I(1,4,5)P_3/Ca^{2+} and the DAG/PKC signalling discussed in Section 1.1 [27, 69]. In particular, PI(4,5)P_2 and PI(3,4,5)P_3 were found to sit at the base of a network of interactions known as the PI3-K pathway [66].

PI3-K signalling is complex, and here the discussion will be limited to the activity of class IA PI3-Ks, which transmit downstream signals from receptor tyrosine kinases [68]. Class IA PI3-Ks are responsible for converting PI(4,5)P_2 to PI(3,4,5)P_3 by phosphorylation of the lipid substrate at the 3′ position of the inositol ring. Protein kinase B (PKB/Akt) then binds to PI(3,4,5)P_3 through its PH domain, and goes on to promote cell survival, proliferation and growth. Phosphatase and tensin deleted on chromosome ten (PTEN) is a tumour suppressor that regulates PI3-K signalling by dephosphorylating PI(3,4,5)P_3 to recover PI(4,5)P_2, thereby opposing the activity of the kinase [70]. A schematic of PI3-K signalling showing the interplay between some of the key participants in the signalling cascade is sketched out in Figure 1.6.

As PI3-K signalling orchestrates so many cellular processes related to growth, repair and survival (Figure 1.6), abnormal PI3-K signalling has been identified as a causal agent in the development of cancer. Oncogenic mutations in PI3-K lead to an increase in PI(3,4,5)P_3 levels and unrestrained activation of Akt signalling [71], and although large scale exomic sequencing has now identified the most frequent mutations that lead to tumourigenesis [72] the molecular mechanism remains unclear. As PI3-K drives cell proliferation and the tumour suppressor PTEN acts to prevent it, mutations that impair the phosphatase activity of PTEN can also lead to tumourigenesis. An investigation into the molecular mechanism of disease for a selection of oncogenic PTEN loss of function mutations forms part of Chapter 6.

1.4 Experimental biophysics and protein-lipid interactions

What follows is a brief overview of the principal experimental techniques used to study protein-lipid interactions, along with some specific examples of their application to the

membrane binding domains discussed above. Some of these techniques have been used to complement work described in this thesis, notably solution NMR and monolayer penetration experiments (Chapter 3), while evidence from other experimental methods is referred to when discussing simulation results in later chapters. Several of these have been reviewed in more detail by Scott *et al.* [73].

1.4.1 Nuclear magnetic resonance spectroscopy

The field of biomolecular nuclear magnetic resonance (NMR) spectroscopy is enormously broad and the technique has found applications in many areas of experimental biophysics, including the study of protein conformational dynamics [74], protein folding [75] and protein-protein interactions [76]. One NMR technique that can be applied to protein-lipid (or protein-ligand) interactions, if the structure of the protein is known, is NMR chemical shift mapping [77]. In chemical shift mapping, the NMR spectrum of a protein in solution is acquired and the resonances are assigned. Then, detergent micelles are titrated into the solution over a range of concentrations. The changes in chemical shift are monitored, and those residues that exhibit the largest perturbations are deemed to interact with the detergent micelles to the greatest extent. These can then be mapped onto the structure of the protein, which then provides information on the location of the likely lipid binding site. This technique has been used to assess binding of GRP1-PH binding to PI(3,4,5)P_3-containing micelles [78] and a similar method was used to determine the molecular mechanism of PX domain binding to PI(3)P-containing micelles [79].

1.4.2 Solid state NMR spectroscopy

While solution NMR has been widely applied to the study of membrane proteins, there is a practical limit to the size of the proteins that can be studied. Large proteins tumble too slowly, and so the anisotropic interactions are not averaged on the NMR time scale. This broadens the spectral linewidths, making the NMR spectra difficult to interpret.

Membrane proteins, which need to be encapsulated by a detergent micelle or bicelle to solubilise the protein for solution phase NMR, are even more problematic in this respect as the bound detergent molecules increase the effective size and mass of the protein.

Solid state NMR (ssNMR) provides an alternative method to study membrane proteins and protein-lipid interactions. In contrast to conventional solution state NMR, ssNMR is performed using a solid sample. The acquisition of high resolution spectra is not dependent upon rapid, isotropic tumbling in solution [80], but instead can be obtained by mechanically spinning the sample at the so-called magic angle (∼54.7°) to orientationally average the interactions and reduce line broadening. One of the advantages of ssNMR is that it typically allows for more versatility in the sample composition than would be the case for analogous solution NMR studies, allowing for greater flexibility when investigating the effect of different lipid compositions on membrane proteins [81]. Related to this, there is some evidence that membrane protein conformations can be altered by the local bilayer environment and membrane curvature [82]. Solid state NMR with a phospholipid bilayer may arguably therefore provide a closer approximation to a physiological membrane environment than solution NMR using detergent micelles, though with the development of new technologies such as lipid nanodiscs membrane mimetic environments compatible with solution NMR now exist [83]. The ssNMR technique has been applied to membrane-binding proteins at lipid bilayer surfaces, for example in the case of cytochrome *c* [84] and the PH domain of PLCδ1 [85].

1.4.3 Electron paramagnetic resonance spectroscopy

Electron paramagnetic resonance (EPR) spectroscopy is a technique related to NMR that relies on the spins of unpaired electrons rather than nuclear spins. These spins are incorporated into proteins by site directed spin labelling (SDSL). Residues of interest are mutated to cysteines by site directed mutagenesis, and these are then covalently coupled to thiol specific spin labels. The lineshape of the EPR spectrum is very sensitive to the

mobility and dynamics of the spin label, so the local environment of the spin label can be inferred from the EPR spectrum. Docking to lipid vesicles can therefore be detected by EPR, as the spin label mobility changes with the decrease in global tumbling observed when an isolated protein domain binds to a larger lipid vesicle. The docking geometry of the protein domain can also be determined by using continuous wave EPR (CW EPR) power saturation measurements in combination with the addition of molecular O_2, which has an unpaired electron and is a paramagnetic relaxing agent. Molecular oxygen preferentially partitions into the lipid hydrophobic environment, and adding O_2 increases the microwave power needed to saturate the EPR signal for residues inserted into the membrane. These techniques have been applied to determine the membrane docking geometries of C2 domains from PLA_2 [86] and PKCα [87], and also GRP1-PH [88].

1.4.4 Fluorescence resonance energy transfer

Fluorescence resonance energy transfer (FRET) is used to measure the proximity of two species, each of which is tagged with a fluorophore. Excitation of one fluorophore (the donor) can result in non-radiative energy transfer to the other fluorophore (the acceptor) if the separation between the donor-acceptor pair is small. Hamman *et al.* [89] described a simple FRET based assay to probe the interaction between a GFP tagged PH domain and PI(3,4,5)P_3-containing lipid vesicles. The vesicles were doped with octadecylrhodamine (OR), which is a lipophilic FRET acceptor for GFP. Efficient FRET occurs only when the PH domain binds to PI(3,4,5)P_3 at the vesicle surface. Sato *et al.* [90] used a more elaborate FRET based technique to measure PI(3,4,5)P_3 binding of GRP1-PH in living cells by engineering a chimaeric construct at the surface of the plasma membrane (Figure 1.7). Cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP) act as the donor and acceptor molecules respectively, as the fluorescent emission peak of CFP overlaps with the excitation wavelength of YFP. Through conformational changes of the fusion construct upon PI(3,4,5)P_3, CFP is brought sufficiently close to YFP for efficient

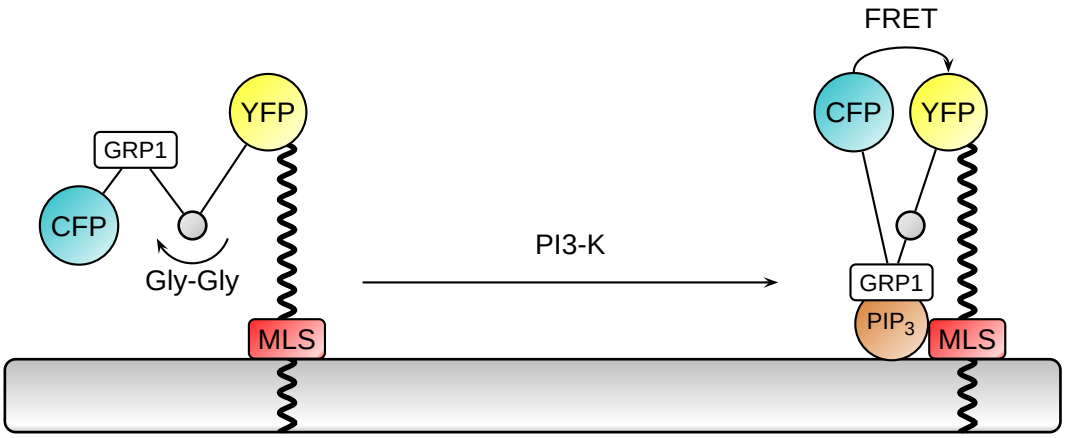

Figure 1.7: Experimental setup of Sato *et al.* [90]. CFP, GRP1-PH and YFP are fused to generate a chimaeric indicator protein, which is then tethered to the membrane through a rigid α helical linker and a membrane localisation sequence (MLS). After PI3-K activation, GRP1-PH binds to PI(3,4,5)P_3 and a conformational change occurs in the chimaeric indicator, facilitated by the inclusion of a flexible diglycine motif in the linker. This brings CFP and YFP into close proximity, allowing for efficient FRET.

FRET to occur.

1.4.5 Surface plasmon resonance

Surface plasmon resonance (SPR) is growing in popularity as a technique to characterise protein-membrane binding [91] and has been used, for example, to study the membrane binding properties of PLA_2 [92] and of C1 [93] and C2 domains [94]. Novel SPR techniques are also being developed in order to study integral membrane proteins [95]. In SPR, polarised laser light is directed up through a prism, with the upper surface of the prism covered by a thin layer of gold film. A membrane mimetic surface, typically either a supported lipid monolayer or immobilised liposomes, is formed on top of the gold film. At a critical angle the beam of light undergoes total internal reflection and produces an evanescent wave that excites surface plasmons, collective oscillations of mobile electrons, in the gold film at the boundary between the gold surface and the lipid monolayer. The surface plasmons absorb the energy of the incident light, with a corresponding decrease in intensity of the reflected light. The critical angle is extremely sensitive to the refractive index of the solution on the upperside of the gold layer, as this changes the nature

of the interface along which the surface plasmons propagate. The refractive index of the solution is altered by proteins binding to the lipid monolayer, which in turn results in a shift in the critical angle, and so the extent of protein adsorption to the monolayer can be assessed by monitoring the value of this critical angle at which the drop in intensity is observed.

1.4.6 Total internal reflection fluorescence microscopy

Total internal reflection fluorescence microscopy (TIRFM) is related to SPR in that it also exploits the phenomenon of total internal reflection and evanescent waves to assess binding of proteins to cell membranes. The technique uses proteins tagged with a fluorophore and observes the resulting fluorescence after illumination with laser light. Ordinary fluorescence microscopy will excite all of the fluorophores in the sample, and will not discriminate between membrane-bound proteins and those in the rest of the sample. The advantage of TIRFM is that the evanescent wave produced at the critical angle is used to excite the fluorophores. Away from the interface the evanescent wave decays rapidly, and so TIRFM can be used to selectively illuminate only those proteins bound at the surface, eliminating the background fluorescence from the other proteins in solution. TIRFM has been used to study GRP1-PH binding to PI(3,4,5)P_3 embedded in a lipid bilayer [96], and single particle tracking of the GRP1-PH domains bound at the surface was then used to estimate the lateral diffusion coefficient of the protein. It can also be used quantitatively to obtain binding kinetics [97]. One drawback of this technique is the difficulty in determining whether or not the protein has dissociated from the lipid bilayer, as the lack of fluorescence emission may be due to protein dissociation or may simply be caused by photobleaching of the dye label. This effect can be minimised by careful control of the power of the laser.

1.4.7 Monolayer penetration studies

Monolayer surface tension experiments are carried out using a Langmuir-Blodgett trough. The trough is filled with a buffer solution and a lipid monolayer is spread onto the surface of this subphase until a desired surface pressure, π_0, is reached. The lateral pressure of the monolayer is measured, and the change in this surface pressure, $\Delta\pi$, is then monitored as protein is injected into the subphase. Increases in surface pressure are interpreted as evidence of protein-lipid interactions, as the insertion of the protein into the lipid monolayer forces the lipids apart and increases the surface pressure. The greater the increase in surface pressure the greater the degree of monolayer penetration of the protein. This technique has been used to characterise membrane insertion of GRP1-PH [98, 78].

1.5 Molecular simulations applied to biological systems

All of the experimental techniques described above are able to provide insights into the nature of protein-lipid interactions, either *in vivo* or *in vitro*. However, with the advent of new computational techniques and improved computing power, molecular simulations offer the prospect of observing, *in silico*, events in atomic detail.

The first simulation of a protein, bovine pancreatic trypsin inhibitor, was performed by McCammon *et al.* in 1977 [99] and the first simulation of a lipid bilayer followed soon afterwards [100]. Later, simulations of hydrated bilayers became possible [101, 102], moving onto mixed lipid systems [103] as computational power increased. Simulations of the interaction between a protein and a bilayer were first reported for the transmembrane proteins gramicidin [104, 105] and bacteriorhodopsin [106]. The first reported simulation of a peripheral protein, PLA_2, interacting with lipids appeared in 1996 [107].

Since then, MD simulation has become established as a powerful tool for investigating protein-lipid interactions [108], and MD simulations are now routinely used to comple-

ment the experimental biophysical techniques described previously. Simulation studies reported to date reflect the diverse array of lipid binding domains discussed above, and include MD simulations of membrane-binding modules such as Bin/Amphiphysin/RVS (BAR) domains [109, 110], C2 domains [111, 112, 87], γ carboxyglutamic acid rich (GLA) domains [113], PH domains [114] and PX and FYVE domains [115].

In the next section, recent simulation work on lipid binding domains is briefly reviewed. N-BAR domains are much larger than the other proteins listed above (Figure 1.4), and studying large protein-membrane complexes using molecular simulation remains challenging. This is due to the size of the system and also the time scales required to observe effects such as membrane remodelling that are characteristic of N-BAR domains. A review of computational work on N-BAR domains is therefore deferred until Chapter 2, where it is used as an illustrative example in the discussion of CG models for molecular simulation.

1.6 Previous computational studies of lipid binding domains

1.6.1 Bioinformatics

Although some PIP binding domains such as the PH domain are structurally very conserved they often exhibit very little sequence homology, and this diversity of peripheral proteins occasionally makes it difficult to determine the likelihood of a particular protein binding to the membrane purely from its amino acid sequence. Bhardwaj *et al.* [116] developed a prediction algorithm based on machine learning that was able to distinguish between proteins that bound to the membrane and those that did not solely on the basis of sequence and structure. This was achieved by assessing each of the proteins using a set of criteria, which included the net charge and the distribution of positive charge across the protein surface. Bhardwaj *et al.* used their algorithm to predict the binding propensities of four C2 domains, and found that it was able to accurately predict which of the

domains would bind to the membrane. They confirmed their results using SPR experiments. The prediction server has been integrated with the Membrane Targeting Domains Resource (MeTaDoR) [117], which provides detailed information on the structure and sequence of a wide variety of membrane targeting domains.

Where possible, MeTaDoR links to another resource, the Orientation of Proteins in Membranes (OPM) database [118]. While this was initially set up to catalogue the membrane orientations of transmembrane proteins [119], it has now expanded to include peripheral and monotopic membrane proteins [120]. In the OPM database, the free energy of each peripheral protein in the membrane is given by the sum of the free energy of transfer of a rigid protein from water to the hydrocarbon core of the membrane and the ionisation energy of the charged residues. Although the proteins remain rigid, the authors do rearrange the side-chain rotamers to remove charged amino acid residues from the hydrophobic core of the membrane. The orientation and position of each peripheral protein in the membrane is optimised by minimising this free energy with respect to three parameters: the tilt of the protein relative to the membrane normal; the rotation of the protein about the membrane normal and the depth of insertion. The resulting orientations are generally in good agreement with experiment, and the authors found that most peripheral proteins do penetrate the membrane surface to some degree. They also concluded that hydrophobic interactions are usually the main factor in determining the orientation of a protein at the surface of a membrane, demonstrating that although electrostatic interactions may dominate at long range, at short range other factors can also become important.

1.6.2 Continuum electrostatics calculations

Although Lomize *et al.* [120] concluded that protein orientation in membranes was controlled by mainly hydrophobic effects, electrostatic interactions still have a substantial in-

fluence on protein-membrane interactions. Continuum electrostatics calculations[2] have been performed on lipid-binding domains to determine their orientation on the surface of the membrane. Diraviyam *et al.* [121] used Poisson-Boltzmann electrostatics calculations to study the membrane-associated states of the FYVE domain and interestingly found that, in addition to the simple protein-ligand interaction, the negatively charged PIP may also be able to act as an 'electrostatic switch' by masking the positive charge on nearby residues and allowing them to insert more deeply into the hydrophobic membrane. Their work suggested a dual role for the PIP in promoting membrane binding. Singh and Murray [122] built homology models of all of the structures of PLC and characterised their binding properties using continuum electrostatics calculations. Some of the members of the PLC family have been studied experimentally while others have not, allowing Singh and Murray to validate their method based on experimental results and simultaneously make predictions regarding the binding properties of the less well-studied isoforms.

While the PH and FYVE domains discussed above are specific for one or other of the PIP species, other domains are much less discerning when choosing lipids to target. C1 and C2 domains for example normally display a wider range of lipid selectivity and, particularly in the case of the former, will generally bind in a non-specific fashion to any negatively charged membrane surface [123]. Interestingly, the binding of some C2 domains is mediated by Ca^{2+} ions, which bind to the surface of the domain in well defined calcium binding loops comprised primarily of aspartate residues that are able to coordinate the positively charged ions [124]. The Ca^{2+} ions are thought to screen the negative charges on the calcium-binding loops, allowing the domain to bind to the membrane, though others have suggested that they act as 'inorganic bridges' connecting the protein and the membrane. The binding affinity of some C2 domains is therefore $[Ca^{2+}]$-dependent, and the concentration of Ca^{2+} ions can be used to regulate the ability

[2]Chapter 2 discusses the Poisson-Boltzmann equation and describes how continuum electrostatics calculations are carried out in practice.

of C2 domains to bind to the membrane[3].

Murray and Honig performed electrostatics calculations on several C2 domains [125] and suggested that many of the experimentally observed binding properties could be accounted for by non-specific electrostatic interactions between the domain and the membrane. Interestingly, they also observed that the presence of bound Ca^{2+} ions can promote membrane binding in two ways: by providing positive charge that is then attracted to negatively charged membranes, and also by reducing the electrostatic desolvation penalty incurred upon binding by neutralising clusters of negatively charged lipid headgroups.

1.6.3 Molecular dynamics simulations

Although bioinformatics and continuum electrostatics approaches have met with considerable success when applied to the interaction of peripheral proteins with membranes, these methods are inherently somewhat limited as they are based on purely *static* structures. Although structure is inextricably linked to function, protein dynamics can also be functionally relevant in a wide variety of processes, from catalysis [126] to signalling [127]. Molecular dynamics (MD) simulations offer one way to investigate protein dynamics and flexibility. Although computationally they are more expensive than the methods using static structures described above, they are able to capture a broad range of time varying phenomena in atomic detail.

An early study on the PH domain of Akt used homology modelling and docking to generate a bound protein-ligand complex. The authors used MD simulations to probe the dynamics of the hydrogen bonding network and an empirical approach to predict the binding affinity for the isolated headgroups of various PIP species [128]. Later, another study again used a docking approach to study the PH domain of a member of the Akt family of proteins. This time, in tandem with experimental studies, docking was used

[3]Chapter 6 discusses C2 domains in more detail, in particular focussing on the C2 domain of PTEN, which unusually exhibits Ca^{2+}-independent binding.

to assess whether or not a benzene polyphosphate compound - an analogue of the isolated I(1,3,4,5)P_4 headgroup with different regiochemistry - was able to adopt the same binding mode as its inositol-based counterpart [129]. In a similar vein, a docking approach derived from NMR data was recently presented by Dancea *et al.* [130], which was used to predict how PX and FYVE domains might position themselves when bound to PIP species contained within lipid micelles. The docking poses were generated from NMR data and refined using a combination of simulated annealing and restrained MD. Rosen *et al.* recently presented extensive MD simulations of the headgroup of I(1,4,5)P_3 bound to Akt in solution [131], though these were conducted in the absence of a lipid bilayer or other membrane mimetic entity.

Dynamic simulations of the behaviour of PIP binding domains in the presence of lipid environments over longer timescales have also been reported. Several experimental studies [132, 98] have previously detected non-specific interactions between bound PH domains and the surrounding membrane lipids, so Psachoulia and Sansom [114] performed simulations of *holo* PH domains from PLCδ1 isolated in solution, as well as those that were embedded in detergent micelles and bound to lipid bilayers, in an effort to understand the influence of a hydrophobic environment on their binding properties. In agreement with experimental results, they found that the PH domain bound to the ligand more tightly in the presence of a lipid bilayer, suggesting that these secondary interactions with the surface of the bilayer reinforced the primary interaction with the PI(4,5)P_2 molecule.

Later, Psachoulia and Sansom extended their work to PX and FYVE domains, again looking at their behaviour in solution and in a lipid bilayer [115]. Their simulations agreed well with prior biophysical studies on these domains but also highlighted additional interactions with the bilayer not observed in experiments, which were mainly conducted using detergent micelles. The different behaviours exhibited by the protein in micellar and lipid bilayer environments emphasised the importance of the substrate in

protein-membrane interactions.

Dynamical approaches have also been applied to C2 domains, notably by Manna *et al.* [112] and Jaud *et al.* [111]. Manna *et al.* used atomistic MD to investigate the binding properties of a C2 domain, beginning with the domain bound to the membrane in the experimentally determined binding orientation. Interestingly, they found that while the addition of the PIP ligand stabilised the protein-membrane interaction to some degree, binding was primarily dependent on the presence of Ca^{2+} ions and negatively charged lipids in the bilayer.

Jaud *et al.* [111] performed MD simulations on the C2 domain of PLA_2 complexed with Ca^{2+} ions and bound to the surface of a membrane. The initial position of the ions was assigned using data gleaned from EPR spectroscopy. Perhaps their most interesting finding was that, in agreement with the results of the electrostatics calculations performed by Murray and Honig [125], the calcium ions formed interactions exclusively with the C2 domain and the surrounding water molecules and did not interact with the lipid headgroups. Therefore, rather than acting as inorganic bridges mediating the interaction between the two organic species, the Ca^{2+} ions instead bind solely to the C2 domain in another example of an electrostatic switch mechanism. The C2 domain with no Ca^{2+} bound is negatively charged and in the 'off' state, since electrostatic repulsion prevents it from binding to the membrane. Upon binding of Ca^{2+} the charge on the calcium-binding loops is screened and the protein switches to the 'on' state and is now able to bind to the membrane.

In a related study, Ohkubo and Tajkhorshid [113] performed MD simulations of another domain dependent upon the concentration of Ca^{2+} ion for successful binding to the membrane. The GLA domain (Figure 1.4) is found in some blood clotting proteins, and in this study the authors used the GLA domain from blood coagulation factor VIIa and studied its Ca^{2+}-dependent binding properties. The GLA domain of factor VIIa binds seven Ca^{2+} ions, and the authors found that these ions played two distinct roles. Some

of the ions, as above, did not interact with the surface of the membrane, but in this case the authors discuss a possible mechanism whereby these ions 'brace' the structure of the protein, maintaining the rigidity of the hydrophobic ω-loop crucial for membrane insertion. Those ions that do interact with the lipid headgroups were thought to be involved in a bridging mechanism, similar to that postulated above for the C2 domain.

While experimental studies had previously determined that the binding of both C2 and GLA domains was Ca^{2+}-dependent, the underlying cause of this dependency was not immediately clear. The simulation studies detailed above have offered new insights by allowing the authors to directly observe the behaviour of the ions by probing these binding events at the atomic level. The electrostatics calculations coupled with the MD simulations therefore reveal several possible roles for the Ca^{2+} ions in regulating the binding affinity of these domains.

A recent innovation from Tajkhorshid and co-workers [133] is their so-called highly mobile membrane-mimetic (HMMM) model. This is a biphasic approximation to a lipid bilayer that represents the hydrophobic core of the membrane with an organic solvent layer and the headgroups by short tailed phospholipids. The HMMM model eschews all atomic detail in the hydrophobic core in favour of an extremely coarse grained description that allows for rapid insertion of proteins into the membrane. While this has been shown to enhance sampling by increasing membrane penetration and diffusion rates, questions remain regarding its wider applicability. For example, the HMMM model does not reproduce important membrane mechanical properties such as compressibility.

1.7 Thesis outline

This thesis describes computational studies on a variety of membrane binding domains involved in cell signalling, building on previous work carried out in the Sansom group [114, 115]. In particular, this thesis attempts to address four related questions.

First, can simulations be used to model lipid binding of a peripheral protein and do

the results compare favourably with experimental data (Chapter 3 and Chapter 5)? Second, is it possible to use simulations to probe the targeting behaviour of a lipid binding domain as it attempts to locate its cognate ligand in the complex environment of the cell membrane (Chapter 4)? Third, are simulations able to describe the influence of non-specific anionic lipid interactions, in addition to the PIP ligand, on membrane targeting and binding (Chapter 4 and Chapter 6)? Fourth, can these methods be extended beyond isolated lipid binding domains to larger, more complex proteins with multidomain architectures (Chapter 6)? Looking ahead, the ultimate goal of these studies would be to develop a multiscale protocol to model the entire signalling event, incorporating membrane association, binding, penetration and eventual dissociation.

These four related questions are addressed in four results chapters, with the thesis organised as follows. Chapter 3 describes equilibrium and non-equilibrium molecular dynamics (MD) simulations of the PH domain from the general receptor for phosphoinositides 1 (GRP1-PH) bound to PI(3,4,5)P_3 in a lipid bilayer. This work was in collaboration with colleagues in the USA, who performed NMR spectroscopy and SPR experiments on GRP1-PH. The high affinity and specificity of GRP1-PH for PI(3,4,5)P_3 make it an attractive target for study, and this, coupled with the additional experimental data provided by colleagues in the USA, allows GRP1-PH to serve as a useful initial test case to validate results from simulations.

Another study combining experiment and simulation is discussed in Chapter 5. Here, some of the computational techniques developed to investigate GRP1-PH in Chapter 3 are applied to a novel structure of the PH domain from kindlin-1. This is the first reported structure of kindlin-1 PH, which was recently determined using X-ray crystallography by colleagues in Oxford. However, little is known about its lipid binding properties, and this chapter describes the results of a combined experimental and computational structure-function study of kindlin-1 PH.

Chapter 4 and Chapter 6 focus purely on computational work, and exploit 'multiscale'

simulation approaches to explore different stages of membrane binding. Chapter 4 revisits the case of GRP1-PH, this time detailing Brownian dynamics (BD) simulations of the protein, which have been adapted to explore the initial encounter event between GRP1-PH and a model of the cell membrane. Chapter 6 considers the more complex case of the PTEN tumour suppressor, which comprises two domains. The behaviour of wild type and mutant PTEN at membrane surfaces is investigated using both CG and atomistic MD simulations. However, before these results are presented, the next chapter briefly reviews the theoretical basis and practical implementation of some of the techniques used in this thesis.

Chapter 2

Theory and Methods

COMPUTER simulations are beginning to play an increasingly important role in studies of biological molecules [134]. Simulations have now become established as a standard biophysical tool alongside existing experimental techniques, and are routinely used to investigate the behaviour of an enormous variety of systems in molecular detail. However like all techniques, simulations do have their limitations.

This chapter briefly describes some of the theoretical background underpinning a range of modern molecular simulation techniques. After setting out the theory, the discussion then turns to more practical matters and outlines strategies for implementing such simulations.

2.1 Molecular dynamics simulations

The aim of molecular dynamics (MD) simulations is to model the dynamical behaviour of a collection of atoms. By monitoring the time evolution of a set of observable quantities,

MD simulations can be used as 'computer experiments' and provide insights into the structural and dynamic characteristics of a system at the atomic level. Supposing there are N atoms in a system, then given the coordinates and momenta of all of the particles at a time, t, the coordinates and momenta at a later time, $t + \Delta t$, can be obtained by solving Newton's equations of motion. For N atoms in a classical[1] MD simulation, these are a set of $3N$ coupled, second order differential equations given by:

$$\begin{aligned} F_i &= m_i \ddot{r}_i \\ &= -\nabla_i V_N(r^N) \end{aligned} \tag{2.1}$$

Where m_i is the mass of the i^{th} atom, F_i is the force on this atom and $\ddot{r}_i$ is its acceleration, the second derivative of the atomic position with respect to time. The quantity $V_N(r^N)$ is the potential describing the interaction of the atoms, and F_i is given by the derivative of this potential. If this potential is continuous then the atom trajectories cannot be solved exactly, and so the set of differential equations shown above must be solved numerically using finite difference methods.

2.1.1 Integrating the equations of motion

One method for solving the equations of motion is due to Verlet [135], and is described in some detail by Hansen and McDonald [136]. If the coordinates of atom i at a time t are r_i then performing a Taylor expansion forwards and backwards in time about r_i gives the following expression for the forward expansion:

$$r_i(t + \Delta t) = r_i(t) + \Delta t\, \dot{r}_i(t) + \frac{\Delta t^2}{2!} \ddot{r}_i(t) + \mathcal{O}(\Delta t^3) \tag{2.2}$$

[1]The term 'classical' here refers to the assumption that the Born-Oppenheimer approximation holds and that electronic degrees of freedom are therefore neglected. The Born-Oppenheimer approximation assumes that the motion of the heavy atomic nuclei and that of the comparatively low mass electrons is effectively separated, such that the electronic configuration can instantaneously readjust to any nuclear motions. This has the effect that the overall energy in the electronic ground state is only dependent upon the nuclear positions, and hence in a classical MD simulation the position of each atom is defined solely by the position of its nucleus and is represented by a single set of Cartesian coordinates.

The corresponding expression for the backward expansion is:

$$r_i(t-\Delta t) = r_i(t) - \Delta t\, \dot{r}_i(t) + \frac{\Delta t^2}{2!}\ddot{r}_i(t) - \mathcal{O}(\Delta t^3) \tag{2.3}$$

Adding these two expressions gives an estimate for the particle coordinates at time $t+\Delta t$:

$$r_i(t+\Delta t) \approx -r_i(t-\Delta t) + 2r_i(t) + \frac{\Delta t^2}{m_i}F_i(t) \tag{2.4}$$

Subtracting Equation 2.3 from Equation 2.2 yields an estimate for the atom velocity $\dot{r}_i$ at time t:

$$\dot{r}_i(t) \approx \frac{1}{2\Delta t}[r_i(t+\Delta t) - r_i(t-\Delta t)] \tag{2.5}$$

Thus, the Verlet algorithm provides a finite difference method to solve Newton's equations of motion and calculate the atomic trajectories over time. An alternative to the Verlet algorithm is the so-called 'leap frog' algorithm [137], which evaluates velocities at half integer time steps and uses these velocities to compute the new positions. Following the treatment in Frenkel and Smit [138], the velocities at half integer time steps are first defined as follows:

$$\dot{r}\left(t-\frac{\Delta t}{2}\right) = \frac{r_i(t) - r_i(t-\Delta t)}{\Delta t} \qquad \text{and} \qquad \dot{r}\left(t+\frac{\Delta t}{2}\right) = \frac{r_i(t+\Delta t) - r_i(t)}{\Delta t} \tag{2.6}$$

Rearranging the expression on the right hand side of Equation 2.6 gives the new positions:

$$r_i(t+\Delta t) = r_i(t) + \dot{r}_i\left(t+\frac{\Delta t}{2}\right)\Delta t \tag{2.7}$$

From the Verlet algorithm (Equation 2.4) it is then possible to extract an expression for the velocities. First, equating Equation 2.4 with Equation 2.7 yields:

$$-r_i(t-\Delta t) + 2r_i(t) + \frac{\Delta t^2}{m_i}F_i(t) = r_i(t) + \dot{r}_i\left(t+\frac{\Delta t}{2}\right)\Delta t \tag{2.8}$$

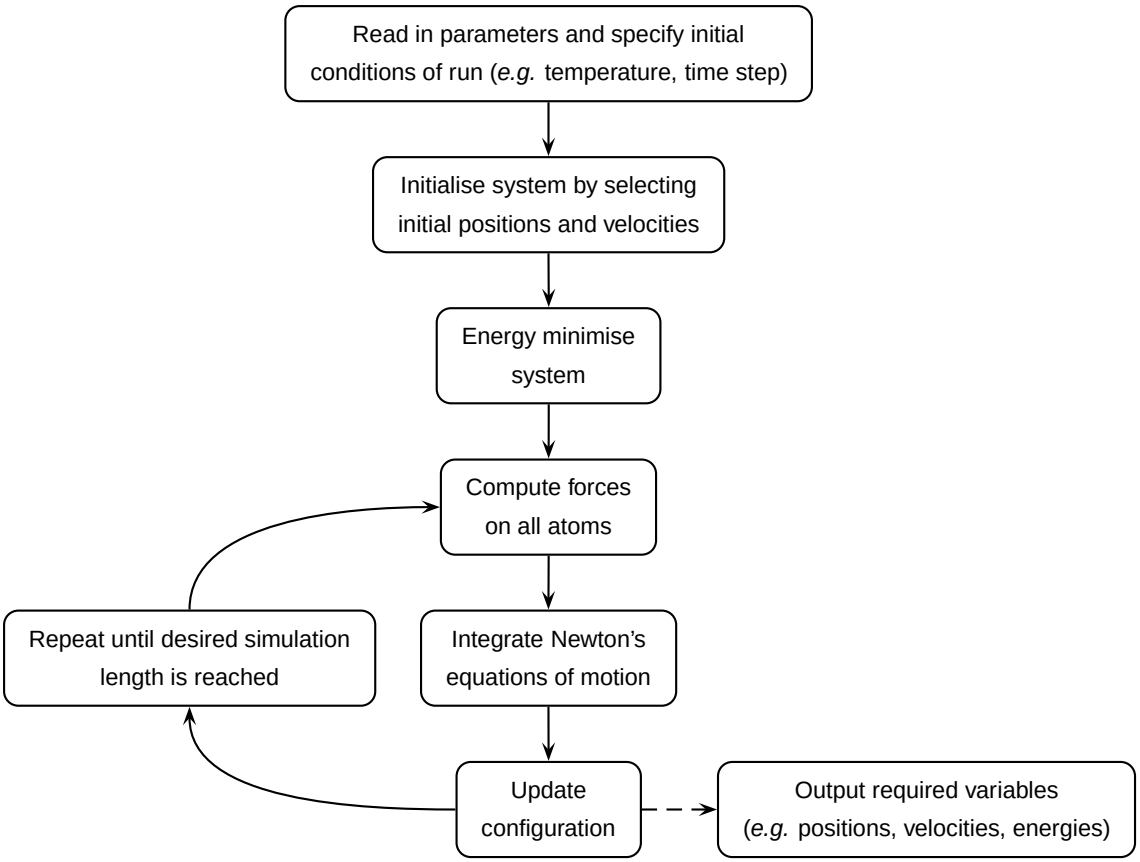

Figure 2.1: Flowchart summarising the molecular dynamics algorithm.

Collecting terms and rearranging, this then simplifies to:

$$\dot{r}_i\left(t+\frac{\Delta t}{2}\right)\Delta t = -r_i(t-\Delta t)+r_i(t)+\frac{\Delta t^2}{m_i}F_i(t) \tag{2.9}$$

Substituting the expression on the left hand side of Equation 2.6 then provides an expression for the velocities:

$$\dot{r}_i\left(t+\frac{\Delta t}{2}\right) = \dot{r}_i\left(t-\frac{\Delta t}{2}\right)+\frac{\Delta t}{m_i}F_i(t) \tag{2.10}$$

The leap frog algorithm is the integrator used for all MD simulations described in this thesis.

Integration of the equations of motion forms the core of a MD simulation program,

and a typical MD workflow is depicted in Figure 2.1. However, as Figure 2.1 shows, before the equations of motion are numerically computed it is necessary to calculate the forces acting between all of the atoms.

2.1.2 Computing the forces

Computing the forces is often the most time-consuming part of the workflow shown in Figure 2.1, although efficient techniques do exist to speed up the calculations (see Section 2.1.6). Forces are computed by calculating the force between all non-bonded atom pairs, plus the forces due to bonded interactions. Atoms interact *via* a series of potentials that can be classed as bonded or non-bonded interactions. For the non-bonded interactions, pairs of atoms (i, j) interact through a short ranged Lennard-Jones (LJ) potential:

$$V_{LJ}(r_{ij}) = 4\epsilon_{ij}\left(\left[\frac{\sigma_{ij}}{r_{ij}}\right]^{12} - \left[\frac{\sigma_{ij}}{r_{ij}}\right]^{6}\right) \tag{2.11}$$

The magnitude of ϵ_{ij} determines the depth of the potential well, σ is the separation at which the potential is zero and r_{ij} is the distance between the atoms. The long ranged electrostatic interaction between a pair of charged atoms is given by the Coulomb potential:

$$V_C(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.12}$$

The charges on the two atoms are represented by q_i and q_j, ϵ_0 is the permittivity of free space and r_{ij} is the separation between the two atoms. Conversely, for the bonded interactions, bond stretching is represented by a harmonic potential:

$$V_b(r_{ij}) = \frac{1}{2}k_{ij}^b(r_{ij} - b_{ij})^2 \tag{2.13}$$

Here, b_{ij} is the reference bond length and k_{ij} is the spring constant of the bond. A

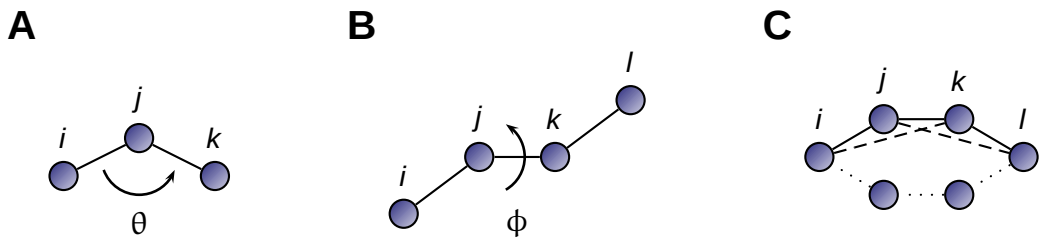

Figure 2.2: Bond angles, proper dihedrals (torsions) and improper dihedrals. **(A)** Bond angles (θ) between triplets of atoms are governed by a harmonic angle potential. **(B)** Proper dihedral angles (bond torsions) between quadruplets of atoms, where the torsion angle ϕ is the angle between the normal to the plane *ijk* and the normal to the plane *jkl*. **(C)** Improper dihedral angles between quadruplets of atoms are designed to keep groups such as aromatic rings planar and prevent inversion of tetrahedral geometries. The improper dihedral angle ξ is defined as the angle between the normal to the plane *ijk* and the normal to the plane *jkl*.

fourth power potential is an alternative, and equally valid, approximation:

$$V_b(r_{ij}) = \frac{1}{4}\tilde{k}^b_{ij}(r^2_{ij} - b^2_{ij})^2 \tag{2.14}$$

Changes in the bond angle between a triplet of atoms (i, j, k) are also represented by a harmonic potential:

$$V_a(\theta_{ijk}) = \frac{1}{2}k^\theta_{ijk}(\theta_{ijk} - \theta^0_{ijk})^2 \tag{2.15}$$

Where θ_{ijk} is the bond angle and θ^0_{ijk} is the reference bond angle (Figure 2.2). In the GROMOS96 force field the angular vibrations are represented by a trigonometric function:

$$V_a(\theta_{ijk}) = \frac{1}{2}k^\theta_{ijk}\left[\cos(\theta_{ijk}) - \cos(\theta^0_{ijk})\right]^2 \tag{2.16}$$

Proper dihedral angles or bond torsions are described by dihedral potentials between quadruplets (i, j, k, l) of four atoms (Figure 2.2). These are commonly implemented through a potential of the form:

$$V_d(\phi_{ijkl}) = k_\phi(1 + \cos(n\phi - \phi_s)) \tag{2.17}$$

Alternatively, a Ryckaert-Bellemans (RB) function can be used for proper dihedral

angles:

$$V_{RB}(\phi_{ijkl}) = \sum_{n=0}^{5} C_n(\cos(\phi - \pi))^n \tag{2.18}$$

A different form of the RB function is used in the OPLS-AA/L force field:

$$V_{RB-OPLS}(\phi_{ijkl}) = V_0 + \frac{1}{2}\sum_{n=1}^{3} V_n(1 + \cos(n\phi_{ijkl})) \tag{2.19}$$

The improper dihedral angles potential, which limits out of plane bending for planar groups and prevents inversion of tetrahedral geometries (Figure 2.2), is given by the following:

$$V_{id} = \xi_{ijkl} = \frac{1}{2}k_\xi(\xi_{ijkl} - \xi_0)^2 \tag{2.20}$$

In practice, these bond oscillations limit the time step that can be used in molecular simulations, as although the fluctuations are of low amplitude the frequency of oscillation is rapid. The time step therefore needs to be correspondingly short in order to be able to monitor these high frequency fluctuations. To overcome this limitation it is possible to implement a constraint algorithm to reset the bond lengths and angles after updating the configuration. This has the advantage of allowing a longer time step to be used, increasing computational efficiency. It has also been suggested that constraints such as these are in any case more representative of the physical behaviour of bond vibrations, at least at room temperature, as the level of bond excitation is anticipated to be low and as a result the bonds will almost exclusively occupy their vibrational ground states. A popular constraint method is the linear constraint solver developed by Hess *et al.* known as the LINCS algorithm [139].

2.1.3 Force fields

In a biomolecular simulation, parameters need to be selected for all of the species present in the system, which may include proteins, ligands, lipids, water and ions. The result is a set of self interactions and cross interactions between species based on the bonded and

non-bonded terms outlined above, and this complete parameter set is known as a force field.

There are now several biomolecular force fields available, but the following are perhaps the most widely used within the simulation community: the Chemistry at Harvard Molecular Mechanics (CHARMM) force field [140]; the Assisted Model Building with Energy Refinement (AMBER) force field [141]; the Groningen Molecular Simulation (GROMOS) force field [142]; and the Optimised Potentials for Liquid Simulations (OPLS) force field [143]. These are all classical force fields, *i.e.* they do not incorporate electronic effects such as polarisability.

Two force fields are used in this thesis. The first, the GROMOS96 43a1 [142] force field, treats bonds according to Equation 2.14, bond angles according to Equation 2.16 and proper dihedral angles according to Equation 2.17. The GROMOS96 43a1 force field is a united atom force field, in that all non-polar hydrogen atoms are combined with the parent carbon atom and are not treated explicitly. The GROMOS96 43a1 force field was originally parametrised for lipids by fitting MD simulations of liquid pentadecane to physical quantities determined by experiment such as the volume and the enthalpy of evaporation [144]. Protein parameters were derived using a similar strategy but simulations were fitted to the enthalpies of solvation in cyclohexane and water [145]. The simple point charge (SPC) water model [146] is recommended for use with the GROMOS96 43a1 force field. In the SPC model, a water molecule is composed of an oxygen atom, which has a combined partial negative charge and a Lennard-Jones site, and two hydrogen atoms with only partial positive charges and no Lennard-Jones component. As the force field neglects polarisation effects the dipole of a SPC water molecule is set at a larger value than the corresponding experimental value to compensate.

GROMOS96 43a1 force field parameters for the I(1,3,4,5)P_4 headgroup were obtained using PRODRG [147]. One report has suggested that charge assignment procedures in PRODRG are inconsistent [148], and so atomic partial charges for the PRODRG

topology were refined by performing geometry optimisation and a subsequent single-point energy calculation on the headgroup using the GAUSSIAN package [149], with the 6-31G** split valence basis set and the Becke three-parameter Lee-Yang-Parr (B3LYP) hybrid functional [150, 151, 151] in a similar vein to Blood *et al.* [152] and Lupyan *et al.* [153].

The second force field, OPLS-AA/L[2] [154], treats bonds according to Equation 2.13, bond angles according to Equation 2.15 and proper dihedrals according to Equation 2.19. In the case of OPLS-AA/L, the four point transferable intermolecular potential (TIP4P) water model [155] is recommended over the SPC water model. The TIP4P model differs from the SPC model in that the oxygen atom is split into two sites, one of which carries the mass and the Lennard-Jones component and another virtual site where the partial negative charge is located (this is the origin of the '4' in TIP4P, which refers to the number of sites in the model).

2.1.4 Energy minimisation

The combination of all the interaction potentials listed in Section 2.1.2 gives rise to a complex potential energy surface, which is a function of the position of all the atoms in the system. For a system of N atoms with $3N$ Cartesian coordinates the resulting potential energy function is a $3N$ dimensional hypersurface. As the configuration of the system evolves over time it collectively explores this hypersurface or potential energy landscape, which is characterised by maxima and minima of the potential energy.

Before a simulation begins it is often necessary to include an energy minimisation step, which is designed to optimise the starting configuration of the system by bringing

[2]OPLS-AA/L is an extension to the original OPLS force field, which was parametrised on the basis of quantum mechanical (QM) calculations. While a detailed exposition of these techniques is beyond the scope of this thesis, briefly, the extended OPLS force field is designated OPLS-AA/L as it is an all atom (AA) force field that is parametrised using a more accurate QM approach. The 'L' in OPLS-AA/L refers to pseudospectral local second order Møller-Plesset perturbation theory (LMP2), which goes beyond the Hartree-Fock level of theory used to parametrise the original OPLS force field by including the effects of electron correlation. The result is an improved set of torsional parameters for peptides and proteins in OPLS-AA/L.

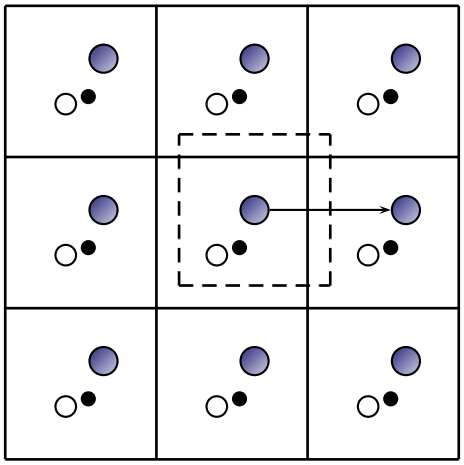

Figure 2.3: Schematic representation of periodic boundary conditions.

it to rest in a nearby local minimum in the potential energy landscape. This has the advantage that, at the local minimum, the derivative of the potential (*i.e.* the force) is zero, and so the simulation can begin at equilibrium.

Two popular energy minimisation techniques are the steepest descent and conjugate gradient methods. These are known as first order techniques because they exploit the first derivative of the potential to identify the downhill path to the nearest local minimum. The steepest descent algorithm has been used throughout this thesis, and corresponds to moving straight downhill for a certain distance (determined by a line search [156]) and then recalculating the first derivative to again find the steepest descent at this new point, then following this direction downhill for a certain distance once more. Eventually, after a suitable number of iterations the algorithm converges on the location of the minimum. An extensive discussion of the conjugate gradient algorithm and higher order energy minimisation methods is presented by Leach [156].

2.1.5 Periodic boundary conditions

In molecular simulations it is often desirable to choose boundary conditions that mimic the presence of an infinite bulk surrounding the model system. This can be achieved by using periodic boundary conditions (Figure 2.3), which treat the system as a lattice made

up of a periodic array of cells. Each molecule then interacts with all other molecules in its periodic cell, its own periodic image in all other cells and the periodic images of all other molecules. While the use of periodic boundary conditions has the advantage of being able to better approximate physical behaviour by simulating systems as if they were in a homogeneous bulk phase, it also has the disadvantage that computing the overall potential energy for all interactions now requires an infinite sum. This can be avoided by applying suitable cut offs and employing an alternative summation method, which will be discussed in the next section.

2.1.6 Treatment of non-bonded interactions

As alluded to above, there are a variety of ways in which the computational burden of calculating the non-bonded interactions at each time step can be alleviated. The simplest method is just to cut off the potential at a given value of r, such that beyond this separation the potential is zero and there is no interaction. This is generally the method applied to the Lennard-Jones interactions as they decay quickly with r (as r^{-6}, Equation 2.11). Often the Lennard-Jones potential is truncated and shifted such that it goes smoothly to zero, as the derivative of the potential still needs to be computed and simply cutting off the potential would complicate this by introducing a discontinuity into the potential function.

While the errors associated with truncating and shifting the Lennard-Jones component of the non-bonded interaction are usually small, the same cannot be said for truncation of the Coulomb or electrostatic component. In general potentials should not be truncated unless the function decays faster than r^{-3} [138], and as the Coulomb potential decays as r^{-1} (Equation 2.12) the interaction is non-negligible even at large separations. Truncation of the Coulomb potential can therefore lead to severe inaccuracies in the force calculation and hence the overall energy [157].

Without truncating the interactions, the full electrostatic energy for a system with

periodic boundary conditions applied is given by:

$$E = \frac{1}{2} \sum_{i,j=1}^{N} \sum_{\mathbf{n} \in \mathbb{Z}^3}{}' \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}L|} \tag{2.21}$$

Where N is the number of particles with charges q_i at positions $\mathbf{r}_i$ in a neutral, cubic simulation box of length L and volume $V = L^3$. The sum over $\mathbf{n}$ takes into account the periodic images of the charges and the prime indicates that for $i = j$ the term $\mathbf{n} = 0$ should be omitted. It is not possible to accurately evaluate this sum as it is only conditionally convergent, and so another approach is required.

A popular technique for treating the electrostatic interactions is the particle mesh Ewald (PME) method. Essentially the PME method allows the Coulomb energy term to be split into two parts, making the conditionally convergent sum easier to work with. Then, the Poisson equation can be solved more efficiently by distributing the charges over a mesh and discretising the problem (for more on the Poisson equation see Section 2.3). Following the treatment of Deserno and Holm [158], the partitoning of the Coulomb energy can be achieved using the following identity:

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r} \tag{2.22}$$

This expression makes it possible to split the conditionally convergent sum in Equation 2.21 into two sums that converge quickly by judicious choice of $f(r)$. The first term collects the rapid variations at small r, and this is summed in real space. The second term accounts for the remaining slow decay at large r, and this is summed in reciprocal space. A traditional choice of function for $f(r)$ is the complementary error function $\mathrm{erfc}(r)$:

$$\mathrm{erfc}(r) = \frac{2}{\pi^{\frac{1}{2}}} \int_r^{\infty} dt \exp(-t^2) \tag{2.23}$$

The Ewald energy is now decomposed into four terms:

$$E = E_r + E_k + E_s + E_d \tag{2.24}$$

E_r is the contribution from real space, E_k is the contribution from reciprocal space, E_s is the self energy and E_d is the dipole correction. These quantities are given by the following:

$$E_r = \frac{1}{2}\sum_{i,j}{}' \sum_{\mathbf{m}\in\mathbb{Z}^3} q_i q_j \frac{\mathrm{erfc}(\alpha|\mathbf{r}_{ij}+\mathbf{m}L|)}{|\mathbf{r}_{ij}+\mathbf{m}L|} \tag{2.25}$$

$$E_k = \frac{1}{2L^3}\sum_{\mathbf{k}\neq 0}\frac{4\pi}{k^2}\exp\left(-\frac{k^2}{4\alpha^2}\right)|\tilde{\rho}(\mathbf{k})|^2 \tag{2.26}$$

$$E_s = -\frac{\alpha}{\sqrt{\pi}}\sum_i q_i^2 \tag{2.27}$$

$$E_d = \frac{2\pi}{(1+2\epsilon')L^3}\left(\sum_i q_i \mathbf{r}_i\right)^2 \tag{2.28}$$

The Fourier transformed charge density $\tilde{\rho}(\mathbf{k})$ in the expression for E_k (Equation 2.26) is defined as:

$$\tilde{\rho}(\mathbf{k}) = \int_{V_b} d^3r\, \rho(\mathbf{r})\exp(-i\mathbf{k}\cdot\mathbf{r}) = \sum_{j=1}^{N} q_j \exp(-i\mathbf{k}\cdot\mathbf{r}_j) \tag{2.29}$$

E_k and, by extension, the Coulomb energy E can be evaluated using a discrete fast Fourier transform (FFT). It is then necessary to assign charges to grid points, solve the Poisson equation on the mesh and calculate and assign forces back to the particles.

2.1.7 Temperature and pressure coupling

A conventional molecular dynamics simulation is conducted with a constant number of N particles moving within a constant volume V. The overall energy of the macrostate of the system, E, is also constant, and this statistical ensemble of states is termed the

microcanonical or NVE ensemble.

However, it is usually desirable to perform simulations at constant temperature, T, to better mimic experimental conditions. This can be achieved by coupling the system to a thermostat to control T, generating the alternative statistical ensemble termed the canonical or NVT ensemble. It can also be helpful to run simulations at constant pressure, P, rather than at constant volume, again to faciliate comparison with experimental results. This is made possible through the use of a barostat to control P, which in addition to the thermostat produces the isothermal-isobaric or NPT ensemble.

A commonly used thermostat is the one developed by Berendsen *et al.* [159]. The Berendsen thermostat works by exploiting the fact that, at thermal equilibrium, there is a relationship between the temperature and the particle velocity given by the equipartition principle:

$$\langle v_\alpha^2 \rangle = \frac{k_B T}{m} \tag{2.30}$$

This can be used as a working definition of the instantaneous temperature, $T(t)$, at a time, t, in order to monitor the temperature fluctuations over time:

$$k_B T(t) = \frac{1}{N} \sum_{i=1}^{N} m v_{\alpha,i}^2(t) \tag{2.31}$$

The Berendsen thermostat works by scaling the velocities at each time step according to the following relation:

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T(t)} - 1 \right) \right]^{\frac{1}{2}} \tag{2.32}$$

The velocity scaling parameter λ is determined by the reference (*i.e.* desired) temperature T_0, the instantaneous temperature $T(t)$ and τ_T, which is the coupling parameter. The Berendsen thermostat therefore effectively acts by coupling the system to a heat bath of temperature T_0, where τ_T regulates the strength of the coupling[3].

[3]It has been shown that in fact the Berendsen thermostat does not reproduce the canonical ensemble [160], and instead produces what is known as a weak coupling ensemble. This is a hybrid that lies somewhere between the microcanonical and canonical ensembles depending on the size of the coupling

Pressure control can be achieved by a similar method, and although there are various algorithms almost all are based on the treatment first devised by Andersen [161]. While the derivation is complex, it essentially amounts to allowing the volume to fluctuate with its average value being determined by the balance between some externally imposed pressure, P, and the internal stresses in the system. For more details on the Andersen extended ensemble approach readers are referred to Andersen's original paper [161] and further work by Parrinello and Rahman who extended Andersen's method to account for changes in shape as well as volume [162]. Both methods are discussed in more detail by Nosé and Klein [163].

2.1.8 Steered molecular dynamics simulations

Many events, such as protein-ligand dissociation, occur over physiological time scales that are inaccessible to biomolecular simulation. A relevant example is the case of a PH domain dissociating from its cognate ligand at the membrane surface. Hammond *et al.* measured the lifetime, τ, of protein-membrane complexes for a variety of PH domains and found that τ was typically between 2 to 7 seconds [164]. Equilibrium simulations are not able to probe these extended time scales, but one way of circumventing this limitation is to employ some form of non-equilibrium enhanced or rare event sampling. In these methods an element of bias is introduced into the simulations to increase the likelihood of observing the event of interest. Various sophisticated approaches to this problem have been developed, and traditionally have been applied to conformational transitions in proteins. Examples include conformational flooding [165], transition path sampling [166], metadynamics [167], and dynamic importance sampling [168].

Leaving these more elaborate techniques aside, a straightforward way of perturbing the system and accelerating the simulation is simply to apply an external force along a predefined reaction coordinate. This technique, known as steered or force probe molec-

parameter τ_T.

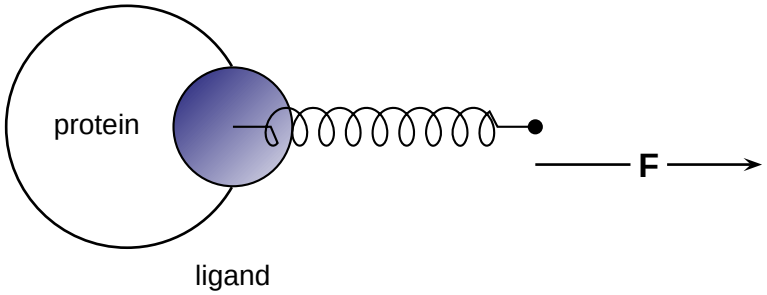

Figure 2.4: Schematic of steered MD simulation setup. A virtual spring is attached to the ligand and a force is applied along the reaction coordinate. The force causes the spring to retract and the ligand is forcibly extracted from the protein binding site.

ular dynamics simulation, has been successfully applied to protein unfolding [169, 170] and ligand unbinding [171, 172]. A virtual spring is attached to the molecule of interest and the spring is retracted along some reaction coordinate (Figure 2.4). The force profile over the course of the simulation can then be extracted using Hooke's law, $F = -kx$, and this can be used to obtain information regarding the unfolding or dissociation pathway. Although the exact shape of the force profile is sensitive to the stiffness of the spring and the speed at which the spring is retracted, the steered molecular dynamics (SMD) techique is nonetheless a simple and rapid method for interrogating the dynamics of a system. Indeed, it was shown by Jarzynski [173] that by combining several SMD trajectories and computing the work done over the course of these non-equilibrium simulations it was possible to recover equilibrium free energy differences between the start and end points of the simulations. Calculating this free energy profile, usually referred to as the potential of mean force (PMF), from SMD simulations using Jarzynski's equality is virtually as efficient as extracting the PMF from more traditional umbrella sampling methods [174].

2.1.9 Coarse grained molecular dynamics simulations

In contrast to the enhanced sampling techniques described above, an alternative strategy to overcome the limitations on accessible time scales is to resort to some level of 'coarse

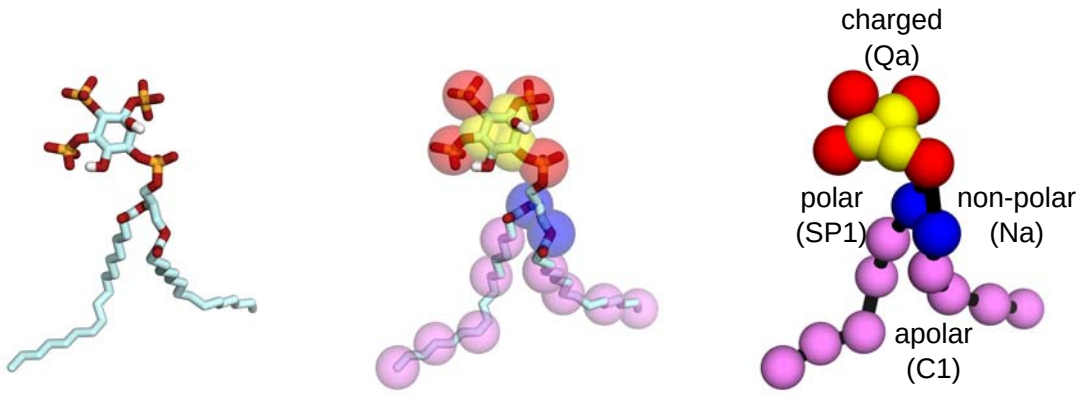

Figure 2.5: Conversion from an atomistic to a CG representation for PI(3,4,5)P_3. The atomistic representation of PI(3,4,5)P_3 (left) is converted to a CG representation by an approximately 4:1 mapping of CG particles onto the atomic sites. Black lines indicate the soft harmonic bonds between the resulting CG beads. Phosphate groups are represented by (negatively) charged Q-type particles (red). The inositol ring is represented by three, ring variant polar P-type particles (yellow). The two glycerol esters are represented by non-polar N-type particles (blue). Finally, the hydrocarbon tails are represented by apolar C-type particles (purrple). In this case, the coarse graining procedure reduces the number of interaction sites from 71 to 18, a reduction of almost a factor of four.

graining'. In coarse grained (CG) simulations, atomic detail is sacrificed in exchange for the ability to simulate larger systems for longer periods of time. The underlying principle of coarse graining is to merge groups of atoms into a smaller number of CG sites, which then interact through a series of effective interactions. Having fewer interaction sites greatly speeds up the simulation while still retaining some molecular and structural detail. The level of abstraction in CG models varies widely. Some models are comparatively high resolution, for example the MARTINI model [175, 176] for lipids and proteins developed by Marrink and co-workers maps on average four heavy atoms to a single CG interaction site (Figure 2.5 and Figure 2.6). Other models are lower resolution and are designed to reproduce the essential physics of a system but not necessarily all of the detailed interactions.

The range and applicability of CG models is nicely illustrated by a series of studies on the BAR domain by the groups of Voth and Schulten. Both groups aimed to simulate deformation and tubulation of membranes by N-BAR domains, which is a process that occurs over long time scales but also over large lengthscales, requiring some level of coarse grained treatment.

Rather than simulating the protein and membrane in atomic detail, Voth and co-

workers used a mesoscopic approach in two papers by Ayton *et al.* [177, 178]. The membrane was modelled as a continuum elastic sheet with material properties such as a bending modulus, and the model was parametrised using prior atomistic simulations of the system. In the field-theoretic approach of Ayton *et al.* the elastic energy of the membrane is given by a Helfrich-like Hamiltonian [179] and the protein is not modelled explicitly; rather, its influence on the shape of the membrane is represented by a spontaneous curvature field, the magnitude of which is equivalent to the concentration of N-BAR domains on the surface of the membrane.

Meanwhile, Schulten and co-workers adopted a different approach and used a four-scale description of the system in a paper by Arkhipov *et al.* [180]. Arkhipov *et al.* used atomistic simulations and their own version of an elastic continuum model to study N-BAR domains, but also used two CG approaches: a residue based model and a shape based model. Both of these models are reduced representations of the full atomic structure. In the residue based model each amino acid is represented by one or two beads interacting *via* effective potentials that reproduce experimentally observed quantities, which is an approach similar to that adopted by several other CG models such as MARTINI (see also [181, 182, 183]). In the shape based model the representation is reduced still further, and hundreds of atoms are now condensed into a single bead with the array of beads approximately reproducing the overall shape and the moments of inertia of the protein, as described in an earlier paper [184]. Each of the representations in this hierarchy is parametrised on the basis of the results obtained using the higher resolution model beneath it. Thus, the CG models are built upon the results of the atomistic simulations, and the mechanical properties of the elastic membrane were optimised to reproduce the curvatures observed in the CG simulations.

The MARTINI force field is used for the CG simulations described in this thesis. This force field was originally parametrised for lipids by comparing simulation results with free energies of hydration, evaporation and partitioning between water and a selection

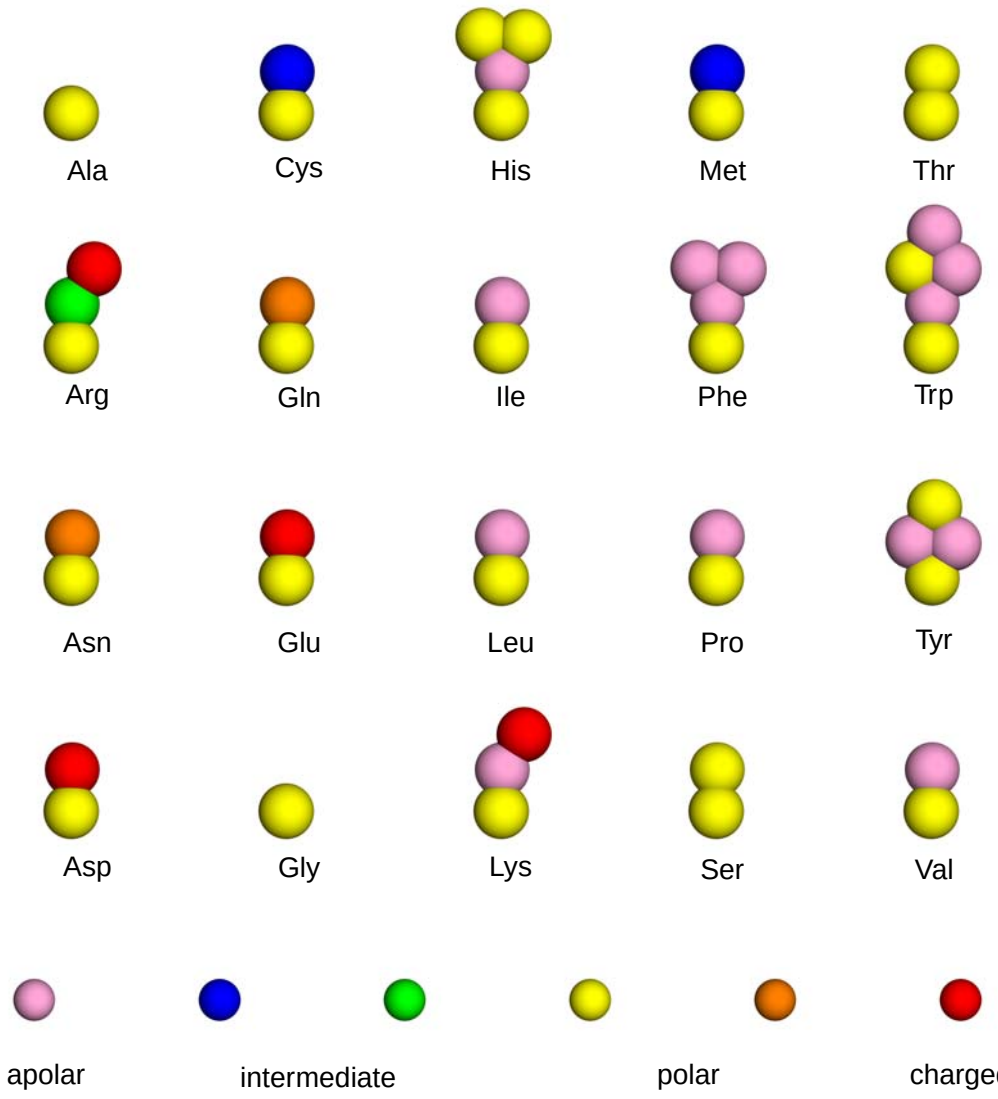

Figure 2.6: Coarse grained representation of amino acid residues in the MARTINI model. Coarse grained beads are coloured according to particle type, ranging from apolar to charged. The charges and polarity vary between the representations of the amino acids in an attempt to account for their different polarities and hydrogen bonding propensities.

of organic solvents [175]. A four to one mapping of heavy atoms to CG interaction sites is used, and each CG particle is assigned a particle type depending on whether it is polar (P), non-polar (N), apolar (C) or charged (Q). Beyond the four principal particle types additional labels determine particle subtypes, distinguished by hydrogen bonding capabilities (d, donor; a, acceptor; da, both donor and acceptor; 0, neither donor nor acceptor) or polarity, which ranges from 1 (low polarity) to 5 (high polarity). An example

mapping is shown in Figure 2.5, where the atomistic form of PI(3,4,5)P_3 is converted to a CG representation.

The model was extended to proteins in a later paper [176], introducing ring-like particles for amino acid sidechains such as tryptophan that incorporate homocyclic and heterocyclic groups. Figure 2.6 shows the complete mapping for all 20 amino acids. The four to one mapping described previously is usually insufficient to capture the geometry of cyclic groups and so higher resolution mapping is typically required, up to two to one in some cases. Amino acid sidechains were parametrised by calculating free energies of partitioning between the oil/aqueous phases in a two phase butane/water system and comparing the results to experiment.

In the implementation of MARTINI in GROMACS, short ranged Lennard-Jones and long ranged Coulomb interactions are shifted and truncated at r_{cut}. Cutting off the electrostatic interactions in this manner increases computational efficiency, as the long range electrostatic interactions no longer need to be computed. However, as discussed in Section 2.1.6 this can introduce simulation artefacts.

All MD simulations were carried out using GROMACS [185, 186]. Specific simulation parameters can be found in each chapter of the thesis under the section entitled 'Methods'.

2.1.10 Analysis tools for molecular dynamics simulations

The root mean square deviation (RMSD) of the overall globular fold of a protein is calculated by a least squares fit of the structure at $t = t_2$ to a reference structure at $t = t_1$. The RMSD is a measure of the average distance between a collection of N atoms at a time t and the same collection of atoms in a reference configuration.

$$\text{RMSD}\,(t_2, t_1) = \left[\frac{1}{M} \sum_{i=1}^{N} m_i \left| \mathbf{r}_i(t_2) - \mathbf{r}_i(t_1) \right|^2 \right]^{\frac{1}{2}} \tag{2.33}$$

The total mass is $M = \sum_{i=1}^{N} m_i$ and $\mathbf{r}_i(t)$ is the position of atom i at time t. In contrast, the root mean square fluctuation (RMSF) is a measure of the deviation of the position of a single atom i about the time-averaged position of the same atom over a given time interval. Whereas the RMSD is designed to assess the overall departure of the fold from some reference structure, the RMSF can be used to monitor the fluctuation of individual amino acid residues over the course of the MD trajectory:

$$\text{RMSF} = \left[\frac{1}{(t_2 - t_1)} \sum_{t=t_1}^{t=t_2} |\mathbf{r}_i(t) - \bar{\mathbf{r}}_i|^2 \right]^{\frac{1}{2}} \tag{2.34}$$

Here $\bar{\mathbf{r}}_i$ is the time averaged position of the atom over the time interval $(t_2 - t_1)$.

Radial distribution functions (RDFs) are used to assess the local structure surrounding a given atom. For example, in the case of a transmembrane protein spanning a heterogeneous lipid bilayer, the RDF could be used to quantify the degree of lipid clustering around the target protein relative to the bulk. The RDF is computed using the following:

$$\begin{aligned} g_{AB}(r) &= \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{\text{local}}} \\ &= \frac{1}{\langle \rho_B \rangle_{\text{local}}} \frac{1}{N_A} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{\delta(r_{ij} - r)}{4\pi r^2} \end{aligned} \tag{2.35}$$

$\langle \rho_B \rangle$ is the density of atom type B at a distance r around atom A, while $\langle \rho_B \rangle_{\text{local}}$ is the time-averaged density of all atoms of type B around atom A up to a defined radius, r_{max}. The latter, $\langle \rho_B \rangle_{\text{local}}$, can effectively be thought of as the 'background' density, with r_{max} usually taken to be half the length of the simulation box.

2.2 Stochastic dynamics

Classical MD techniques describe the dynamics of each atom according to Newton's equations of motion, with the behaviour of the i^{th} atom governed by $m_i \ddot{r}_i = F_i$. When performing simulations of larger systems using classical MD it is therefore necessary to treat

the dynamics of all molecules, including solvent molecules, explicitly. This can often be computationally demanding, and at least in the case of protein-protein interactions (PPIs) the behaviour of the solvent is often of comparatively little interest[4]. In simulations of PPIs the principal areas of interest are typically the protein dynamics, the nature of the association pathway and the interactions between the amino acid residues in the bound complex. It is possible to model the dynamic and electrostatic effects arising from the solvent molecules without including them explicitly, and this will be the focus of the following sections.

2.2.1 Langevin dynamics

For the case of one or more large molecules immersed in a solvent, an alternative approach is to 'integrate out' the solvent motions by replacing them with additional terms in the equation of motion of the large molecule(s). Additional terms are incorporated into Newton's equation of motion for the large molecule to compensate for the removal of the explicit solvent molecules. One term represents the solvent friction, and another stochastic term accounts for the random displacements arising from the Brownian motion of the solvent. In a sense, this could loosely be thought of as an alternative coarse graining strategy, similar to those described in Section 2.1.9. Here, rather than merging several atoms together to form a CG particle, instead all of the explicit solvent molecules are combined and replaced with a set of representative effective interactions.

For a large, single molecule in a viscous medium, the equation of motion becomes:

$$m\ddot{r} = -\zeta\dot{r} + F(r) + \sigma S(t) \tag{2.36}$$

[4]This is, however, not always the case for studies of PPIs, and in particular proper treatment of solvation is often essential when considering protein-ligand interactions. Water molecules form a hydration shell around the ligand in solution, which can influence the free energy of binding to the protein due to the associated desolvation penalty when the ligand is stripped of its surrounding water molecules upon binding in the active site. Protein-ligand complexes also frequently incorporate 'structural' water molecules. A comprehensive analysis of water molecules at protein-ligand interfaces in ∼400 high-resolution crystal structures conducted by Lu *et al.* [187] revealed that 76% of these water molecules bridge the interaction between the protein and the ligand in some fashion.

This is the Langevin equation [188]. Here, ζ is the solvent friction and $S(t)$ is the stochastic force due to the random 'kicks' from the solvent described above with coupling coefficient σ. $F(r)$ is some external force, for example gravity in the case of simulations of a colloidal dispersion.

2.2.2 Brownian dynamics (BD)

Most proteins diffusing in solution will interact with the surrounding solvent molecules to such an extent that the protein motion will be 'overdamped' by the solvent viscosity. The momentum of the protein decays so swiftly upon collisions with solvent molecules that its behaviour is almost non-inertial, and the protein motion then has the characteristics of Brownian motion (a random walk), with the forces on the right hand side of Equation 2.36 becoming dominant. In this case, the Langevin equation above can be simplified by averaging the equation over a time Δt that is much longer than the momentum relaxation time, *i.e.*:

$$\Delta t \gg \frac{m}{\zeta} \tag{2.37}$$

Over this average the acceleration term, $\langle m\ddot{r}\rangle$, is zero and the other terms in the equation adopt time averaged forms:

$$\langle \zeta \dot{r}\rangle = \zeta \frac{\Delta x}{\Delta t} \qquad \langle F(r)\rangle = F \qquad \langle S(t)\rangle = \frac{1}{\sqrt{\Delta t}} S \tag{2.38}$$

Equation 2.36 can thus be recast as:

$$\zeta \frac{\Delta x}{\Delta t} = F + \frac{\sigma}{\sqrt{\Delta t}} S \tag{2.39}$$

Rearranging gives:

$$\Delta x = \frac{\Delta t}{\zeta} F + \frac{\sigma \sqrt{\Delta t}}{\zeta} S \tag{2.40}$$

The Stokes-Einstein relation between the friction coefficient and the diffusion coeffi-

cient [189] is:

$$D = \frac{k_B T}{\zeta} \tag{2.41}$$

With $\zeta = 6\pi\eta a$, where η is the solvent viscosity and a is the radius of the suspended molecules. Invoking this, and noting that $\sigma/\zeta = \sqrt{2D}$, yields:

$$\Delta x = \frac{D\Delta t}{k_B T} F + \sqrt{2D\Delta t}\, S \tag{2.42}$$

Or alternatively:

$$\Delta x = \frac{D\Delta t}{k_B T} F + R \tag{2.43}$$

Here the stochastic term $R = \sqrt{2D\Delta t}\, S$ has the properties that $\langle R \rangle = 0$ and $\langle R^2 \rangle = 2D\Delta t$. Equation 2.43 is a simplified form of the Ermak-McCammon algorithm for the case of a single Brownian molecule moving in one dimension. It is, however, relatively straightforward to extend this and apply it to the case of two proteins co-diffusing in three dimensions in solution. The one dimensional displacement Δx is replaced by a *relative* three dimensional displacement vector $\Delta \mathbf{r}$ between the two molecules. Similarly, a relative translational diffusion coefficient between the two proteins is used, and the stochastic term R is replaced by a random vector $\mathbf{R}$ that satisfies $\langle \mathbf{R} \rangle = 0$ and $\langle \mathbf{R}^2 \rangle = 6D\Delta t$. Equation 2.43 then becomes:

$$\Delta \mathbf{r} = \frac{D\Delta t}{k_B T} \mathbf{F} + \mathbf{R} \tag{2.44}$$

This is the form of the algorithm implemented in the Simulation of Diffusional Association (SDA) software, designed to perform BD simulations of protein-protein association. SDA was originally developed by Gabdoulline and Wade [190], and is a widely used successor to earlier programs designed to simulate enzyme-substrate encounters (the University of Houston Brownian Dynamics (UHBD) program [191, 192]) and the association of electron transfer proteins (Macrodox [193]). In SDA, rotational motions

are dealt with in an analogous fashion to the translational motions (Equation 2.44):

$$\Delta \mathbf{w}_i = \frac{D_{iR}\Delta t}{k_B T}\mathbf{T}_{ij} + \mathbf{W}_i \tag{2.45}$$

The torque on the protein is given by $\mathbf{T}_{ij}$ and $\mathbf{W}$ is a random rotation angle that satisfies $\langle \mathbf{W}_i \rangle = 0$ and $\langle {\mathbf{W}_i}^2 \rangle = 6 D_{iR}\Delta t$.

In their original paper [194], Ermak and McCammon reported a much more detailed derivation of a general form of their equation that allows for an arbitrary number (N) of molecules and that also incorporates hydrodynamic interactions through either the Oseen or Rotne-Prager-Yamakawa tensors [195, 196]. The final displacement equation obtained by Ermak and McCammon (using their notation) is:

$$r_i(\Delta t) = r_i^0 + \sum_j \frac{\partial D_{ij}^0}{\partial r_j}\Delta t + \sum_j \frac{D_{ij}^0 F_j^0}{k_B T}\Delta t + R_i(\Delta t) \tag{2.46}$$

The derivation required to obtain this displacement equation is rather lengthy and for reasons of space will not be outlined here. Readers are referred to the original paper for a detailed derivation of the algorithm and a comparison to the corresponding Fokker-Planck description[5]. However, a brief comparison of the simplified (Equation 2.43) and general (Equation 2.46) forms of the Ermak-McCammon algorithm is instructive.

Previously, the Langevin formulation (Equation 2.36) and the simplified Ermak-McCammon algorithm (Equation 2.43) used a low level treatment of friction, diffusion and viscous effects. These quantities were taken to be isotropic and so ζ and D are scalars. The more general form of the Ermak-McCammon algorithm in Equation 2.46 treats diffusion as a tensor property. A typical protein will exhibit some shape anisotropy so the diffusion constants along its three principal axes are likely to be different. To account for this, the diffusion constant is replaced by a diffusion tensor. Equation 2.46 also includes the effect of hydrodynamic coupling between all N molecules, a factor that has been

[5]Further information on stochastic processes in general and diffusion processes in particular can be found in Robert Mazo's monograph on Brownian Motion [197]

overlooked thus far. When a large solute molecule moves through a solvent it displaces solvent molecules, which can in turn have a small effect on the motion of other nearby solutes, and so the motions of the two solute molecules become correlated. Equation 2.46 includes this effect through the $3N \times 3N$ diffusion tensor $\mathbf{D}$, which describes the hydrodynamic interactions between the N molecules with their three translational degrees of freedom. The tensor is made up of a set of 3×3 submatrices, each describing the diffusion behaviour of one of the N proteins, which gives a total of $3N \times 3N$ elements in the complete tensor:

$$\mathbf{D} = \begin{bmatrix} D_{11} & D_{12} & \cdots & D_{1N} \\ D_{21} & D_{22} & \cdots & D_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ D_{N1} & D_{N2} & \cdots & D_{NN} \end{bmatrix} \tag{2.47}$$

Here, each individual 3×3 D_{ij} tensor is defined by:

$$D_{ii} = \frac{kT}{6\pi\eta a}\,\mathbf{I} \quad \text{and} \quad D_{ij} = \frac{kT}{8\pi\eta r_{ij}} \left(\mathbf{J} + \frac{1}{r_{ij}^2}\mathbf{r}_{ij}\mathbf{r}_{ij} \right) \tag{2.48}$$

In the submatrix, $\mathbf{I}$ is the three dimensional unit tensor and $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ is the position of molecule i relative to molecule j. With no hydrodynamic coupling the only non-zero terms are for $D_{i=j}$. Thus, computing the correlated random displacements, $\langle \mathbf{R}_i \mathbf{R}_j \rangle = 6D_{ij}\Delta t$, at each time step is straightforward. However, explicit inclusion of hydrodynamic coupling leads to correlated motion that manifests itself as non-zero off-diagonal terms in the diffusion tensor. In this case, in order to compute the correlated random displacements the matrix must first be diagonalised, making the calculations much more taxing.

As the inclusion of hydrodynamic interactions is so computationally expensive they are often neglected in BD simulations, though there is increasing evidence of their importance [198, 199, 200]. However, great strides have recently been made in developing

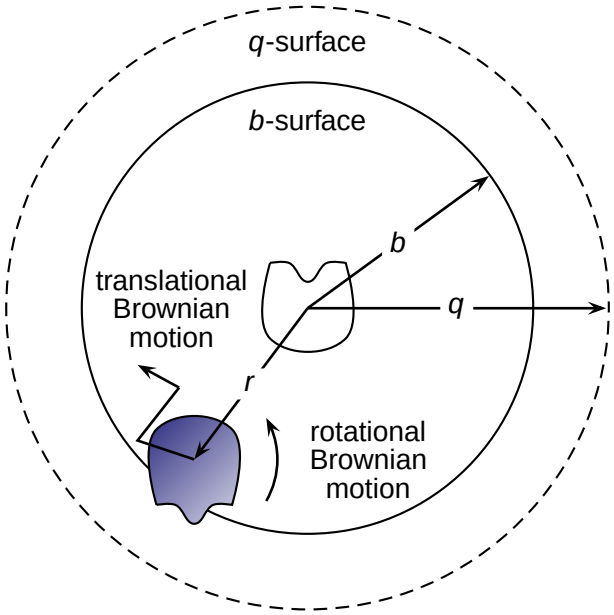

Figure 2.7: Schematic picture of BD simulation and diffusional association. The partitioning between the anisotropic (b) region and the isotropic (q) region is at $r = b$.

more efficient algorithms to reduce the computational cost of calculating the hydrodynamic interactions (see *e.g.* [201]).

2.2.3 Calculating rate constants from Brownian dynamics simulations

It is also possible to compute association rates between two proteins using BD simulations. While the work described in this thesis will not make direct use of this technique (for reasons discussed in Chapter 4) it is included here for completeness.

In order to calculate rate constants it is necessary to make use of a framework first developed by Northrup, Allison and McCammon [202] in 1984, hereafter referred to as the NAM method. The NAM method partitions the configuration space into two regions, denoted as b and q (Figure 2.7). The inner region, b, is finite and is the region closest to the reference protein. Here the diffusion is governed by a complicated set of non-centrosymmetric interactions and must computed numerically. In contrast, the outer q region is infinite but is sufficiently spatially separated from the reference protein that the diffusion behaviour can be dealt with analytically, as any intermolecular forces are

considered approximately centrosymmetric at this distance. The ensemble reactive flux should also be isotropic at this distance, *i.e.* the steady state rate at which mobile reactants outside the b surface strike the surface should have no angle dependence and should be the same everywhere.

The result is that BD trajectories only need to be calculated explicitly for the inner region, where computation of a swarm of trajectories is required to extensively sample the association pathways. When these trajectories reach the outer region they are truncated and any contribution they make to the overall rate is just given by an analytic correction factor.

The overall rate constant, k, can therefore be decomposed into two parts:

$$k = k_D(b)p \tag{2.49}$$

The $k_D(b)$ component is the rate constant governing the rate at which molecules at separations $r > b$ strike the b surface. If there are no forces or hydrodynamic interactions for $r > b$ this is just given by the Smoluchowski result[6] [203]:

$$k_D(b) = 4\pi D b \tag{2.50}$$

Now the complicated asymmetric element of the problem has been wholly isolated in the quantity p, and this can be evaluated explicitly by consideration of reactant motion in the inner and outer regions. Again, the derivation is slightly involved and readers are

[6]Alternatively if forces and hydrodynamic interactions do exist, but are still centrosymmetric, then the following more general expression can be used:

$$k_D(b) = \left(\int_b^\infty dr \left[\frac{\exp[U(r)/k_B T}{4\pi r^2 D(r)} \right] \right)^{-1}$$

referred to the original paper for additional details [202], but it can be shown that:

$$k = k_D(b)\left[\frac{\beta}{1-(1-\beta)\Omega}\right] \tag{2.51}$$

Where $\Omega = k_D(b)/k_D(q)$ and β is the fraction of reactive trajectories. It is therefore possible to obtain an expression for the overall rate constant by an appropriate division of the diffusion space.

2.3 Poisson-Boltzmann electrostatics

In addition to the hydrodynamic interactions with the solvent molecules described above, at high salt concentrations biological macromolecules also interact with the excess salt in the solution. Immersion in an electrolyte solution leads to the buildup of oppositely charged counterions around proteins and also at the surface of the cell membrane, leading to the formation of an electrical double layer. The electrical double layers screen the Coulomb interaction between proteins and the membrane surface with the result that the interaction is in fact an 'effective' one [204].

The phenomenon of charge screening was first investigated by Gouy [205] and also independently by Chapman [206] in the early part of the 20th Century. They each showed how the electrical double layer could be described by the Poisson-Boltzmann equation, a combination of the Poisson equation and a Boltzmann distribution. The former describes the electrostatic interactions between a biomolecule and the additional salt ions, while the latter describes the equilibrium between the two. Several alternative derivations of the Poisson-Boltzmann equation have been documented, and for completeness two of them will be outlined here. The first is a more traditional approach, similar to that used by Gouy and Chapman, that considers the electrical double layer at a uniformly charged plane. The second method considers an arbitrary array of point charges fixed in space and surrounded by an electrical double layer. This geometry is perhaps more relevant for

applications in molecular simulation.

For the first derivation, the treatment of Dill and Bromberg is followed [207]. Beginning with the Poisson equation:

$$\nabla^2\psi(x) = -\frac{\rho(x)}{\epsilon(x)} \tag{2.52}$$

Here, $\nabla^2\psi(x)$ is the Laplacian of the electrostatic potential, $\epsilon(x)$ is the dielectric constant in a polarisable medium and $\rho(x)$ is the charge density of the salt ions. At a plane surface, the mean field concentration of mobile positively charged salt ions at a distance x from the plane is given by:

$$n_+(x) = n_\infty e^{-ze\psi(x)/k_BT} \tag{2.53}$$

The concentration of positive ions in the bulk, far away from the surface where $\psi(\infty) = 0$, is given by n_∞. The corresponding expression for the negative ions is:

$$n_-(x) = n_\infty e^{+ze\psi(x)/k_BT} \tag{2.54}$$

The mobile ions also contribute to the overall electrostatic potential. Their contribution is given by:

$$\begin{aligned}\rho(x) &= \sum_i z_i e n_i(x) \\ &= ze[n_+(x) - n_-(x)]\end{aligned} \tag{2.55}$$

Now, substituting Equation 2.53 and Equation 2.54 into Equation 2.55 yields:

$$\nabla^2\psi(x) = \frac{zen_\infty}{\epsilon(x)}(e^{ze\psi(x)/k_BT} - e^{-ze\psi(x)/k_BT}) \tag{2.56}$$

This can be expressed in terms of the hyperbolic sine function:

$$\nabla^2\psi(x) = \frac{2zen_\infty}{\epsilon(x)}\sinh\left(\frac{ze\psi(x)}{k_BT}\right) \tag{2.57}$$

This is the non-linear version of the Poisson-Boltzmann equation. In the case where the potential is small, $ze\psi(x)/kT \ll 1$, the approximation $\sinh(x) \approx x$ can be used. The Poisson-Boltzmann equation then becomes:

$$\nabla^2\psi(x) = \frac{2zen_\infty}{\epsilon(x)}\left(\frac{ze\psi(x)}{k_BT}\right) = \kappa^2(x)\psi(x) \tag{2.58}$$

Where κ is defined as:

$$\kappa^2(x) = \frac{2z^2e^2n_\infty}{\epsilon(x)k_BT} \tag{2.59}$$

This is the linearised form of the Poisson-Boltzmann equation, which is also known as the Debye-Hückel equation [208]. The quantity $1/\kappa(x)$ is the Debye length, which is the characteristic thickness of the electrical double layer and can be thought of as the distance required for effective screening by mobile charge carriers. The behaviour of the mobile ions at distances greater than $1/\kappa(x)$ away from the surface is influenced to only a limited extent by the surface charge. As a practical example of this, when using SDA with the NAM b/q-surface formalism described above, it is recommended to set the radius of the b surface to at least ten times the Debye length. This is done to maintain the validity of the NAM method by ensuring all electrostatic interactions are completely screened out.

An alternative derivation can be found in an article by Baker [209]. Rather than a uniform surface charge, the derivation outlined by Baker yields a form of the Poisson-Boltzmann equation that can then be used to compute the charge distribution around an array of point charges. This is particularly useful when considering the electrostatic properties of irregularly shaped biomolecules with non-uniform charge distributions.

Written in a slightly different form, the Poisson equation is:

$$-\nabla \cdot \epsilon(r)\nabla\psi(r) = \rho(r) \tag{2.60}$$

The variables are the same as in Equation 2.52. In a biomolecular system two types of charge distribution are considered. First, there are the atomic partial charges of the protein that are represented by a fixed charge distribution:

$$\rho_f(r) = \frac{4\pi e_c^2}{k_B T} \sum_{i=1}^{N} Q_i \delta(r - r_i) \tag{2.61}$$

The N atomic partial charges of the protein with magnitudes Q_i are represented by delta functions located at atom centres r_i. Second, the contributions of counterions are modelled as a mean field Boltzmann distribution, such that the mobile charge distribution is given by:

$$\rho_m(r) = \frac{4\pi e_c^2}{k_B T} \sum_{j}^{n} c_j q_j \exp(-q_j\psi(r) - V_j(r)) \tag{2.62}$$

This accounts for n counterion species with charges q_j, bulk concentrations c_j and steric potentials V_j. Steric potentials prevent the overlap between the biomolecule and the counterions. For a one-to-one electrolyte this simply reduces to:

$$\rho_m(r) = \kappa^2(r)\sinh[\psi(r)] \tag{2.63}$$

Combining Equation 2.60, Equation 2.61 and Equation 2.62 yields:

$$-\nabla \cdot \epsilon(r)\nabla\psi(r) + \kappa^2(r)\sinh[\psi(r)] = \frac{4\pi e_c^2}{k_B T} \sum_{i=1}^{N} Q_i \delta(r - r_i) \tag{2.64}$$

This is an alternative form of the Poisson-Boltzmann equation. Again, for small potentials the approximation $\sinh(r) \approx r$ can be used. The linearised form of this Poisson-

Boltzmann equation then becomes:

$$-\nabla \cdot \epsilon(r)\nabla\psi(r) + \kappa^2(r)\psi(r) = \frac{4\pi e_c^2}{kT}\sum_{i=1}^{N} Q_i \delta(r - r_i) \tag{2.65}$$

The Poisson-Boltzmann equation can therefore be used to compute the electrostatic potential around a biomolecule if the charge distribution and salt concentration are known. In practice the complexity of these systems means that an analytical solution is not possible and the non-linear PB equation must be solved numerically. Several software packages are now available to do this: examples include the Adaptive Poisson-Boltzmann Solver (APBS) [210], DelPhi [211, 212] and UHBD [191, 192]. The various numerical methods used to solve the non-linear PB equation are documented in detail elsewhere [213], but briefly, they typically involve partitioning space into a mesh of grid points and solving the non-linear PB equation at each grid point. This generates an approximate solution for the electrostatic potential, and the more closely spaced the grid points the better the approximation to the true value. Discretising the problem in this way makes the calculations tractable, but for larger systems with a dense array of grid points the calculations can still be computationally expensive.

Continuum electrostatics calculations typically do not incorporate the internal dynamics of the proteins: rather, the structures remain rigid while the calculations are performed. Poisson-Boltzmann theory also has its limitations as, for example, it ignores ion-ion correlations and instead uses a mean-field approximation [136], which can have serious consequences when divalent ions are present in the system [214].

APBS has been used throughout this thesis to calculate electrostatic potentials. In APBS calculations the salt concentration is set explicitly at the outset and is a key input parameter, and hence the ionic strength has a direct impact on the computed electrostatic potential. It is therefore straightforward to incorporate the effect of salt concentration into BD simulations using SDA by simply calculating electrostatic potentials at different ionic strengths using APBS. The output of APBS is an electrostatic potential grid that

can be inputted directly into SDA[7]. The influence of ionic strength on the electrostatic potential and, by extension, the interaction forces computed in Equation 2.67 is therefore already bound up in these pre-calculated electrostatic potential grids before the BD simulation even begins.

2.3.1 Further details of BD simulations using SDA

SDA performs BD simulations of associating proteins using experimentally determined structures, which are then converted from their PDB format to the PQR format using PDB2PQR [215]. PDB2PQR assigns charges and atomic radii to each of the atoms in the protein structure, and the resulting PQR file then forms an input for calculating the electrostatic map using APBS. The electrostatic maps are calculated at the outset and remain unchanged throughout the simulation as it would otherwise be too computationally costly to recalculate each map for every time step. The initial electrostatic potential is instead superimposed onto the position and orientation of the protein at each time step, so that the protein and its electrostatic potential effectively undergo synchronous bulk rotations and translations. If the structure of the bound complex is known then SDA is able to calculate the association rate.

The forces between the diffusing proteins are computed as finite-difference derivatives of the free energy of interaction between the two proteins:

$$\Delta G = \frac{1}{2}\sum_{iB} \Phi_{eA}\, q_{iB} + \frac{1}{2}\sum_{jA} \Phi_{eB}\, q_{jA} + \sum_{iB} \Phi_{edsA}\, {q_{iB}}^2 + \sum_{jA} \Phi_{edsB}\, {q_{jA}}^2 \tag{2.66}$$

Electrostatic desolvation grids are calculated according to the protocol developed by Elcock *et al.* [216]. The unfavourable electrostatic desolvation term is approximated by

[7]For historical reasons, it is necessary to first convert the output file format of APBS to the legacy format used by UHBD using the tools provided with SDA.

the following expression:

$$\Delta G_{eds} \approx \alpha \frac{\epsilon_s - \epsilon_p}{\epsilon_s (2\epsilon_s + \epsilon_p)} \sum_{ij} (1 + \kappa r_{ij})^2 \exp(-2\kappa r_{ij}) \frac{{q_i}^2 {a_j}^3}{{r_{ij}}^4} \tag{2.67}$$

ϵ_p and ϵ_s are the dielectric constants of the protein and solvent, set to 2 and 80 respectively. As in previous studies, the scaling factor α is set to 1.67. The Stokes-Einstein relations are often used to estimate the translational and rotational diffusion constants for biological macromolecules:

$$\overbrace{D = \frac{k_B T}{6\pi\eta a}}^{\text{translational}} \quad \text{and} \quad \overbrace{D_R = \frac{k_B T}{8\pi\eta a^3}}^{\text{rotational}} \tag{2.68}$$

However, these assume a spherical geometry. Here, the HYDROPRO program [217] has been used, which constructs a hydrodynamic bead model of the proteins based on the non-hydrogen atoms to calculate these quantities more accurately. The effective charge method [218] was used to assign partial charges to the proteins.

2.4 Molecular docking

Molecular docking is a commonly used technique for predicting the interaction of ligands with biomolecular targets, and is particularly useful in the field of computer aided drug design (CADD) when the objective is to attempt to identify ligands that are able to bind to a target biomolecule [219]. If the structure of the target is known (from X-ray crystallography, NMR spectroscopy or homology modelling) then computational docking can expedite this lead identification process by rapidly screening a large number of candidate drug molecules against a target in a process known as virtual screening. The resulting protein-ligand complexes are evaluated and ranked by a scoring function to eliminate the weaker drug candidates and produce a series of 'hits' that are then earmarked for further investigation. These lead compounds can then be tested by experimental assays,

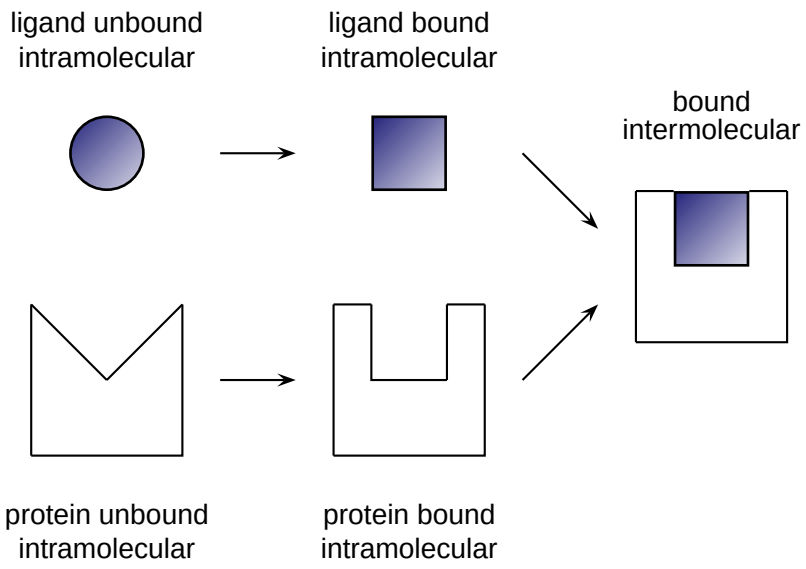

Figure 2.8: Stepwise evaluation of ligand binding using AutoDock.

and synthesis of more potent analogues by chemical modification of the scaffold can then be attempted, a procedure known as lead optimisation.

Docking studies are typically performed by either docking whole molecules into protein binding sites or by 'growing' a new molecule in the active site by utilising a fragment based approach. Docking and virtual screening are now firmly established stages of the drug development pipeline and are widely used in industry and academia. Popular docking programs include the freely available AutoDock [220, 221] and commercial software such as GOLD (CCDC [222]), Glide (Schrödinger, LLC [223]) and LigandFit (Accelrys [224]).

All molecular docking described in this thesis was performed using AutoDock. AutoDock evaluates binding in a stepwise fashion (Figure 2.8). First, AutoDock estimates the intramolecular energetics for the transition from the protein unbound state to its conformation in the bound state, and separately for the ligand transition from its unbound conformation to its bound state. Next, the intermolecular energetics of combining the protein and the ligand in the bound conformation are computed. The AutoDock force field contains six pairwise evaluations, V, and an estimate of the conformational entropy

lost upon binding, ΔS_{conf}:

$$\begin{aligned}\Delta G = &(V^{L-L}_{\text{bound}} - V^{L-L}_{\text{unbound}}) \\ &+ (V^{P-P}_{\text{unbound}} - V^{P-P}_{\text{unbound}}) \\ &+ (V^{P-L}_{\text{bound}} - V^{P-L}_{\text{unbound}} + \Delta S_{\text{conf}})\end{aligned} \tag{2.69}$$

Here, L is the ligand and P is the protein target in a docking run. Each of these pairwise evaluations includes terms to account for the van der Waals interactions, hydrogen bonding, electrostatics and desolvation, given by:

$$\begin{aligned}\Delta G = &\overbrace{W_{\text{vdW}} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right)}^{\Delta G_{\text{vdW}}} + \overbrace{W_{\text{Hbond}} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right)}^{\Delta G_{\text{Hbond}}} \\ &+ \underbrace{W_{\text{elec}} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}}}_{\Delta G_{\text{elec}}} + \underbrace{W_{\text{desolv}} \sum_{i,j} (S_i V_j + S_j V_i) \exp\left(\frac{-r_{ij}^2}{2\sigma^2} \right)}_{\Delta G_{\text{desolv}}}\end{aligned} \tag{2.70}$$

The weighting constants, W, are included to calibrate the energy function, which has been benchmarked against experimentally determined binding affinities. AutoDock allows for the inclusion of rotatable bonds within the ligand to increase the flexibility of the small molecule and improve the chances of obtaining a favourable docking pose. It also provides functionality to assign flexibility to amino acid sidechains near the binding site, providing additional degrees of freedom when attempting to generate a suitable protein-ligand complex. The details of the bond torsions assigned to the various molecules used in docking are discussed in Chapter 5.

Chapter 3

Computational studies of membrane-bound GRP1-PH

In this first results chapter, simulations of the PH domain from GRP1 bound to PI(3,4,5)P_3 are described. GRP1-PH is an attractive candidate for study as it binds to PI(3,4,5)P_3 with unusually high specificity and high affinity, and owing to these properties a great deal of experimental data on GRP1-PH are available. This makes GRP1-PH a useful test case for validation of some of the computational techniques used in this thesis, before moving on to apply them to more complex examples of membrane-binding proteins in later chapters. This study, along with experimental data from collaborators in the USA, was published in *Structure* [78].

3.1 Introduction

The PH domain is a highly structurally conserved domain of approximately 100 amino acid residues [35, 36]. This structural similarity is, however, somewhat remarkable given the low level of sequence identity among different PH domains [37]. Many PH domains bind to PIPs, a comparatively rare class of target lipid possessing a large negative charge, which arises from the presence of a variable number of phosphate group substituents on the inositol ring. Different patterns of phosphorylation at the 3′, 4′, and 5′ positions of the ring afford a total of seven distinct PIP species. The low level of sequence homology between PH domains gives rise to a correspondingly broad spectrum of behaviour, so while some PH domains are extremely specific and recognise only one of the seven species of PIP, others are virtually unable to discriminate between them [38, 132].

The PH domain of GRP1 (GRP1-PH) is one of the most selective PH domains, reversibly binding to PI(3,4,5)P_3 with pronounced specificity and high affinity [225]. Indeed, the fidelity of GRP1-PH for its ligand has led to its use as a biosensor to monitor the concentration of PI(3,4,5)P_3 in PI3-K signalling [226] or *in vivo* intracellular I(1,3,4,5)P_4 dynamics [227]. The concentration of PI(3,4,5)P_3 in the cytoplasmic leaflet of the cell membrane is ordinarily very low, and so recruitment of GRP1 to the membrane surface is dependent upon the activity of class IA PI3-Ks, which are responsible for generating elevated levels of PI(3,4,5)P_3 by phosphorylating the inositol headgroup of the more abundant PIP species PI(4,5)P_2 [66].

GRP1 is a member of the cytohesin family of proteins, which are small, phosphoinositide-dependent guanine nucleotide exchange factors (GEFs) that promote activation of ADP ribosylation factor (Arf) GTPases. Like other members of the Ras superfamily, Arf GTPases regulate several cellular processes including endocytosis and cytokinesis by switching between their active (GTP-bound) and inactive (GDP-bound) forms [228], and they are thought to be particularly important in mediating the invasivity of tumour cells in breast

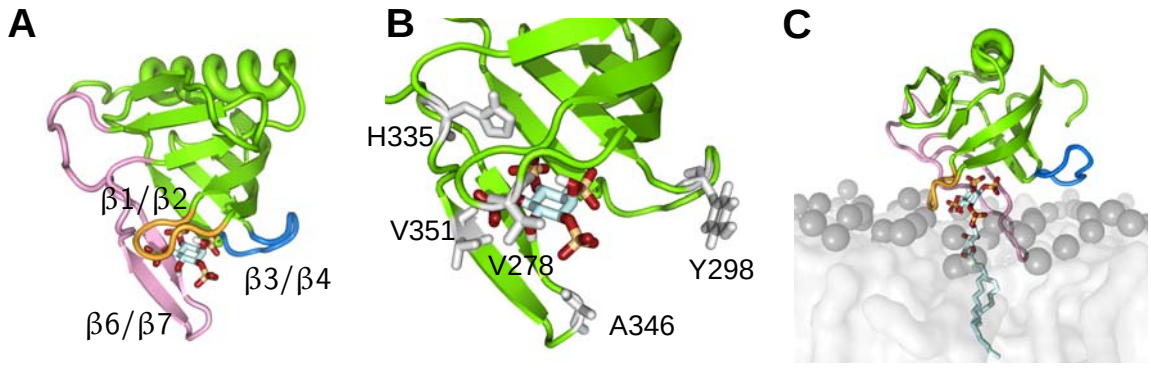

Figure 3.1: Structural features of GRP1-PH. **(A)** The three binding loops, β1/β2 (orange), β3/β4 (blue), and β6/β7 (pink), and their positions relative to the I(1,3,4,5)P_4 headgroup. **(B)** The geometry of the I(1,3,4,5)P_4 binding site, with the key amino acid residues referred to in the text shown as stick representations in white. **(C)** The initial setup for the MD simulations, with the centres of the phosphate groups shown as grey spheres and the lipid bilayer depicted as a translucent white surface. Water molecules and ions are omitted for clarity. In all cases, the protein is shown as a green ribbon diagram, and the I(1,3,4,5)P_4 headgroup is shown as a stick model.

and skin cancer [229]. While the catalytic activity of GRP1 has been attributed to its Sec7 domain, recent experimental studies [230] have suggested that this activity is autoinhibited in the cytosol and that the protein only becomes catalytically competent when bound to the lipid bilayer. This underscores the importance of the GRP1 PH domain, since it is the high specificity and affinity of GRP1-PH for PI(3,4,5)P_3 that is primarily responsible for driving the recruitment of its host protein to the surface of the plasma membrane. However, it is possible that when bound to the membrane, nonspecific interactions with the lipid molecules in close proximity may play a role in anchoring GRP1-PH to the lipid bilayer. It is the nature and strength of these additional protein-lipid interactions that are the subject of this chapter.

The canonical PH domain comprises a seven stranded β barrel capped at one end by an amphipathic α helix, while the other end is flanked by three loops (β1/β2, β3/β4, and β6/β7) that together form the PIP binding site (Figure 3.1). The precise sequence of amino acid residues in these loops is a major factor in determining the specificity of the domain, and it has been observed that even small changes in this binding motif such as the addition or deletion of a single glycine residue can have a profound effect on the binding properties of the domain [231, 232]. An unusual structural feature of GRP1-PH is the β6′/β6″ hairpin insertion, which lies within the extended β6/β7 loop.

Comparative studies between different PH domains have suggested that the presence of this insertion in GRP1-PH is the factor that determines its pronounced PI(3,4,5)P_3 specificity, as it 'deepens' the ligand binding pocket and provides additional coordination for the 5′-phosphate [233]. The position of the hairpin insertion relative to the I(1,3,4,5)P_4 headgroup in the crystal structure of GRP1-PH suggests that in the membrane-bound state this loop may also interact with the lipid bilayer. However, this remains only a conjecture as, although the binding of GRP1-PH to soluble inositol phosphates is now relatively well characterised, much less is known about the nature of the membrane-associated complex. In particular, the extent of the interaction between GRP1-PH and the lipid bilayer remains unclear. While SPR experiments by Manna *et al.* [132] and later by He *et al.* [98] detected some membrane penetration, when Knight and Falke monitored the diffusion of individual GRP1-PH domains using single particle tracking in TIRFM experiments, they found that, within error, the diffusion constant of the bound GRP1-PH-PI(3,4,5)P_3 complex was the same as that of a free lipid [96]. This was interpreted as indicating that there was comparatively little interaction between the bound PH domain and the adjacent membrane, since such an interaction would presumably exert a viscous drag force on the bound complex and slow its diffusion rate. Here, the interaction between GRP1-PH and the membrane is examined, focusing on the relative importance of the specific, PI-mediated interactions and the nonspecific interactions with the lipid molecules of the adjacent bilayer[1].

3.2 Methods

To generate a suitable initial configuration of the GRP1-PH-PI(3,4,5)P_3 complex at the membrane surface, it was necessary to place the crystal structure of the protein incorporating the I(1,3,4,5)P_4 headgroup at an appropriate position and orientation relative

[1]This project was part of a collaboration with Professor Tatiana Kutateladze and colleagues at the University of Colorado Denver School of Medicine, and Professor Robert Stahelin and colleagues at Indiana University School of Medicine and the University of Notre Dame.

to the lipid bilayer (Figure 3.1). The orientation of PIP headgroups at the surface of lipid bilayers has recently been studied extensively by Pastor and co-workers [234] using MD simulations and finite-difference Poisson-Boltzmann calculations. They found that the I(1,3,4,5)P_4 headgroup sampled a range of tilt-twist (θ-ϕ) angles at the membrane surface, but that some orientations were preferred. These were used as a guide when positioning the *holo* GRP1 PH domain relative to the surface of the membrane, matching the orientation of its bound I(1,3,4,5)P_4 headgroup to the preferred θ-ϕ angles determined by Pastor and co-workers (Figure 3.2).

MD simulations were performed using GROMACS version 3.3.3 [185] with the GROMOS96 43a1 forcefield [142]. Bond lengths and angles were constrained using the LINCS algorithm [139] and a timestep of $\Delta t = 2$ fs was used, writing atomic positions every 10 ps. The neighbour list was updated every ten steps. The 1.5 Å resolution structure of GRP1-PH complexed with I(1,3,4,5)P_4 (PDB 1FGY) [4] was used in the simulations. Force field parameters for the I(1,3,4,5)P_4 headgroup were generated as described in Chapter 2. The charge on the I(1,3,4,5)P_4 headgroup was assumed to be -7, although *in vivo* the net charge of PI(3,4,5)P_3 may vary substantially [17]. McLaughlin and co-workers point out that the net charge of PI(4,5)P_2, and by extension PI(3,4,5)P_3, is also likely to be influenced by the interaction with PIP binding proteins because protons bound to the headgroup may be displaced upon protein binding [8]. Despite this, when Lai *et al.* [87] performed simulations of a membrane-bound C2 domain from protein kinase Cα (PKCα) in complex with PI(4,5)P_2 they did not observe any change in the stability of the protein-membrane complex upon altering the net charge on PI(4,5)P_2.

Simulations were performed under periodic boundary conditions and temperature was kept constant at 296 K by coupling the system to a heat bath using a Berendsen thermostat [159] with $\tau_T = 0.1$ ps. Pressure was maintained at 1 atm using a Parrinello-Rahman barostat [162] and [163] and semi-isotropic pressure coupling with $\tau_p = 1$ ps and a compressibility of 4.6×10^{-5} bar^{-1}. Long-range electrostatics were treated using

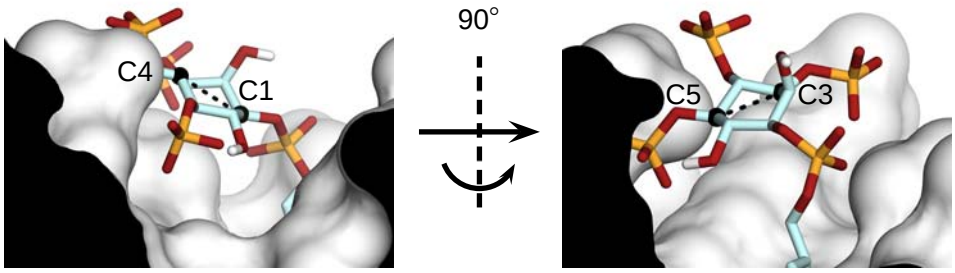

Figure 3.2: Initial positioning of PI(3,4,5)P_3 in the lipid bilayer. The I(1,3,4,5)P_4 headgroup was positioned such that the C1-C4 vector lay at an angle of approximately 40° to the membrane surface, while the C3-C5 vector lay at -17.5° to the membrane surface, in line with the results from simulation studies performed by Pastor and co-workers to determine PIP headgroup orientations at the surface of lipid bilayers [234].

the particle mesh Ewald (PME) method [235] with a short range real space cut off of 10 Å. To obtain the initial membrane-bound conformation of the protein, the I(1,3,4,5)P_4 headgroup in complex with GRP1-PH from the crystal structure was overlaid onto a phosphatidylcholine headgroup of a lipid molecule close to the centre of a preformed POPC bilayer comprising 174 lipid molecules. The I(1,3,4,5)P_4 headgroup was positioned such that the C1-C4 vector lay at an angle of approximately 40° to the membrane surface, while the C3-C5 vector lay at -17.5° to the membrane surface, within the ranges determined in a recent computational study of PIP headgroup orientations in lipid bilayers [234]. The protein-ligand complex was then exchanged for the PC headgroup and solvent molecules were added using the GROMACS tools, employing the SPC model for water [236]. After addition of the requisite number of ions to neutralise its net charge, the system was energy minimised for up to 1000 steps using the steepest descent algorithm and then sequentially equilibrated for a total of 3 ns. During the first 1 ns of equilibration, isotropic position restraints were applied to the protein and the ligand allowing the water and lipids to relax around the static complex. Restraints on the protein side chains were then removed, and equilibration continued for a further 1 ns, before continuing for a final 1 ns with all restraints removed bar those on the protein Cα atoms and the ligand atoms. Production runs were then performed for 100 ns to allow the protein to adopt

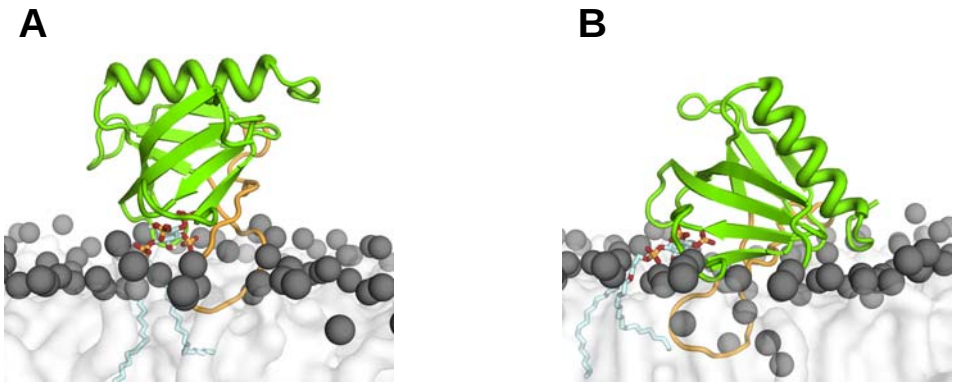

Figure 3.3: Simulation snapshots of membrane-bound GRP1-PH. **(A)** Snapshot at the beginning of a simulation (0 ns) with GRP1-PH bound to PI(3,4,5)P_3 positioned at the surface of the lipid bilayer. **(B)** Snapshot after 100 ns, showing the insertion of the $\beta 6/\beta 7$ loop coloured in orange. The aliphatic lipid tails are shown as a white translucent surface and the lipid phosphate headgroups are shown as grey spheres. Water molecules and ions are omitted for clarity.

its preferred orientation at the bilayer surface, and subsequent SMD simulations were initiated from the conformation achieved after 50 ns of equilibrium MD simulation.

SMD simulations were carried out using the GROMACS pull code. A stiff, harmonic spring of force constant 500 pN $Å^{-1}$ was attached to the centre of mass of the protein, though the protein remained free to rotate about its centre of mass. Another, identical spring was attached to the centre of mass of the lipid molecules in the bilayer to maintain the integrity and planarity of the bilayer as the protein was pulled away. The spring attached to the protein was then retracted parallel to the z axis at a rate of 1.0 Å ns^{-1} over a period of 40 ns, or at a rate of 0.5 Å ns^{-1} over a period of 80 ns.

Lateral diffusion constants were estimated by first removing the centre of mass motion of the combined protein and lipid bilayer and then calculating the mean square displacement (MSD) of the protein in the xy plane, such that the motion of the protein was measured relative to the centre of mass of the membrane. The value of D was extracted by fitting to the MSD over the first 50 ns of the simulation trajectory where it is approximately linear using the GROMACS tools, and subsequently using the Einstein

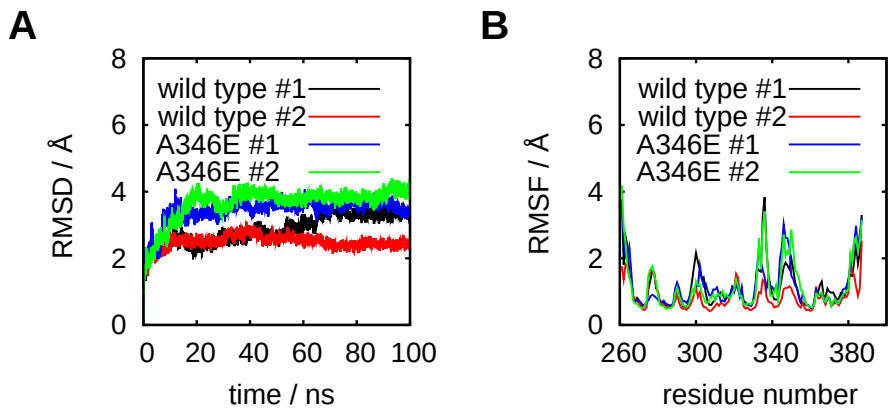

Figure 3.4: Conformational drift of bound GRP1-PH during the equilibrium MD simulations. **(A)** RMSD and **(B)** RMSF of the protein Cα atoms over the 100 ns trajectories. The wild type simulations are shown in black and red, while the simulations of the A346E mutant are shown in blue and green.

relation in two dimensions [136].

$$D = \lim_{t\to\infty} \frac{\langle |r_i(t) - r_i(0)|^2 \rangle}{4t} \tag{3.1}$$

3.3 MD simulations of membrane-bound GRP1-PH-PI(3,4,5)P_3

To explore the extent of the interaction between GRP1-PH and the lipid bilayer, two 100 ns equilibrium MD simulations of the membrane-bound GRP1-PH-PI(3,4,5)P_3 complex were performed (Figure 3.3) using two different sets of initial velocities. Analysis of the Cα RMSD of the protein demonstrated that GRP1-PH experienced relatively low levels of conformational drift over the trajectories (Figure 3.4). The simulations revealed an array of contacts between the protein and the lipids of the adjacent bilayer, showing that the $\beta1/\beta2$, $\beta3/\beta4$, and $\beta6/\beta7$ loops flanking the binding site are well positioned to interact with the membrane. To assess the degree of membrane penetration by each of the three loops, the non-polar protein-lipid interactions were analysed (Figure 3.5 and Figure 3.6), which show that the residues within the $\beta1/\beta2$ and $\beta3/\beta4$ loops form transient non-polar contacts with the POPC carbon tails over the course of the trajectory. The $\beta6/\beta7$ loop on the other hand experiences a much more sustained interaction with the

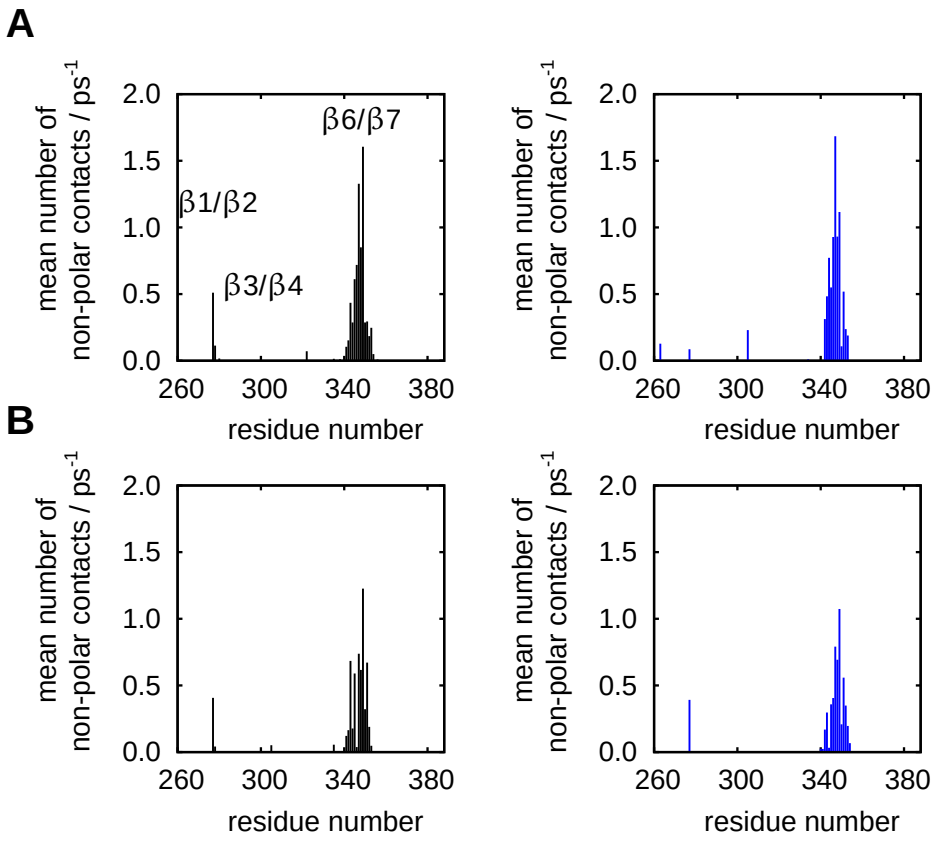

Figure 3.5: Membrane penetration of GRP1-PH observed during the MD simulations. **(A)** Mean number of non-polar protein-lipid contacts per residue per ps over the 100 ns for the two wild type protein simulation (one simulation in black and the second simulation in blue). **(B)** The same measure for the two A346E mutant simulations. Non-polar contacts are defined as the number of POPC tail carbon atoms within 4 Å of a protein heavy atom.

lipid acyl chains throughout the simulation, indicative of a greater degree of penetration into the hydrophobic core of the membrane. Thus, the extension to the core β barrel provided by the hairpin insertion in the GRP1 PH domain not only deepens the PI(3,4,5)P_3 binding pocket but also appears to provide a framework for more extensive interactions with the surrounding membrane lipids. Several residues in GRP1-PH interact directly with the I(1,3,4,5)P_4 headgroup in the crystal structure [4], and this extensive network of interactions is preserved throughout both simulations of membrane-bound GRP1-PH. Analysis of the hydrogen bonding interactions between the amino acid residues lining the binding pocket of GRP1-PH and the phosphate groups of PI(3,4,5)P_3 detected hydrogen bonds between PI(3,4,5)P_3 and the following amino acid residues: K273; G276; R277; V278; T280; K282; R284; Y295; R305; K343; and H355 (Figure 3.7). This is in excellent

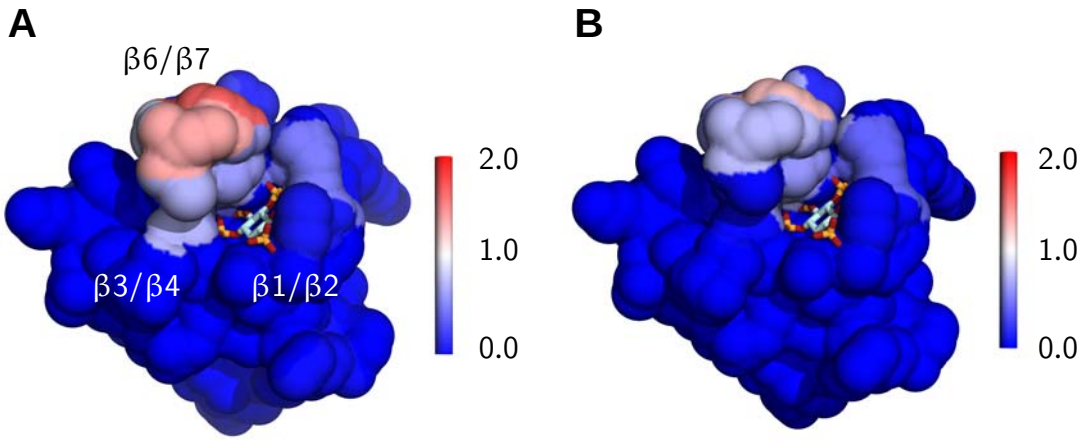

Figure 3.6: Non-polar contacts from Figure 3.5 projected onto the molecular surface of GRP1-PH for **(A)** the wild type protein and **(B)** the A346E mutant respectively, illustrating the decrease in penetration depth of the β6/β7 loop upon mutation. The colour scale corresponds to the plots in Figure 3.5 and shows the mean number of non-polar protein-lipid contacts per residue per ps, varying from 0 (blue) to 2 (red). Non-polar contacts are defined as the number of POPC tail carbon atoms within 4 Å of a protein heavy atom.

agreement with the hydrogen bonding pattern found in the crystal structure, with only the hydrogen bond donated by N354 to the 5′-phosphate not detected. This indicates that the binding mode adopted by the soluble I(1,3,4,5)P_4 headgroup is preserved in the membrane-associated state when GRP1-PH is bound to PI(3,4,5)P_3.

3.3.1 Experimental verification using NMR spectroscopy

To test the prediction that it is these regions of GRP1-PH that interact with the lipid bilayer, He and Kutateladze investigated the association of the PI(3,4,5)P_3-bound protein with membrane mimetic DPC micelles using NMR spectroscopy. The ^{1}H, ^{15}N HSQC spectra of the ^{15}N-labelled PH domain in complex with PI(3,4,5)P_3 at a 1:2 protein to lipid ratio were collected as DPC micelles were titrated in. Substantial changes in amide resonances of R267, L272, C292, L293, F296, E303, C342, K343, T344, E345, V351, E352, G353, and K373 were observed (Figure 3.7), indicating that these residues are directly or indirectly involved in the interaction with the micelles. The majority of the most perturbed residues were located in the β1 and β3 strands and in the β3/β4 and β6/β7 loops of the PH domain. The magnitude of the perturbation in chemical shift was

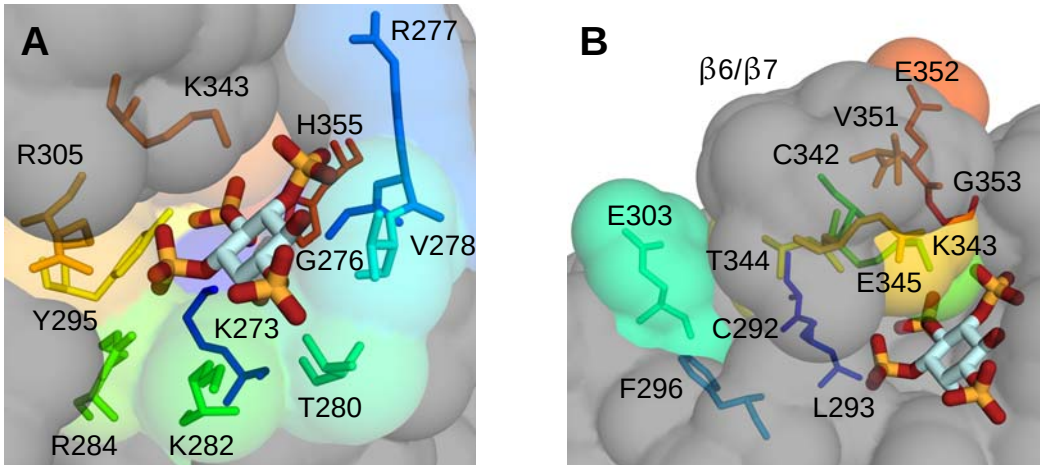

Figure 3.7: PI(3,4,5)P_3 hydrogen bonding and NMR micelle contacts. **(A)** Hydrogen bonds between the amino acid residues lining the binding pocket of GRP1-PH and PI(3,4,5)P_3 occuring in simulations of bound GRP1-PH. **(B)** Residues experiencing medium to large perturbations in their amide resonances in chemical shift mapping NMR spectroscopy when DPC micelles were titrated in. The magnitude of the chemical shift change was particularly large in the region around the β6/β7 loop. The grey solvent-accessible surface of GRP1-PH is shown in the same orientation as Figure 3.6.

particularly large in the region corresponding to the β6/β7 loop, suggesting that of the three loops the β6/β7 loop interacts with the lipid environment to the greatest extent, in good agreement with the MD simulations. The NMR studies also indicate that monolayer penetration of GRP1-PH is pH dependent, as the magnitude of the NMR perturbations in the PI(3,4,5)P_3-bound GRP1 PH domain induced by DPC micelles is increased by lowering the pH. This builds on previous work [98], which established that protonation of the H355 residue increases the binding affinity of the protein.

3.4 SMD simulations of dissociation of GRP1-PH from a membrane

On the basis of the results from NMR spectroscopy, the relative strength and pH dependence of the interactions between each of the three binding loops and the PI(3,4,5)P_3-containing lipid bilayer were investigated using SMD simulations to probe the likely mechanism of dissociation of the complex. Thus, the bound PH domain was 'pulled' away from the ligand and the surrounding lipid bilayer by a force applied parallel to the z axis, perpendicular to the plane of the membrane (Figure 3.8). The sequence in which

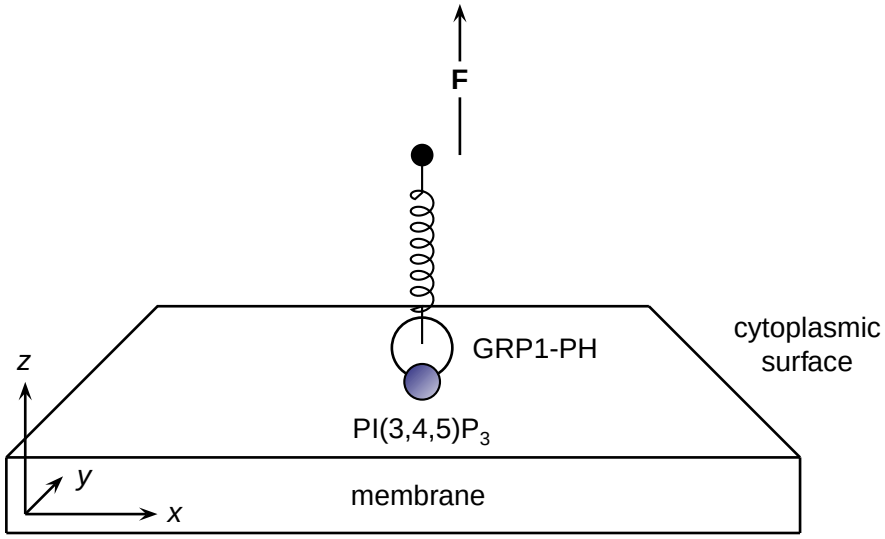

Figure 3.8: Schematic of steered MD simulation setup for dissociation of GRP1-PH. A virtual spring is attached to the pulled group and a force is applied along the reaction coordinate. The force causes the spring to retract and the pulled group is forcibly dissociated from the surface of the membrane.

the protein-ligand and protein-lipid contacts are broken as the protein is pulled away allows for exploration of the relative contribution of each of the loop regions in sustaining the complex.

By performing two SMD simulations, one with the side chain of H355 in a neutral state and another with the side chain in a protonated state, it is possible to partially mimic the effect of a change in pH and assess how this may affect the interaction between the protein and the ligand or the membrane. Each pulling simulation was performed twice, once at a pulling speed of 1.0 Å ns^{-1} over a period of 40 ns and once at a pulling speed of 0.5 Å ns^{-1} over a period of 80 ns, in order to assess the reproducibility of the results. This gave a total of four SMD simulations performed under different conditions. The β6/β7 loop that penetrated the lipid bilayer to the greatest extent in the equilibrium MD simulations is sufficiently tightly bound to the lipid bilayer that it is able to extract POPC lipids upon dissociation of the complex in the SMD simulations (Figure 3.9). This effect is even more pronounced when H355 is protonated.

The electrostatic and van der Waals contributions to the potential energy of the com-

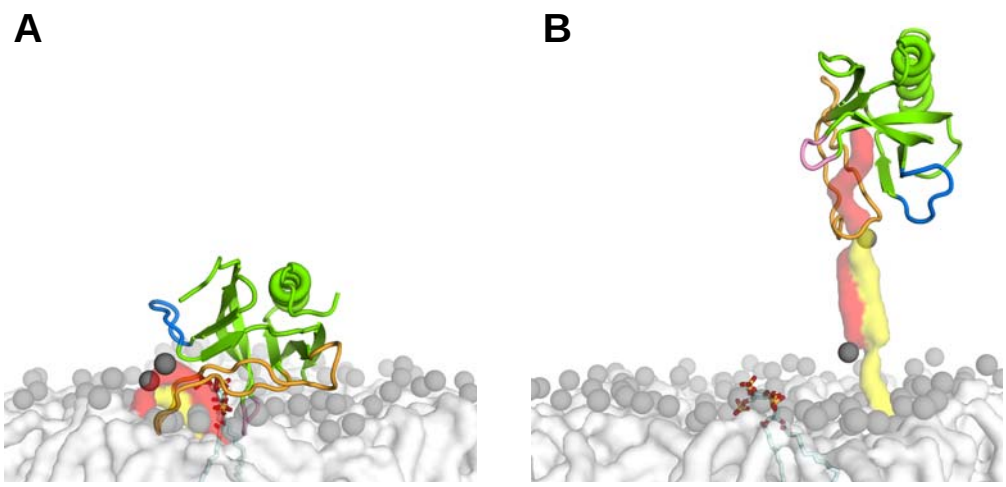

Figure 3.9: Simulation snapshots of GRP1-PH being pulled from the bilayer surface. **(A)** Snapshot at t = 0 ns for the neutral H355 case. **(B)** Snapshot at t = 36 ns, again for the neutral H355 case. The interaction between the β6/β7 loop and the membrane lipids is sufficiently strong to completely remove a lipid from the bilayer (shown in red) at this pulling velocity, while another lipid (shown in yellow) is partially extracted.

plex over the course of the SMD simulations were extracted and are shown in Figure 3.10 for the faster pulling simulations, with a pulling rate of 1.0 Å ns^{-1} for 40 ns, and in Figure 3.11 for the slower pulling simulations with a pull rate of 0.5 Å ns^{-1} for 80 ns. The lipid bilayer provides the dominant contribution to the van der Waals component of the potential energy in both cases. The van der Waals component remains non-zero at the end of both the simulations despite the protein having been pulled away from the membrane due to the extraction of lipids from the bilayer as discussed above, which remain bound to the protein as it is pulled away. Interestingly, in the case where H355 was in a neutral state, the protein-PI(3,4,5)P_3 interaction and the protein-POPC interaction both contribute in approximately equal measure to the overall electrostatic interaction between the protein and the membrane. This is likely due to multiple interactions between the protein and the phosphatidylcholine headgroups in the lipid bilayer. However, when H355 is protonated, the protein-PI(3,4,5)P_3 interaction dominates. As before, the component of the electrostatic potential energy does not fall to zero over the course of the simulation, again as a result of the extraction of lipid molecules from the bilayer. Pulling the protein over the same distance at a faster rate gives similar results.

The considerable contribution to the van der Waals and electrostatic potential energy

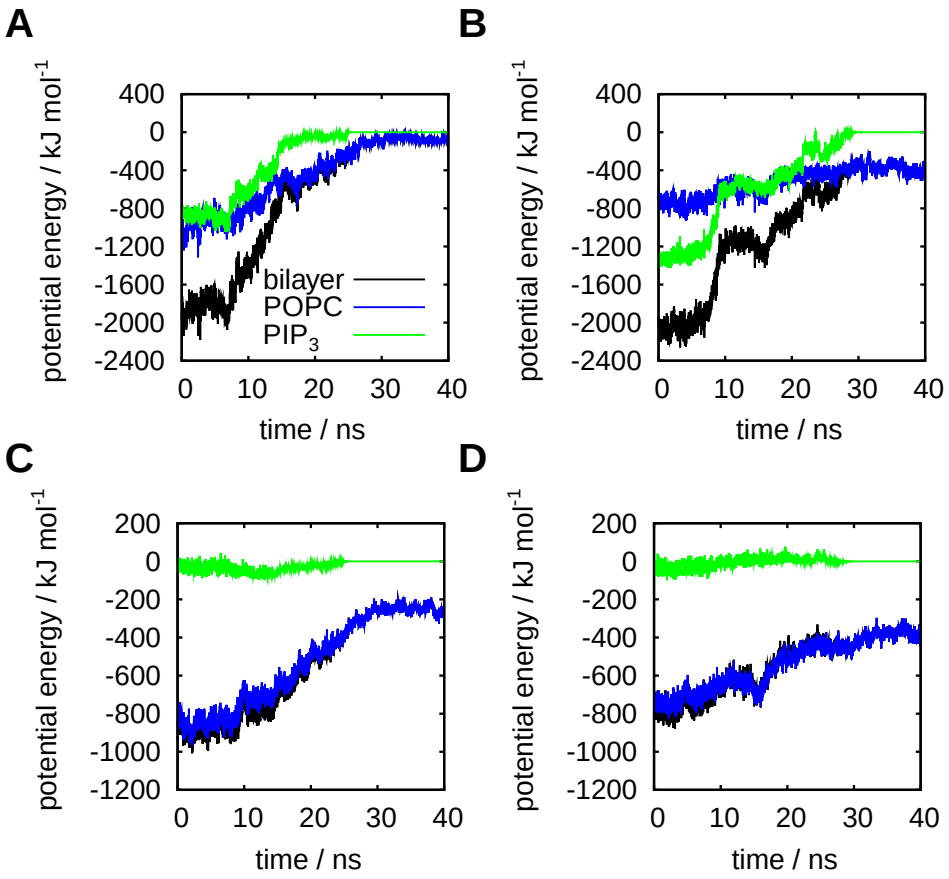

Figure 3.10: Short ranged components of the electrostatic and van der Waals contributions to the potential energy of the protein-membrane complex calculated from the shorter SMD simulations, duration 40 ns and a pulling rate of 1.0 Å ns^{-1}. The electrostatic potential energy is shown for **(A)** the system with neutral H355 and **(B)** the system with H355 in a protonated state. The van der Waals potential energies are shown in **(C)** for neutral H355 and in **(D)** for H355 in a protonated state. The traces are qualitatively similar in shape to those calculated for the slower SMD simulations, 80 ns duration and a pulling rate of 0.5 Å ns^{-1} (Figure 3.11).

of the protein-membrane complex provided by the POPC lipids suggests that the lipid bilayer plays an important supporting role in the membrane interaction of GRP1-PH. Since the H355 residue lies within the β6/β7 loop, which experienced the most sustained interaction with the lipid bilayer during the MD and SMD simulations, it is possible that protonation of this residue plays a dual role in increasing the binding affinity of the complex by not only reinforcing the specific interaction between H355 and $PI(3,4,5)P_3$, but also by extending the residence time of the β6/β7 loop in the lipid bilayer through more extensive electrostatic interactions with the headgroups of the surrounding lipids. In this respect, the protonated H355 residue could be said to act as a 'latch' mechanism, which prolongs the interaction between the loop and the membrane as the protein is

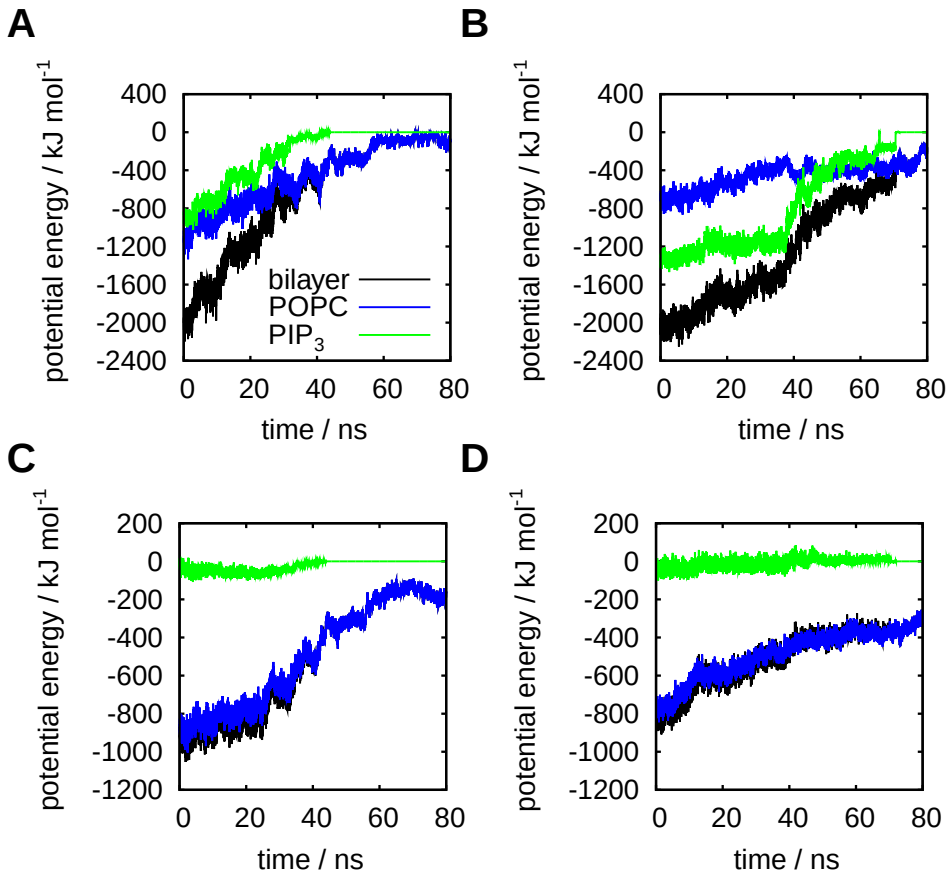

Figure 3.11: Short ranged components of the electrostatic and van der Waals contributions to the potential energy of the protein-membrane complex calculated from the longer SMD simulations, duration 80 ns and a pulling rate of 0.5 Å ns^{-1}. The electrostatic potential energy is shown for **(A)** the system with neutral H355 and **(B)** the system with H355 in a protonated state. The van der Waals potential energies are shown in **(C)** for neutral H355 and in **(D)** for H355 in a protonated state. The traces are qualitatively similar in shape to those calculated for the faster SMD simulations, 40 ns duration and a pulling rate of 1.0 Å ns^{-1} (Figure 3.10).

pulled away.

Similar SMD simulations were also performed for the isolated protein with the I(1,3,4,5)P_4 headgroup in solution, with no bilayer present. Again, simulations were carried out for the two cases of neutral and protonated H355. Comparison with the previous simulations in which the membrane was present shows that the inclusion of the lipid bilayer appears to prolong the contacts between the PI(3,4,5)P_3 headgroup and the protein (Figure 3.12), again pointing toward a potential 'dual recognition' mechanism where the lipid bilayer complements the specific interaction between GRP1-PH and PI(3,4,5)P_3.

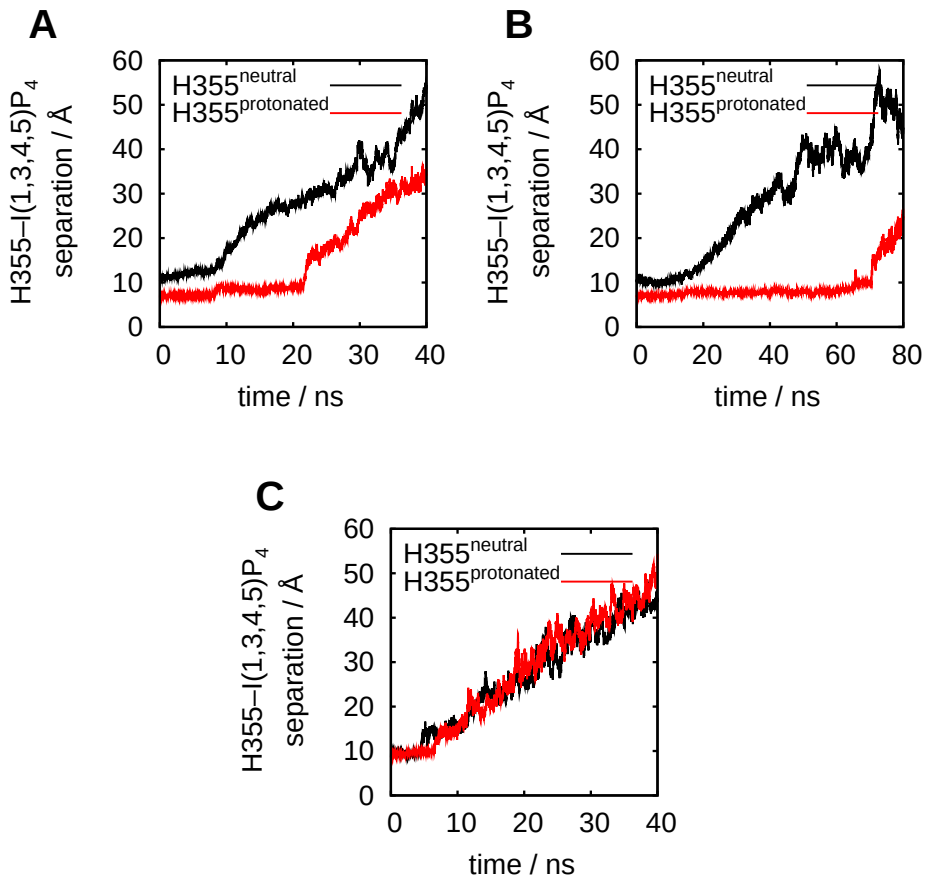

Figure 3.12: Interactions of the protein with both PI(3,4,5)P_3 and the lipids in the membrane are enhanced by protonation of H355. **(A)** Distance between the centre of mass of the I(1,3,4,5)P_4 headgroup and the centre of mass of H355 in the faster membrane-bound SMD simulations for both the neutral H355 case (black) and the protonated H355 case (red). The protonated H355 residue appears to act as a 'latch', prolonging the interaction between the protein and the I(1,3,4,5)P_4 headgroup as the protein is pulled away. **(B)** Distance between the centre of mass of the I(1,3,4,5)P_4 headgroup and the centre of mass of H355 in the slower membrane-bound SMD simulations. **(C)** The same measure is shown, this time for the SMD simulations carried out with the protein and isolated I(1,3,4,5)P_4 headgroup in solution, with no bilayer present. The 'latching' behaviour seen in the membrane-bound simulations is still observed, but is a much weaker effect in the absence of the lipid bilayer, suggesting that protonation also favours a more substantial interaction with the membrane.

3.5 Membrane penetration of mutant GRP1-PH

3.5.1 Monolayer penetration studies

To further define the role of the three loop regions in membrane binding, Xue and Stahelin replaced the hydrophobic residues of each loop (V278 of β1/β2; Y298 of β3/β4; A346 and V351 of β6/β7) with a glutamate residue and the corresponding mutant proteins were evaluated by monolayer penetration experiments in a Langmuir-Blodgett

trough. The wild type PH domain penetrated up to 28 mN m^{-1} at pH 7.4 as previously reported [98], and the V351E mutation displayed a similar degree of penetration. However, the V278E, Y298E, and A346E mutants all exhibited reduced penetration, with π_c values between 23 and 24 mN m^{-1} for all three mutations.

3.5.2 MD simulations

The β6/β7 loop penetrated most deeply into the bilayer in the MD simulations of the wild type protein. As A346 lies at the base of this loop (Figure 3.1), the A346E mutant tested in the monolayer penetration studies was thought to be most likely to result in observable reduction in membrane penetration. Simulations of the A346E mutant were therefore conducted to examine how this might alter the degree of membrane insertion. Again, a 100 ns simulation of the mutant complex was performed and this simulation was also repeated using different initial velocities. The array of contacts between GRP1-PH and $PI(3,4,5)P_3$ observed for the wild type simulation is conserved in the mutant. However, the non-polar contacts between the β6/β7 loop and the lipid tails in the mutant become less frequent in comparison to the wild type simulations (Figure 3.5 and Figure 3.6).

Subsequent experiments by the Falke group used continuous wave electron paramagnetic resonance (CW EPR) spectroscopy to determine the membrane penetration depth and docking geometry of membrane-bound GRP1-PH [88]. Falke and colleagues used site-directed mutagenesis to couple spin-labels to a variety of sites selected for Cys mutagenesis, and assessed the spectral changes upon binding to membranes. Their model of membrane-bound GRP1-PH is in very good agreement with the results presented here, and in particular they too observed membrane penetration by the β6/β7 loop. Their EPR docking model revealed that V278, R284 and R305 contacted the bilayer, while V351 did not, in agreement with the results from the monolayer penetration experiments performed by Xue and Stahelin.

3.6 Discussion

Results from the MD simulations indicate that the GRP1 PH domain is able to penetrate the membrane, with the three loops flanking the binding site together forming a dynamic yet sustained array of contacts with the membrane lipids. However, the MD simulations do suffer from some limitations, which are discussed in more detail below.

The docked structure is based on the GRP1-PH crystal structure superimposed onto a POPC lipid bilayer, with the headgroup orientation matched to that observed in previous MD simulations of PIP lipids. This superimposition of the ligand-bound crystal structure onto a lipid bilayer also determined the initial orientation of the protein relative to the bilayer surface. However, the crystal structure of the protein bound to a *soluble* inositol phosphate may not be representative of the membrane-bound conformation. The protein has to first approach the membrane and then bind to the surface *in vivo*, a process thought to be dominated by electrostatic interactions (see Chapter 4), and this may involve conformational changes not captured by our static superimposition of the crystal structure onto the PI(3,4,5)P_3-containing membrane. This may be particularly true of the three unstructured flanking loops. Nonetheless the good agreement with experimental data is encouraging, as this suggests that the correct overall orientation of the protein relative to the bilayer has been determined.

3.6.1 Lateral diffusion constants of GRP1-PH

To test whether or not the positioning of the protein relative to the surface of the lipid bilayer was appropriate, single molecule[2] lateral diffusion constants for GRP1-PH were estimated from the two dimensional MSD of the PH domain from the two equilibrium simulations of the membrane-bound wild type protein (Figure 3.13), which yielded val-

[2]Recently, Knight *et al.* extended their studies of GRP1-PH to look at multimeric GRP1-PH domains [237]. The authors again used TIRFM, this time to explore the collective diffusion at the membrane surface of multiple GRP1-PH domains tethered together.

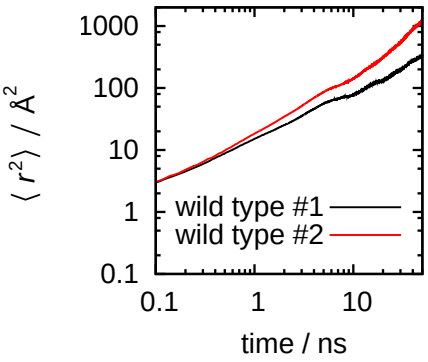

Figure 3.13: Two dimensional MSD of GRP1-PH. The mean square displacement, $\langle r^2 \rangle$, was calculated for lateral displacements of the protein in the *xy* plane, and is approximately linear over the first half of the trajectory.

ues of $D = 1.4 \times 10^{-8} \pm 0.3 \times 10^{-8}$ cm^2 s^{-1} and $6.3 \times 10^{-8} \pm 2.2 \times 10^{-8}$ cm^2 s^{-1}. These values compare reasonably well with the experimental value of $D = 3.0 \times 10^{-8} \pm 0.3 \times 10^{-8}$ cm^2 s^{-1} obtained by Knight and Falke, again suggesting that the simulations are able to reproduce experimentally observed behaviour.

3.6.2 Simulation length and the stability of the bound complex

Kinetic binding parameters determined from previous experimental studies [96] and [238] indicate that the value of k_{off} for wild type GRP1-PH in mixed PC/PI(3,4,5)P_3 lipid bilayers ranges from 0.38 to 1.60 s^{-1}, depending upon experimental conditions and membrane composition. These data suggest that the mean lifetime of the complex, $1/k_{off}$, is of the order of seconds, and the MD simulations of the bound complex are therefore comparatively short. While no distinctive changes in the global conformation of the protein were detected over this period (Figure 3.4), this does not preclude the possibility that the degree of membrane insertion may change over longer timescales. In particular, it should be noted that in the crystal structure the β3/β4 loop is twisted away from the binding cleft and, by extension, the membrane surface (Figure 3.1). As a result, relatively little membrane penetration is observed for this loop in the simulations. Although there is no evidence of this loop flexing and inserting more deeply into the bilayer during any of the simulations, this remains a possibility.

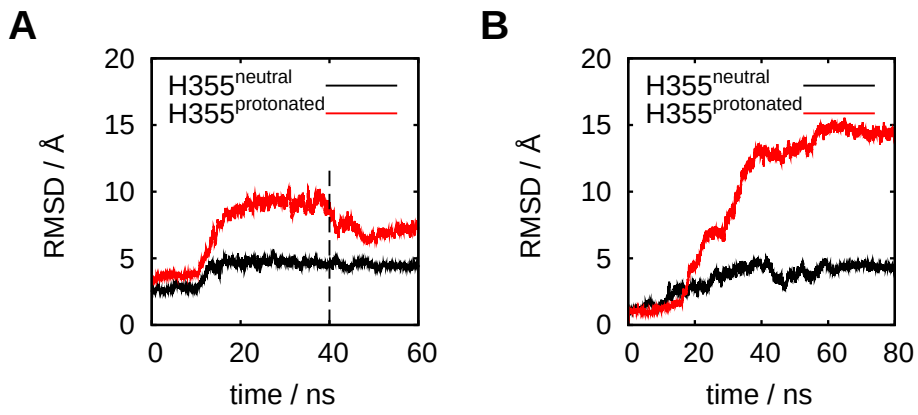

Figure 3.14: RMSD of the protein Cα atoms over the course of the SMD trajectories. **(A)** RMSD of the protein Cα atoms for the faster SMD trajectory, duration 40 ns and a pulling rate of 1.0 Å ns^{-1}, and the resulting effect on the RMSD when the force was removed (indicated by a dashed line) and the system was simulated under equilibrium conditions for a further 20 ns. There is considerable distortion induced in the secondary structure during the SMD simulations, particularly in the case where H355 is protonated, reflecting the higher affinity binding to the membrane. Removing the applied force and continuing the simulation leads to partial recovery of the secondary structure. **(B)** RMSD of the protein Cα atoms for the slower SMD trajectory, duration 80 ns and a pulling rate of 0.5 Å ns^{-1}. Distortion is still induced in the protein structure even at this slower pulling rate.

3.6.3 Pull rate and reaction coordinate definition in SMD simulations

The pulling rates achievable in SMD (0.5 and 1.0 Å ns^{-1} in this case) are four orders of magnitude faster than experimental atomic force microscopy (AFM) rates [239]. This fast pulling rate does induce some distortion in the protein structure (Figure 3.14), and the protein is capable of removing lipids from the bilayer at this velocity (Figure 3.9). However, subsequent removal of the force does lead to a partial recovery of the secondary structure (Figure 3.14). The approach used here is similar to that used to investigate the force required to remove individual lipid molecules from a bilayer [240] and the pulling rate is comparable to or slower than that used in more recent studies of partitioning of anesthetics into lipid bilayers [241] and the unbinding of protein-ligand complexes [171, 172]. To assess the reliability of the SMD results, SMD simulations were run at two different speeds (0.5 Å ns^{-1} and 1.0 Å ns^{-1}) for different lengths of time (80 and 40 ns) such that the distance moved by the protein over the course of the simulation remained the same. Distortion in the protein secondary structure is still observed for slower pulling (Figure 3.14), and performing the same energy analysis yields similar results (Figure 3.10

and Figure 3.11).

By pulling the protein along the z axis perpendicular to the surface of the lipid bilayer, the dissociation pathway of the protein-membrane complex has effectively been defined *a priori*. This pathway of detachment of the protein from the surface of the membrane may not be the preferred dissociation mechanism *in vivo*, and it is likely that other dissociation pathways may exist. Here then, SMD simulations have simply been used to study the notional response of the protein-membrane complex to an applied force. Accordingly, the discussion is restricted to qualitative observations about the relative strength and pattern of interactions between different regions of the protein and the lipid bilayer. It is freely acknowledged that attempting to extract more quantitative information, such as free energies, would require much more extensive computations in order to reliably estimate these values. For a complex system such as this, it is expected that obtaining sufficiently well-converged potentials of mean force in order to draw meaningful conclusions would be extremely difficult, owing to the inevitable problems attaining adequate levels of sampling. This is discussed at greater length in Chapter 7. SMD therefore provides a computationally straightforward way to elicit what is essentially a qualitative picture of how different regions of the protein may interact with the lipid bilayer.

A computational study of the behaviour of PI(4,5)P_2 in lipid bilayers published recently [153] found that the work required to extract a PI(4,5)P_2 molecule from a DPPC bilayer using SMD simulations was approximately three times greater than that required to remove a DPPC molecule from the same bilayer. The authors ascribed this behaviour to the formation of a lipid microdomain around the PI(4,5)P_2 molecule and rationalised this result by noting that the primary role of PI(4,5)P_2 is thought to be the provision of a stable, membrane-bound anchor for signalling proteins at the plasma membrane. The observation that the bound GRP1-PH-PI(3,4,5)P_3 complex is able to extract POPC lipids when removed from the membrane surface during the SMD simulations, but that the PI(3,4,5)P_3 molecule remains within the lipid bilayer, appears to be consistent with this

result.

3.6.4 Anionic lipid clustering

Several authors have reported that the presence of anionic lipids such as PS or other PIP species increases the binding affinity of the bound complex [98, 238] and also that these anionic lipids exert a drag force, slowing the lateral diffusion of the complex at the membrane surface [96]. Several candidate arginine and lysine residues close to the PI(3,4,5)P_3 binding site may be responsible for sequestration of additional anionic lipids, though this possibility has not been explored here, partly due to the difficulties in observing large scale lipid motions in comparatively short atomistic simulations. The phenomenon of anionic lipid sequestration is discussed in more detail in Chapter 6, where CG MD simulations are used to investigate anionic lipid clustering around the C2 domain of PTEN.

Taken together, the results presented in this chapter provide clear evidence that nonspecific interactions play a role in binding the GRP1 PH domain to the membrane surface and sustaining the bound complex. This is consistent with the membrane insertion behaviours displayed by other lipid-binding modules such as PX domains [12], which inserts a hydrophobic loop into the membrane to anchor itself to the surface. Membrane insertion of the PH domain of PLCδ1 has also been detected in solid state NMR [85] and monolayer penetration experiments [242], though the authors of the latter paper suggested that this may be because this domain exhibits different membrane binding modes depending upon the composition of the lipid bilayer.

The work described in this chapter suggests a dual recognition mechanism of interaction whereby the protrusion of the PI(3,4,5)P_3 headgroup from the bilayer surface facilitates initial interactions between the PH domain and the target membrane, which are then reinforced through 'deeper', nonspecific contacts with the adjacent bilayer. This is in keeping with the behaviour of the PH domain from PLCδ1 in simulations of the pro-

tein bound to the I(1,4,5)P_3 headgroup and bound to PI(4,5)P_2 in lipid bilayers. Here the PH domain bound to the ligand more tightly in the presence of a lipid bilayer, suggesting that the non-specific interactions with bilayer lipids reinforced the specific interaction between PLCδ1 and PI(4,5)P_2 [114].

The mutation of residues within the PI(3,4,5)P_3 binding site has been shown to considerably reduce the binding affinity of GRP1-PH [232, 98] for its phosphoinositide ligand. However, the non-specific interactions with the bilayer lipids observed in this chapter suggest that the bilayer may play an important role in stabilising the bound GRP1-PH-PI(3,4,5)P_3 complex. This, coupled with the observed reduction in membrane penetration in the mutants described above, raises the possibility that mutations in other regions could also lead to a reduction in binding affinity and a corresponding loss of function. The idea that mutations may perturb the interaction of a protein with a lipid bilayer is explored further in Chapter 6.

In summary, membrane penetration and hydrophobic interactions appear to be important in sustaining the GRP1-PH-PI(3,4,5)P_3 complex at the membrane surface. However, the initial diffusional encounter between GRP1-PH and the membrane is thought to be governed primarily by electrostatic effects, and the influence of electrostatics on this initial association event is the subject of the following chapter.

Chapter 4

Electrostatic mechanism of target lipid recognition by GRP1-PH

In the previous chapter, MD simulations were combined with biophysical methods to investigate the protein-lipid interactions in the membrane-bound state of GRP1-PH. Here, the scope of this work is expanded to consider the initial encounter between GRP1-PH and the membrane, which occurs prior to complex formation at the lipid bilayer surface. The role of electrostatics and lipid composition in successful target selection, membrane association and complex formation is explored using a multiscale approach, combining BD simulations with subsequent MD simulations. This study has been published in *PLoS Computational Biology* [243].

4.1 Introduction

The PH domain of GRP1 binds to PI(3,4,5)P_3 with high affinity and specificity, and binding is likely to be achieved through electrostatic interactions between the positively charged GRP1-PH binding site and the negatively charged PI(3,4,5)P_3 headgroup (Figure 4.1). Aside from the specific interaction between GRP1-PH and PI(3,4,5)P_3, there also exists the potential for a non-specific interaction with monovalent anionic lipids present in the bilayer [244]. A putative membrane targeting mechanism proposed by Falke and co-workers [238], among others, suggests that the protein first associates with the membrane surface *via* weak nonspecific interactions with background anionic lipids. This transient binding increases the residence time of GRP1-PH at the membrane surface, and subsequent exploration of the surface by lateral diffusion in two dimensions allows GRP1-PH to locate its target lipid PI(3,4,5)P_3. Diffusion in three dimensions in the crowded environment of the cell is slow [245]: for example, when measuring values of the diffusion coefficient of GFP, Konopka *et al.* found that the *in vivo* value in the *E. coli* cytoplasm is only around 7% of its *in vitro* value [246]. However, facilitated diffusion at the membrane surface is thought to increase the on-rate, as the protein need now only execute a two dimensional search for its target. This reduction in dimensionality is thought to dramatically speed up the rate of binding [247, 248]. However, the proposed link between reduced dimensionality and kinetic enhancement is somewhat controversial, with several authors instead attributing the increased reaction rates to local concentration effects arising from spatial confinement of the proteins in a small volume adjacent to the membrane [249, 250].

Work in the previous chapter used atomistic MD simulations to look at the detailed interactions of the membrane-bound GRP1-PH-PI(3,4,5)P_3 complex. However, as protein-membrane association is thought to be governed primarily by electrostatic interactions [252], here BD simulations have been used to investigate the binding of GRP1-PH to a

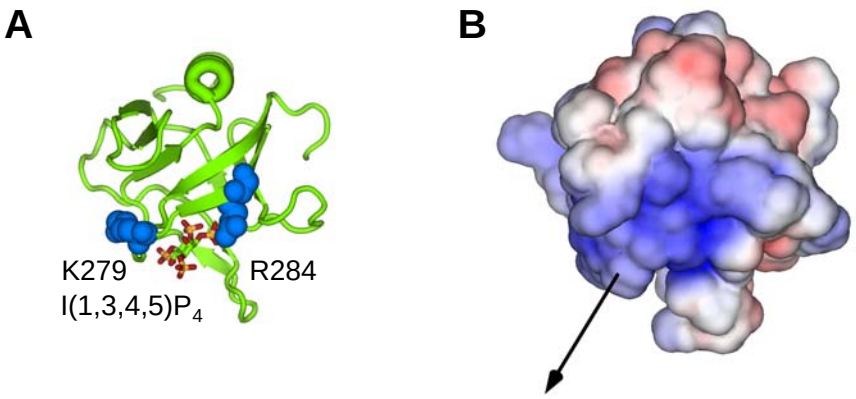

Figure 4.1: (A) Crystal structure of GRP1-PH (PDB 1FGY) [4] bound to I(1,3,4,5)P_4. The I(1,3,4,5)P_4 headgroup is shown as a stick model and the sidechains of two basic residues (K279 and R284) are displayed as blue van der Waals spheres. **(B)** Electrostatic potential of GRP1-PH calculated using APBS [210], with areas of positive and negative potential coloured blue and red respectively at a contour level of $\pm$ 5 *kT/e* and projected onto the solvent-accessible surface. The molecular dipole moment is shown by a black arrow, and was calculated using the Protein Dipole Moments Server (`http://bioinfo.weizmann.ac.il/dipol/`) [251]. By convention, the arrow indicating the direction of the dipole moment points from negative to positive. GRP1-PH is shown in the same orientation in both cases.

PI(3,4,5)P_3-containing membrane.

For electrostatically facilitated processes such as the binding of barnase and barstar [253], BD simulation methods have been able to accurately reproduce experimental association rates in aqueous solution [190, 254], and they have been used to model a variety of processes ranging from enzyme-substrate encounters [255] to protein folding within the crowded environment presented by the bacterial cytoplasm [256]. BD simulations have also been successfully applied to more complex situations involving, for example, electron transfer between associating proteins [257] and membrane proteins [258], in particular the association of cytochrome *c* with membrane-bound cytochrome *c* oxidase [259, 260]. There have also been studies of multiple proteins binding to charged membranes using simplified representations of cytochrome *c* [261] or lysozyme [262].

Association events in solution are typically thought of as being two-step processes, with the diffusing solute molecules first interacting through electrostatic interactions at long range, approaching closely and then forming an initial encounter complex. The second step involves the relaxation and conformational rearrangement of the two solutes

into the final bound complex [263]. Protein-membrane association might therefore be expected to proceed in a similar fashion.

One of the limitations of BD simulations run using SDA is that the simulations only include the rigid-body rotations and translations of the diffusing molecules, with no explict solvent molecules or conformational dynamics. While BD simulations are therefore useful to model the initial association event and generate encounter complexes, at close range conformational flexibility not captured in the BD simulations becomes important in forming the bound complex. A multiscale simulation strategy has therefore been adopted, where BD simulations are employed to model the initial encounter between the protein and the membrane, before subsequently switching to atomistic MD simulations to model the formation of the membrane-bound GRP1-PH-PI(3,4,5)P_3 complex.

The aim of this chapter is to study the membrane targeting mechanism of GRP1-PH. Of particular interest is how, through electrostatic competition between target (PI(3,4,5)P_3) and decoy (anionic) lipids, the presence of the anionic background may help or hinder target selection by luring GRP1-PH towards or away from PI(3,4,5)P_3. This is a specific example of the more general problem of encounter and recognition within the crowded environment of the interior of the cell [264, 265, 266, 267].

4.2 Methods

4.2.1 Lipid bilayer models

The model of a PI(3,4,5)P_3-containing membrane was taken from previous atomistic MD simulations [78] (see also Chapter 3), replicated in the xy plane to generate a square bilayer patch of approximately 1560 POPC lipids with a single PI(3,4,5)P_3 molecule in its centre. The final dimensions of the bilayer were approximately 220 Å $\times$ 220 Å $\times$ 40 Å. Atomic partial charges on the lipids were identical to those used previously [78].

In the first instance, BD simulations were performed with a zwitterionic POPC mem-

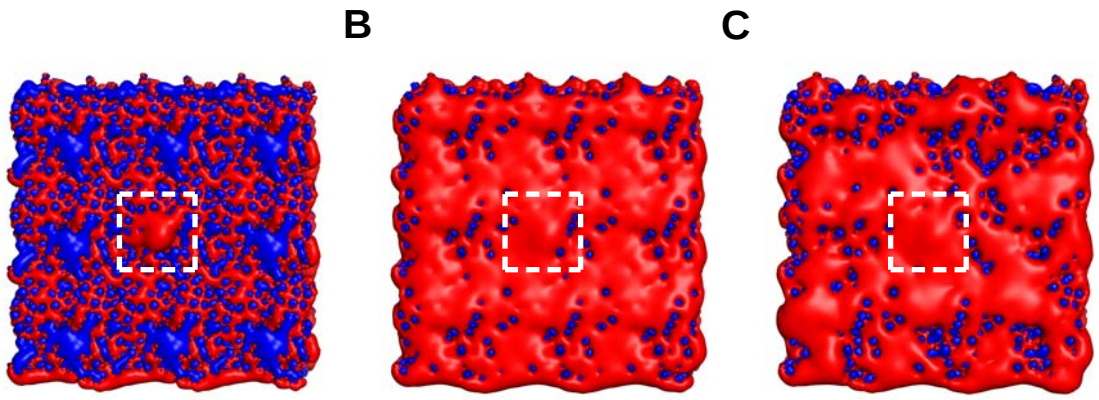

Figure 4.2: Lipid bilayer models. **(A)** Electrostatic potential of a pure POPC bilayer projected onto the solvent-accessible surface. The large negative charge associated with the PI(3,4,5)P_3 molecule (red) is visible in the centre and the periodicity of the system is also evident. **(B)** Electrostatic potential for the case where each lipid is assigned a fractional charge of -0.4 *e*, with the negatively charged PI(3,4,5)P_3 molecule now effectively masked by the background charge distribution. **(C)** Electrostatic potential where 40% of the lipids are assigned a charge of -1.0 *e*, showing the heterogeneous nature of the charge distribution. In all cases electrostatic potential isocontour surfaces are shown at -1kT/e (red) and +1kT/e (blue).

brane, with only the PI(3,4,5)P_3 molecule carrying a net negative charge (Figure 4.2). To mimic the presence of anionic lipids in BD simulations of charged membranes, negative charges were assigned to the PC lipid headgroups. Initially, identical, fractional negative charges were placed on all of the nitrogen atoms to generate an even charge distribution across the surface (Figure 4.2). In this case, the diffusing protein effectively interacts with the average charge distribution of the lipid bilayer.

The topography of the electrostatic potential at the membrane surface is dependent upon the distribution of the lipids. The lipid bilayer is dynamic and lipids undergo two dimensional lateral diffusion in the plane of the membrane, and so the charge distribution is likely to fluctuate over time. The lateral diffusion constant of POPC lipids at 300 K is 1.7×10^{-8} cm^2 s^{-1} [268]. However, the calculated diffusion constant of GRP1-PH at 300 K using HYDROPRO [217] is 1.0×10^{-6} cm^2 s^{-1}, almost two orders of magnitude larger[1].

[1]Care must be taken when attempting to directly compare three dimensional diffusion in bulk solvent with two dimensional lateral diffusion in a viscous lipid bilayer. For a protein diffusing in two dimensions in a viscous membrane, diffusion is not completely free as the membrane is bounded by no-slip conditions at the solvent interface. This induces finite size effects and means that there is no exact solution for the viscous flow (Navier-Stokes) equations in two dimensions and hence no two dimensional equivalent of the translational Stokes-Einstein expression (Equation 2.68). Saffman and Delbrück [269] developed an approximation to quantify lateral diffusion within membranes, but recently deviations from the Saffman-Delbrück theory have been reported in both experimental [270] and simulation [271] studies. Corrections

This indicates that the lipids may be much less mobile than the protein. Applying the Einstein relations for the two dimensional case and the three dimensional case give mean square displacements of $\langle r^2 \rangle = 4Dt$ and $\langle r^2 \rangle = 6Dt$ respectively, so for a 50 Å diffusional motion this yields time scales of approximately 30 μs and 40 ns for the lipids and the protein respectively. The large discrepancy between the time scales for motion of the protein and the lipids suggests that the consequences of assuming the interaction of the protein with the average charge distribution of the lipid bilayer should be examined more thoroughly.

To explore this further, an alternative approach was adopted in which single negative charges were assigned to a subset of the headgroups selected randomly (Figure 4.2), thus generating a less even distribution of decoy negative charges. (Pseudo)random numbers were generated using the implementation of the Mersenne twister algorithm in Perl [274]. In this case, the protein interacts with a discrete distribution of charges, which is more in keeping with faster diffusion of the protein relative to the lipids. The fractional charges explored in the first set of simulations ranged from -0.2 *e* to -1.0 *e*, while the subset of lipids randomly assigned a charge of -1.0 *e* ranged from 20% to 100% in the second approach.

Estimates of the concentration of anionic lipids in the cell membrane vary widely, since anionic lipids are distributed asymmetrically across the cytoplasmic and extracellular leaflets of the lipid bilayer, with the majority of the anionic lipids being found in the former [10]. Typical physiological values can range from 10-20% depending on the cell type, so the range of anionic lipid concentrations used in the BD simulations is perhaps larger than would be found *in vivo*. However it is possible to synthesise vesicles and bilay-

to the Saffman-Delbrück model have been proposed [272], with the discrepancies linked to the propensity of transmembrane proteins to induce membrane deformation. However, it is not immediately clear how the mobility of a protein tethered at a bilayer surface should be treated, as the overall diffusional behaviour of a (peripheral) protein-lipid complex would presumably be influenced by the two dimensional diffusion of the protein in bulk solvent coupled with the two dimensional diffusion of the lipid in the membrane. An interesting paper by Honig and colleagues [273] considered a similar problem on the dimensionality of binding affinities, and outlined a scheme to transform such affinities from three dimensions to two.

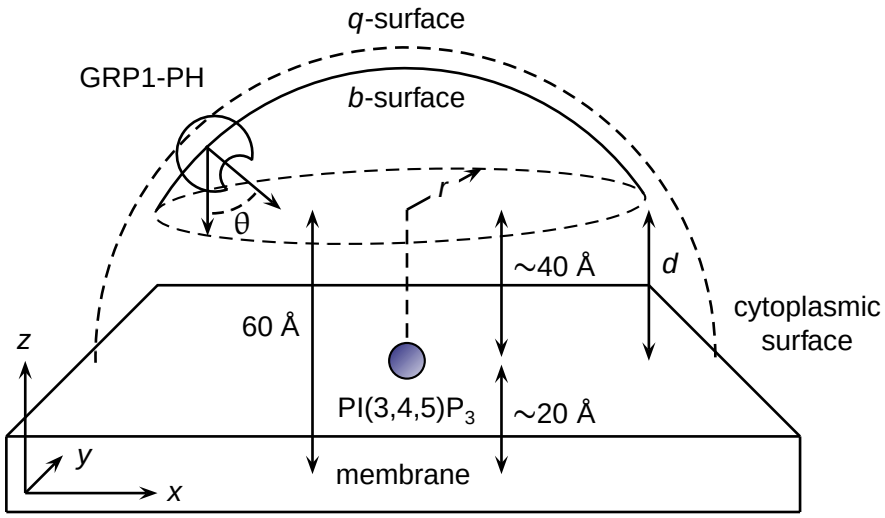

Figure 4.3: Schematic of BD simulation setup, showing the (r, d, θ) coordinate system and illustrating how the spherical b- and q-surfaces are truncated to hemispheres to address the system geometry and the lipid asymmetry. See also Figure 2.7.

ers comprised wholly of anionic lipids: for example, 100% POPS [275, 276, 41] and so the results from the simulations conducted at artificially high anionic lipid concentrations may nonetheless be relevant for *in vitro* studies.

Another alternative method for exploring slower lipid diffusion is to generate lipid configurations using CG MD simulations of a lipid bilayer and then extract representative snapshots for use in the BD simulations. This has an advantage over the simple random assignment described above in that it is able to explicitly account for more complex lipid bilayer phenomena such as lipid demixing.

4.2.2 BD simulations

The SDA source code was modified to truncate the b-sphere, such that all trajectories begin on the hemispherical open surface given by $r^2 = x^2 + y^2 + z^2$, $z > 60$ Å, ensuring that the protein always lies at least 40 Å distant from the surface of the bilayer at the start of the trajectory (Figure 4.3). Rotational diffusion of the lipid bilayer was switched off. The Debye length of the system at an ionic strength of 0.1 M is approximately 10 Å,

and accordingly the radius of the b-surface was chosen to be 100 Å. The bilayer patch was approximately 220 Å wide, and so the q-surface was set to 105 Å. This resulted in the termination of any trajectory that came close to the edge of the bilayer, thus limiting edge effects. This choice of the q-surface also avoided the unphysical situation whereby the protein could in principle diffuse around the periphery of the bilayer to the extracellular side during a trajectory. This was also important as the lipid composition of the cytoplasmic and extracellular leaflets of the lipid bilayer is substantially different, with a much lower abundance of anionic lipids found in the latter.

The truncation of the b-sphere coupled with the choice of the q-surface is likely to invalidate the NAM method for computing reaction rates, since the ensemble reactive flux is no longer isotropic and spherically symmetric. As discussed in Section 2.2.3, computing the rate constants using the NAM method relies upon partitioning the diffusion space into b and q regions such that the q region is far enough removed from the reference protein that the intermolecular forces are centrosymmetric at this separation. However, here it is necessary to set q to a sufficiently low value so that any trajectories that reach the perimeter of the bilayer are terminated. Increasing q to a large enough value to make intermolecular forces truly centrosymmetric at that separation would require an extremely large bilayer to compensate. Aside from these practical considerations associated with setting the parameter q, there are more fundamental issues with respect to the calculation of rates. The hemispherical geometry of the system, with the membrane acting as an additional boundary condition, makes it difficult to envision the precise form of the analytic correction factor that would be required.

The focus of this study was to investigate the behaviour of the protein as it explored the bilayer surface, as opposed to explicitly calculating reaction rates. As a result of this, and the difficulties imposed by the unusual system geometry described above, the NAM method was not used here. Admittedly, computing reaction rates would have allowed for more extensive validation of the simulations, as for example it may then have been

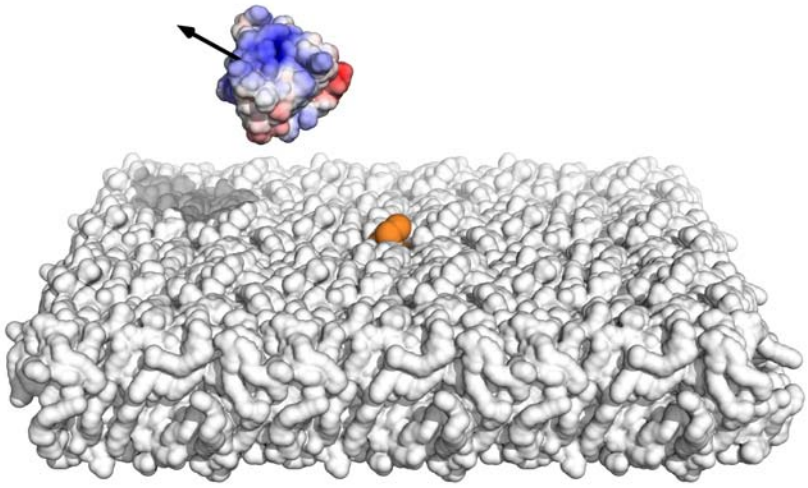

Figure 4.4: Simulation snapshot from a typical BD trajectory, showing the protein diffusing relatively close to the surface of the membrane. The lipid bilayer is depicted as a white solvent-accessible surface with the PI(3,4,5)P_3 molecule shown as orange van der Waals spheres at the centre. GRP1-PH appears in the same representation as in Figure 4.1.

possible to compare the calculated rates with experimentally determined on-rates.

An ensemble of 5000 BD simulations was performed for each system, with a typical snapshot shown in Figure 4.4. All simulations started with the protein centre randomly positioned on the hemispherical open surface and the protein randomly oriented relative to the model membrane (Figure 4.3).

4.2.3 Coordinate system

The position and orientation of GRP1-PH are specified by three coordinates (Figure 4.3): a radial coordinate, r, in the xy plane; an axial coordinate, d, along z; and an angular coordinate θ relative to the initial orientation of the bound complex from the manual docking of GRP1-PH to a PI(3,4,5)P_3-containing bilayer described in the previous chapter. It is helpful to distinguish between translational and orientational aspects of electrostatic steering, and so a distinction is made between radial translational steering along r, axial translational steering (the value of d along the z axis) and finally orientational steering arising from changes in θ. In this coordinate system, a combination of favourable

steering in all three cases is required to yield a 'productive' encounter complex. A counterexample could be a scenario in which GRP1-PH closely approaches the PI(3,4,5)P_3 ligand (*i.e.* good translational steering) but with its molecular dipole moment in the wrong orientation for binding (*i.e.* poor orientational steering). To use the terminology of Schreiber and co-workers, a distinction is therefore made between so-called 'fruitful' and 'futile' encounters [277].

The distributions of the position and orientation of GRP1-PH were evaluated for each time course from the ensemble of 5000 simulations and subsequently combined and normalised to yield an overall distribution.

4.2.4 Histogram construction

To visualise how the influence of membrane charge affects the distribution of GRP1-PH across the surface of the membrane, it is instructive to plot histograms of the radial position, r, the axial position, d, and orientation, θ, of GRP1-PH over the different BD trajectories. The difficulty associated with attempting to plot histograms of r is that the bin width between r and $r+dr$ is not constant, and so the area of the bin is proportional to r. This has the effect of overrepresenting the larger distances in the histogram. One way of remedying this is to reweight the data by some appropriate factor to compensate for this effect [278]. In plane polar coordinates the Jacobian factor is $2\pi r$, and so the data in each of the bins of the histogram is rescaled according to:

$$f(r_i) = \frac{1}{N}\left(\frac{A}{2\pi r_i}\right) p(r_i) \tag{4.1}$$

Similarly, the orientation of the protein can be represented by a unit vector rotating in space. Again, attempting to plot histograms of the distribution of angles that this vector makes with the z axis is problematic since the areas of the spherical segments between θ and $\theta + d\theta$ are not equal, which again distorts the distribution. In spherical polar coordinates the Jacobian factor is $r^2\sin(\theta)$, and so the histograms of the protein

orientation are rescaled using the following:

$$g(\theta_i) = \frac{1}{N}\left(\frac{p(\theta_i)}{\sin(\theta_i)}\right) \tag{4.2}$$

These problems are not encountered when plotting histograms of d, as the bin widths between z and $z + dz$ along the z axis remain constant.

4.2.5 Molecular dipole moment calculations

Molecular dipole moments were calculated using the Protein Dipole Moments Server (`http://bioinfo.weizmann.ac.il/dipol/`) [251]. The server calculates dipole moments by considering only the protein heavy atoms (C, N, O and S) and placing the geometrical centre at the average coordinates of these heavy atoms. After assignment of partial charges the dipole moment, **D**, in Debye is calculated as:

$$\mathbf{D} = 4.803 \sum_i \mathbf{r}_i q_i \tag{4.3}$$

The conversion factor 4.803 accounts for the change from Ångström electron charge units (Åe) to Debye. By convention, the arrow indicating the direction of the dipole moment points from negative to positive.

4.3 Radial and axial translational steering of GRP1-PH

Initially, radial and axial translational steering of GRP1-PH towards PI(3,4,5)P_3 were investigated to assess how easily GRP1-PH was able to locate its target lipid PI(3,4,5)P_3 when the surrounding membrane contained an increasing background density of negatively charged lipids. It might be expected that increasing this negative surface charge density would disrupt GRP1-PH targeting by masking the position of the negatively charged PI(3,4,5)P_3 (Figure 4.2).

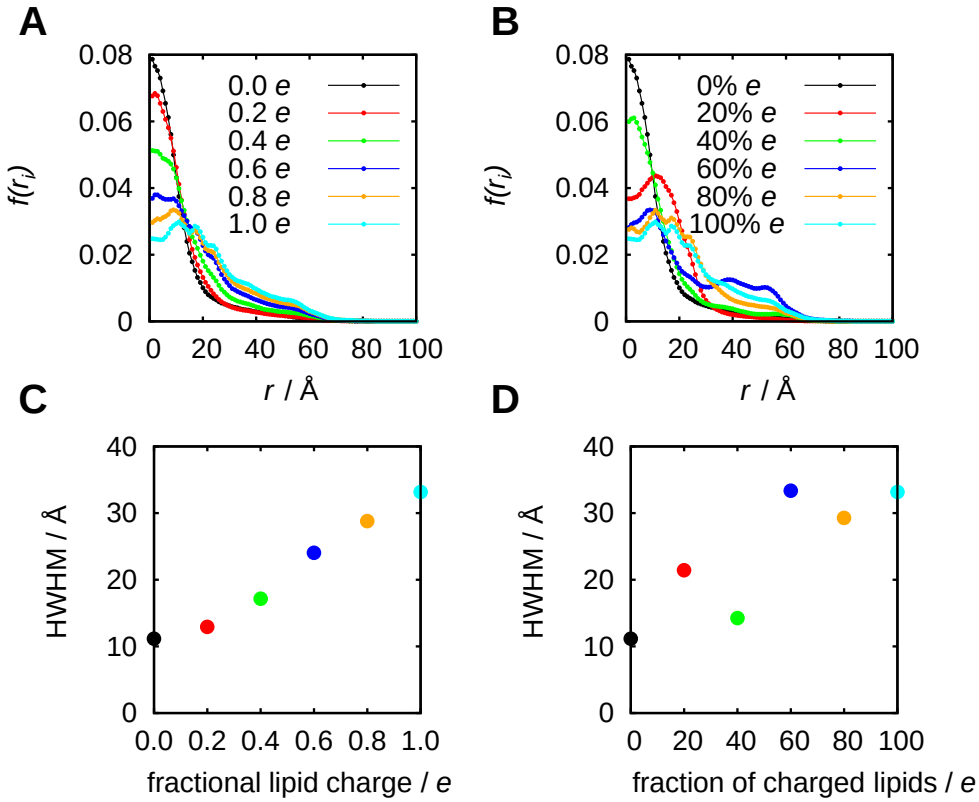

Figure 4.5: Radial translational steering of GRP1-PH for **(A)** the uniform distribution of fractional charges and **(B)** the non-uniform, random distribution of integer charges. In each case the distribution of radial positions, r, of GRP1-PH over the course of the 5000 BD simulations is shown for each ensemble. The centre of mass of the 'target' PI(3,4,5)P_3 headgroup is located at r = 0. **(C)** and **(D)** show the HWHM for the uniform fractional charges and the non-uniform, random integer charges respectively.

With only PI(3,4,5)P_3 present in a POPC bilayer (0.0 e) GRP1-PH spends the majority of the trajectory diffusing close to its target lipid, evidenced by the large peak at small values of r (Figure 4.5). However, when the surface charge is increased (-0.2 to -1.0 e) the peak height diminishes and the maximum shifts to larger values of r, which suggests a reduction in radial translational steering with the protein much less likely to closely approach the PI(3,4,5)P_3 molecule. To semi-quantitatively evaluate the degree of targeting, the peak half-width at half maximum (HWHM) was extracted by fitting the data to a Gaussian distribution centred at $r = 0$:

$$f(x) = a \exp\left(-\frac{x^2}{2c^2}\right) \tag{4.4}$$

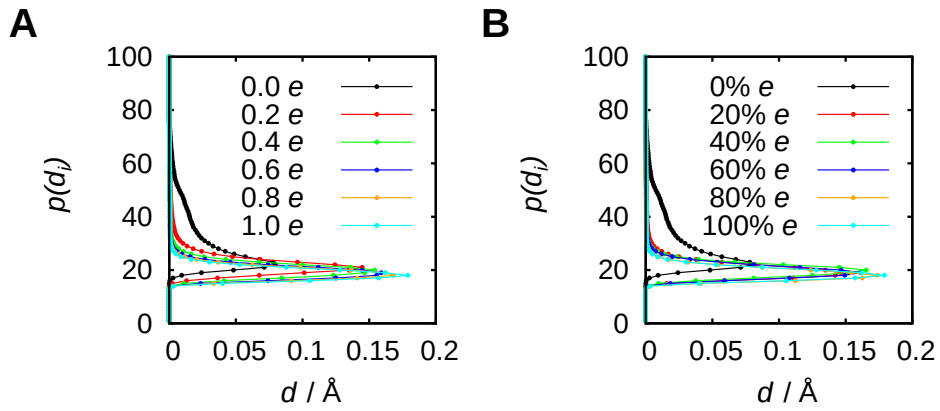

Figure 4.6: Axial translational steering of GRP1-PH for **(A)** the uniform distribution of fractional charges and **(B)** the non-uniform, random distribution of integer charges. Distributions of axial positions, d, along z for GRP1-PH are shown. The z coordinate of the centre of mass of PI(3,4,5)P_3 is at d = 0.

The HWHM is given by $\sqrt{2\ln 2}\,c$, and the resulting values are plotted in Figure 4.5. Low values correspond to narrow distributions of r and efficient PI(3,4,5)P_3 radial translational steering, while higher values correspond to wide distributions of r and comparatively poor PI(3,4,5)P_3 steering. The surface charge density, which in this case is uniform, is therefore able to modulate GRP1-PH steering with the HWHM monotonically increasing with rising charge density.

When the same set of simulations was repeated for the non-uniform charge distribution with randomly placed integer negative charges, the behaviour of GRP1-PH was somewhat different. The peak is no longer necessarily centred at $r = 0$, and shoulders in the peaks are generally more pronounced when compared with the uniform charge distribution. This is reflected in the HWHM, which now increases non-monotonically as the charge density increases. Radial translational steering therefore seems to be governed not only by the macroscopic net charge on the lipid bilayer, but also by microscopic influences arising from the distribution of charged lipids.

Axial translational steering of the protein also changes with membrane charge. Figure 4.6 shows the distribution of d positions of GRP1-PH along the z axis normal to the bilayer. As anticipated, with increasing surface charge density GRP1-PH spends more time closer to the surface of the bilayer. In contrast to the radial translational steering,

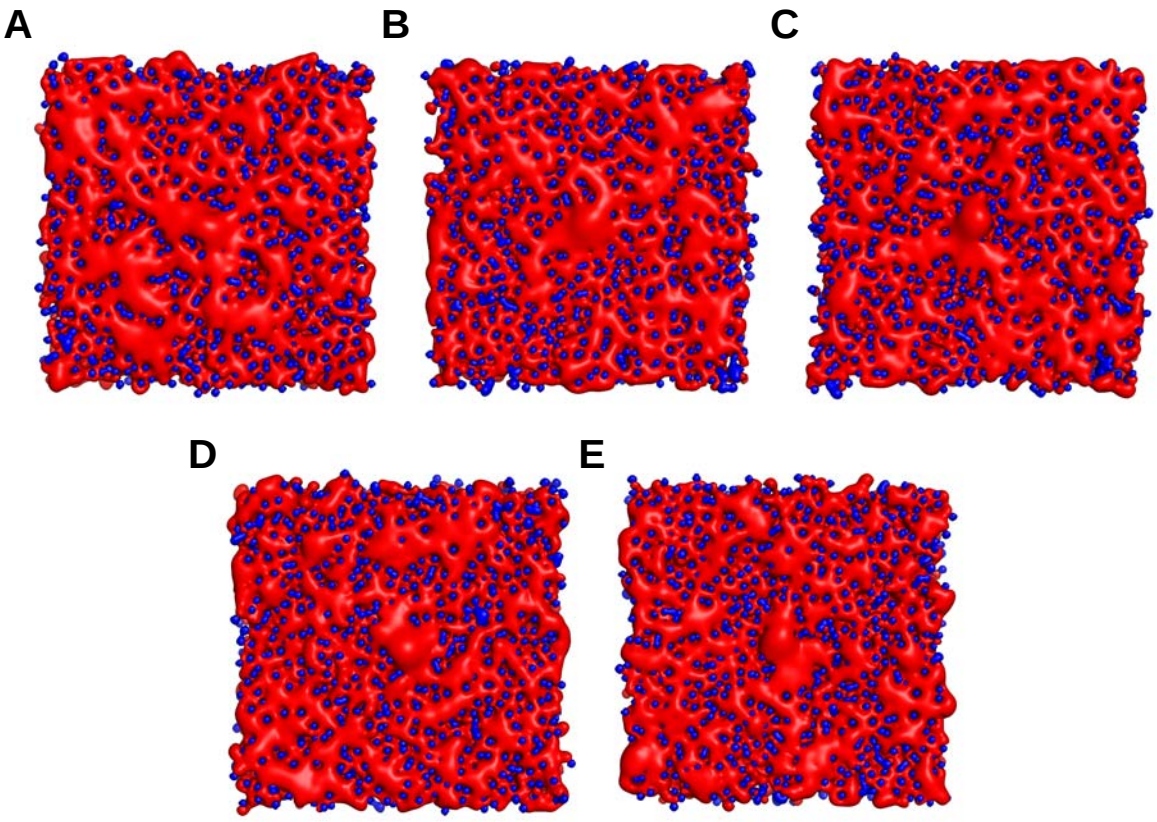

Figure 4.7: Lipid bilayer models based on CG MD simulations. Electrostatic potentials calculated from lipid distributions obtained from CGMD simulations of a bilayer containing 20% negatively charged lipids. Electrostatic potentials were calculated using snapshots taken from the simulation at **(A)** 100 ns, **(B)** 200 ns, **(C)** 300 ns, **(D)** 400 ns and **(E)** 500 ns. Electrostatic potential isocontour surfaces are shown at contour levels of -1 *kT/e* (red) and +1 *kT/e* (blue).

axial translational steering appears to be relatively insensitive to the charge distribution, and in this case the net charge on the bilayer seems to be the dominant effect. A -0.2 e surface charge on the cytoplasmic leaflet would correspond to an overall bilayer composition of 10% anionic lipids, close to that observed experimentally [11]. It is perhaps significant that, at this approximately physiological concentration of anionic lipids, GRP1-PH spends a greater proportion of its time diffusing close to the bilayer surface with lower values of d along z while largely retaining its ability to target PI(3,4,5)P_3 along r.

As the extent of radial translational steering behaviour observed above appeared to be dependent upon the fixed, instantaneous lipid configuration used in the BD simulations,

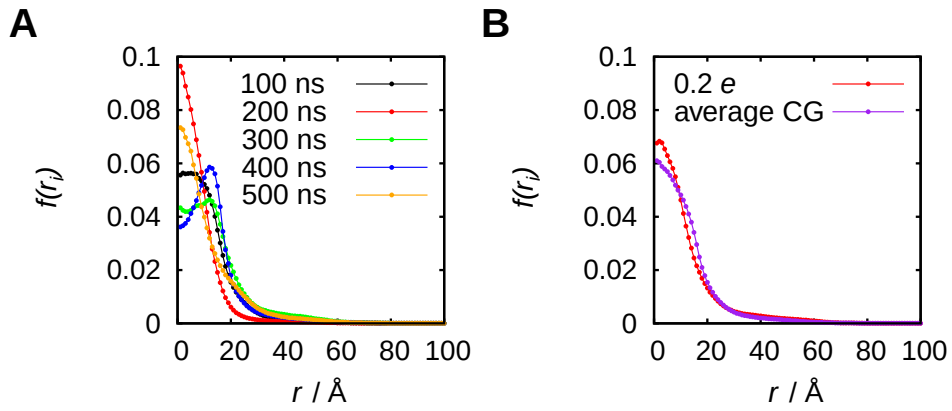

Figure 4.8: Radial translational steering using alternative lipid configurations generated by CG MD simulations. **(A)** Distributions in r for five lipid configurations using 20% anionic lipids extracted from CG MD simulations at intervals of 100 ns. **(B)** Average r profile (purple) for the five distributions shown in **(A)**, compared with the r profile from the BD simulations with each lipid assigned a fractional charge of 0.2 e (red).

lipid clustering may also modulate radial translational steering of GRP1-PH domains to PI(3,4,5)P_3. To probe the sensitivity of GRP1-PH targeting to the distribution of anionic lipids, a 0.5 μs CG MD simulation was performed of a mixed lipid bilayer containing 20% anionic lipids and configurations were extracted at intervals of 100 ns (Figure 4.7). After conversion to an atomistic representation using a fragment based approach [279], these five configurations were then used as an input for a set of BD simulations. Despite the fact that the anionic lipid concentration is the same in each snapshot, and the net charge is therefore also identical between snapshots, variations in the degree of radial translational steering are nonetheless observed for the different lipid configurations (Figure 4.8). This again suggests that steering is influenced not only by the concentration of anionic lipids but also by the distribution of these lipids over the surface. Interestingly, by taking the average of the r distributions over these five sets of BD simulations, a profile is generated similar to that observed for the single set of BD simulations using the fractional charge distribution where each lipid is assigned a fractional charge of -0.2 e (Figure 4.8). This suggests that the fractional charge distribution is a reasonably good model of the time averaged behaviour of the system, allowing for lipid dynamics on the sub-microsecond timescale.

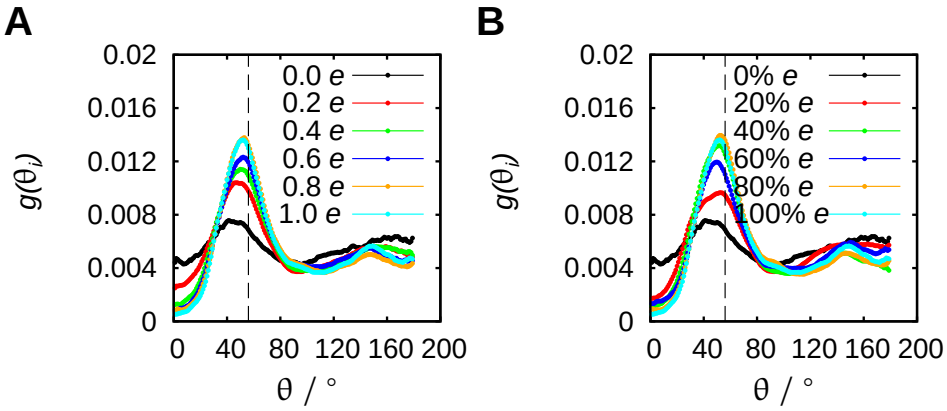

Figure 4.9: Orientational steering of GRP1-PH for **(A)** the uniform distribution of fractional charges and **(B)** the non-uniform, random distribution of integer charges. The dotted line at $\theta \approx 56^{\circ}$ shows the angle the molecular dipole moment makes with respect to the z axis in the membrane-bound complex.

4.4 Orientational steering of GRP1-PH

GRP1-PH carries a molecular dipole moment, and the vector is directed towards the binding cavity of the protein (Figure 4.1). In order to monitor the orientation of the protein over the BD trajectories the angle, θ, between an arbitrary unit vector on GRP1-PH and the z axis was computed over the ensemble of BD simulations. The reference value of θ, $\theta = 0^{\circ}$, corresponds to the protein orientation seen in the docked GRP1-PH-PI(3,4,5)P_3 complex observed in structural and MD simulation studies [78] discussed in the previous chapter.

The distribution of θ as a function of negative surface charge shows that a surface charge of -0.2 e or more has a substantial effect on the orientation of GRP1-PH (Figure 4.9). The distribution of orientations shifts towards values of θ corresponding to alignment of the GRP1-PH molecular dipole moment, which lies at an angle of $\sim 56^{\circ}$ to the reference vector, with the membrane normal (Figure 4.9). The membrane charge therefore appears to influence not only translational steering of the protein but also its orientation. Although increasing surface charge seems to promote orientational steering of GRP1-PH into a favourable orientation for binding, results discussed in the previous section indicate that this would be accompanied by some loss of radial translational steer-

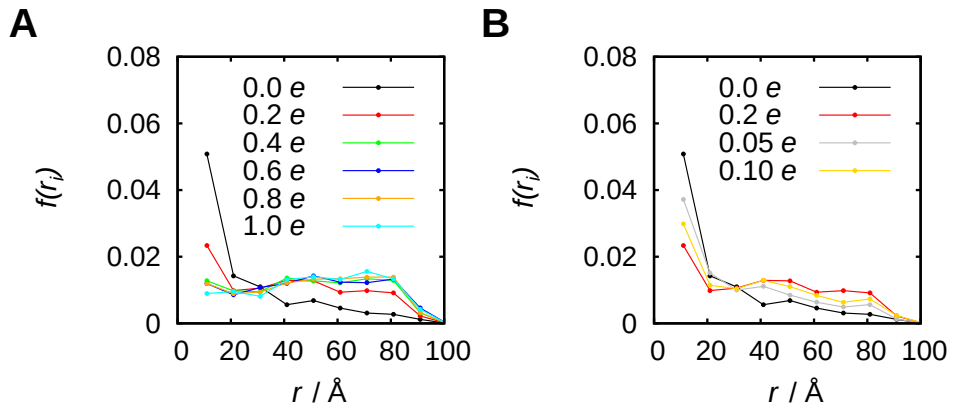

Figure 4.10: Radial first encounter positions of GRP1-PH. **(A)** First encounters of GRP1-PH with the surface of the bilayer over each ensemble of trajectories. The first encounter was defined as the first frame in which the *z* coordinate of GRP1-PH was less than 20 Å. **(B)** shows additional data for BD simulations using bilayer systems with lipids assigned fractional charges of 0.05 *e* and 0.10 *e*.

ing.

To explore the interplay between radial translational and orientational steering further, the distribution of first encounter positions between the bilayer and the PH domain was analysed (Figure 4.10). At -0.2 *e* there was clear radial translational steering of the PH domain to the PI(3,4,5)P_3 molecule. Increasing the surface charge beyond -0.2 *e* to - 0.4 *e* or above resulted in the complete loss of steering, as for values greater than -0.2 *e* the first encounter with the bilayer is typically with a non-PI(3,4,5)P_3 lipid rather than with the PI(3,4,5)P_3 molecule itself, suggesting that steering is disrupted at higher charge densities. This electrostatic competition between PI(3,4,5)P_3 and non-PI(3,4,5)P_3 lipids shows that there is a payoff between translational and orientational steering at higher charge densities. It therefore appears that optimal enhancement of steering occurs at uniform surface charge densities of 0.2 *e* with an overall membrane composition of approximately 10% anionic lipids, close to the physiological value.

Physiological levels of anionic lipids seem to be optimal to both enhance orientational steering and to localise GRP1-PH proximal to the surface of the bilayer without sacrificing its ability to locate PI(3,4,5)P_3 within the bilayer plane. The protein is steered into the correct orientation for binding by the higher surface charge density than would be the

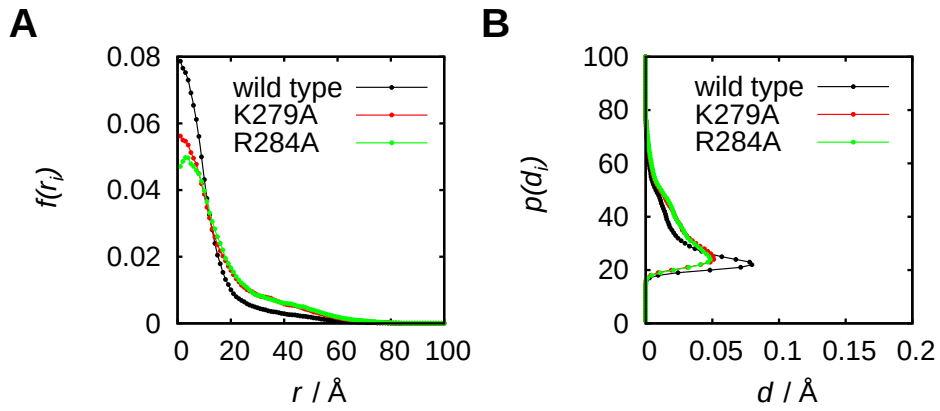

Figure 4.11: Effect of mutation on GRP1-PH translational steering. Distributions of **(A)** r and **(B)** d are shown for wild type GRP1-PH (black) and for two mutants, K279A (red) and R284A (green), using a purely zwitterionic bilayer with a single PI(3,4,5)P_3 molecule at r = 0, d = 0.

case for a zwitterionic membrane, but the charge density is not yet high enough to mask the position of PI(3,4,5)P_3.

4.5 Effect of mutation on target lipid recognition

BD simulations were performed for two mutants of GRP1-PH (R284A and K279A) which have previously been shown to reduce binding of the protein to *soluble* inositol phosphates [232]. It was therefore of interest to see whether or not these mutations also influenced steering of the PH domain to PI(3,4,5)P_3 in a lipid bilayer.

The mutant R284A has been shown experimentally to almost completely abolish binding, such that the ratio $K_D^{\mathrm{wt}}/K_D^{\mathrm{mut}}$ is virtually zero [232]. In the same study, the K279A mutant also affected the binding affinity, though less dramatically, reducing the ratio $K_D^{\mathrm{wt}}/K_D^{\mathrm{mut}}$ to approximately 0.5. To test whether or not these experimentally observed reductions in binding affinity could be detected in the simulations, both of these charged → nonpolar mutations were performed *in silico* and BD simulations of both mutants were subsequently carried out.

Both mutants show a modest reduction in radial translational steering (Figure 4.11), with the effect more pronounced in the case of the R284A mutant, in agreement with the

experimental results. The mutants also appear to spend a shorter period at the membrane surface relative to the wild type, which again would contribute to the observed reduction in the ratio $K_D^{\mathrm{wt}}/K_D^{\mathrm{mut}}$. Although complete abolition of binding is observed experimentally for the R284A mutant, from the simulations it is only possible to detect a somewhat weak effect, with slightly reduced radial and axial translational steering. The BD simulations are not able to capture the detailed sidechain specific protein lipid interactions present in the docking complex, or indeed any conformational changes as a result of the mutation. It is likely that the overall abolition of binding is as a result of a combination of several effects, not all of which are taken into account in the BD simulations. Despite the qualitative nature of the results for the two mutants, the fact that the simulations are able to distinguish between the mutants is encouraging, indicating that the observed reduction in steering is not just as a result of reducing the overall net charge on the protein.

4.6 Subsequent MD simulations of bound complex formation

In order to implement a multiscale approach to simulating membrane binding of GRP1-PH, BD simulations were combined with subsequent MD simulations. Similar combined approaches have been successful in studying, for example, DNA-enzyme interactions [280], in which swarms of BD simulations were used to map out the protein-DNA encounter pathways and the protein search pattern along the DNA. MD simulations were then used to model the binding interaction. In contrast to this 'serial' approach, other authors have utilised a 'parallel' BD-MD approach, such as that employed by Marcos *et al.* [281]. They used a parallel BD-MD method in which they performed MD simulations to study the internal motions of the protein, a dehydrogenase enzyme, while simultaneously running separate rigid body BD simulations of a crowded protein solution to investigate the bulk diffusive motion. They then generated hybrid BD-MD trajectories by mapping the internal motions of the protein onto the bulk rotational diffusion and the translational diffusion of the centre of mass (such that $\mathbf{r}^{\mathrm{m}}(t) = \mathbf{r}_{\mathrm{MD}}(t) \cdot \mathbf{R}^{\mathrm{m}}_{\mathrm{rot}}(t) + \mathbf{R}^{\mathrm{m}}_{\mathrm{cofm}}(t)$).

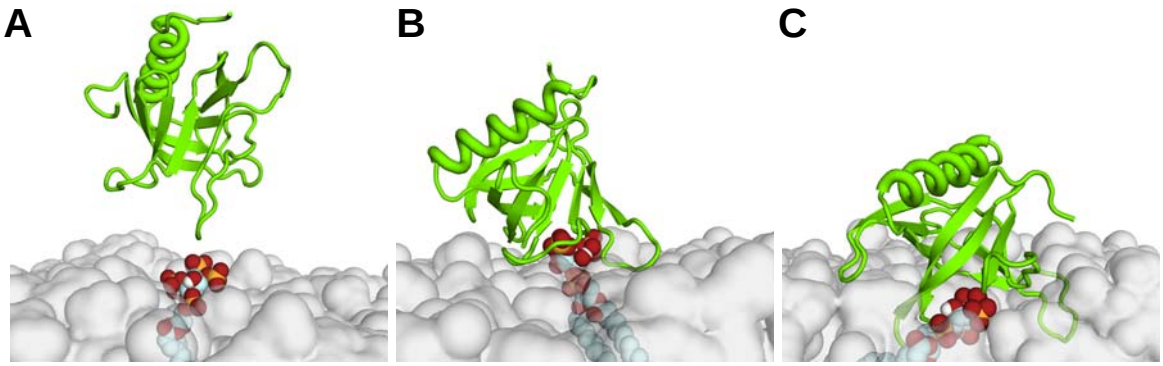

Figure 4.12: MD simulations of bound complex formation. **(A)** shows an encounter complex extracted from the BD simulations (r = 9 Å, d = 18 Å and θ = 27°). **(B)** shows the bound complex formed after 100 ns of MD simulation, and for comparison **(C)** shows the complex generated after 100 ns of MD simulation using the manually docked complex discussed in Chapter 3 (Figure 3.1) as a starting point. See also Figure 4.14.

Optimal encounter complexes from the BD simulations, in which beneficial translational and orientational steering were observed, were used as initial configurations for atomistic MD simulations to explore the conformational changes involved in complex formation. A suitable configuration for binding is likely to be one with small values of r, d and θ simultaneously, resulting in GRP1-PH being in close proximity to PI(3,4,5)P_3 with its binding cavity and also its molecular dipole moment oriented towards the ligand. A simple search of the trajectories was performed in order to locate a configuration satisfying these requirements. One such optimal configuration ($r = 9$ Å, $d = 18$ Å, $\theta = 27°$) was extracted and two 20 ns MD simulations were carried out to test whether or not the protein was able to bind to its target lipid from this position (Figure 4.12). In both MD simulations GRP1-PH rapidly approached the PI(3,4,5)P_3 molecule in the lipid bilayer and bound to the lipid, with the separation between the centre of mass of the protein and that of the I(1,3,4,5)P_4 headgroup falling to 15 Å in both cases (Figure 4.13). This is in good agreement with the centre-to-centre separation of 13 Å found in the ligand-bound crystal structure (PDB 1FGY). One of these simulations was extended to 100 ns to ensure that the complex was stable, and the centre-to-centre separation remained constant throughout the simulation.

To investigate the geometry of the complex, the minimum distances between each

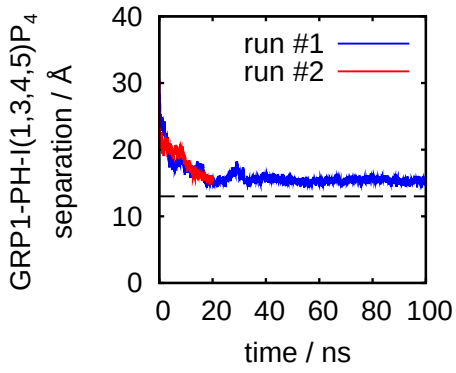

Figure 4.13: GRP1-PH-I(1,3,4,5)P_4 separation after conversion from BD to atomistic MD. GRP1-PH rapidly binds to PI(3,4,5)P_3 with a characteristic GRP1-PH-I(1,3,4,5)P_4 separation of ∼15 Å, in good agreement with the value of 13 Å from the X-ray crystal structure (PDB 1FGY). This occurs in two independent MD simulations (blue and red) and the complex remained stable when one of the simulations (blue) was extended to 100 ns duration.

amino acid residue of the protein and each of the phosphorus atoms of the I(1,3,4,5)P_4 headgroup in the crystal structure were mapped. This revealed a characteristic protein-ligand interaction 'fingerprint' (Figure 4.14). After the extended 100 ns MD simulation, when GRP1-PH has located the membrane-bound PI(3,4,5)P_3 within the first 20 ns of simulation, it binds *via* a set of amino acid residues similar to that found in the crystal structure. Performing the same fingerprinting analysis on the final frame of the 100 ns simulation showed a protein-ligand interaction fingerprint that agreed well with that observed in the crystal structure (Figure 4.14). As a further test, the same analysis using the final frame of the manually docked complex derived from one of the 100 ns simulations of GRP1-PH-PI(3,4,5)P_3 in Chapter 3 converges on the same interaction pattern (Figure 4.14).

Using appropriate encounter complexes from the BD simulations as initial configurations for atomistically detailed MD simulations, which include explicit solvent molecules and intermolecular motions, leads to formation of a GRP1-PI(3,4,5)P_3 complex at the membrane surface that accurately reproduces the geometry of the bound complex from the crystal structure as well as that generated from the manually docked MD simulations from Chapter 3. This combined BD-MD technique may be transferable, providing a means to model the membrane binding modes of lipid recognition proteins involved in

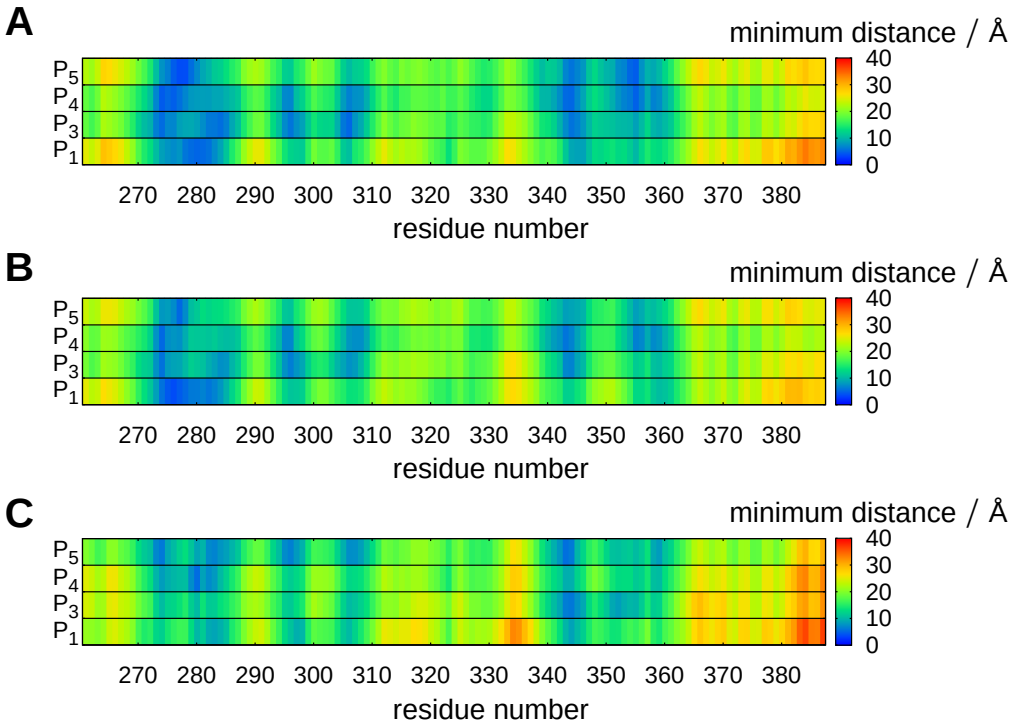

Figure 4.14: Protein-ligand interaction fingerprints in the bound complex. **(A)** Interaction fingerprint in the crystal structure (PDB 1FGY) [4], showing the minimum distances between the centres of mass of each amino acid residue and each phosphorus atom around the inositol headgroup. Interaction fingerprints are shown for the bound complexes generated by **(B)** BD simulation and a subsequent 100 ns MD simulation and **(C)** manual docking followed by a 100 ns MD simulation as described in Chapter 3 (see also Figure 4.12).

membrane function [12] and disease [282].

4.7 Molecular dipole moments of PH domains

As an aside, the molecular dipole moment vector of GRP1-PH points from the centre of mass of the protein towards the location of the bound PIP headgroup in the X-ray crystal structure. This, coupled with the observation that increasing surface charge leads to enhanced alignment of the dipole moment with the membrane normal, suggests that the orientation of the molecular dipole moment may play an important role in successful PH domain targeting. Evaluation of the molecular dipole moments for a variety of other PH domains suggests that this orientation is a conserved feature of PH domains (Figure 4.15). Thus, it is likely that the GRP1-PH targeting behaviour observed here may be conserved across the PH domain family.

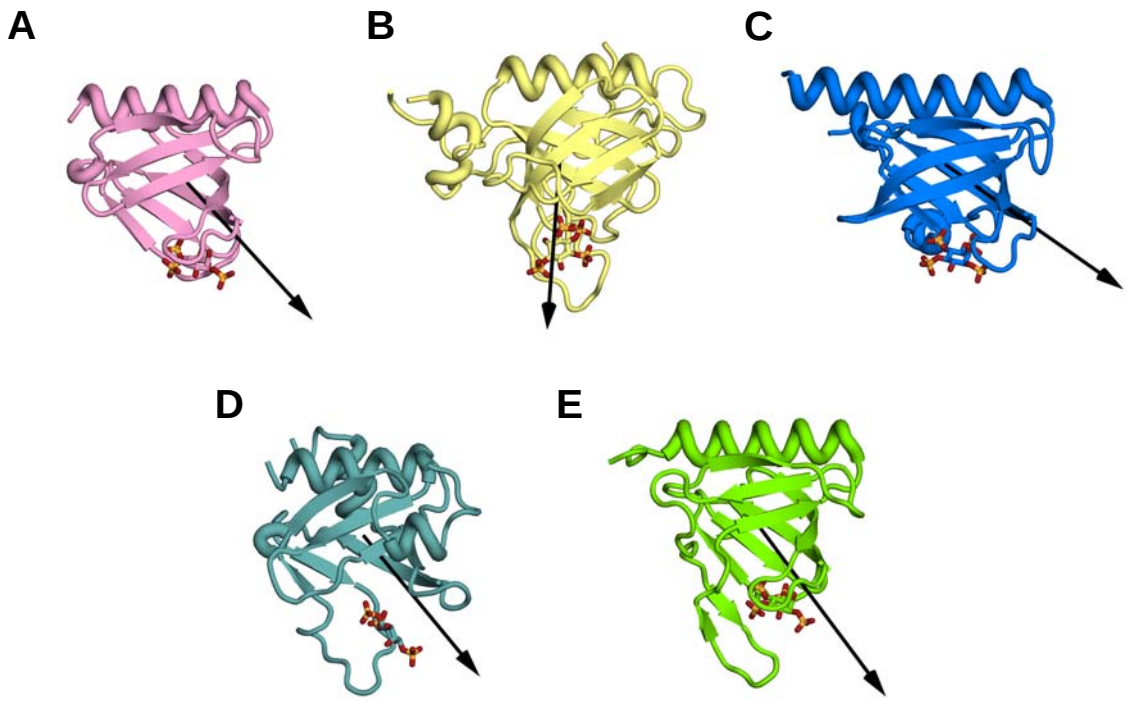

Figure 4.15: Molecular dipole moments for the *holo* forms of a selection of PH domains. In all cases the molecular dipole moment was calculated in the absence of the ligand using the Protein Dipole Moments Server [251] and is shown by a black arrow. PH domains from the following proteins are shown **(A)** DAPP1 (PDB 1FAO) [233]; **(B)** BTK (PDB 1B55) [283]; **(C)** PKB/Akt (PDB 1H10) [284]; **(D)** PLCδ1 (PDB 1MAI) [285] and finally **(E)** GRP1 (PDB 1FGY) [4].

This concept bears some similarities to the mechanism of electrostatic funnelling in acetylcholinesterase (AChE) proposed by Sussman and co-workers [286]. In AChE the dipole is precisely aligned with the axis of a gorge running through the centre of the protein, and the active site of the enzyme lies at the bottom of this gorge. The dipole controls substrate ingress by drawing positively charged substrates to the mouth of the gorge and channelling them along it towards the active site. The role of the molecular dipole moment in determining the location of ligand binding by PH domains will be discussed in more detail in the next chapter, which investigates the PIP binding properties of the previously uncharacterised PH domain from kindlin-1.

4.8 Discussion

4.8.1 Membrane complexity in physiological systems

The test system used in this study, with a $PI(3,4,5)P_3$ to lipid ratio of approximately 1:1000, is likely to be globally representative of mammalian cell plasma membranes. However, it is difficult to estimate the typical physiological concentration of $PI(3,4,5)P_3$, which varies according to the level of cell stimulation. $PI(3,4,5)P_3$ is generated from $PI(4,5)P_2$, and the concentration of the latter is approximately 1% of the total cell membrane lipids [9]. Assuming that even at the peak of cell stimulation the concentration of $PI(3,4,5)P_3$ will be less than 1%, then the $PI(3,4,5)P_3$ concentration present in the simulations is of the correct order of magnitude. This is perhaps something of a simplification given the importance of localisation and gradients of $PI(3,4,5)P_3$ in cell signalling and dynamics [287]. The differences in PIP composition between the apical and basal regions of plasma membranes, which have higher concentrations of $PI(4,5)P_2$ and $PI(3,4,5)P_3$ respectively in epithelial cells [288], need also to be considered.

The work described in this chapter, which uses a comparatively simple membrane model, indicates that lipid composition can have a substantial influence on the ability of GRP1-PH to locate its cognate ligand $PI(3,4,5)P_3$. From a biological perspective, perhaps the main limitation of this work is that the simulations mimic *in vitro* biophysical studies using simplified bilayer compositions. Current lipidomics studies [11] are beginning to reveal the spatial and temporal complexities of membrane lipid composition within living cells. Physiological cell membranes typically incorporate not just PC and PS but can also include PE, phosphatidylglycerol (PG), cholesterol, ceramide and other PIP species such as $PI(4,5)P_2$, as well as many classes of integral membrane proteins (see also Chapter 1). How peripheral proteins are able to locate their target ligands *in vivo* within this highly dynamic and heterogeneous background remains an open question.

4.8.2 Lack of membrane fluidity

It is important to consider the limitations of the current model. The BD simulations treat the bilayer as a static entity lacking any internal dynamics. This is likely to be sufficient to capture longer range steering interactions, but at short range factors such as protein conformational changes and membrane fluidity can also have an influence on the binding process. For example, the elastic membrane is able to deform in order to accommodate the protein, and often proteins insert a hydrophobic anchor or an amphipathic helix into the membrane to attach themselves to the surface [289]. Some proteins are able to increase their binding affinity by inducing lipid demixing in order to sequester negatively charged lipids such as PI(4,5)P_2 and/or PS when bound [290, 291], since the fluidity of the membrane allows lipid molecules to undergo lateral diffusion in the plane of the membrane [292, 293, 294]. This phenomenon was recently demonstrated experimentally in studies of the interaction of syntaxin-1A with PI(4,5)P_2 [63]. STED microscopy was used to show that electrostatic interactions between PI(4,5)P_2 and the basic residues in the juxtamembrane region of syntaxin-1A can lead to co-localisation of PI(4,5)P_2 microdomains and syntaxin-1A clusters at the membrane surface.

The degree of lipid demixing, the extent of domain formation and hence the effect on the binding affinity of the protein is therefore dependent upon the composition of the membrane [295]. Indeed, in simulation studies that utilised a flexible polyelectrolyte as a model for an unstructured protein adsorbing to a membrane, several different authors found that adsorption to a fluid membrane with mobile surface charges was enhanced over the frozen equivalent [296, 297, 298]. Perhaps surprisingly, Dias *et al.* [296] and Fleck *et al.* [299] found that for sufficiently mobile and disordered charge distributions, a negatively charged polyion was capable of adsorbing to net neutral and even net negatively charged membranes. These effects can also result in the emergence of correlated diffusion between the lipids and the protein [300].

Several authors have also noted that the ability of the membrane to deform [301]

or the protrusion of lipid molecules from the surface [302] can be important in protein-membrane binding as this allows the membrane to 'embrace' the protein as it approaches the surface, altering the binding energetics. The ability of flexible membranes to 'wrap' around adsorbing particles has previously been shown in both theoretical treatments of colloidal particles binding to flexible membranes [303, 304, 305, 306, 307] and dissipative particle dynamics [308] and Monte Carlo [309] simulation studies.

4.8.3 Neglect of hydrodynamic effects

The BD simulations presented here also neglect any effect of hydrodynamic interactions on the association process. While several groups have illustrated the potential problems that may arise when neglecting hydrodynamic interactions in simulations of protein-protein association in solution [198, 199, 200], the simulations described in this chapter are concerned with a protein diffusing close to a surface. Work on colloidal particles near surfaces and interfaces has uncovered intriguing hydrodynamic behaviour in such systems: for example, the interaction between two like-charged spheres spheres near a hard wall can actually be *attractive*, which was shown to be the result of non-equilibrium hydrodynamic effects [310, 311]. The effect of a hard wall or a vapour-liquid interface has been explored theoretically and numerically by Nägele, Wajnryb and co-workers [312, 313, 314] among others, but the influence of a membrane - essentially a fluctuating, fluid-fluid interface - on the motion of approaching proteins remains an open question.

There are a variety of other competing physicochemical factors that influence protein-membrane association that are not fully taken into account here. As the protein closely approaches the membrane there is an electrostatic desolvation penalty, as solvent-exposed charged groups on the protein begin to experience the low dielectric constant of the membrane, and *vice versa*. This manifests itself as a short ranged repulsive interaction, the so-called Born repulsion [315]. The protein itself also loses bulk translational and

rotational degrees of freedom upon binding to the surface, with an associated entropic penalty [316]. Conversely, this close approach also displaces the ions localised in the electrical double layers, increasing the entropy of the system and driving association [317, 318, 319], while another factor favoring association is the minimisation of exposed hydrophobic surface area [320].

Notwithstanding these limitations, the modelling work presented in this chapter has clarified some of the mechanistic details of protein-lipid association, and in particular how the ability of the protein to locate its target lipid is modulated by increasing background charge. Strikingly, the results suggest that an approximately physiological composition of anionic lipids is optimal in simultaneously enhancing both translational and orientational steering of the protein to its target lipid. It remains to be seen how the introduction of greater membrane complexity may influence this behaviour.

This concludes the discussion of GRP1-PH. In the next chapter, some of the methods developed over the previous two chapters are applied to a novel PH domain structure. In contrast to GRP1-PH, the PIP binding properties and functional role of this PH domain are as yet unknown.

Chapter 5

Structural studies of the kindlin-1 PH domain

THIS chapter describes simulation work on the kindlin-1 pleckstrin homology (PH) domain. While the work described in the previous two chapters focussed on the application of computational techniques to a well characterised signalling protein (GRP1), in contrast this chapter is concerned with a signalling protein for which scant structural or functional data are available.

5.1 Introduction

Integrins are a class of proteins involved in cell-cell and cell-matrix adhesion, and they are important for a host of biological processes including tissue repair and embryogenesis. Integrins switch between an active, extended conformation with high affinity for ligands, and an inactive, low affinity, compact conformation. Activation can occur by either an 'outside in' mechanism (mediated by, for example, fibronectin) or by an 'inside out' mechanism, with the latter primarily initiated by talin binding to the cytoplasmic

tails of the integrin. Recently however, kindlins have been identified as co-activators of integrins, potentially working in tandem with talin [321].

The integrin dimer is formed from non-covalently bound α and β subunits, which comprise a large extracellular domain, a transmembrane domain and, most importantly for inside out activation, a short cytoplasmic tail. The β subunit typically contains two well-defined binding sites, a membrane proximal NPxY motif and a membrane distal NxxY motif. The F3 subdomain of talin is known to bind to the NPxY motif, which results in enhanced integrin activation by disrupting the interaction between the two integrin subunits (Figure 5.1). Conversely, kindlin-1 and kindlin-2 are thought to interact with the NxxY motif.

Talin and kindlin both incorporate a FERM (F for 4.1 protein, E for ezrin, R for radixin and M for moesin) domain [322], a protein module responsible for membrane localisation. The FERM domain typically comprises three subdomains (F1, F2 and F3) that together form a compact cloverleaf arrangement [323, 324], but talin and kindlin are unusual as in both cases the F1 domain is preceded by another domain (F0), and F1 itself contains a large insert not observed in other FERM domains [325]. Furthermore, talin is distinct from other FERM domains in that the subdomains adopt a linear rather than cloverleaf configuration [326, 327] (Figure 5.1). The similarity of talin to the kindlins suggests that the FERM domains of the kindlins may also be arranged in a linear fashion.

However, despite all of their similarities, there are also some striking structural differences between the kindlin family of proteins and talin. For example, the F1 insertion loop is much longer in the kindlins than in talin, comprising approximately 100 amino acid residues (variable between the three kindlin isoforms) as opposed to $\sim$40 residues in talin [325]. Kindlins are also unusual in that they possess an additional PH domain inserted in the F2 subdomain [325], but the physiological role of this PH domain is not yet known.

The kindlin family consists of three members in mammals: kindlin-1; kindlin-2 and

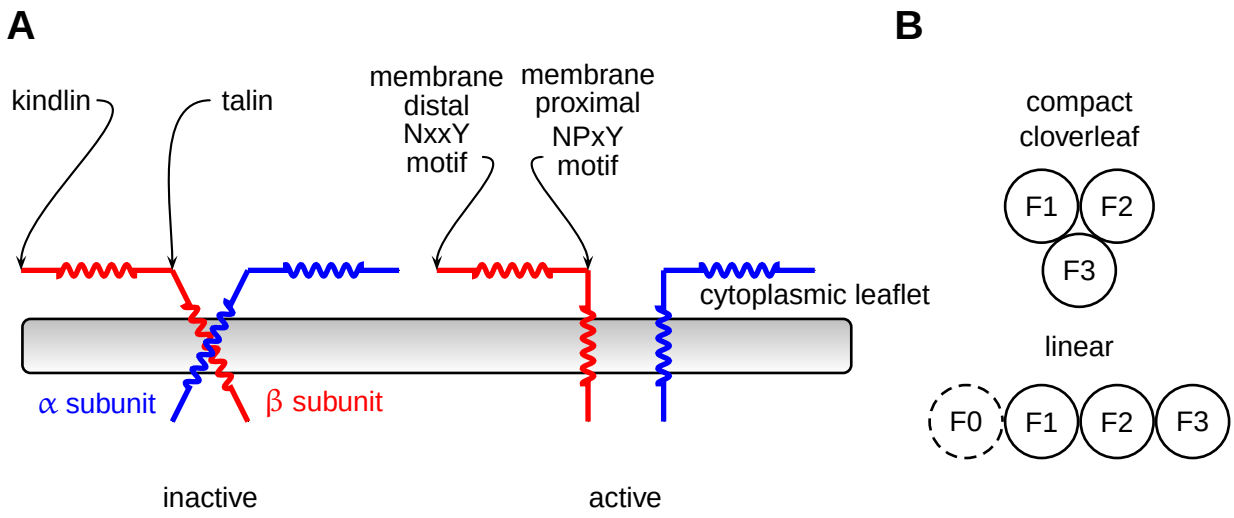

Figure 5.1: Schematic of integrin activation by talin and integrin. **(A)** Talin and kindlin bind to the membrane proximal NPxY and membrane distal NxxY motifs respectively. Binding to the β tail shifts the integrin into a high affinity activated state. **(B)** FERM domains are conventionally arranged in a compact cloverleaf geometry, but the subdomains of talin are arranged in a linear fashion.

kindlin-3. The focus of this chapter is on kindlin-1, which is expressed in epithelial cells. Kindlin-2 is expressed in virtually all tissues, while kindlin-3 is expressed almost exclusively in haematopoietic tissue. Loss of function mutations in human kindlin-1 can result in Kindler syndrome, symptoms of which include skin blistering, skin atrophy [328] and in some rare cases cutaneous squamous cell carcinoma [329]. Loss of kindlin-2 in mouse models causes embryonic lethality, while loss of kindlin-3 leads to severe bleeding and fatal anaemia due to inactive platelet integrins. Mutations in human kindlin-3 result in leukocyte adhesion deficiency syndrome III (LAD3) [330, 331], and this can lead to life threatening bleeding disorders such as Glanzmann's thrombasthaenia [331, 332].

The PH domain of kindlin-2, in common with several other PH domains, binds to PI(3,4,5)P_3 with high affinity [333, 334], and kindlin-2 mutations abrogating PI(3,4,5)P_3 binding result in impaired integrin activation [333]. This suggests that the kindlin-2 PH domain may be important in membrane binding and integrin activation by the full length kindlin-2. Thus, by extension, one might expect that the PH domain of kindlin-1 would also be involved in these processes, despite the different roles of kindlin-1 and kindlin-2 in, for example, keratinocytes [335].

Recently, the full length kindlin-1 protein has been cloned, overexpressed and purified by Gilbert and co-workers[1]. Yates and Gilbert were able to crystallise and solve the structure of the fragment corresponding to the PH domain using X-ray crystallography. Biophysical and biochemical data on the binding properties and ligand specificity of kindlin-1 PH remains, however, relatively scarce. In this chapter, the overall aim is to utilise this new structural information to attempt to characterise the binding properties of kindlin-1 PH using molecular modelling. A related aim is the use of simulation to explore, *in silico*, the behaviour of kindlin-1 PH and generate testable hypotheses that can be verified experimentally by Gilbert and co-workers.

5.2 Methods

5.2.1 MD simulations

MD simulations were performed with GROMACS version 4.0.5 [186], using either the united atom GROMOS96 43a1 forcefield [142] or the all atom OPLS-AA/L forcefield [143] with a timestep of $\Delta t = 2$ fs. The protein was placed in a cubic box with sides of length 60 Å and solvent molecules were subsequently added. The SPC [146] water model was used in conjunction with the GROMOS 43a1 forcefield, while the TIP4P [155] water model was used with the OPLS-AA/L forcefield. Chloride ions were added to neutralise the net charge of the system. Temperature was maintained at 310 K by coupling to a Berendsen thermostat [159] with a coupling constant of $\tau_T = 0.1$ ps. Pressure was kept constant at 1 atm using a Parrinello-Rahman barostat [162, 163], with isotropic pressure coupling, a coupling constant of $\tau_p = 1.0$ ps and a compressibility of 4.6×10^{-5} bar^{-1}. Bond lengths and angles were constrained using the LINCS algorithm [139]. Electrostatic interactions were treated with the PME [235] method using a cutoff of 10 Å.

[1]This project was part of a collaboration with Dr Robert Gilbert and colleagues in the Division of Structural Biology, Nuffield Department of Clinical Medicine, University of Oxford and with Professor David Calderwood and colleagues at Yale University School of Medicine

5.2.2 Molecular docking

AutoDock version 4.2 [220, 221] was used for ligand docking. The ligand structures used for molecular docking were extracted from high resolution X-ray crystal structures deposited in the PDB. The headgroups of six of the seven physiological PIP isoforms were used in the docking calculations: I(1,3)P_2 (PDB 1JOC) [33]; I(1,4)P_2 (PDB 1I9Z) [336]; I(1,3,4)P_3 (PDB 1Z2P) [337]; I(1,3,5)P_3 (PDB 1ZVR) [338]; I(1,4,5)P_3 (PDB 1MAI) [285] and I(1,3,4,5)P_4 (PDB 1UNQ) [339]. Although in total there are seven naturally occurring PIP isoforms in eukaryotic cells, it appears that no structures with I(1,5)P_2 bound have yet been reported. The absence of structural data for PI(5)P binding domains is perhaps not surprising, as PI(5)P was identified only recently and its role and effectors have not yet been clearly determined (see Chapter 1). Structures were prepared for docking by adding Gasteiger partial charges [340] and ligand torsions using AutoDockTools. The protein was kept rigid and grid maps were calculated using 80 × 80 × 80 grid points with a spacing of 0.375 Å. Grids were centred such that they encompassed the whole kindlin-1 PH domain.

The Lamarckian genetic algorithm was used for ligand conformational searching and in general the docking parameters were based on those used in [341]. For each run 100 trial dockings with a population size of 300 were performed. Translation step ranges were 1.5 Å and rotation step ranges were 35°. Elitism, mutation rate, crossover rate and local search rate were set to 1, 0.02, 0.8 and 0.06 respectively, with 25 million energy evaluations and 27000 generations. Starting positions and conformations were randomised, and clusters were evaluated using a tolerance of 1.5 Å RMSD. To validate the docking results blind dockings using SwissDock (`http://www.swissdock.ch`) [342] with the accurate/default parameter preset [343] were also performed. Ligands were converted to the SYBYL MOL2 format prior to docking using PRODRG [147].

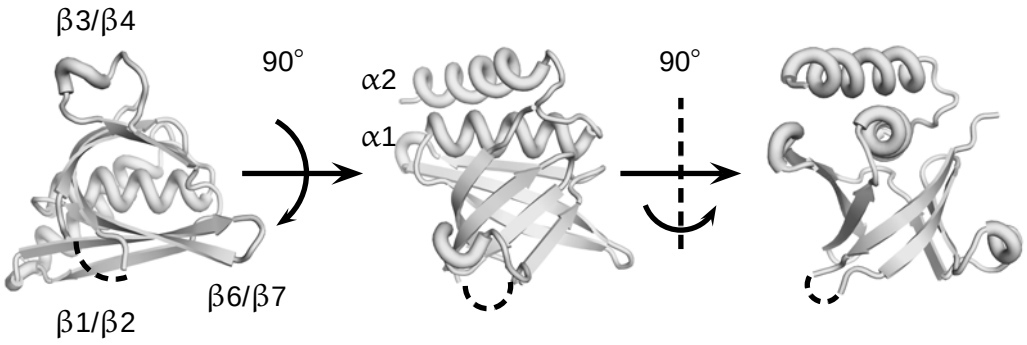

Figure 5.2: Structure of kindlin-1 PH at 2.1 Å resolution from the Gilbert group. The missing region (residues 382-388) from the β1/β2 loop is shown as a dashed line.

5.2.3 Binding pocket volume calculations

Ligand binding pocket volumes were calculated using the POcket Volume MEasurer (POVME) [344]. A pocket encompassing region was defined by placing a sphere of radius 10 Å at the geometrical centre of the binding pocket, and the sphere was filled with equispaced points 1 Å apart. Points overlapping with protein atoms were systematically deleted, leaving only the cavity populated with grid points. Volume calculations were then performed on this subset of grid points.

5.3 Structure of kindlin-1

The kindlin-1 PH domain comprises 129 amino acid residues and adopts the archetypal PH fold, with a seven stranded β barrel open at one end and the other end capped by an α-helix (Figure 5.2). In this respect it shares many similarities with other PH domains, but somewhat unusually the capping helix (denoted α1) is predominantly made up of amino acid residues with hydrophobic sidechains. The capping helix is conventionally amphipathic, but the apolar nature of the α1 helix in kindlin-1 PH allows for stacking of an additional helix (α2) on top of α1. This feature is conserved in kindlin-2 and kindlin-3.

As weak electron density was observed for residues 382-388 in the β1/β2 loop, this

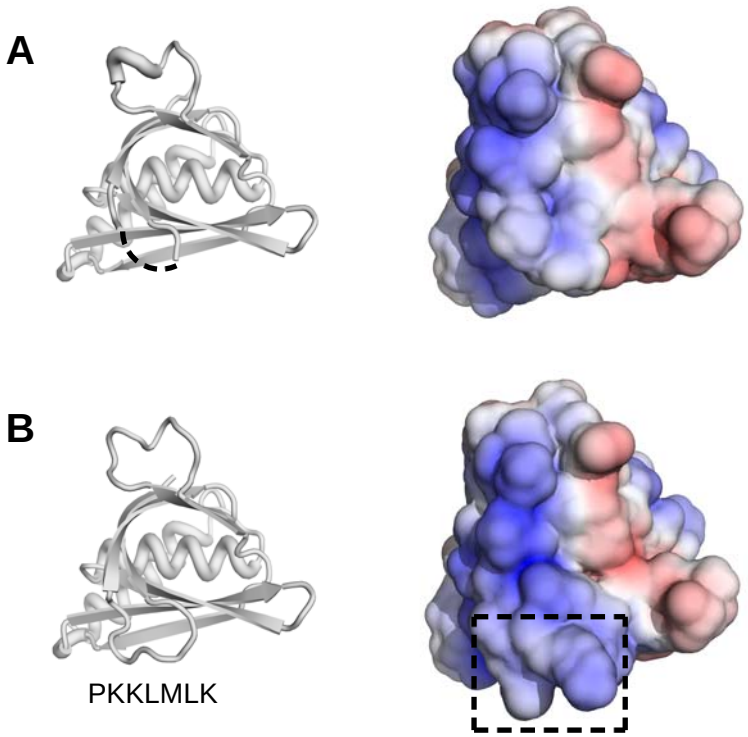

Figure 5.3: Structures and corresponding electrostatic potentials of kindlin-1 PH. **(A)** Crystal structure of kindlin-1 PH (left panel) and its electrostatic potential projected onto the solvent-accessible surface (right panel). The missing loop is shown as a dashed line. **(B)** Top scoring model derived from *de novo* loop modelling using MODELLER (left panel) with the corresponding electrostatic potential (right panel). Electrostatic potentials are coloured from -5*kT/e* (negative, red) to +5*kT/e* (positive, blue).

region of the protein was not built. These residues were modelled using *de novo* loop modelling before conducting the MD simulations. MODELLER [345] was used in combination with the MMTSB toolset [346] to place the missing region in the crystal structure, which comprises residue numbers 382-388 with sequence PKKLMLK. A total of 200 models were generated. These were clustered based on the conformations adopted by the flexible loop formed by residues 382-388. The top cluster was extracted on the basis of the average MODELLER score, and the highest scoring model from this cluster was used as a starting point for the MD simulations (Figure 5.3).

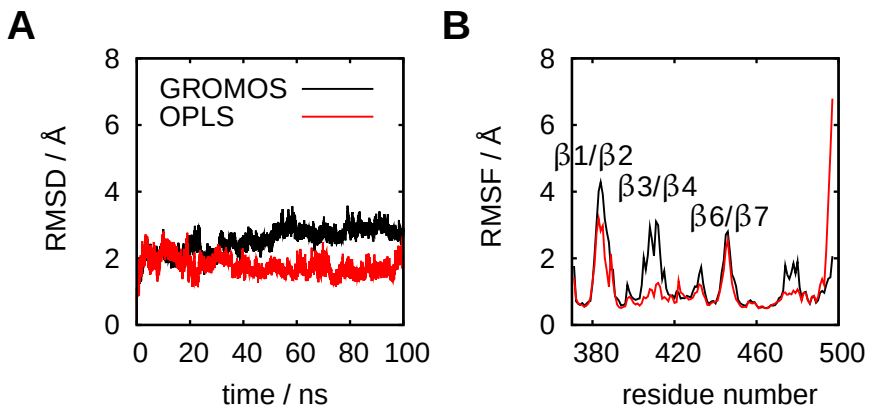

Figure 5.4: Conformational dynamics of the kindlin-1 PH domain during the MD simulations. **(A)** Cα RMSD calculated as a function of time over the MD simulations using the GROMOS96 43a1 forcefield (black) and the OPLS-AA/L forcefield (red). **(B)** RMSF of the protein Cα atoms calculated for each residue over the course of the MD simulations.

5.4 MD simulations of wild type kindlin-1

Initially, MD simulations of 100 ns duration were conducted for solvated kindlin-1 PH after model building and addition of the missing fragment of the β1/β2 loop. To assess the reproducibility of the results, simulations were carried out using two different forcefields: the united atom GROMOS96 43a1 forcefield and the all atom OPLS-AA/L forcefield. The RMSD and RMSF of the Cα atoms over the course of the simulation trajectories are shown in Figure 5.4. The Cα RMSD indicated that the protein experienced little conformational drift over a period of 100 ns. The Cα RMSF suggested that, of the three loops flanking the open end of the β barrel, the β1/β2 loop appears to be the most flexible, with the maxima in the two RMSF traces (aside from the N- and C-termini) located at residues 383 (OPLS-AA/L) and 384 (GROMOS96 43a1), perhaps explaining why this region was disordered in the crystal.

Interestingly, although the R380–E416 salt bridge spanning the entrance to the binding pocket was typically preserved during the simulations, transient switching to an alternative conformation involving the breaking of R380–E416 and the formation of a different salt bridge, R380–E450 (Figure 5.5) was also observed. This transition between a partially closed state and an open state was apparent in both simulations using either the

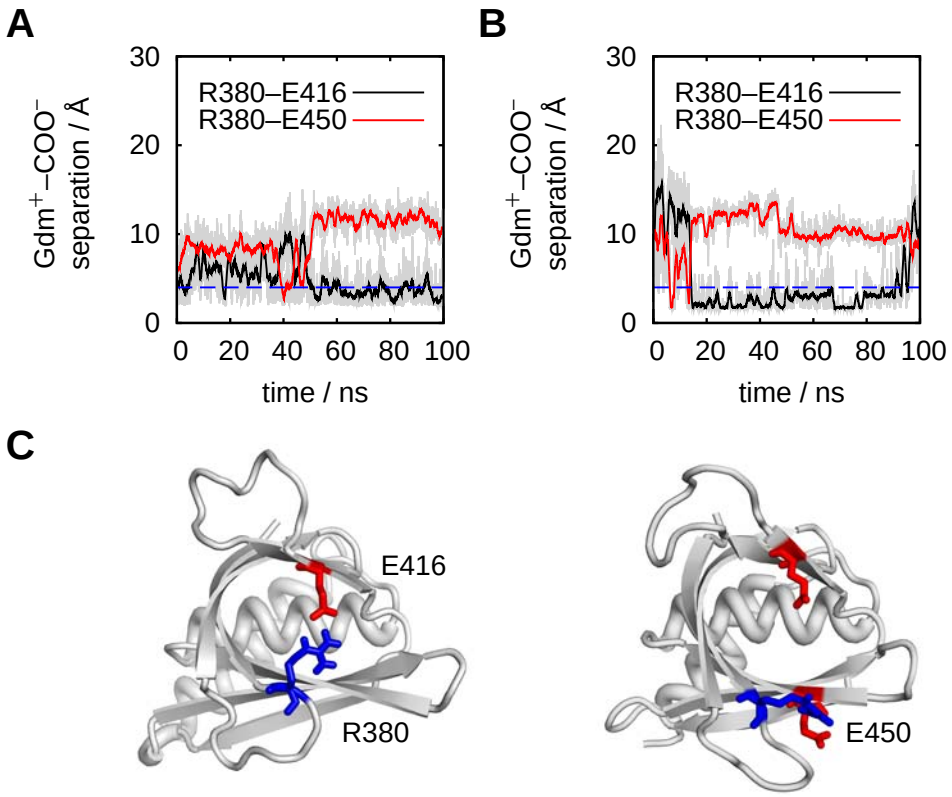

Figure 5.5: Salt bridge switching dynamics in kindlin-1 PH during the MD simulations. For **(A)** the GROMOS96 43a1 forcefield and **(B)** the OPLS-AA/L forcefield, running averages taken every 100 ps of the separation between the centres of mass of the R380 guanidinium cation and the carboxylate anion of both E416 (black) and E450 (red) are shown. The dotted blue line shows the separation at which a salt bridge is formed, following the working definition of Kumar and Nussinov [347]. **(C)** Snapshots from the GROMOS96 43a1 simulation showing the R380–E416 salt bridge closed (left panel) and open due to the formation of the alternative R380–E450 salt bridge (right panel).

GROMOS96 43a1 forcefield or the OPLS-AA/L forcefield. The dynamics of salt bridge formation appear to be coupled in that breaking of one salt bridge is concomitant with formation of the other.

MD simulations of the E416A mutant, designed to permanently disrupt the R380–E416 interaction, were also performed. On mutation, the distribution of separations between R380 and E416 shifts to larger values under both forcefields (Figure 5.6), leading to a more exposed binding pocket in the mutant protein. Note that it is necessary to adopt a slightly different metric when measuring the residue separation in the mutant and comparing it with the wild type, as Glu and Ala have sidechains of different lengths. The measurement was therefore taken from the sidechain of R380 to the Cα atom of the

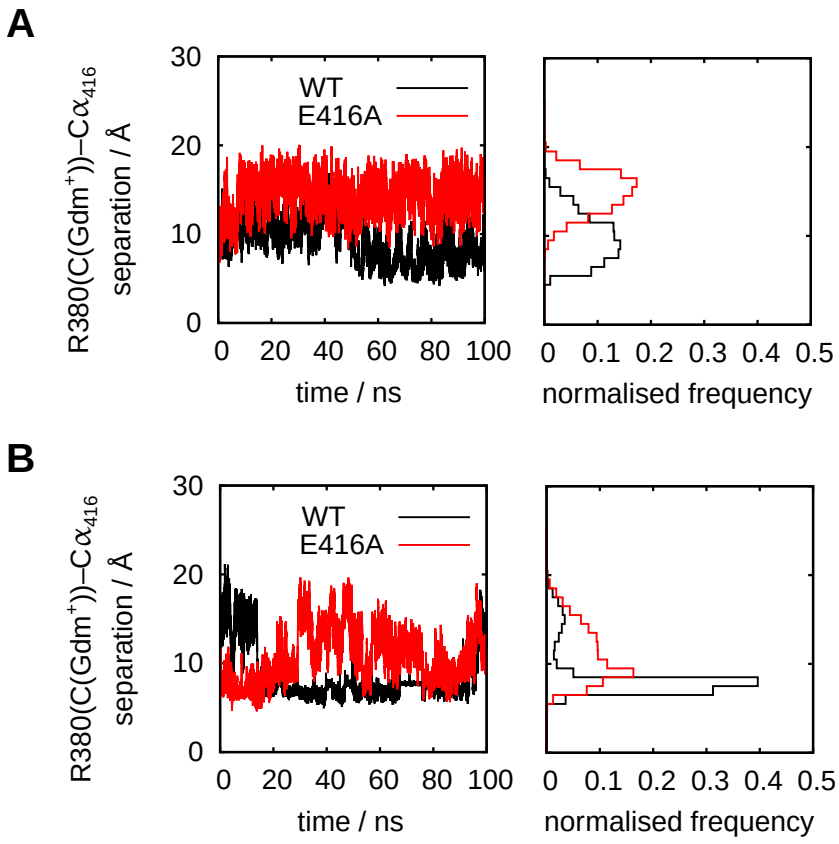

Figure 5.6: Breaking of the R380–E416 salt bridge in the E416A mutant. **(A)** The left panel shows the separation between the central carbon atom of the R380 guanidinium cation (C(Gdm^+)) and the Cα atom of the residue at position 416 over the 100 ns MD trajectory for the wild type protein (black) and the E416A mutant protein (red) using the GROMOS96 43a1 forcefield. The right panel shows normalised frequencies of R380(C(Gdm^+))–416(Cα) separations over the MD trajectories. Separations were binned at intervals of 0.1 Å. **(B)** The same metrics are shown for the simulations carried out using OPLS-AA/L.

residue at position 416 instead. The switch to the alternative, open conformation and the truncation of the Glu sidechain to Ala leaves the putative binding pocket much more exposed when compared to the crystal structure (Figure 5.7).

5.5 Docking of phosphoinositides to kindlin-1

The MD simulations indicated that the binding pocket of kindlin-1 PH was dynamic and possibly gated by the presence of a salt bridge. To probe the accessibility of the pocket in different states, several blind, rigid body docking runs were carried out against multiple targets: the closed conformation of the wild type protein where the R380–E416 salt

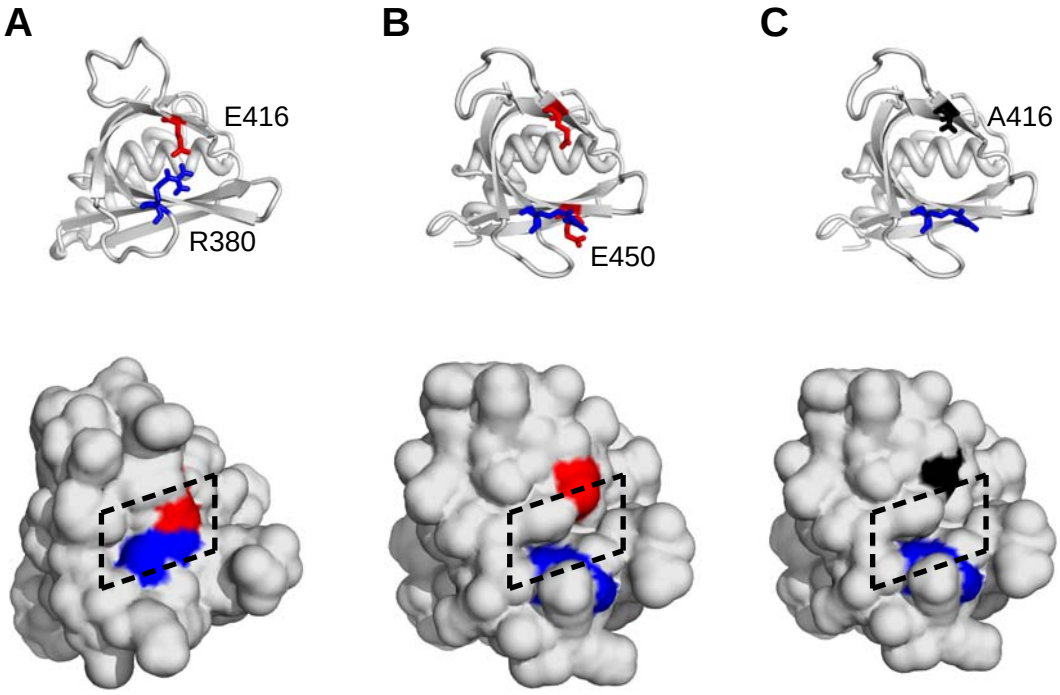

Figure 5.7: Pocket accessibility of kindlin-1 PH. **(A)** The crystal structure with the R380–E416 salt bridge intact and blocking entry to the pocket. **(B)** A conformation with the R380–E416 salt bridge broken and the alternative R380–E450 salt bridge formed. **(C)** The open conformation in the E416A mutant is more accessible owing to both (i) the breaking of the salt bridge and (ii) the charged → non-polar mutation and shortening of the sidechain of the residue at position 416. In all cases the upper panel shows the ribbon representation of the secondary structure, while the lower panel shows the corresponding solvent-accessible surface.

bridge remained intact; an open conformation of the wild type protein with the salt bridge broken and finally an open conformation of the E416A mutant protein. In each case, each of the six PIPs was docked with kindlin-1 PH in an effort to identify a putative lipid binding site.

When molecular docking was performed using the closed state the results were generally poor. Clustering the conformations resulted in a large array of small, spatially separated clusters and little consensus emerged as to the location of any possible binding site (Figures 5.8 and 5.9). Indeed, even when the number of energy evaluations was increased by a factor of 500 to 10 billion, a likely binding site failed to emerge. In the absence of a predicted binding site, the open conformation observed in the MD simulations was considered as an alternative. Molecular docking using this open conformation extracted from the simulations led to a greater number of docking runs converging around

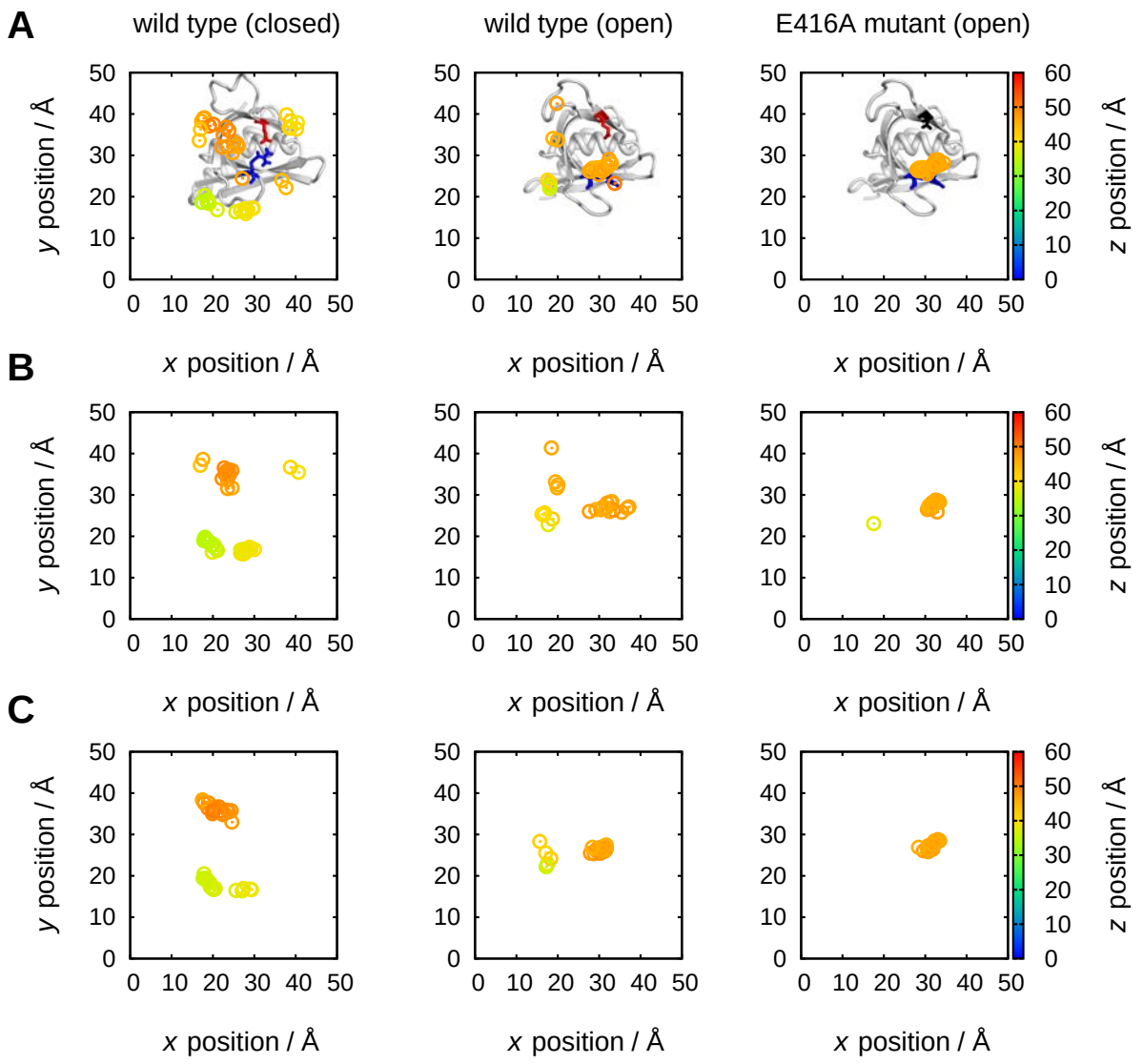

Figure 5.8: Positions of conformational clusters after 100 docking runs using AutoDock for **(A)** I(1,3)P_2, **(B)** I(1,4)P_2 and **(C)** I(1,3,4)P_3. The left panel shows results for docking to the closed conformation (R380–E416 salt bridge intact), the central panel shows results for the open conformation (R380–E416 salt bridge broken) and right panel shows results for the E416A mutant in the open conformation. See also Figure 5.9.

a single likely binding site. Clustering of these results produced a smaller number of larger clusters, and most of the docking runs placed the ligand in or near the central cavity (Figures 5.8 and 5.9). Docking using the open rather than the closed conformation approximately halved the number of conformational clusters and multi-member clusters. This higher level of consensus resulted in a reasonably consistent prediction of a putative binding site for all six PIP headgroups, lined by R380, K382, K390 and

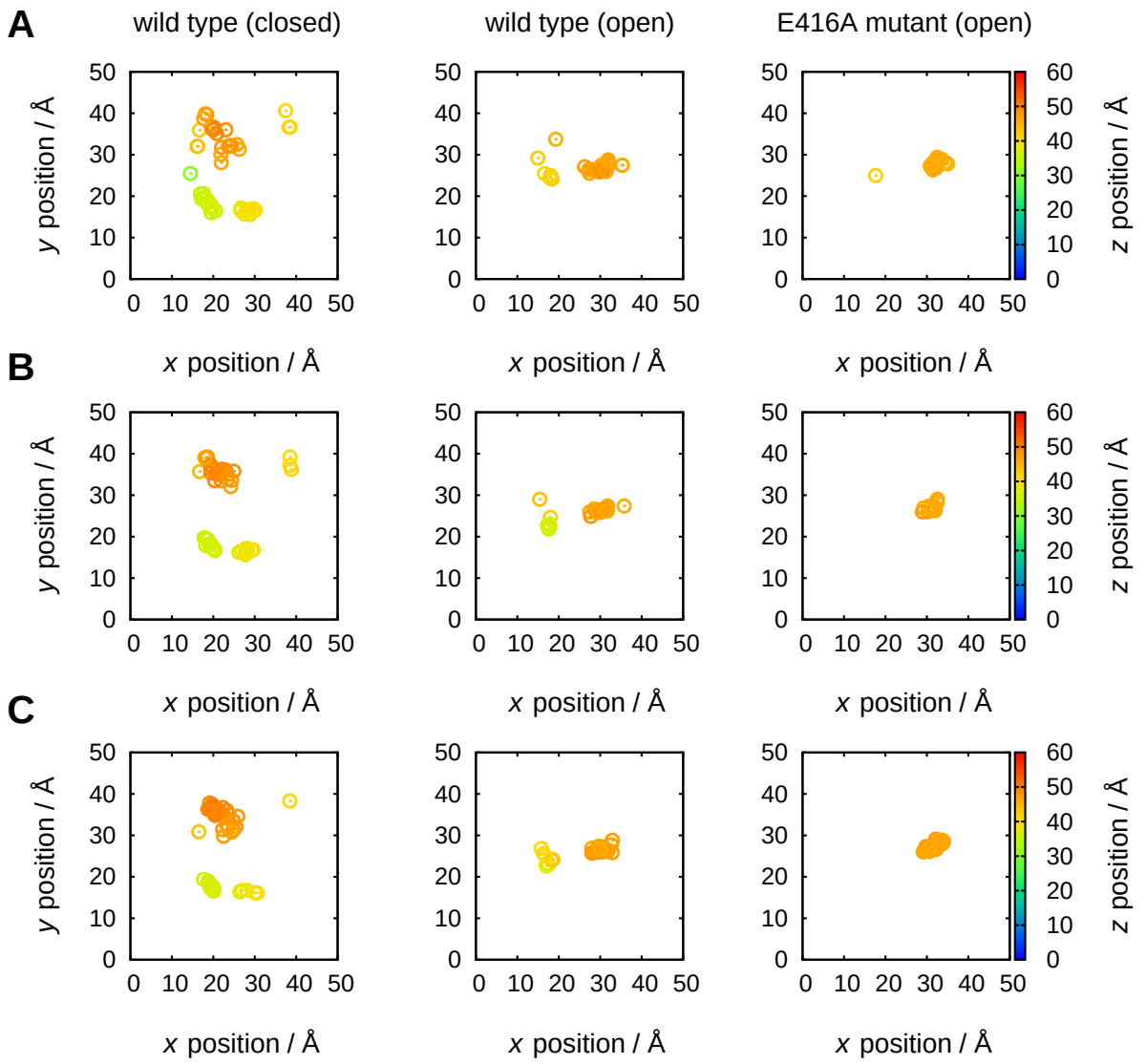

Figure 5.9: Positions of conformational clusters after 100 docking runs for **(A)** I(1,3,5)P_3, **(B)** I(1,4,5)P_3 and **(C)** I(1,3,4,5)P_4. The left panel shows results for docking to the closed conformation (R380–E416 salt bridge intact), the central panel shows results for the open conformation (R380–E416 salt bridge broken) and right panel shows results for the E416A mutant in the open conformation. See also Figure 5.8

N449 (Figure 5.10). Performing similar docking runs using another docking program (SwissDock) yielded qualitatively similar results.

This effect was even more pronounced when performing docking runs with the E416A mutant in the open state. It therefore appears that PIP binding is enhanced in the E416A mutant for two reasons. First, the breaking of the R380–E416 salt bridge, which otherwise would block the entrance to the binding pocket. Second, the replacement of the

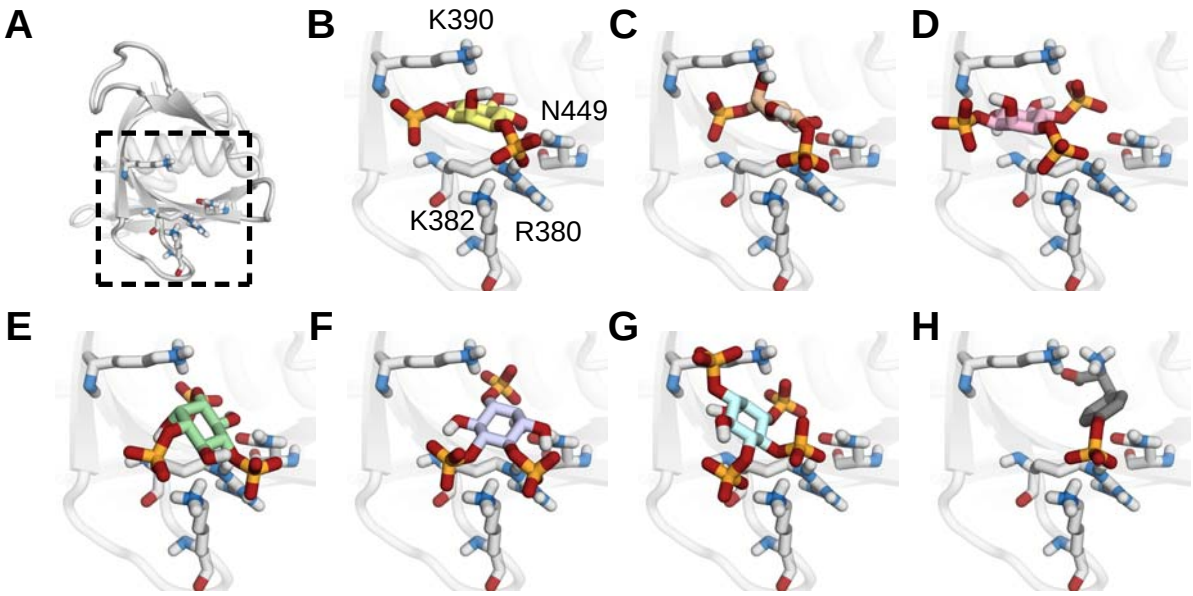

Figure 5.10: Lowest energy docking poses from the largest conformational cluster for ligand docking to kindlin-1. The location of the putative binding site in kindlin-1 PH is shown in **(A)**. Docking poses are shown for **(B)** I(1,3)P_2 (yellow); **(C)** I(1,4)P_2 (brown); **(D)** I(1,3,4)P_3 (pink); **(E)** I(1,3,5)P_3 (green); **(F)** I(1,4,5)P_3 (blue); **(G)** I(1,3,4,5)P_4 (cyan) and **(H)** p-Tyr (grey).

sentry glutamate residue, E416, by alanine in this open state, which appears to further increase the accessibility of the binding site for negatively charged lipid headgroups by truncation of the sidechain and the charged → non-polar nature of the residue change. This stepwise enhancement of PIP binding appears to be confirmed by the estimated free energies of binding predicted by AutoDock, which suggest that docking to the E416A mutant (open) PH domain is more favourable than docking to the wild type (open), which in turn is an improvement over the wild type (closed) PH domain. In general, this trend is observed for all six PIP species, and the more favourable binding energies (Figure 5.11) are accompanied by an improved level of clustering (Figure 5.12).

The location of this glutamate residue in kindlin-1 PH is of particular note as several previous studies have identified that the occurrence of an acidic sentry sidechain is a widespread structural feature of PI(3,4,5)P_3 binding PH domains such as Akt [348, 349] and GRP1 [350]. In these PH domains, the sentry glutamate is thought to exclude PI(4,5)P_2 from the binding pocket, thus favouring PI(3,4,5)P_3 binding. This may offer a clue as to why kindlin-1 PH does not bind strongly to PI(4,5)P_2, though interestingly it

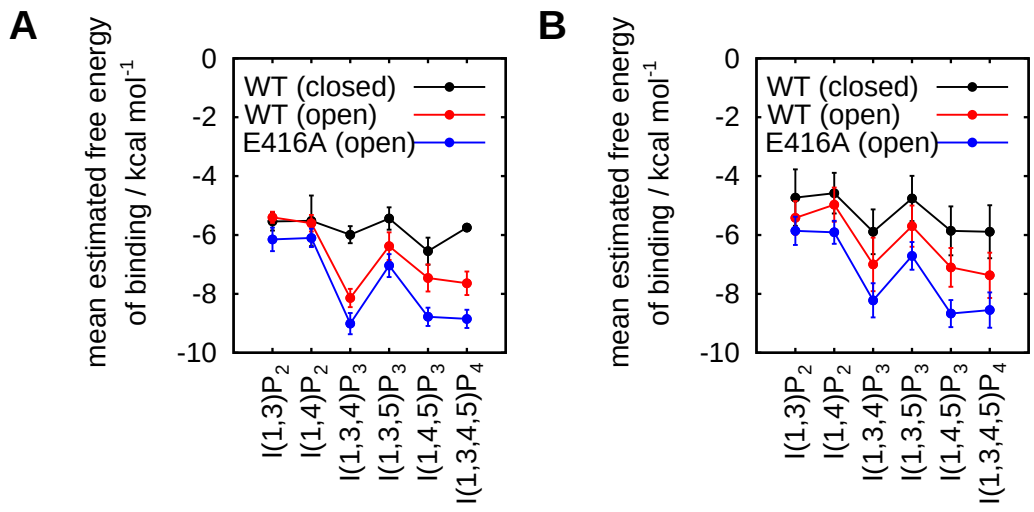

Figure 5.11: Estimated free energies of binding for PIP docking using AutoDock. **(A)** Mean estimated free energy of binding for all 100 docking runs using AutoDock for each of the six PIP species. Results are shown for docking to the wild type (closed) PH domain (black), the wild type (open) PH domain (red) and the E416A mutant (open) PH domain (blue). Error bars are plus or minus one standard deviation from the mean. Lines are guides for the eye. A more negative value is indicative of more favourable binding. **(B)** Mean estimated free energy of binding for the members of the largest cluster after docking based on a cluster tolerance of 1.5 Å RMSD. Similar trends to those in panel **(A)** are observed.

does not appear to be a PI(3,4,5)P_3 binding domain either.

5.5.1 Experimental verification

On the basis of the results from the MD simulations and the docking runs, Yates and Gilbert performed site directed mutagenesis of the isoform specific glutamate at position 416 to permanently disrupt the salt bridge, predicting that this disruption would lead to an improved ligand binding module and broaden the range of PIP lipids that it is able to recognise.

Yates and Gilbert used the mutant protein in lipid overlay assays alongside the wild type protein. Both proteins exhibited preferential binding to PI monophosphate species and PS. However, the mutant also displayed additional lipid recognition of PI(3,5)P_2. To test this observation further Yates and Gilbert incubated purified kindlin-1 PH domain and the E416A mutant with membranes containing different quantities of PIPs, which gave the same result. The wild type kindlin-1 PH domain bound moderately to the lipids, with signal observed at 6.25 pM quantities of PI(3)P, PI(4)P and PI(5)P. In contrast, the

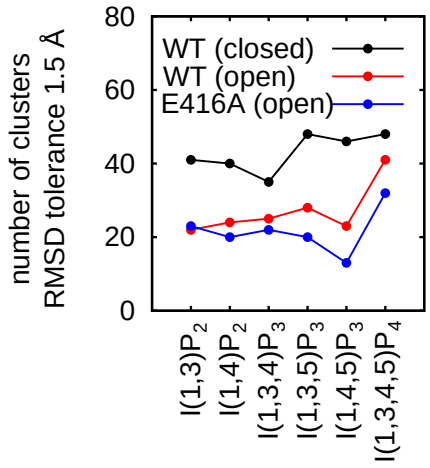

Figure 5.12: Number of conformational clusters after 100 docking runs for each of the six PIP species based on a cluster tolerance of 1.5 Å RMSD. A lower number of clusters suggests a greater degree of consensus as to the location of the binding site. Lines are guides for the eye.

E416A mutant bound strongly to the same lipids with signal observed at 1.56 pM for PI(3)P, PI(4)P and PI(5)P. Moreover, the mutant exhibited moderate binding to PI(3,5)P_2, whereas the wild type PH domain demonstrated undetectable binding. This suggests that the liberated arginine from the salt bridge may coordinate the additional phosphate in PI(3,5)P_2. While Yates and Gilbert were unable to quantify the binding by deriving a K_D from these measurements, against a common background density they were able to show quantitatively that I(5)P and I(3)P were bound more strongly than I(4)P and that the E416A mutant binds with higher affinity than the wild type protein. The E416A mutant therefore exhibits broadened lipid specificity and higher affinity binding to PI monophosphates, in good agreement with the predictions from the computational work.

It is tempting to speculate that physiological interactions undergone by kindlin-1 may produce similar conformational changes to those generated in the salt bridge mutant and allow PIP binding. For example, Yates and Gilbert observed that the protein tends to be dimeric in solution, and dimerisation could in principle lead to the conformational change observed in the MD simulations. A possible link between salt bridge opening and dimerisation was demonstrated when the removal of the salt bridge in the E416A mutant by Yates and Gilbert was observed to change the dimerisation properties. This

could perhaps be caused by some distortion of the topography of the β barrel surface in the open conformation.

5.5.2 A note on estimated binding affinities from docking

At this point, it should be noted that while docking programs are typically able to generate reasonable ligand docking poses, they are frequently unable to predict ligand binding affinities with any accuracy [351]. Ideally, docking programs should be able to predict binding poses for a variety of ligands and calculate their respective binding affinities, thus allowing ligands to be ranked in order of affinity. However, as Figure 5.11 shows, it is not possible to infer the experimentally observed PI monophosphate specificity of kindlin-1 PH from the docking runs. Furthermore, the lipid binding assay suggests that among the PI monophosphates, the binding affinities exhibit the following trend: I(5)P > I(3)P > I(4)P. The similarity of the predicted free energies of binding for I(3)P and I(4)P suggests that it would be unlikely that this sequence could be correctly predicted on the basis of the docking results alone.

The difficulties in predicting binding affinities chiefly arise from the large number of conformational degrees of freedom in the protein-ligand complex, and also the fact that the free energy of binding is often the difference between two very large numbers [352]. As a result, caution is required when evaluating binding free energies and interpreting the results shown in Figure 5.11. While the absolute values are unlikely to be correct, the estimated free energies of binding do at least show a clear trend for all PIP species of increasing affinity as the binding pocket becomes more accessible.

5.6 Comparison of kindlin-1 to kindlin-2 and kindlin-3 isoforms

The PH domains of the three kindlin isoforms have a high sequence identity of between 50-60% (see Figure 5.16). The solution structures of the PH domains of kindlin-2 and kindlin-3 have been solved using NMR spectroscopy, and so a comparison between the

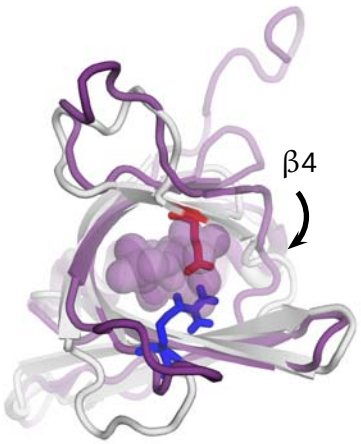

Figure 5.13: Binding pocket constriction in kindlin-1 PH compared with kindlin-2 PH. A structural alignment of kindlin-1 PH and kindlin-2 PH in PyMOL shows the binding pocket of kindlin-1 PH (white) is more constricted than in kindlin-2 PH (purple). The R380–E416 salt bridge in the crystal structure draws in the β4 strand by ∼6 Å in kindlin-1 PH. The bound PI(3,4,5)P_3 in the NMR structure of kindlin-2 PH (PDB 2LKO [334]) is shown as a collection of transparent van der Waals spheres, illustrating the likely steric hindrance associated with binding of bulkier PIPs such as PI(3,4,5)P_3 to kindlin-1 PH.

three domains is instructive. Simulations of kindlin-2 (PDB 2LKO [334]) and kindlin-3 (PDB 2YS3 (to be published)) were performed using the solution structures of these proteins as a starting point. In each case the lowest energy conformer from the NMR ensemble was used. The simulation parameters used were identical to those used for kindlin-1, except that the simulation box was expanded from 60 $Å^3$ to 80 $Å^3$ to allow for the fact that these solution structures are slightly larger (by ∼10 amino acid residues) than the crystal structure of kindlin-1.

Although no experimental data associated with the NMR structure for kindlin-3 is yet available, there is substantial biochemical and functional data for kindlin-2 [334]. The authors of the study in which the NMR structure was reported presented data suggesting that kindlin-2 PH binds to the PI(3,4,5)P_3 headgroup, I(1,3,4,5)P_4, with high affinity. They also show that kindlin-2 PH binds to the headgroup of PI(3,4,5)P_3 more strongly than to that of PI(4,5)P_2 by chemical shift mapping, but did not characterise the interaction of kindlin-2 PH with PI monophosphates using NMR.

Previous work on kindlin-2 using biochemical PIP binding assays [333] demonstrated a strong preference for PI(3,4,5)P_3 and very limited PI monophosphate binding. Taken

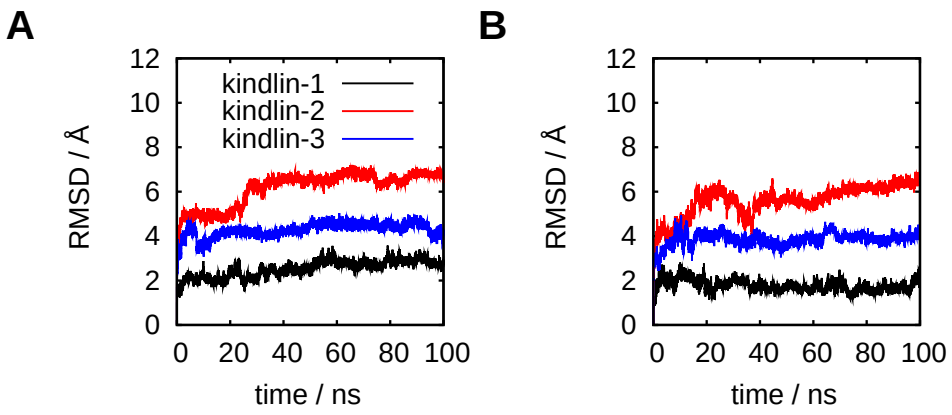

Figure 5.14: Conformational dynamics of the kindlin isoforms. **(A)** Cα RMSD calculated as a function of time over the 100 ns MD simulations using the GROMOS96 43a1 forcefield for kindlin-1 PH (black), kindlin-2 PH (red) and kindlin-3 PH (blue). **(B)** The same measure for the simulations using the OPLS-AA/L forcefield.

together, the two studies by Liu *et al.* [334] and Qu *et al.* [333] point towards a trend in PIP binding affinities for kindlin-2, with PI triphosphates binding with higher affinity than PI diphosphates, which in turn are more tightly bound than PI monophosphates. In contrast, the results suggest quite the opposite is the case for kindlin-1 PH with this trend completely reversed. Kindlin-1 PH exhibits stronger binding to PI monophosphates, somewhat weaker binding to PI(3,5)P_3 in the E416A mutant and virtually no detectable interaction with PI(3,4,5)P_3.

In their report of the NMR structure of kindlin-2, Liu *et al.* hypothesised that the binding pocket was conserved among the kindlin PH domains [334]. The structure of kindlin-1 shows this conjecture is not necessarily true, as the R380–E416 salt bridge constricts the binding pocket in kindlin-1, drawing in the β-sheet wall by ∼6 Å (Figure 5.13). To investigate this MD simulations of kindlin-2 and kindlin-3 were conducted, after first removing any ligands present in the NMR structures, to compare their dynamics with those of kindlin-1. The Cα RMSD and the binding pocket volume calculated using POVME were found to be much larger for kindlin-2 than for kindlin-1. To calculate pocket volumes, the 100 ns MD trajectories for kindlin-1, kindlin-2 and kindlin-3 were processed and snapshots extracted every 1 ns, giving a total of 100 conformations for each isoform. All conformations were then superimposed *via* a structural alignment us-

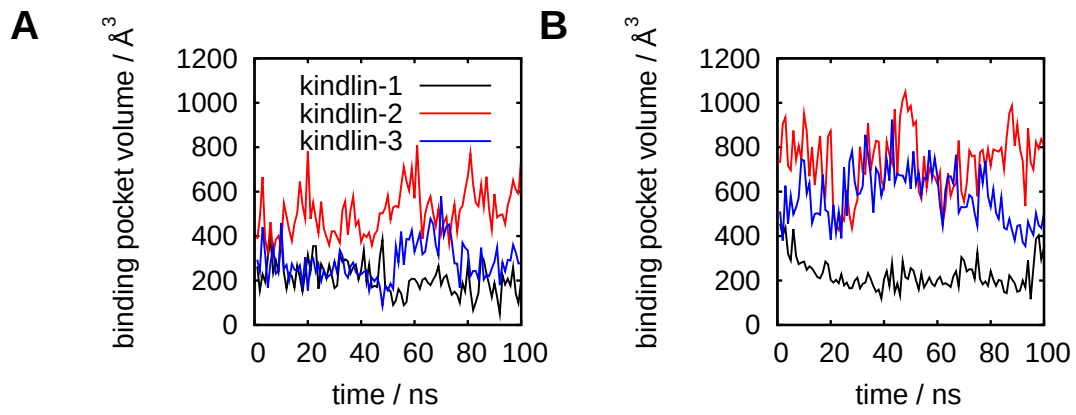

Figure 5.15: Binding pocket volume of kindlin isoforms. **(A)** The binding pocket volumes of the three kindlin isoforms over the 100 ns GROMOS96 43a1 MD trajectory are plotted for kindlin-1 PH (black), kindlin-2 PH (red) and kindlin-3 PH (blue), sampling from the trajectory and computing the pocket volume every nanosecond. **(B)** The same measure for the simulations using the OPLS-AA/L forcefield.

ing PyMOL. While it might be expected that simulations based on NMR structures would display slightly larger conformational drift, the larger pocket volume of kindlin-2 when compared to kindlin-1 may help to rationalise the trends in PIP binding affinities described above. Presumably the more constricted binding pocket in kindlin-1, which has a consistently smaller volume over the 100 ns simulation, is less able to accommodate larger PIP lipids but is still able to bind PI monophosphates.

Interestingly, a crystal structure of the *apo* form of the kindlin-2 PH domain reported very recently [353] suggests an induced fit mechanism for I(1,3,4,5)P_4 binding. The authors compared the *apo* crystal structure with the *holo* NMR structure, showing expansion of the core β barrel in the latter, perhaps due to I(1,3,4,5)P_4 binding. This apparent conformational change observed on ligand binding may perhaps help to reconcile the discrepancies between the binding pocket volumes of *apo* kindlin-1 and *holo* kindlin-2 described above. The simulations used the NMR structure of kindlin-2 with the ligand manually removed. It is therefore possible that over sufficiently long simulation timescales the binding pocket may collapse due to removal of the ligand and eventually adopt a similar conformation to that seen in the newly-reported crystal structure of kindlin-2. Regrettably however, at the time of writing the deposited crystal structure of this protein has not yet been released and so at this stage a detailed comparison is not

possible.

5.7 PH domains and non-canonical phosphoinositide binding

The pleckstrin homology domain is the 11th most common module in the human genome [354], and despite a common scaffold the various PH domains exhibit diverse functionalities. Interestingly, while most PH domains bind PIPs at the opening of the central β barrel bounded by the three variable loops, a subset of PH domains bind PIPs in a non-canonical manner. These PH domains are characterised by their unusual PIP binding mode, with the ligand positioned between the β1/β2 and β5/β6 loops outside the central cavity.

Here a limited phylogenetic analysis based on a simple sequence alignment has been performed (Figure 5.16) to illustrate some similarities between PH domains, though a more comprehensive analysis has been performed by Yates and Gilbert [355]. Inspection of the phylogenetic tree (Figure 5.17) shows that the three kindlin isoforms cluster closely together, as expected owing to their high sequence similarity. The PH domains from ArhGAP9 and β-spectrin branch from the same common scaffold, and their close proximity is particularly interesting in view of the fact that they both bind PIPs in a non-canonical fashion. As both the ArhGAP9/β-spectrin and the kindlin branches emanate from a single node, the possibility exists that the proteins within this clade are in some way related. In light of this, the structure of kindlin-1 and the docking results were re-examined to see whether there was any evidence of non-canonical binding.

Figure 5.18 shows the dipole moments for the PH domains of ArhGAP9 and β-spectrin, which both bind ligands non-canonically, compared with that for the kindlin-1 PH domain. The dipole points between the β1/β2 and β5/β6 loops in all three cases, and in the first two cases the coordinates of the bound PIP ligand lie along the molecular dipole moment vector. If the close correspondence between the location of the binding site and the direction of the molecular dipole moment is, as it seems to be, a commonly

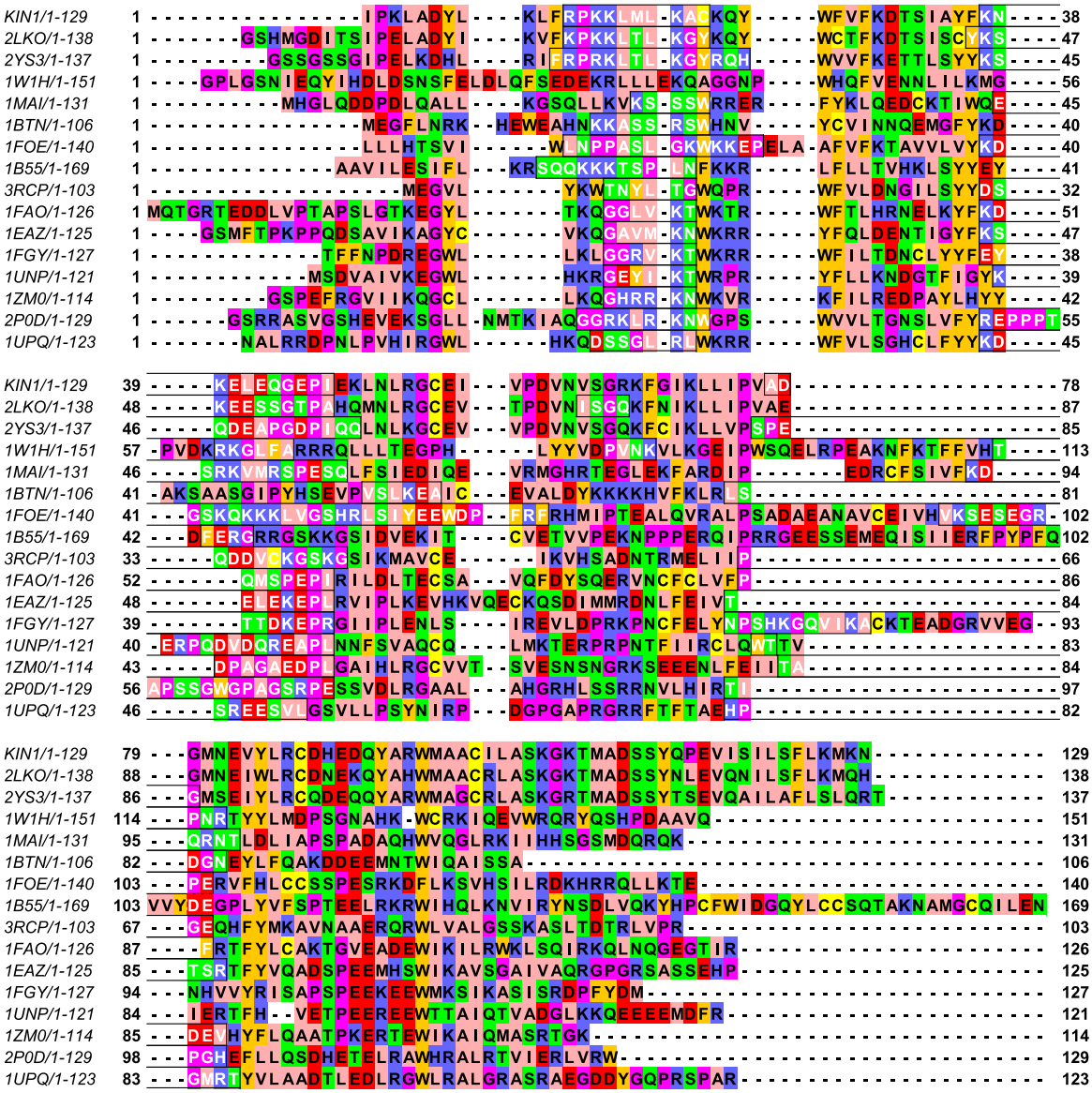

Figure 5.16: Multiple sequence alignment of 16 PH domains generated using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) software [356]. Sequences are, in order: kindlin-1 (KIN1/1-129); kindlin-2 (2LKO/1-138) [334]; kindlin-3 (2YS3/1-137) [to be published]; PDK (1W1H/1-151) [357]; PLCδ1 (1MAI/1-131) [285]; β-spectrin (1BTN/1-106) [358]; Tiam1 (1FOE/1-140) [359]; BTK (1B55/1-169) [283]; FAPP1 (3RCP/1-103) [52]; DAPP1 (1FAO/1-126) [233]; TAPP1 (1EAZ/1-125) [360]; GRP1 (1FGY/1-127) [4]; PKB/Akt (1UNP/1-121) [339]; Pleckstrin (1ZM0/1-114) [361]; ArhGAP9 (2P0D/1-129) [362] and PEPP1 (1UPQ/1-123) [to be published]. The regions of each sequence corresponding to the three variable loops in the PH domains are highlighted and indicated by white text. Reading left to right along the sequences the loops occur in the order β1/β2, β3/β4 and finally β6/β7. The multiple sequence alignment was visualised using Jalview [363, 364] and is coloured according to the physicochemical properties of the residues using the Zappo colour scheme. Residues with aliphatic sidechains are shown in pink; aromatic sidechains in orange; basic sidechains in blue; acidic sidechains in red; hydrophilic sidechains in green; conformationally special sidechains in magenta; and finally cysteine residues are shown in yellow. The assistance of Alan Robinson (Medical Research Council Mitochondrial Biology Unit, Cambridge) in the preparation of this figure is gratefully acknowledged.

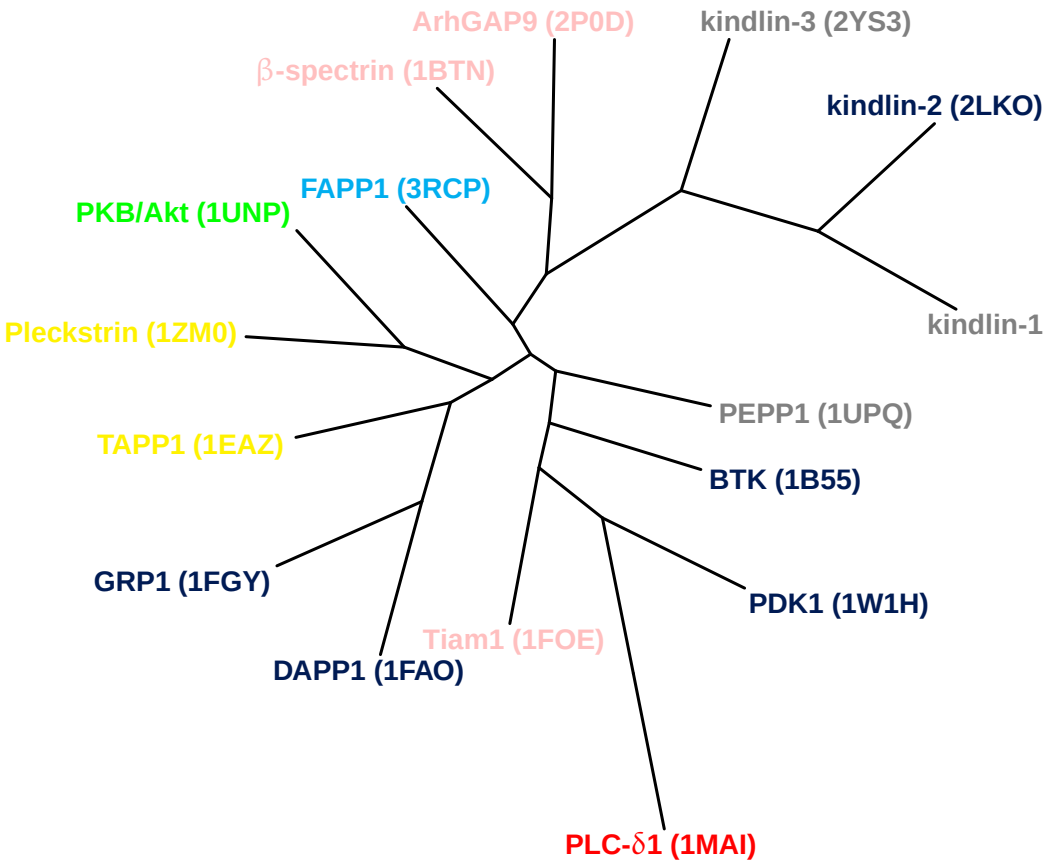

Figure 5.17: Phylogenetic tree showing interrelationships between 16 PH domains generated using PHYLIP [365]. SEQBOOT was used to bootstrap the alignment shown in Figure 5.16 with 100 bootstrap replicates performed. PROTDIST was then used to obtain distance matrices for all 100 replicates, using the Jones-Taylor-Thornton amino acid substitution matrix to generate a set of pairwise distances for every pair of the 16 sequences. FITCH [366] was then used to estimate phylogenies and infer the 100 unrooted trees from these distance matrices using the Fitch-Margoliash criterion. Sequence input order was randomised with a jumble parameter of 10. Finally, CONSENSE was used to calculate the majority rule consensus tree from the 100 replicates. The resulting tree was drawn with DRAWTREE. PH domains are coloured according to their PIP binding properties: PI(4)P binding domains (cyan); PI(3,4)P_2 binding domains (yellow); PI(4,5)P_2 binding domains (red); dual PI(3,4)P_2 and PI(3,4,5)P_3 binding domains (green); and PI(3,4,5)P_3 binding domains (dark blue). Domains that bind PIPs in a non-canonical fashion are shown in pink, while those with as-yet unidentified binding properties are shown in grey.

observed feature among PH domains (see also Figure 4.15) then one might predict on the basis of the molecular dipole alone that kindlin-1 had a non-canonical binding site[2].

Several PH domains that bind PIP ligands in a non-canonical fashion have been identified, and in addition to β-spectrin and ArhGAP9 the PH domains of Tiam1 [362] and

[2]By restricting this examination of electrostatic properties to the protein molecular dipole moment higher order multipoles such as quadrupoles, octopoles and so on, and indeed the detailed molecular electrostatic potential itself, are all ignored. However, it does serve as a crude but useful guide.

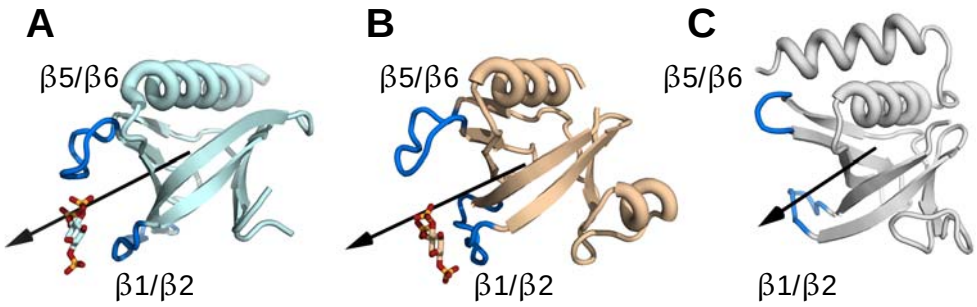

Figure 5.18: PH domains with non-canonical lipid binding sites. Several PH domains have been identified with non-canonical binding sites lying between the β1/β2 and β5/β6 loops. **(A)** The PH domain from ArhGAP9 (cyan) and **(B)** the PH domain from β-spectrin (brown) are two such examples. **(C)** The PH domain from kindlin-1 (white) is shown in approximately the same orientation. The β1/β2 and β5/β6 loops are shown in blue and the direction of the molecular dipole moment is indicated by a black arrow. See also Figure 4.15, which shows PH domains that possess a canonical lipid binding site and also Figure 5.19 showing the locations of the conformational clusters for PIP docking to kindlin-1.

more recently Slm1 [367] also exhibit this behaviour. Figure 5.19 shows the centres of mass of the docking clusters for all six PIP species (docking to the closed crystal structure) superimposed onto the kindlin-1 structure showing its molecular dipole moment. Some of the clusters are indeed coincident with the molecular dipole moment vector, and on the basis of this it seems reasonable to suggest that an additional and/or alternative binding site could be present in kindlin-1.

5.8 Discussion

The weak interaction of the wild type kindlin-1 PH domain with PIPs, coupled with its specificity for PI monophosphates, suggests that the kindlin-1 PH domain may bind to another ligand altogether. PI monophosphates are found in the Golgi (PI(4)P) and the early endosome and multivesicular body (PI(3)P) [28] (Figure 1.3) but not at the plasma membrane where kindlins act. With this in mind, it may be that the PH domain of kindlin-1 actually recognises phosphotyrosines (p-Tyr) and not PI monophosphates as it would be unlikely to encounter them *in vivo*. Repeating the docking runs described above using p-Tyr as a ligand produced similar results to those obtained for the PI monophosphates (Figure 5.10), perhaps not surprisingly as p-Tyr is also singly phosphorylated. Thus, on

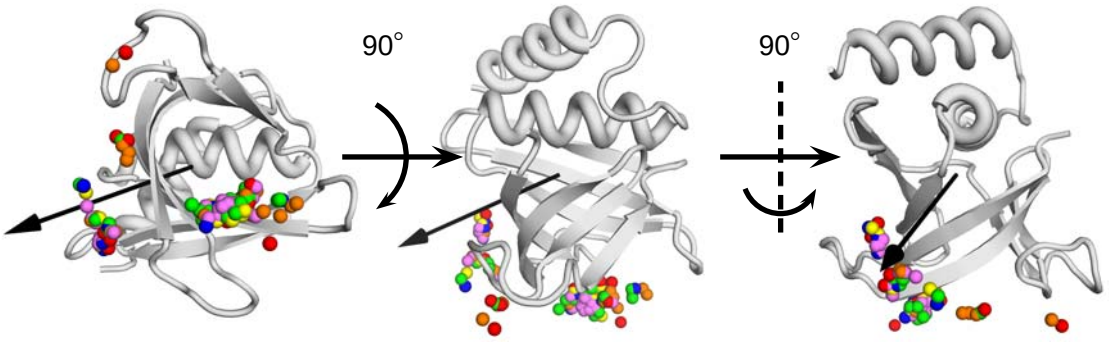

Figure 5.19: Spatial distribution of conformational clusters around the target protein for PIP docking to the open conformation of kindlin-1 PH (white). Cluster centres of mass are shown as coloured spheres for I(1,3)P_2 (red); I(1,4)P_2 (orange); I(1,3,4)P_3 (yellow); I(1,3,5)P_3 (green); I(1,4,5)P_3 (blue) and I(1,3,4,5)P_4 (purple). The molecular dipole moment is shown as a black arrow.

the basis of the docking results, it is reasonable to suggest that kindlin-1 PH is capable of binding to p-Tyr. However, the errors associated with predicting free energies of binding from docking discussed above should be borne in mind, and so it is not possible to infer specificity or state with any certainty whether or not p-Tyr binding is stronger than PI monophosphate binding to kindlin-1 PH.

The positively charged surface presented by the kindlin-1 PH domain, together with its binding to the β integrin NxxY motif through the F3 subdomain, may perhaps be sufficient to localise it to the membrane [354] without the need for PIP lipid binding. For example, it has been reported that a point mutation abolishing PI(3,4,5)P_3 binding by kindlin-2 does not affect focal adhesion (FA) localisation [333]. If PIP/p-Tyr binding by the kindlin-1 PH domain is not necessary for successful binding to integrins then the precise functional role of the PH domain in kindlin-1 remains unclear. Another possibility, as discussed in Chapter 1, is that perhaps kindlin-1 PH is simply one of the 90% of PH domains that do not exhibit strong PIP binding [39].

With these rather ambiguous results in hand, at first glance one might think that the PH domain in kindlin-1 could just be an evolutionary relic, given that talin for example is capable of binding to membranes and activating integrins without such a domain. Interestingly though, functional studies performed by Brahme and Calderwood [355] suggest

that kindlin-1 PH does have some role in integrin activation, but the nature of this role is not entirely clear. As anticipated, integrin activation assays showed that kindlin-1 enhanced activation of the αIIbβ3 integrin in Chinese hamster ovary (CHO) cells when co-expressed with talin. Activation was significantly impaired when an expression construct lacking the PH domain (denoted kindlin-1ΔPH) was used instead of full length kindlin-1. This perhaps indicates a role for the PH domain, as co-expression of talin with the kindlin-1ΔPH construct did lead to slightly enhanced activation relative to that observed for talin alone. However, when expressed alone, kindlin-1 PH was unable to activate αIIbβ3 integrin and indeed when co-expressed with talin it did not significantly enhance activation.

In summary, the PH domain of kindlin-1 clearly has some part to play in integrin activation, but additional work will be required to identify its functionally relevant physiological ligand and to define its precise role *in vivo*.

Chapter 6

Simulations of the PTEN tumour suppressor

HAVING spent the previous three chapters describing work on the interaction between lipids and single PH domains, this chapter goes on to consider a membrane-binding protein with more than one domain and asks whether or not the methods used previously can be extended and applied to this more complex case. Here, simulation studies of the core of the PTEN tumour suppressor are described. PTEN has a two-domain architecture and binds to PI(3,4,5)P_3 at the plasma membrane. PTEN is larger and has a more elaborate fold than the PH domains investigated previously, and is also one of the most frequently mutated proteins in human cancer [368].

6.1 Introduction

6.1.1 PI3-K signalling and PTEN

The plasma membrane and its constituent PIPs described in the preceding chapters form the bedrock of the PI3-K signalling pathway, which is crucial for cell proliferation and

survival [282]. PI3-K signalling primarily relies upon two lipid species, PI(4,5)P_2 and PI(3,4,5)P_3, which are differentiated only by the presence or absence of a phosphate group at the 3′ position of the inositol ring. Activation of PI3-K is dependent upon biosynthesis of PI(3,4,5)P_3 from PI(4,5)P_2 by the PI3-K enzyme. The newly synthesised PI(3,4,5)P_3 can then provide a stable membrane anchor to recruit other proteins to the surface, for example Akt that binds PI(3,4,5)P_3 through its PH domain.

Akt is an essential downstream effector in PI3-K signalling [66] that regulates cell proliferation and survival by inhibiting apoptosis. Membrane localisation *via* PI(3,4,5)P_3 binding is essential for Akt activation, and so any factors that enhance Akt binding to PI(3,4,5)P_3 can result in pathological membrane localisation and, by extension, uncontrolled activation of Akt. This in turn leads to overall dysregulation of PI3-K signalling and has been linked to the development of several cancers[1]. Pathological membrane localisation can be caused by a mutation in Akt itself [348], or alternatively it can occur as a consequence of elevated levels of PI(3,4,5)P_3. Production of PI(3,4,5)P_3 is regulated by the PTEN tumour suppressor, which negatively regulates PI3-K signalling by converting PI(3,4,5)P_3 back to PI(4,5)P_2 keeping the basal level of PI(3,4,5)P_3 low [371]. Under normal circustances PTEN therefore attenuates the survival signal from Akt, but loss of function mutations in PTEN can lead to overproduction of PI(3,4,5)P_3 and uncontrolled PI3-K activation.

6.1.2 Discovery of PTEN and its role as a tumour suppressor

PTEN was first discovered on chromosome 10q23 in 1997, and the protein was rapidly pinpointed as a putative tumour suppressor after mutant PTEN was detected in a broad range of sporadic tumour types [372, 373]. PTEN mutation occurs at particularly high frequency in endometrial cancer and the aggressive and malignant brain tumour glioblas-

[1]Evasion of apoptosis was highlighted as one of the 'hallmarks of cancer' in the landmark reviews of progress in cancer research by Hanahan and Weinberg in 2000 [369] and more than a decade later in 2011 [370].

toma multiforme [368, 374]. Soon afterwards the phosphatase activity of PTEN was identified [375] and work with the cancer derived mutant G129E in human embryonic kidney (HEK293) cells established that the *lipid* phosphatase activity of PTEN was responsible for its role as a tumour suppressor, as the G129E mutant is lipid phosphatase inactive but retains its ability to dephosphorylate other proteins [376]. Lipid phosphatase mediated tumour suppression and regulation of PI3-K signalling is therefore thought to be the primary role of PTEN.

6.1.3 PI3-K and PTEN as therapeutic targets

With its key role in PI3-K signalling, it might be expected that PTEN would be a major target for small molecule drug discovery. However, PTEN has received comparatively little attention as a potential drug target, chiefly as the loss of phosphatase activity due to mutation is difficult to reverse through direct therapeutic intervention using small molecules[2]. Most research on rescue of PI3-K signalling by restoring the function of defective PTEN has therefore revolved around gene therapy using adenoviruses to induce exogenous overexpression of PTEN, which has been shown to be effective in bladder [380], prostate [381, 382] and colorectal [383] cancers.

Rather than trying to recover the regulatory activity of PTEN, an alternative approach is simply to inhibit PI3-K signalling. Small molecule drug discovery efforts have either focussed on inhibition of Akt or have targeted the PI3-K enzyme itself in an attempt to cripple the kinase [384]. The first generation PI3-K inhibitor wortmannin was extremely effective in the laboratory, but clinically it was found to have poor biological stability and high liver and haematologic toxicity [385]. Despite this, several second generation PI3-K inhibitors including derivatives of wortmannin are now in active development and some such as PX-866 have entered clinical trials[3] [387].

[2]Although interestingly PTEN has also been implicated in insulin signalling, with loss of PTEN leading to insulin hypersensitivity. This raises the prospect of using PTEN inhibition as a possible strategy for treatment of type II diabetes [377, 378, 379]

[3]For more information on PI3-K inhibitors, including their possible role as antiangiogenic agents, read-

In the case of Akt, there are several inhibition strategies, as it is possible to engineer competitive inhibitors for binding of ATP [388], design allosteric inhibitors that prevent phosphorylation [389, 390], or produce small molecules that inhibit PI(3,4,5)P_3 binding. The first two approaches have met with some success, but in spite of encouraging preclinical data for molecules such as perifosine [391], triciribine [392] and a class of sulfonamide derivatives [393] the development of effective PI(3,4,5)P_3 inhibitors has been beset with problems [393]. These molecules typically need to carry a large negative charge to outcompete PI(3,4,5)P_3, and/or a long aliphatic chain to aid localisation to the plasma membrane. The necessity of retaining these features in any candidate compound renders subsequent drug development challenging. In addition, an unintended consequence of the use of PI(3,4,5)P_3 analogues to inhibit Akt is that they may also have off-target drug interactions with PTEN itself, exacerbating the problem of aberrant PI3-K signalling.

6.1.4 Structure of PTEN

The full-length PTEN protein comprises approximately 400 amino acid residues and is subdivided into four main regions: the N-terminal PI(4,5)P_2 binding module (N-PBM); the phosphatase domain; the C2 domain and a ∼50 amino acid residue C-terminal tail (C-TT) (Figure 6.1). A crystal structure of truncated PTEN at 2.1 Å resolution that includes just the phosphatase domain (residues 14-185) and the C2 domain (residues 186-351) was reported in 1999 (PDB 1D5R [371]). The crystal structure of truncated PTEN does not incorporate the N-PBM nor the C-TT, and there is also a missing loop of 31 residues in the C2 domain between P281 and K313.

The canonical C2 domain fold has three loops at the base of the eight stranded antiparallel β sandwich, denoted as Ca^{2+} binding region 1 (CBR1), CBR2 and CBR3. As discussed in Chapter 1, most C2 domains bind to membranes in a Ca^{2+}-dependent man-

ers are referred to the review by Kong and Yamori [386].

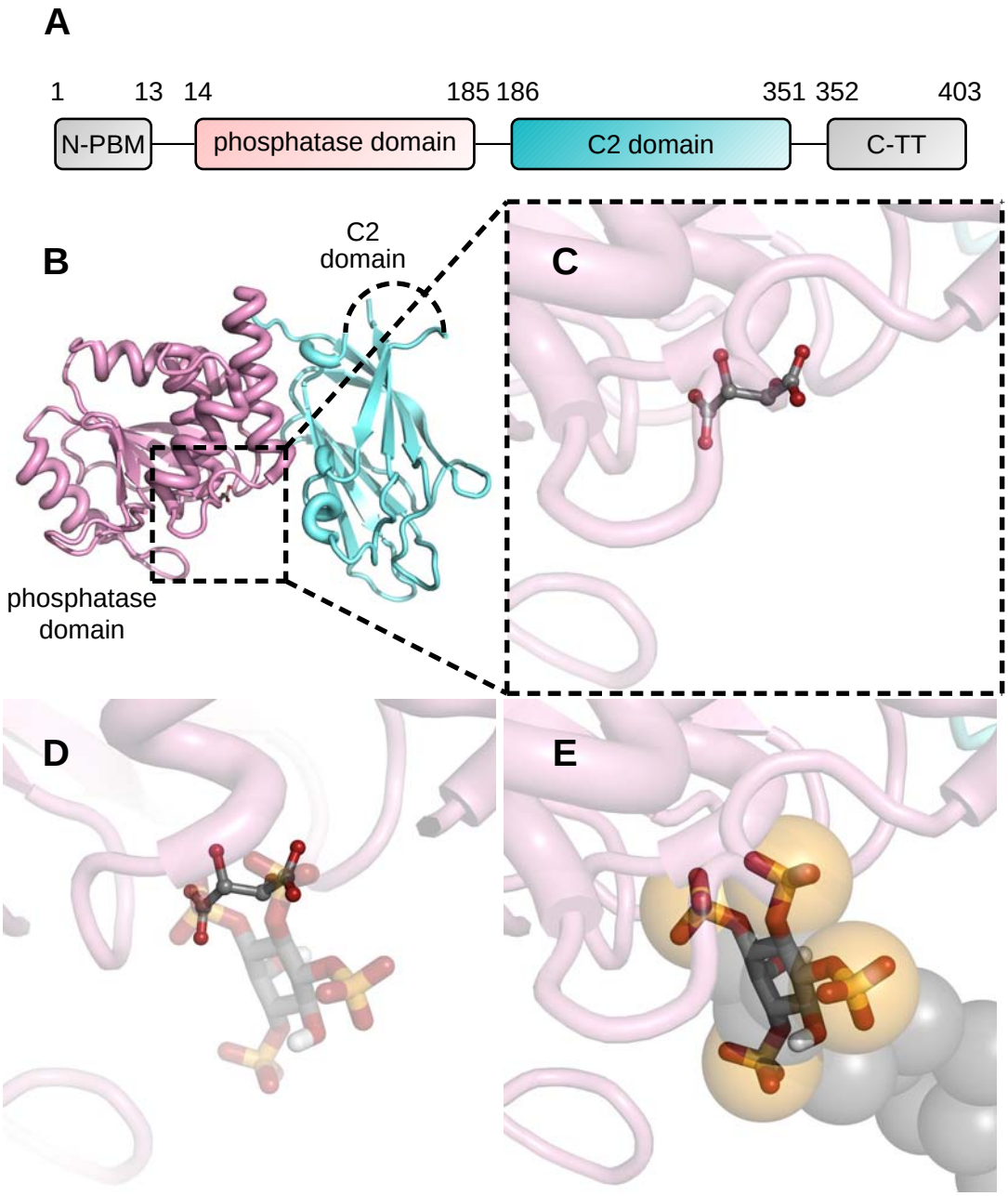

Figure 6.1: PTEN structure and model building. **(A)** Schematic of the domain architecture of PTEN, showing the N-terminal PI(4,5)P_2 binding module (N-PBM) in grey, the phosphatase domain in pink, the C2 domain in cyan and the ∼50 amino acid residue C-terminal tail (C-TT) also in grey. **(B)** The crystal structure of the core domain of the PTEN tumour suppressor (PDB 1D5R) [371]. The phosphatase domain is shown in pink, while the C2 domain is shown in cyan. The missing loop in the C2 domain of the crystal structure between P281 and K313 is shown as a dashed line. **(C)** The location of the bound tartrate (shown as a stick model) suggests a location for the PI(3,4,5)P_3 binding site. **(D)** Modelling the atomistic headgroup of PI(3,4,5)P_3 into the binding site by alignment of the phosphorus atoms at the 3′ and 4′ positions of the inositol ring with the carboxylate carbon atoms of the tartrate, as suggested by Lee *et al.* [371]. **(E)** CG model of PI(3,4,5)P_3 shown as van der Waals spheres overlaid onto the atomistic model.

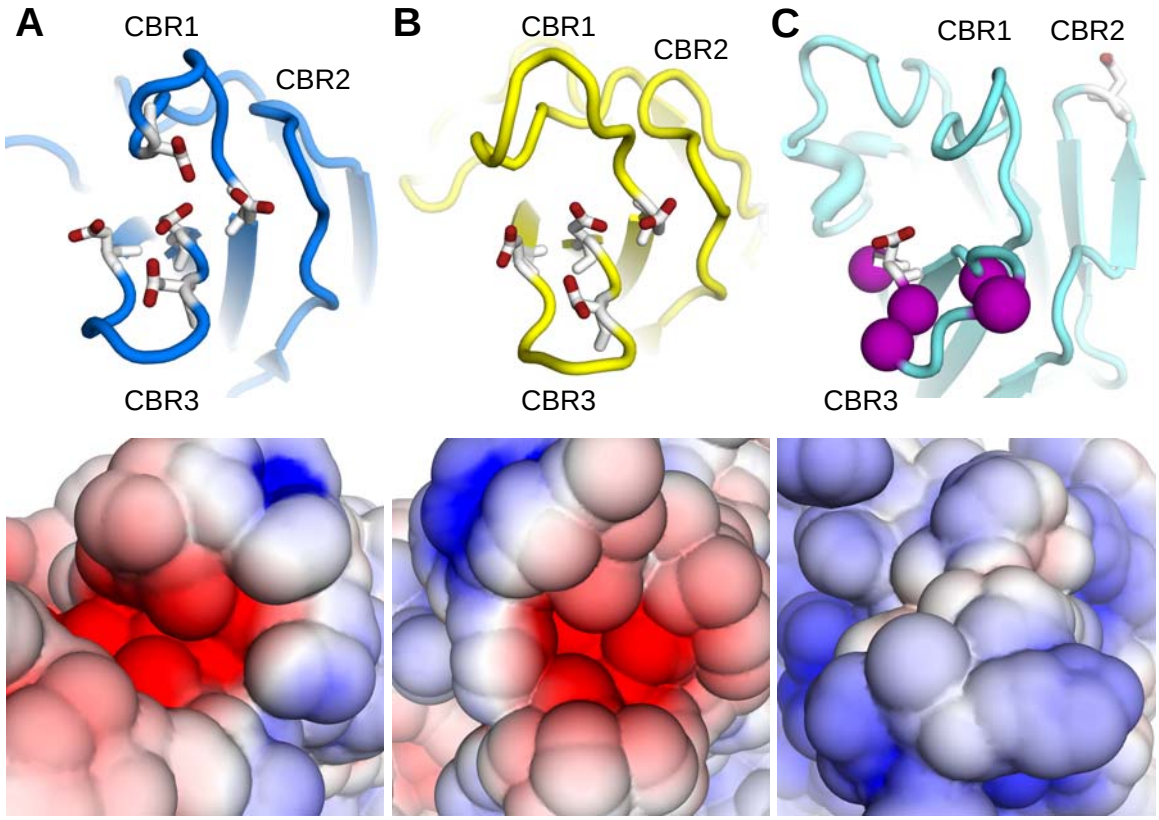

Figure 6.2: Coordination of divalent ions in canonical C2 domains compared with the PTEN C2 domain. C2 domains of **(A)** synaptotagmin III (PDB 1DQV) [394] shown in blue, **(B)** PLCδ1 (PDB 1DJI) [30] shown in yellow and **(C)** PTEN (PDB 1D5R) [371] shown in cyan. Corresponding electrostatic potentials projected onto the relevant molecular surface are shown in the lower panel for each case ($\pm$ 5*kT/e*). The C2 domains of synaptotagmin III and PLCδ1 possess multiple aspartate ligands (shown as stick representations) to coordinate Ca^{2+} ions, but the PTEN C2 domain lacks all but one of these. The Cα atoms of K260, K263, K266, K267 and K269 in CBR3 of the PTEN C2 domain are shown as purple van der Waals spheres.

ner [124], using an array of aspartate ligands within the CBRs that are able to coordinate the metal centre [40]. Examples are shown in Figure 6.2 for the C2 domains of synaptotagmin III and PLCδ1. Unusually, the PTEN C2 domain binds to membranes in a Ca^{2+}-independent fashion [371], as it lacks all but one of these critical aspartate ligands (Figure 6.2). The electrostatic potential around the three Ca^{2+}-binding regions reveals how, in the case of synaptotagmin III and PLCδ1, Ca^{2+} binding is required to mask the large negative electrostatic potential around the Ca^{2+} binding loops. This is an example of an 'electrostatic switch' mechanism of membrane targeting [125], where membrane affinity is controlled by Ca^{2+} binding. In the absence of Ca^{2+} in the unbound state,

electrostatic repulsion between the negatively charged C2 domain and the negatively charged membrane surface prevents lipid binding. When the CBRs are occupied by Ca^{2+}, the positively charged Ca^{2+} ions screen the negatively charged CBRs and allow membrane binding. In contrast, the PTEN C2 domain is predominantly positively charged in the vicinity of the three CBRs, due in part to the cluster of five lysine residues within CBR3 (Figure 6.2). As a result, the C2 domain of PTEN does not rely on an electrostatic switch mechanism to bind to membranes, and so the paucity of Ca^{2+}-binding aspartate residues in the CBRs does not adversely affect its membrane binding affinity.

6.1.5 Aims

Although PTEN is a key component of PI3-K signalling, little is known about its membrane-associated state. Previous computational work on PTEN has included molecular docking and short MD simulations of 3-deoxy-phosphatidylinositol inhibitors to soluble PTEN [395], and finite-difference Poisson-Boltzmann electrostatics calculations to probe the interaction between the C2 domain of PTEN and the membrane [125]. However, no MD simulations of PTEN in the presence of a membrane environment have yet been reported.

This chapter describes the construction of a model of the membrane-bound state of the PTEN tumour suppressor based on the truncated crystal structure. A serial multiscale simulation approach is used to generate the complex and probe its dynamics. The advantage of such an approach is that it allows exploration of the dynamics over multiple timescales, so it becomes possible to investigate complex formation and lipid clustering at a coarse grained level before switching to atomistic resolution to examine the protein-lipid contacts in more detail (Figure 6.3). The aim was to investigate the factors promoting the stability of the PTEN-membrane complex by identifying those regions of the protein that interact with the membrane to the greatest extent, and testing how this interaction was perturbed in a variety of mutants (Table 6.1). There is substantial experimental evidence [371, 396] to suggest that particular regions of the protein

Protein	Lipid composition	Resolution	Length / ns	Multiplicity
wild type	POPC(100%)	CG	500	3
...	...	atomistic	50	1
...	...	...	200	1
...	POPC(80%):POPS(20%)	CG	500	3
...	...	atomistic	50	1
...	POPC(60%):POPS(40%)	CG	500	3
...	...	atomistic	50	1
R161E/K163E/K164E	POPC(100%)	CG	500	3
...	...	atomistic	50	1
...	POPC(80%):POPS(20%)	CG	500	3
...	...	atomistic	50	1
...	POPC(60%):POPS(40%)	CG	500	3
...	...	atomistic	50	1
R335L	POPC(100%)	CG	500	3
...	...	atomistic	50	1
...	POPC(80%):POPS(20%)	CG	500	3
...	...	atomistic	50	1
K260E/K263E/K266E/ K267E/K269E	POPC(80%):POPS(20%)	CG	500	1
...	POPC(60%):POPS(40%)	CG	500	1

Table 6.1: Summary of PTEN simulations. All simulations contained PI(3,4,5)P_3 and 300 lipids, with approximately 10000 water particles (CG) or around 20000 solvent molecules (atomistic) with counterions added as appropriate to neutralise the net charge of the system. This amounts to 13 μs of CG simulation and 0.6 μs of atomistic simulation in total. See also Figure 6.3.

far removed from the PI(3,4,5)P_3 binding site are crucial for robust membrane attachment, and the resulting membrane-bound model agrees well with experimental results concerning protein-lipid interactions.

6.2 Methods

The initial model of membrane-bound PTEN was generated using the crystal structure of truncated PTEN (PDB 1D5R [371]) as a basis. In the atomistic simulations, a harmonic restraint with spring constant 100 pN $Å^{-1}$ was placed between the Cα atoms of P281 and K313 to maintain the separation observed in the crystal structure. In the CG simulations,

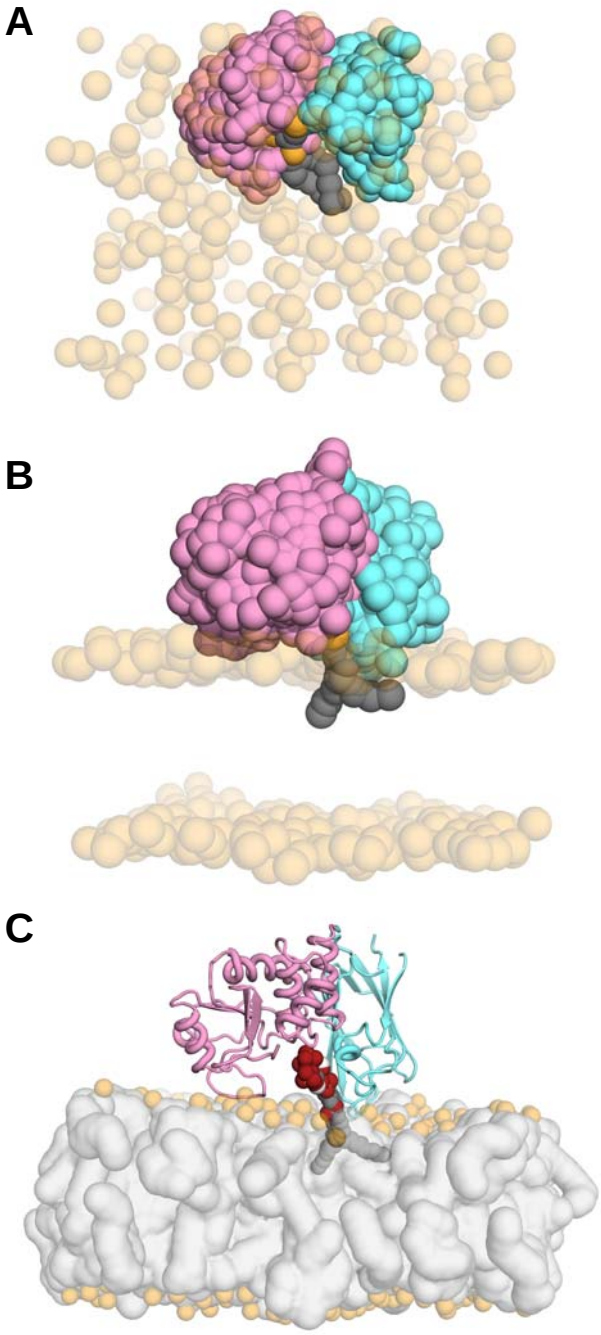

Figure 6.3: PTEN simulation setup and workflow. **(A)** CG self assembly simulations were conducted using the model of PI(3,4,5)P_3 bound to the crystal structure of PTEN. A CG representation of the protein was generated and was placed along with the bound PI(3,4,5)P_3 into a cubic box of dimension 120 $\times$ 120 $\times$ 120 Å with 300 randomly placed lipids. Lipid headgroups are shown as transparent orange van der Waals spheres. **(B)** PTEN-membrane complex assembly after 500 ns of CG MD simulation. **(C)** Subsequent conversion to an atomistic representation using a fragment based protocol [279]. This system was then simulated for 50 ns. See also Table 6.1

a Gaussian network model is employed to maintain the secondary structure and so this additional restraint was not required. This region is far removed from the PI(3,4,5)P_3 binding site and the proposed membrane binding surface, and was not observed to interact with bilayer lipids during any of the simulations (Figure 6.4).

The protein was crystallised in the presence of tartrate, and a tartrate molecule is bound at the active site. The authors of the original paper in which the structure was reported observed that the separation between the two carboxylate carbon atoms of the tartrate molecule is approximately the same as that between the phosphorus atoms at the 3′ and 4′ positions of the inositol ring of PI(3,4,5)P_3. These geometric considerations provided a guide when constructing the initial CG model of PI(3,4,5)P_3-bound PTEN, positioning the CG phosphate particles at the coordinates of the carboxylate carbons (Figure 6.1). All MD simulations were carried out using GROMACS version 4.0.5 [186].

CG self assembly simulations of 500 ns duration were followed by atomistic simulations of 50 ns duration. Each CG system comprised ∼300 lipids and ∼10,000 water particles with monovalent ions added as appropriate to compensate for any net charge, leaving the system neutral overall. As the binding of PTEN to membranes is independent of Ca^{2+}, calcium ions were not included in the simulations. CG simulations were performed using the MARTINI force field [175, 176]. In MARTINI, zwitterionic lipids such as POPC are approximated by a positively charged particle (choline), a negatively charged particle (phosphate), two polar particles (glycerol) and two acyl chains made up of four and five hydrophobic particles respectively. Anionic lipids such as POPS are treated in a similar fashion except that the positively charged particle is replaced with a polar particle to represent the switch from choline to serine. The lipid headgroup then carries an overall net negative charge. Owing to the random initial placement of the lipids and the stochastic nature of the self assembly procedure, anionic lipids may be present in both leaflets of the bilayer whereas in physiological membranes anionic lipids are typically restricted to the cytoplasmic leaflet of the plasma membrane. An analysis of

the 17 CG simulations in which anionic lipids were used gives a mean ratio (upper:lower) plus or minus one standard deviation of 1.103 ± 0.151 for POPS. The lipid distribution across the two leaflets is therefore virtually symmetrical, but there are typically a few more POPS lipids in the upper leaflet than in the lower. This is probably due to interactions with the protein during the self assembly procedure, with the negatively charged POPS lipids slightly favouring the upper leaflet where the protein is located.

For the CG simulations a cut off of 12 Å for both the Lennard-Jones and the electrostatic interactions was used. Simulations were performed under periodic boundary conditions and temperature was kept constant at 323 K by coupling the system to a heat bath using a Berendsen thermostat [159] with $\tau_T = 1$ ps. Pressure was maintained at 1 atm using a Parrinello-Rahman barostat [162, 163] and semi-isotropic pressure coupling with $\tau_p = 1$ ps and a compressibility of 5×10^{-6} bar^{-1}. A time step of $\Delta t = 10$ fs was used, writing atomic positions every 10 ps. The neighbour list was updated every ten steps. The final frame of each CG simulation was converted to an atomistic representation of the system using a fragment based approach [279].

Atomistic simulations were performed with the GROMOS96 43a1 force field [142]. A cut off of 10 Å was used for the Lennard-Jones interactions and electrostatic interactions were treated using the PME [235] approach with a short range real space cut off of 10 Å. Atomistic simulations were performed using the same parameters as for the CG simulations, except that T = 296 K, $\tau_T = 0.1$ ps and compressibility was set to 4.6×10^{-5} bar^{-1}. For the atomistic simulations, bond lengths and angles were constrained using the LINCS algorithm [139] and a timestep of $\Delta t = 2$ fs was used.

Electrostatic potentials were computed on a cubic grid of dimensions 129 Å × 129 Å × 129 Å with a grid spacing of 1 Å using APBS [210] at an ionic strength of 0.1 M and a temperature of 300 K.

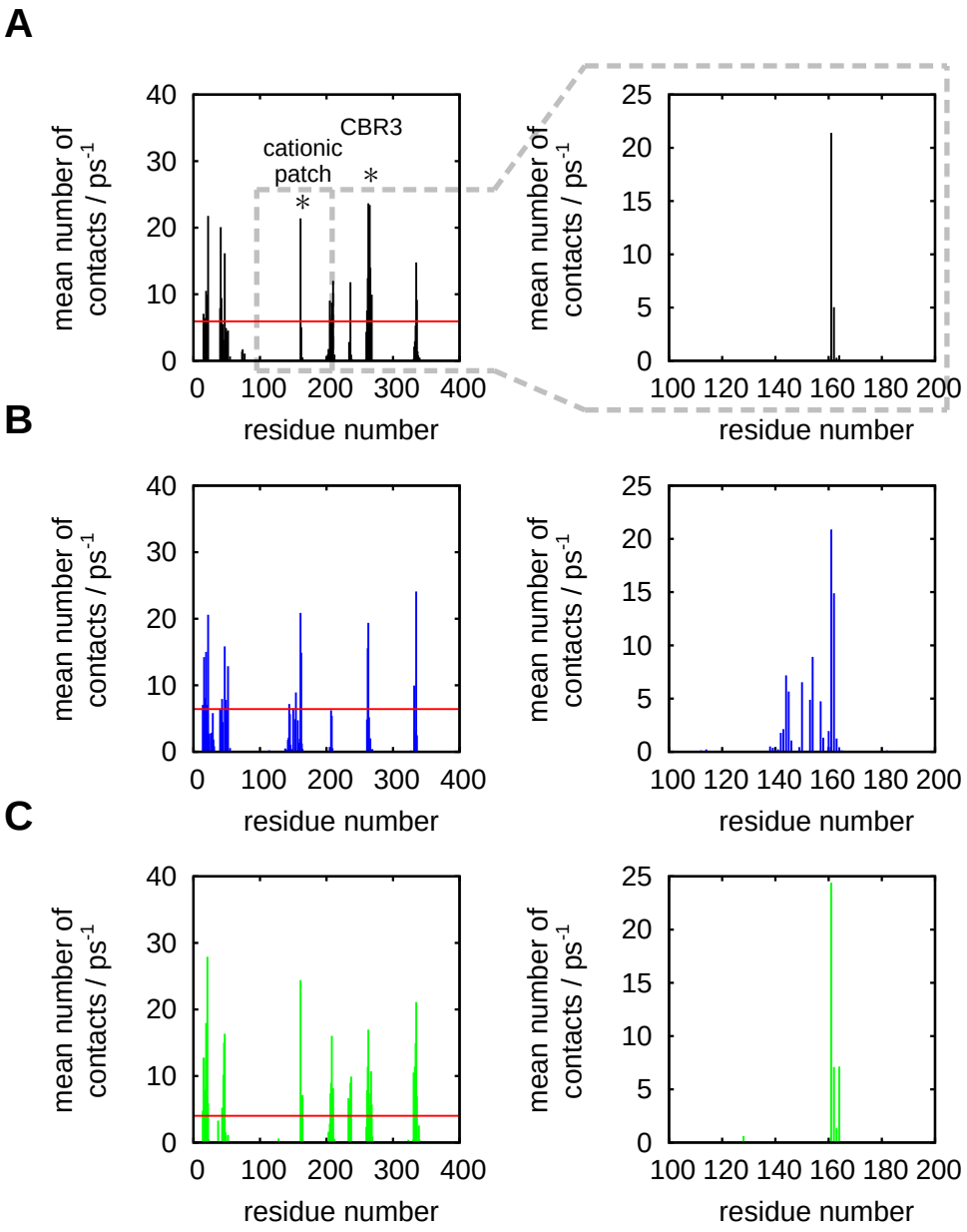

Figure 6.4: Protein-lipid contacts in wild type PTEN. **(A)** The mean number of protein-lipid contacts per residue is shown for the atomistic simulations with a POPC bilayer, **(B)** a POPC(80%):POPS(20%) bilayer and **(C)** a POPC(60%):POPS(40%) bilayer. The region containing R161, K163 and K164 is indicated and expanded in the right panel. Contacts were computed between the protein and lipid heavy atoms using a 4 Å cut off and averaged over the 50 ns trajectory. The horizontal red line shows the average plus one standard deviation for the distribution. Any residue exceeding this number of contacts is deemed to interact with the lipids to a significant extent.

6.3 Membrane-binding surface of PTEN

Multiple sequences of MD simulations were carried out for the wild type and mutant proteins using different compositions of bilayer lipids (Figure 6.3 and Table 6.1). Bilayers

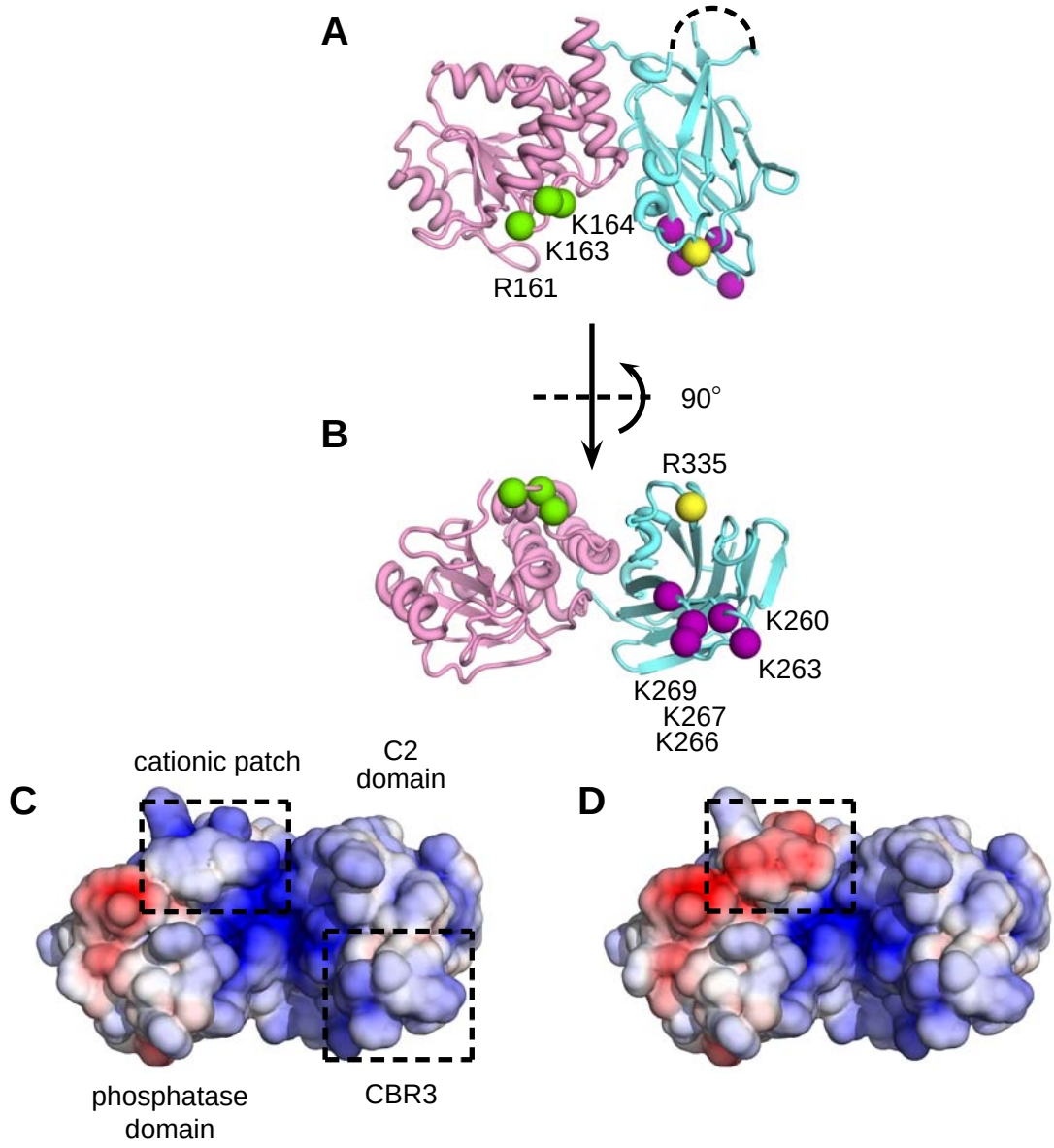

Figure 6.5: PTEN mutants and electrostatic potentials. **(A)** Structure of PTEN with the residues selected for mutation *in silico* shown as van der Waals spheres. The cationic patch (R161, K163, K164) is shown in green, CBR3 (K260, K263, K266, K267 and K269) in purple and R335 is shown in yellow. **(B)** Structure of PTEN rotated by 90°, showing the location of these residues on the membrane binding face of the protein. **(C)** Electrostatic potential of wild type PTEN projected onto its solvent-accessible surface. The regions around the cationic patch (R161, K163 and K164) and CBR3 (K260, K263, K266, K267 and K269) are both highlighted. **(D)** Electrostatic potential of the triple mutant R161E/K163E/K164E PTEN projected onto its solvent-accessible surface. The cationic patch is once again highlighted, illustrating the switch in polarity from positive to negative in the triple mutant. Electrostatic potentials are coloured from -5*kT/e* (red) to +5*kT/e* (blue) and are shown in the same orientation as **(B)**.

were composed of either purely zwitterionic lipids (100% POPC and its CG equivalent) or alternatively two different mixtures of anionic and zwitterionic lipids (POPS(20%):POPC(80%), POPS(40%):POPC(60%) and their CG equivalents). PTEN is able to bind to both purely

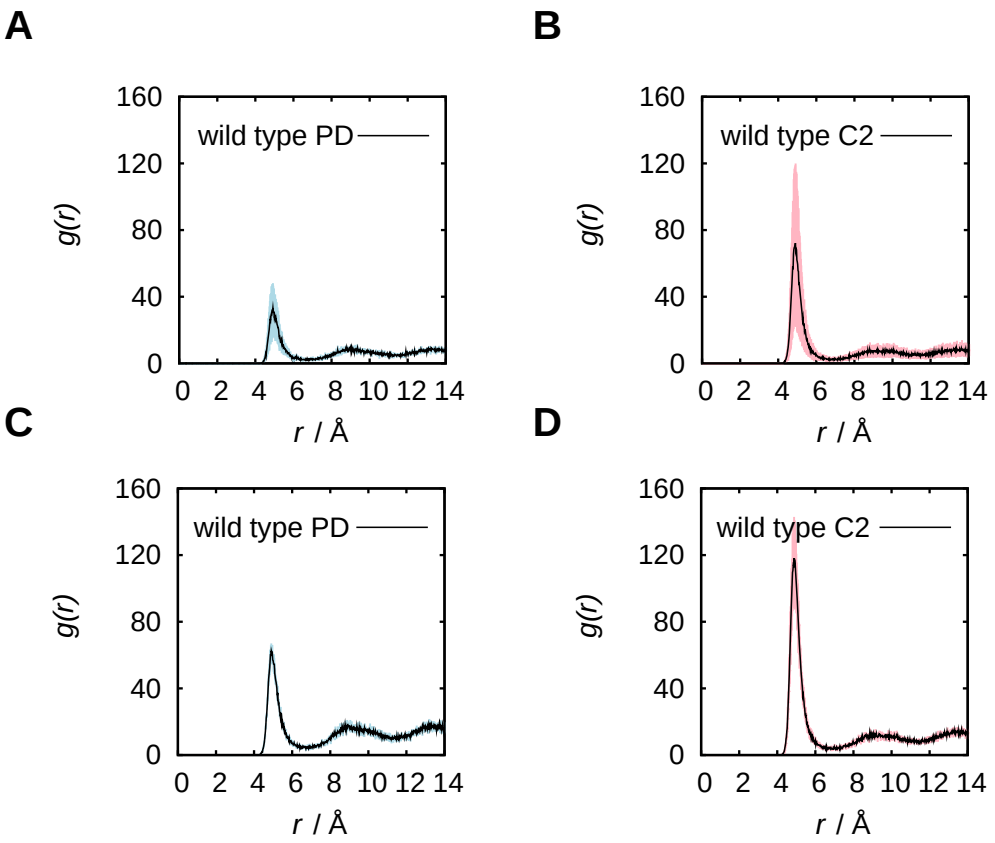

Figure 6.6: Anionic lipid RDFs for the wild type PTEN CG MD simulations. Mean RDFs for POPS calculated from three independent simulations are shown for **(A)** the wild type phosphatase domain and **(B)** the wild type C2 domain using the POPC(80%):POPS(20%) bilayer composition. For the case of the POPC(60%):POPS(40%) bilayer the RDF is again shown for **(C)** the phosphatase domain and **(D)** the C2 domain. The RDFs were computed for POPS lipids in the upper leaflet of the bilayer post self assembly over the time interval 100–500 ns. Shading represents the uncertainty in the RDFs, plus or minus one standard deviation from the mean calculated from three independent simulations.

zwitterionic and mixed zwitterionic/anionic lipid vesicles *in vitro* [396], though its binding affinity for the latter is higher.

After conversion to an atomistic representation and 50 ns of MD simulation, protein-lipid contacts between the heavy (*i.e.* non-hydrogen) atoms were computed over the course of the MD trajectory for each residue (Figure 6.4). In all cases, the strongest interactions between the protein and the lipid bilayer were observed for those regions of truncated PTEN previously shown by experiment to be most important for membrane binding, namely the cationic patch, R161, K163 and K164 [396], and the basic residues in CBR3: K260; K263; K266; K267 and K269 [397] (Figure 6.5). The binding interface is generally well conserved between the three atomistic simulations using different lipid

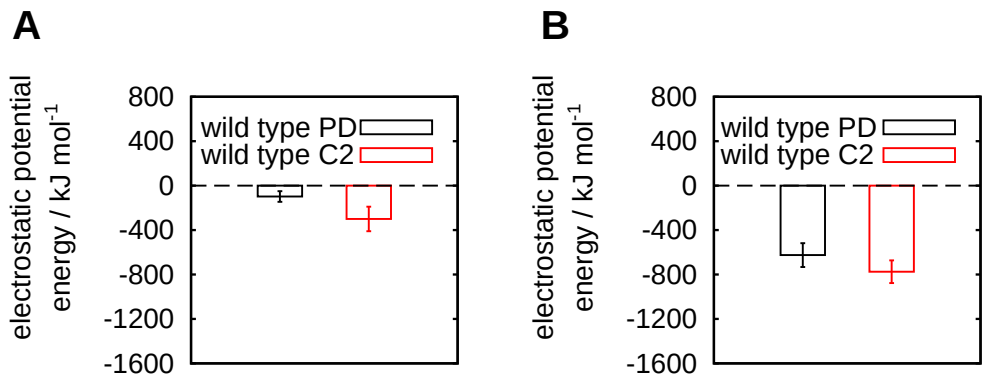

Figure 6.7: Potential energy of interaction between PTEN and POPS. The mean short ranged component of the Coulomb potential energy was computed between the protein and the anionic POPS lipids over the 50 ns atomistic MD trajectory. Interaction energies between the wild type protein and POPS are shown for the simulations using bilayers of composition **(A)** POPC(80%):POPS(20%) and **(B)** POPC(60%):POPS(40%). Error bars are ±1 standard deviation from the mean potential energy over the trajectory.

bilayer compositions, suggesting that the model is reasonably robust. This, coupled with the fact that the model of the PTEN-membrane complex is able to reproduce experimental results, suggests that it could be a reasonable approximation for the physiological membrane-bound state.

6.4 Clustering of anionic lipids

It has been suggested that the C2 domain of PTEN is responsible for positioning the phosphatase domain in an optimal orientation for binding to PI(3,4,5)P_3 [398], and that it enhances the interaction of the phosphatase domain with the membrane [396] possibly by non-specific binding to monovalent anionic lipids [125, 15]. This is likely due to the positive electrostatic potential around the three CBRs of the PTEN C2 domain (Figure 6.5). To investigate this, the radial distribution function (RDF) of the anionic lipids in the upper leaflet of the lipid bilayer for the CG simulations was computed (Figure 6.6). Although clustering of anionic lipids around both domains is observed relative to the bulk, anionic lipids preferentially cluster around and bind to the C2 domain. The uncertainty in the RDFs is admittedly rather high in the case of POPC(80%):POPS(20%), however it is much lower at higher anionic lipid concentrations (Figure 6.6) indicating

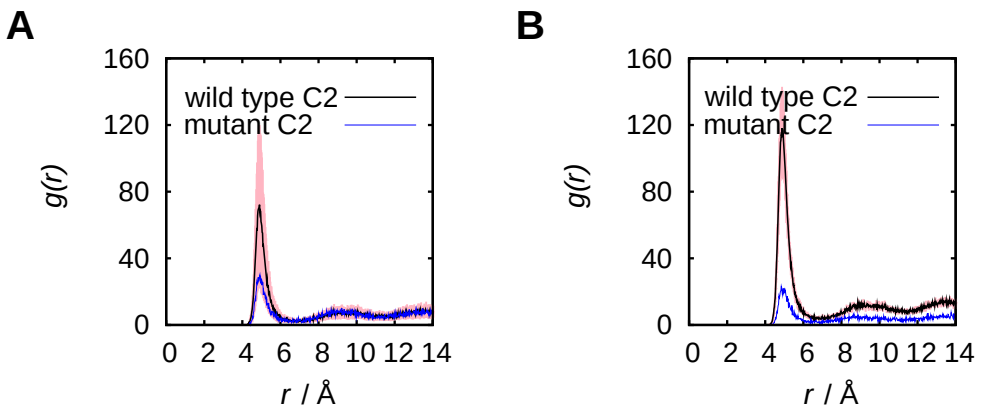

Figure 6.8: CG anionic lipid RDFs for the PTEN C2 domain in the pentuple mutant (K260E/K263E/K266E/K267E/K269E). CG anionic lipid RDFs around the C2 domain are shown for a single 500 ns CG simulation of **(A)** the pentuple mutant with a POPC(80%):POPS(20%) bilayer and **(B)** the pentuple mutant with a POPC(60%):POPS(40%) bilayer. In both cases the pentuple mutant is compared with the corresponding RDFs from the wild type simulations (see also Figure 6.6), and anionic lipid clustering around the C2 domain is greatly reduced in both cases. No substantive changes in the POPS RDFs around the phosphatase domain were observed in the pentuple mutant.

that the elevated levels of anionic lipid clustering around the C2 domain are significant.

The short ranged component of the electrostatic potential energy between the anionic POPS lipids and the protein over the 50 ns atomistic MD trajectories was computed and is shown in Figure 6.7. As expected, the electrostatic interaction energy scales with the fraction of anionic lipids. The electrostatic potential energies also reveal that the interaction energy is higher on average between the C2 domain and the anionic lipids than is the case for the phosphatase domain. Taken together, these results provide evidence that anionic lipids preferentially interact with the C2 domain.

To test the effect of anionic lipid clustering around the C2 domain, the five positively charged residues in CBR3 within the C2 domain were mutated *in silico*. CBR3 has been shown to interact with the bilayer lipids (Figure 6.4), and a limited set of simulations of the resulting charge reversal mutant K260E/K263E/K266E/K267E/K269E show how lipid clustering can be modulated by mutation. In this pentuple mutant, the five-fold substitution of Lys → Glu in the C2 domain substantially reduces the clustering around the C2 domain (Figure 6.8).

Protein	Lipid composition	161+163+164 contacts / ps^{-1}	335 contacts / ps^{-1}
wild type	POPC(100%)	22.2	14.8
...	POPC(80%):POPS(20%)	22.6	24.1
...	POPC(60%):POPS(40%)	32.9	
R161E/K163E/K164E	POPC(100%)	0.1	
...	POPC(80%):POPS(20%)	11.6	
...	POPC(60%):POPS(40%)	0.1	
R335L	POPC(100%)		5.0
...	POPC(80%):POPS(20%)		8.8

Table 6.2: Reduction in protein-lipid contacts in the PTEN mutants. Protein-lipid contacts were computed between the protein and lipid heavy atoms using a 4 Å cut off and averaged over the 50 ns trajectory. Wild type contacts are shown graphically in Figure 6.4, and the table shows the sum of the contacts between residues at positions 161, 163 and 164 with the bilayer lipids, and the same measure for the residue at position 335. The triple charge reversal mutant R161E/K163E/K164E shows a substantial drop in the number of contacts between this region of the protein and the lipids. The charged → non-polar mutation R335L shows a more modest reduction in the number of contacts.

6.5 Mutation reduces protein-lipid contacts

R161, K163 and K164 form a cationic patch at the surface of the phosphatase domain, and previously Das *et al.* [396] established that mutation of these residues to alanine resulted in a 22-fold decrease in membrane binding affinity. Nonetheless, the mutant still had comparable phosphatase activity to wild type PTEN, indicating that this region contributes to membrane binding through non-specific electrostatic interactions with the membrane lipids, rather than by binding to $PI(3,4,5)P_3$ directly. This region was also observed to interact with the lipid bilayer in all three atomistic MD simulations. To test whether or not the model was able to reproduce this reduction in membrane binding the same set of simulations with the *in silico* triple mutant, R161E/K163E/K164E, was performed. The electrostatic potential around the mutant protein as compared with the wild type is shown in Figure 6.5, illustrating how the triple mutant switches the polarity of the cationic patch from positive to negative. The protein-lipid contacts in this region between mutant PTEN and the membrane are substantially reduced for all three lipid

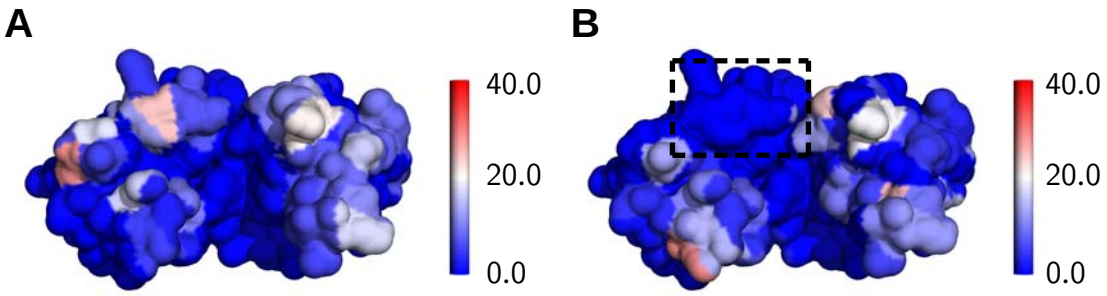

Figure 6.9: Protein-lipid contacts projected onto the solvent-accessible surface of PTEN. The lower (membrane-binding) surface of the protein is shown. Atomistic contacts are shown for **(A)** the wild type protein and **(B)** the triple mutant R161E/K163E/K164E in the POPC(60%):POPS(40%) bilayer system. The location of residues 161, 163 and 164 is indicated by a dashed box. See also Table 6.2 and the electrostatic potentials in Figure 6.5

compositions, in agreement with experimental results (Table 6.2). Figure 6.9 shows the mean number of lipid contacts over the MD trajectory projected onto the solvent-accessible surface of the protein for both the wild type and the triple mutant. Comparison of Figure 6.9 with the electrostatic potentials in Figure 6.5 emphasises the fact that the fall in the number of lipid contacts in this region of the triple mutant is likely to be due to electrostatic effects.

Residues R161, K163 and K164 are located in the phosphatase domain, and anionic lipid clustering around the phosphatase domain is already a comparatively weak effect (Figure 6.6). As a result, the small reduction in the POPS RDF around the phosphatase domain for the R161E/K163E/K164E mutant compared with the wild type is statistically insignificant. This is perhaps due to the fact that the RDF is computed around the whole phosphatase domain and is radially averaged, such that any local density fluctuations are likely to be smoothed out.

To visualise how *local* lipid clustering is modulated in the triple mutant, lipid density maps were computed for the anionic lipids over the course of the CG MD trajectories. Figure 6.10 shows how the local density of POPS changes upon mutation, with a reduction in POPS clustering around the cationic patch observed for both bilayer systems incorporating anionic lipids. Corresponding lipid density maps for the lower leaflet of the bilayer show no such clustering (Figure 6.11), indicating that this local enrichment of POPS rel-

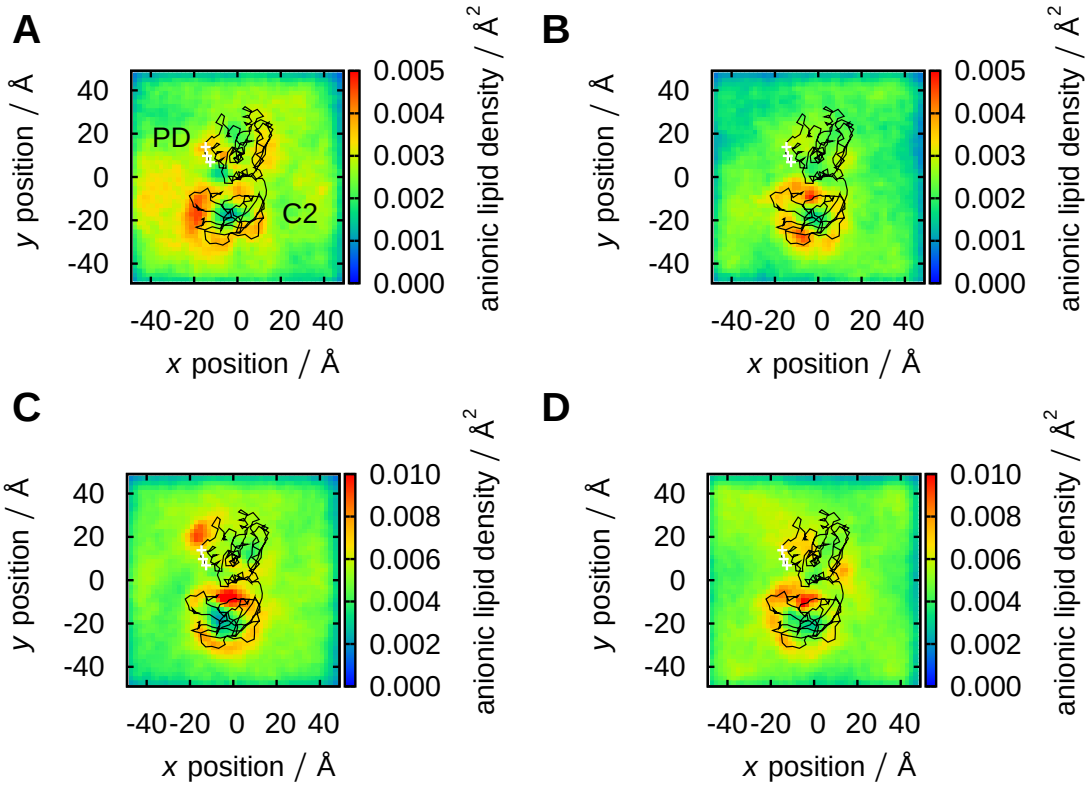

Figure 6.10: Density of POPS in the upper leaflet of the lipid bilayer around PTEN in the system using a bilayer composition of POPC(80%):POPS(20%) for **(A)** the wild type protein and **(B)** the triple mutant R161E/K163E/K164E. Density of POPS in the upper leaflet of the lipid bilayer around PTEN in the system using a bilayer composition of POPC(60%):POPS(40%) for **(C)** the wild type protein and **(D)** the triple mutant. The range of the colour scale in the lower panels has been doubled with respect to that used in the upper panels to reflect the higher POPS concentration. CG trajectories were post-processed to fit the protein to its initial position and orientation in the *xy* plane. Lipid densities were subsequently computed over an interval of 100–500 ns after self assembly of the PTEN-membrane complex had occurred, and distributed into bins of dimension 2 × 2 Å. The resulting plots are average densities computed for the three independent CG simulations in each case. The positions of residues 161, 163 and 164 are indicated by white crosses.

ative to the bulk is almost certainly due to the presence of the protein at the surface, rather than an artefact arising from lipid flow patterns or phase separation. The reduced interaction between the membrane and the phosphatase domain may explain the experimental observation that the membrane binding affinity of the R161E/K163E/K164E mutant is much lower than that of the wild type despite the fact that it remains catalytically competent and able to bind to PI(3,4,5)P_3.

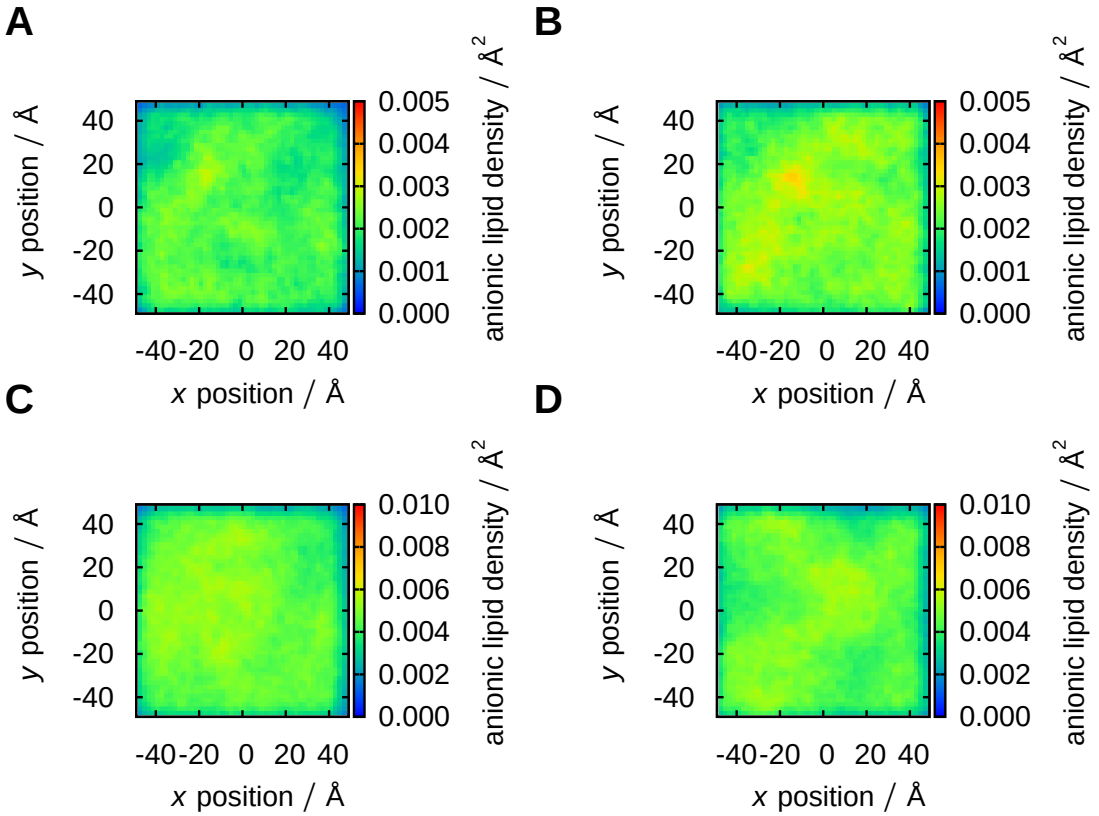

Figure 6.11: Density of POPS in the lower leaflet of the lipid bilayer with a bilayer composition of POPC(80%):POPS(20%) for **(A)** the wild type protein and **(B)** the triple mutant R161E/K163E/K164E. Density of POPS in the lower leaflet of the lipid bilayer with a bilayer composition of POPC(60%):POPS(40%) for **(C)** the wild type protein and **(D)** the triple mutant. The range of the colour scale in the lower panels has again been doubled with respect to that used in the upper panels. Lipid densities were computed over an interval of 100–500 ns after self assembly of the PTEN-membrane complex had occurred, and distributed into bins of dimension 2 × 2 Å. The resulting plots are average densities computed for the three CG simulations in each case. No lipid clustering is evident for the lipids in the lower leaflet of the bilayer. See also Figure 6.10.

6.6 Locations of disease-causing mutations in PTEN

PTEN is one of the most frequently mutated proteins in human cancer [368], so much so that several authors have adopted the alternative nomenclature PTEN/MMAC1 (MMAC1: mutated in multiple advanced cancers) when referring to the protein. As a result there is now a wealth of mutational data available on PTEN. To determine whether or not the model of membrane-bound PTEN was able to shed any light on these data, the contacts observed in the simulations were compared to disease-causing missense mutations extracted from the Human Gene Mutation Database (HGMD) [399] (see Appendix A).

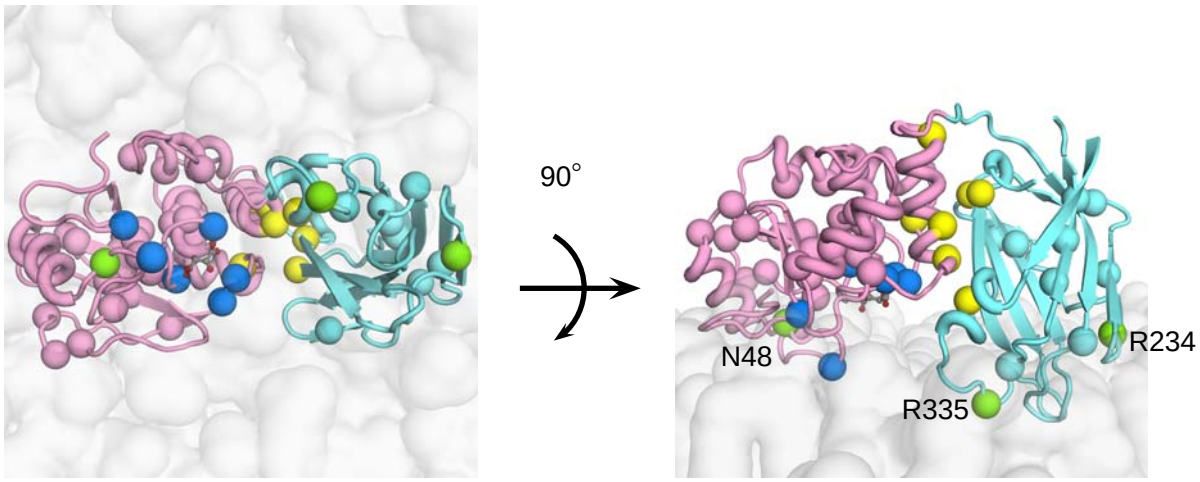

Figure 6.12: Mapping disease-causing mutations onto the PTEN-membrane complex. The X-ray structure of PTEN is superimposed onto the final frame from an atomistic simulation of PTEN in a zwitterionic lipid bilayer, showing how the protein interacts with the membrane. The Cα atoms of each residue for which a missense mutation has been reported (see Appendix A) are shown as van der Waals spheres. Of these residues, those at the interface between the two domains are shown in yellow, while those at the suggested PI(3,4,5)P_3-binding site (the location of the bound tartrate in the crystal structure) are shown in blue. All other mutants are shown in pink or cyan, depending on whether they are located in the phosphatase domain or C2 domain respectively. The three residues that interact with the membrane to the greatest extent (N48, R234 and R335) are coloured green.

Several of the oncogenic and disease-causing mutations in PTEN occur either at the phosphatase/C2 domain interface where they are thought to affect folding, or are hydrolysis defective mutations located at the phosphatase active site (Figure 6.12). However, many mutations occur at locations spatially removed from these regions and it is less clear how these may lead to loss of function and subsequent oncogenesis. The work on membrane-bound GRP1-PH described in Chapter 3 demonstrated the considerable contribution of the surrounding lipid bilayer to the stability of the protein-membrane complex as a whole, in addition to the specific interaction between the protein and its cognate PI(3,4,5)P_3 ligand. With this in mind, one possible explanation for the loss of function associated with these mutations could be that they impair membrane binding in some fashion. It was therefore of interest to map the locations of all disease-causing PTEN mutations reported in the HGMD onto the three dimensional structure of the putative PTEN-membrane complex to determine whether any of these residues could conceivably interact with the membrane (Figure 6.12).

Examining the protein-lipid contacts from the simulations of the PTEN-membrane

complex reveals that three residues listed as sites of disease-causing mutations in the HGMD also interact with the model membrane: N48 [400]; R234 [401] and R335 [402] (Figure 6.12). Of these three residues, R335 was observed to interact with the membrane to the greatest extent across all of the simulations. The mutant R335L, in common with several other germline mutations, has been associated with the inherited cancer syndrome Cowden's disease [403], which results in higher risk of developing breast and thyroid cancer.

The authors of the report describing the R234Q mutation [401] postulated that this mutant may affect the ability of PTEN to associate with other proteins at the membrane surface, owing to its location in CBR2 at the periphery of the protein (Figure 6.12). However, when the novel germline mutation R335L was identified by DNA amplification and sequence analysis [402], the authors did not consider its structural or functional implications. As a result, no mechanism has yet been proposed to describe the molecular origin of disease associated with this mutant. A visual inspection of the PTEN-membrane complex (Figure 6.12) suggests that R335 is unlikely to be involved in any protein-protein interactions, given its position in the loop between the β7 and β8 strands of the eight stranded β sandwich (Figure 6.12) and its proximity to the basic residues in CBR3 that are known to bind to the membrane.

While R335 is unlikely to participate in protein-protein interactions, the simulations suggest that this residue lies in close proximity to the membrane in the PTEN-membrane complex. This, coupled with the charged → non-polar nature of the mutation, makes it reasonable to hypothesise that perhaps the oncogenic nature of the R335L mutant may in part be due to a reduced binding affinity for anionic lipids. To examine this, the simulations for the zwitterionic and 20% anionic lipid systems were repeated with the R335L mutant, and again contacts between the protein and the lipid bilayer were analysed. A modest reduction in protein-lipid contacts for this region of the protein was observed (Table 6.2), providing some support for the hypothesis that a reduction in the interaction

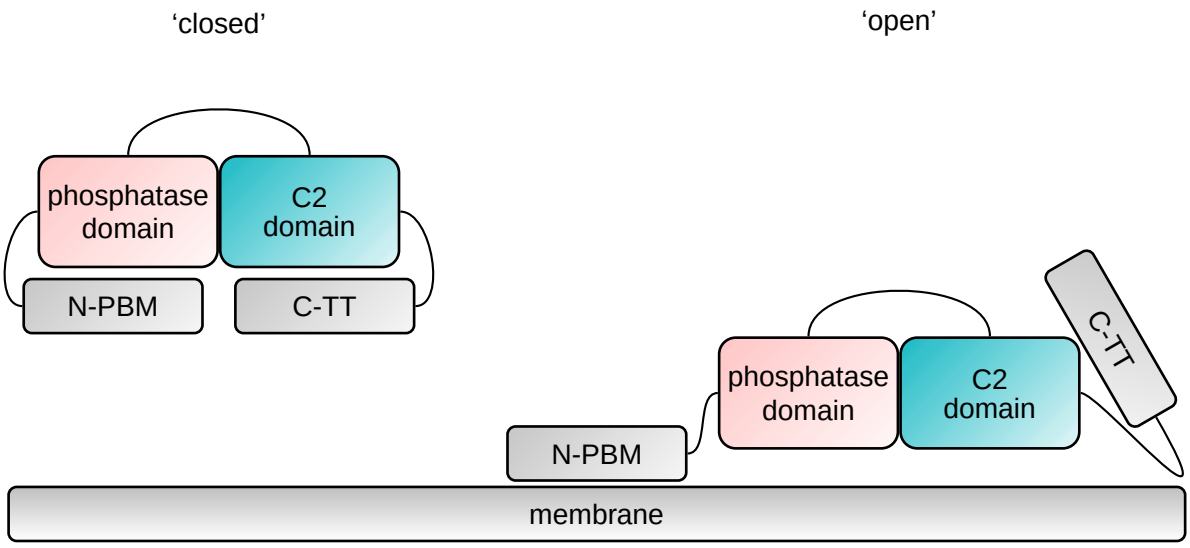

Figure 6.13: Model of closed ⟷ open conformational switching by PTEN proposed by Rahdar *et al.* [397]. In the closed conformation, membrane binding is prevented by self association of the N-PBM and C-TT with the phosphatase domain and C2 domain respectively. Binding of the N-PBM to $PI(4,5)P_2$ coupled with dephosphorylation of the C-TT shifts PTEN into an open conformation capable of binding to the membrane.

between the mutant protein and the membrane contributes to its oncogenicity. Clearly however, further work (both simulation and experiment) would be required to confirm this. Some avenues of further work that could be explored are discussed in Chapter 7.

6.7 Discussion

6.7.1 Membrane interaction of full-length PTEN

Rahdar *et al.* [397] developed a model for the membrane binding of PTEN on the basis of an extensive experimental study that included the N-PBM and the C-TT of PTEN. A schematic of their model is depicted in Figure 6.13. They proposed a mechanism whereby the N-PBM and the phosphorylated C-TT act in an autoinhibitory fashion while the protein is present in the cytosol. This is caused by the N-PBM and the phosphorylated C-TT self associating with the membrane binding surfaces of the phosphatase domain and the C2 domain respectively through electrostatic interactions. In this state the protein is effectively 'closed' and is unable to bind to the membrane. In their model, dephos-

phorylation of the C-TT in tandem with PI(4,5)P_2 binding by the N-PBM is responsible for shifting the protein to an 'open' state. The membrane binding surfaces of the phosphatase domain and C2 domain are then exposed and PTEN is able to fully interact with the membrane.

The lack of the N-PBM and the C-TT in the crystal structure and in the model constructed in this chapter means that it is not possible to capture this closed $\longleftrightarrow$ open transition in the simulations. However, the aim here was simply to model the interaction of the core PTEN fold with a single PI(3,4,5)P_3 molecule at the binding site and a simplified membrane in the absence of more complex lipids such as PI(4,5)P_2. This setup effectively corresponds to the open state described by Rahdar *et al.* without the inclusion of the N-PBM-PI(4,5)P_2 interaction, as the C-TT is not thought to interact with the membrane [404]. However, the isolated phosphatase/C2 domain core of PTEN remains a reasonable basis for a model of the PTEN-membrane complex as the core is catalytically competent and able to metabolise its PI(3,4,5)P_3 substrate in cells [405]. In contrast, as discussed above the N-PBM and C-TT are thought to be involved in membrane targeting and regulation of enzyme activity.

PI(4,5)P_2 binding by the N-PBM is likely to be a contributory factor in membrane binding of PTEN. Expanding the model beyond the truncated PTEN core to incorporate the N-PBM and possibly the C-TT should allow for exploration of the membrane binding role of these regions in the proposed open state. Beginning from this simple model, it should also be possible to introduce more complex membrane compositions that include PI(4,5)P_2 and other lipids identified in recent lipidomics studies [11] to further characterise the membrane-associated state of PTEN under more physiological conditions.

6.7.2 Other PTEN-like domains

Recently, crystal structures of other PTEN-like domains such as auxilin [406] and *Ciona intestinalis* voltage sensing phosphatase (*Ci*-VSP) [407, 408] have been determined. CBR3

of auxilin is much more extended relative to that of PTEN, indicating that it may be able to penetrate the membrane to a greater extent. This apart, auxilin is akin to PTEN in several important respects. The overall fold is extremely similar and CBR3 carries a large positive charge. The latter, combined with a shared lack of aspartate residues in CBR3 for both proteins, suggests that auxilin binding to membranes may operate through a similar Ca^{2+}-independent mechanism. Although auxilin is devoid of catalytic activity with respect to PI(3,4,5)P_3, it is able to bind non-specifically to PI(4)P and PI(4,5)P_2 through its C2 domain, and mutation of arginine and lysine residues to glutamate in CBR3 dramatically reduced its binding affinity for liposomes containing these lipids. This provides further evidence that the C2 domain, and CBR3 in particular, is important for membrane binding in PTEN-like proteins. However, this may not be universally true, as the *Ci*-VSP crystal structure indicates CBR3 carries fewer positively charged residues than the corresponding region in PTEN. CBR3 also packs against the phosphatase domain in *Ci*-VSP instead of adopting the more extended conformation observed in PTEN, again suggesting that its role in membrane binding may be limited.

6.7.3 Ligand migration from the enzymatic cleft

While the model presented here appears able to reproduce the binding interface between PTEN and the membrane surface, unfortunately it is somewhat less successful when considering the protein-ligand interaction between PTEN and PI(3,4,5)P_3. A feature common to all of the simulations is the expulsion of the PI(3,4,5)P_3 ligand from the tartrate binding site of PTEN, which frequently occurs early in the CG MD simulations. Monitoring the separation between the ligand and the catalytic cysteine residue C124 shows the departure of PI(3,4,5)P_3 from this site and its migration over a distance of several Ångströms to alternative sites (Figure 6.14). Furthermore, when switching to atomistic resolution the ligand remains at these sites and the tartrate binding pose observed in the crystal structure is not recovered during the subsequent atomistic MD simulations.

This effect is potentially due to the constricted binding pocket containing the tartrate molecule in the crystal structure. The displacement of the ligand from this pocket is likely due to steric hindrance, as the I(1,3,4,5)P_4 headgroup is bulky compared with the tartrate and it may therefore be difficult to accommodate multiple phosphate groups in the cleft. It is possible that this effect is a direct consequence of the low resolution of the CG model and the larger van der Waals radii possessed by the CG particles, making protein-ligand interactions difficult to treat. It may also be due to the short electrostatic cut off used in the CG model, which does not take into account long range electrostatic interactions. As the interaction between PTEN and PI(3,4,5)P_3 is likely to have a significant electrostatic component, the use of a plain truncated and shifted cut off rather than a more elaborate scheme such as PME may be partly responsible for the escape of the PI(3,4,5)P_3 from the putative binding site. However, previous studies have demonstrated the utility of CG models when investigating protein-ligand interactions. For example, an analogous CG model of PI(4,5)P_2 and an inwardly rectifying potassium (Kir) channel accurately predicted the PI(4,5)P_2 binding site of the channel [409] before the advent of the crystal structure [410].

To assess the dynamic constriction of the binding pocket relative to the crystal structure, binding pocket volume calculations were performed using POVME (see Chapter 5). Protein conformations were extracted at intervals of 1 ns from the atomistic MD trajectories and fitted using a least squares method to the crystal structure. A pocket encompassing region was defined by placing a sphere of radius 15 Å at the position of the sulfur atom of C124 for each snapshot. The resulting volumes are shown in Figure 6.14. The calculated binding pocket volume shows that the cleft is dynamic and prone to collapse during the MD simulations, with the calculated value being on average slightly smaller than that seen in the crystal structure. This may help to explain the difficulties in maintaining a suitable PTEN-PI(3,4,5)P_3 complex in the simulations.

To determine whether or not the ejection of the ligand from the binding site is due

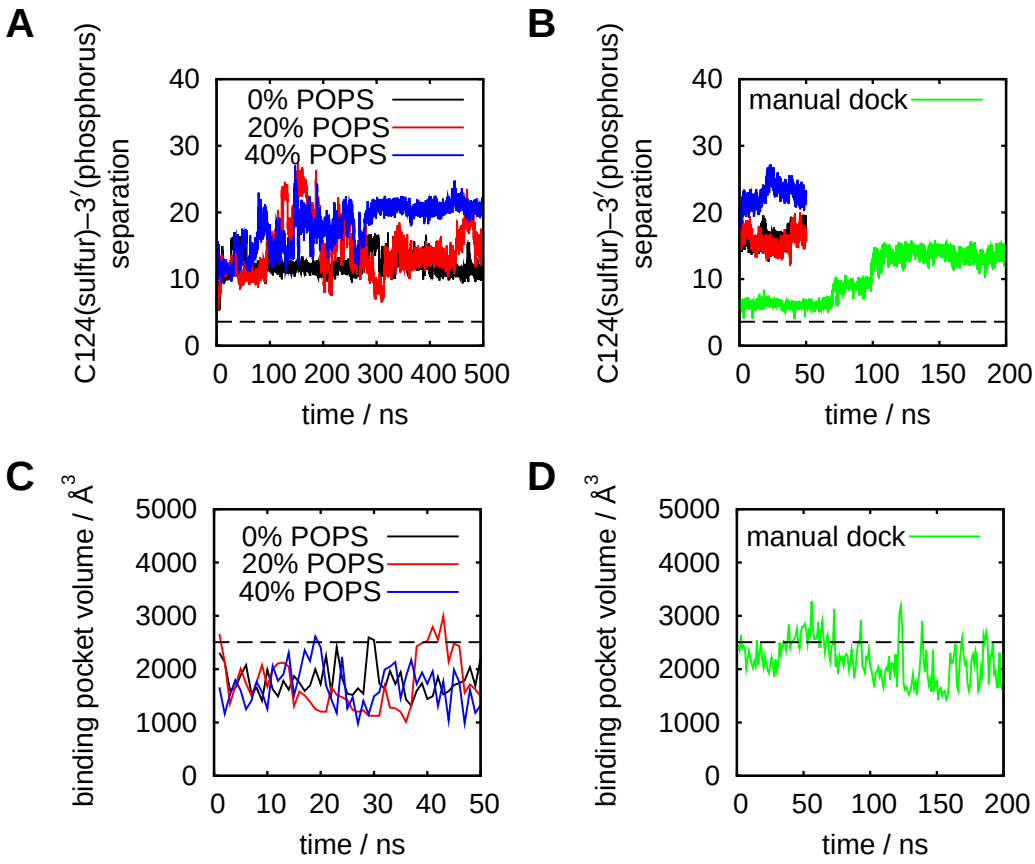

Figure 6.14: Ligand migration and binding pocket collapse in PTEN. **(A)** Separation between the C124 sidechain particle and the 3′ phosphate particle of the I(1,3,4,5)P_4 headgroup over the CG MD simulations. Rapid ejection of the ligand from the binding pocket is observed in all cases. **(B)** Separation between the C124 sulfur atom and the 3′ phosphorus atom of the I(1,3,4,5)P_4 headgroup in the subsequent atomistic simulations, and for the 200 ns atomistic simulation where PTEN was manually positioned at the surface of a PI(3,4,5)P_3-containing lipid bilayer. The horizontal dotted line shows the 3.6 Å reference separation between the C124 sulfur atom and the carboxylate carbon of the tartrate in the crystal structure. **(C)** Binding pocket volume of PTEN calculated by placing a 15 Å radius pocket encompassing sphere at the sulfur atom of C124 at 1 ns intervals. The dotted line indicates the calculated binding pocket volume from the crystal structure, 2506 Å^3. Mean values plus or minus one standard deviation from the mean over the trajectories are 1779 ± 293 Å^3, 1728 ± 514 Å^3 and 1676 ± 396 Å^3 in the case of the POPC(100%), POPC(80%):POPS(20%) and POPC(60%):POPS(40%) bilayer systems respectively. **(D)** Pocket volume calculated for the 200 ns atomistic simulation where PTEN was manually positioned. Mean values plus or minus one standard deviation from the mean are 2172 ± 386 Å^3.

solely to the CG component of the MD simulations, two other approaches to positioning PI(3,4,5)P_3 in the binding site were explored. In the first, the tartrate molecule was removed from the crystal structure of PTEN and blind docking using I(1,3,4,5)P_4 was performed using AutoDock. Identical docking parameters to those described in Chapter 5 were used, except that the grid was centred at the position of the sulfur atom of C124. The centre of mass of the lowest energy cluster, which is also the largest cluster,

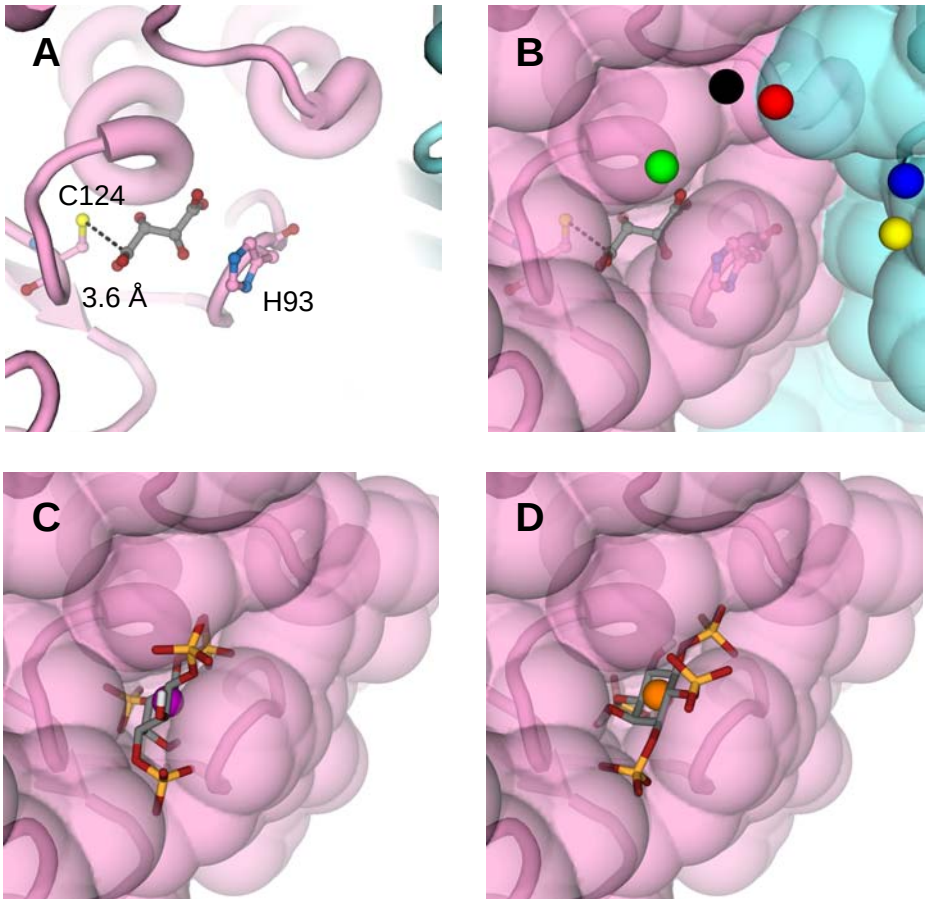

Figure 6.15: Structural details of ligand migration and molecular docking. **(A)** The tartrate molecule in the crystal structure, showing the 3.6 Å separation between the C124 sulfur atom and the carboxylate carbon of the tartrate. H93 and C124 line the binding pocket and are shown as ball-and-stick representations. **(B)** The solvent-accessible surface of PTEN highlights the constriction of the binding pocket. Spheres show the final positions of the centre of mass of the I(1,3,4,5)P_4 headgroup after 500 ns CG MD simulation and 50 ns atomistic MD simulation using a POPC(100%) bilayer (black), a POPC(80%):POPS(20%) bilayer (red) and a POPC(60%):POPS(40%) bilayer (blue). The green sphere shows the centre of mass position after manual positioning of PTEN at the surface of a PI(3,4,5)P_3-containing atomistic lipid bilayer and 200 ns of MD simulation. The position of the centre of mass of the lowest energy, largest conformational cluster after docking using AutoDock is shown as a yellow sphere. **(C)** Lowest energy configuration of the I(1,3,4,5)P_4 headgroup after docking to the isolated phosphatase domain using AutoDock. The centre of mass is indicated by a purple sphere. **(D)** Largest conformational cluster based on a tolerance of 1.5 Å RMSD for docking of the I(1,3,4,5)P_4 headgroup to the phosphatase domain, with the centre of mass position represented by an orange sphere.

is spatially separated from the active site (Figure 6.15) and no docking pose reproduced the predicted pose based on the position of the tartrate in the PTEN crystal structure.

Given the propensity for anionic lipid clustering around the C2 domain, it is perhaps not surprising that the negatively charged PI(3,4,5)P_3 ligand tends to migrate towards the C2 domain in the simulations. Molecular docking using the I(1,3,4,5)P_4 headgroup also revealed that the vast majority of the predicted docking poses interacted with the

C2 domain. Interestingly, Wang *et al.* [395] were able to obtain a binding pose similar to that predicted by Lee *et al.* [371] from the crystal structure by removing the C2 domain and docking to the phosphatase domain alone. Following the approach of Wang *et al.*, docking was repeated using just the phosphatase domain, and this results in a predicted binding mode that is more consistent with the position of the tartrate in the crystal structure (Figure 6.15). Wang *et al.* equilibrated their complexes using short MD simulations of 2 ns duration after docking. However in light of the results presented here, which show long term conformational drift of the PTEN-PI(3,4,5)P_3 complex, it is unclear whether these complexes would be stable over longer time scales. Indeed, even over a period of 2 ns the authors observed substantial changes in hydrogen bonding patterns in the MD simulations compared with their docked structures.

In the second approach, after removal of the tartrate the PTEN protein was manually positioned at the surface of a zwitterionic lipid bilayer containing PI(3,4,5)P_3, using the same method used for GRP1-PH in Chapter 3. After 200 ns of atomistic simulation the ligand eventually departs from the binding site, though more slowly than was the case for the CG simulations. This might be expected owing to the greater number of interaction sites and the rougher potential energy surface in the atomistic MD simulations. Interestingly, no excursions from the binding site were observed for the atomistic simulations of PI(3,4,5)P_3 bound to GRP1-PH described in Chapter 3, despite the fact that they were up to 100 ns in length.

6.7.4 Summary

Collectively, these results suggest that PI(3,4,5)P_3 is unable to form a long-lived complex with PTEN at the site formerly occupied by the tartrate. Mutational data strongly suggests that this is the location of the PI(3,4,5)P_3 binding site, as mutations in either the cysteine (C124S [376]) or the histidine (H93A [398]) residues flanking the tartrate inactivate the phosphatase. However, tartrate is much smaller than the PI(3,4,5)P_3 head-

group, which is decorated with bulky phosphate groups. It may therefore be the case that some form of induced fit mechanism operates in the phosphatase domain, and that the PI(3,4,5)P_3 headgroup cannot be accommodated in the constricted active site observed in the crystallographic conformation. The CG MD simulations are unlikely to be able to capture this mechanism as the protein's capacity for conformational change is limited by the presence of the Gaussian network model, which restrains its secondary structural elements.

Although PI(3,4,5)P_3 is jettisoned from the predicted binding site, it always remains bound to the protein within the archway-like cavity formed by the packing of the C2 domain against the phosphatase domain in all of the simulations. While PI(3,4,5)P_3 undeniably binds to the phosphatase domain, the fact that anionic lipids were observed to cluster around the C2 domain makes it tempting to speculate that perhaps the C2 domain also plays a role in PI(3,4,5)P_3 binding. It could be that the C2 domain assists the phosphatase domain in accessing its lipid substrate by locally enriching the concentration of anionic lipids such as PI(3,4,5)P_3 *via* electrostatic interactions. This would explain the preferential interactions with the C2 domain in the docking runs and the MD simulations, as it could perhaps be the case that the role of the C2 domain is to localise PI(3,4,5)P_3 proximal to the active site in the phosphatase domain, thus funnelling the substrate into the binding cleft and facilitating the phosphatase activity of PTEN. The mechanistic details of how PTEN binds to membranes are still not clear, and it is not yet known how the lipid substrate is able to access the binding site. As both domains of PTEN appear able to recognise anionic lipids, substrate ingress could perhaps be achieved by transient dissociation of the phosphatase domain from the surface while the C2 domain keeps the protein bound to the membrane. Experimental data would be required to assess the contribution of the C2 domain to the activity of the phosphatase domain, and to determine whether or not it is responsible for enrichment of PI(3,4,5)P_3 and enhanced phosphatase activity. This could most easily be achieved by comparing the activity of the isolated phosphatase

domain to that of the phosphatase/C2 core, but unfortunately when this was attempted by Das *et al.* the isolated phosphatase domain was found to exhibit low protein stability [396] making the experiment challenging to perform.

It may well be the case that the simulations are simply not long enough to observe long-lived complex formation. Shaw and colleagues carried out MD simulations of dasatinib and Src kinase [411], in an attempt to model *ab initio* the binding of the drug to its target. They began with Src kinase immersed in a solvent box surrounded by several dasatinib molecules and simply allowed the simulation to run. Shaw and colleagues found it necessary to run multiple simulations for tens of μs, and in some - but not all - they eventually converged upon a protein-ligand complex with the drug in a similar site to the crystallographic binding pose.

Despite the limitations of the model when considering the protein-ligand interaction, it does at least appear to be robust with respect to the protein-lipid interactions and the location of the PTEN-membrane interface. More elaborate techniques, or perhaps longer simulations, may be necessary to treat the PTEN-PI(3,4,5)P_3 interactions with more accuracy, and clearly to model effects such as dephosphorylation of the 3′ phosphate higher level treatments would be required. For example, recent advances in density functional theory (DFT) methods implemented in hybrid quantum mechanical/molecular mechanical (QM/MM) simulations have allowed for simulation of phosphate hydrolysis by alkaline phosphatases [412, 413]. It is conceivable that this could eventually be adapted to investigate the membrane binding mechanism and phosphatase activity of PTEN through an expanded multiscale scheme, incorporating successive simulations at CG, atomistic and QM levels of resolution.

Chapter 7

Conclusions

THE preceding chapters have attempted to tackle the problem of simulating protein-lipid interactions and protein-membrane association using a variety of computational techniques, complemented by experimental results where possible. In this final chapter, some of the main conclusions from this thesis are briefly reviewed and possible opportunities for further work are identified.

7.1 Electrostatics in protein-membrane binding

Throughout this thesis, electrostatic effects have been repeatedly shown to influence many aspects of protein-membrane association. These manifest themselves in many different ways, from the disruption of GRP1-PH targeting by non-specific interactions with non-PI(3,4,5)P_3 lipids described in Chapter 4 to anionic lipid clustering around the positively charged C2 domain of PTEN seen in Chapter 6. The fact that GRP1-PH responded to alternative lipid configurations by executing different search patterns at the mem-

brane surface is of particular note, as it reinforces the idea that the exact distribution of lipids can modulate the electrostatic effects arising from the presence of the membrane. This suggests that modelling the membrane as a uniformly charged slab would be insufficient to capture the subtle steering phenomena observed in Chapter 4. Furthermore, the observed preferential accumulation of anionic lipids around the PTEN C2 domain suggests that lipid redistribution can also be a cooperative effect, with the membrane readjusting to accommodate the presence of a bound peripheral protein and optimise the favourable electrostatic interactions. This is in line with results from other simulation studies [296, 297, 298] and with the complementary PI(4,5)P_2 microdomain formation and syntaxin-1A clustering observed by van den Bogaart *et al.* [63], though such effects cannot be observed in the BD simulations where a static membrane was used.

An interesting observation that emerged from the BD simulations is that an anionic lipid concentration of around 20% is optimal to enhance both axial translational steering and orientational steering of GRP1-PH to PI(3,4,5)P_3, without sacrificing its ability to locate PI(3,4,5)P_3 through radial translational steering in r. This is noteworthy because typical anionic lipid concentrations in the cytoplasmic leaflets of mammalian plasma membranes are also approximately 20%, and this makes it tempting to speculate that perhaps this concentration has been evolutionarily tuned in some way to promote successful protein-ligand targeting at the surface of plasma membranes.

Most PH domains bind their PIP ligands in a canonical fashion at the binding site formed by the three variable loops at the base of the β barrel. Interestingly, an analysis of the electrostatic properties of a variety of PH domains revealed that the location of this binding site was often coincident with the direction of the molecular dipole moment vector measured from the centre of mass of the domain (Chapter 4). This raises the prospect of an electrostatic funnelling mechanism, acting in reverse to that observed in AChE, that is able to steer PH domains into the correct orientation to bind to their PIP ligand by alignment of the molecular dipole moment.

This close correspondence between the direction of the molecular dipole moment vector and the location of the PIP binding site was also observed in PH domains that bind their PIP ligands in a non-canonical manner (Chapter 5). As this property appeared to be reasonably well conserved between PH domains, it provided some support for the existence of a secondary, or perhaps alternative, binding site in kindlin-1 PH after the docking results proved to be somewhat inconclusive. The idea of dipole moments influencing protein behaviour is not unprecedented, and aside from electrostatic funnelling in AChE, there is evidence that dipole moments of α helices stabilise protein secondary structure [414], guide folding proteins towards their native structures [415] and influence protein-protein aggregation [416]. Molecular dipole moments aligned with PIP binding sites may therefore be a general feature of the PH domain family, and a comprehensive analysis of the 109 PH domain structures in the PDB could confirm whether or not this is indeed the case. If so, this may have implications for a general mechanism of PIP binding, allowing for prediction of ligand binding propensities based on structure alone. Indeed, as PH domains are so structurally conserved, an ambitious goal might be to attempt to make predictions about binding properties based solely on sequence data.

7.2 Membrane penetration and non-polar interactions

Although electrostatic interactions may dominate protein-membrane binding in many cases, it also seems clear that non-specific, non-polar interactions with bilayer lipids can contribute to the stability of membrane-bound complexes. Support for this was provided by the observed penetration of GRP1-PH into zwitterionic lipid bilayers in Chapter 3, the extent of which was strikingly evident when forcible dissociation of the protein from the surface of the membrane was attempted using SMD simulations. Further evidence for this was provided in Chapter 6, where numerous interactions were observed between PTEN and zwitterionic lipids in the bilayer. These results suggested that, far from being a passive substrate that simply serves to accommodate ligands, the membrane plays an

important supporting role in sustaining bound protein-ligand complexes. This points towards a dual recognition model of membrane binding, where the interaction of the protein with its cognate ligand is supplemented by additional non-specific interactions with the bilayer lipids.

7.3 Membrane encounter, recognition and dissociation

This thesis has also addressed the general problem of encounter and recognition in the context of lipid targeting by GRP1-PH using BD simulations. The complexity of the observed association behaviour also raises questions about the nature of the eventual *dissociation* mechanism from the membrane surface, as GRP1-PH exhibits reversible binding to PI(3,4,5)P_3. While dissociation and association are unlikely to occur *via* identical and reversible routes, the complex association pathways arising from the BD simulations of GRP1-PH binding to PI(3,4,5)P_3 suggest that defining a dissociation coordinate along the z axis in the SMD simulations presented in Chapter 3 is likely to be an oversimplification.

7.4 Further work

One possible avenue for further work would be to explore the energetics of this dissociation using PMF calculations. However, a traditional one dimensional PMF, such as that used when investigating the free energy profile for passage of ions through an ion channel, may not be sufficient in this case. As BD simulations have demonstrated that the anionic lipid composition and configuration in the membrane influences association pathways, dissociation pathways may also be more complicated than a conceptually simple 'vertical takeoff' from the bilayer surface. This may require a multidimensional PMF calculation, with reaction coordinates not only in z, but also in r and possibly θ. As electrostatic interactions have been shown to be important in protein-membrane association, the truncated and shifted cut off used for the Coulomb potential in the CG MD simula-

tions may be too crude, and atomistic simulations with a more elaborate treatment of long range electrostatics may be necessary. The combination of multiple reaction coordinates, an atomistic level of detail and a comprehensive treatment of electrostatics would however make calculating this multidimensional PMF extremely challenging.

The work on PTEN described in Chapter 6 was an attempt to apply multiscale techniques to a larger, more complex protein with a multidomain architecture. An alternative application of these techniques might be to investigate the association of two proteins at the surface of the membrane. An obvious choice, given the focus on GRP1-PH in this thesis, would be the interaction between the small GTPase Arf6 and GRP1-PH, which serves as an Arf guanine nucleotide exchange factor. Arf is myristoylated at the N-terminus and is localised at the plasma membrane through this lipid tail, and so PI(3,4,5)P_3-bound GRP1-PH and Arf6 together execute the now familiar two dimensional diffusion at the membrane surface that facilitates binding. The diffusion, encounter and binding process could be modelled by CG MD simulations, switching to atomistic resolution after formation of the bound complex to probe the protein-protein interaction in more detail.

To test the wider applicability of multiscale simulations of peripheral membrane proteins, a limited number of CG MD simulations were performed to investigate protein-protein association of GRP1-PH and Arf6 (data not shown). While the results from the simulations were promising, there are several problems that still need to be overcome. Simulations were successful in generating bound complexes, but as diffusion rates are enhanced in the CG model it was found that the GRP1-PH-Arf6 complex formed rather quickly. This rapid association of the two proteins presents a problem when attempting to study complex formation, as the location of the protein-protein interface appeared to be at least partly dependent upon the inital separation and relative orientation of the proteins. Rotational autocorrelation functions showed that protein reorientation before complex formation is limited. To alleviate this problem, switching to larger box sizes was considered, though this is computationally expensive due to the presence of additional

lipids and solvent. Another option was to increase the ionic strength of the solution, in the hope that this would lead to a greater degree of ionic screening between the proteins. A limited number of exploratory simulations provided hints that increasing the salt concentration does indeed slow down the rate of association, increasing the time available for the proteins to reorient themselves at the membrane surface prior to formation of the complex. However, this effect is only observed when the Na^+ and Cl^- concentrations are increased substantially, to values of the order of 10 M. With physiological measurements of, for example, the Na^+ concentration in cells giving values of the order of 0.01 M [417, 418], questions remain about the physiological relevance of these simulations when the salt concentration is around 1000 times greater than that observed *in vivo*.

7.5 Towards simulations of cell signalling events

A common recurring theme in this thesis has been the need to adopt a multiscale approach to simulate processes that occur over long length and time scales. This is reflected in Chapter 4, where BD simulations were used to assemble the protein-membrane complex by targeting the protein to a rigid membrane surface, before switching to atomistic MD simulations to probe the protein-lipid contacts. A multiscale approach was also used in Chapter 6, but in contrast CG MD simulations were used to generate the protein-membrane complex by dynamic self assembly of the lipids around the protein, before again converting to atomic resolution to investigate the detailed protein-lipid interactions. Clearly, both of these techniques have their limitations. The BD simulations incorporate a thorough treatment of electrostatics but lack conformational flexibility and membrane fluidity, while in many ways the CG MD simulations are quite the opposite, with rigorous treatment of the electrostatics sacrificed for enhanced sampling of the dynamics.

However, despite the shortcomings of some of the techniques described in this thesis, overall the results have demonstrated that using a variety of methods as part of a multi-

scale strategy can provide complementary information about the behaviour of different aspects of a given system. Looking ahead, an ambitious, serial multiscale study of a signalling protein interacting with a lipid bilayer might first use BD simulations to model the initial diffusional association with the membrane surface. This could then be followed by CG MD simulations to incorporate lipid mobility and limited protein flexibility, allowing the encounter complex to relax and the lipids to rearrange themselves to maximise the favourable interactions with the protein. Switching to atomistic resolution MD simulations would then provide details on the specific protein-lipid and protein-ligand interactions, and perhaps a QM/MM stage might follow that would allow modelling of a catalytic step, for example phosphate hydrolysis. Finally, SMD simulations could be used to remove the protein from the surface and return it to the cytosol. Taken together, these five stages would effectively be representative of a complete signalling event, though clearly downstream effects after the protein departed from the membrane surface would also be important.

Ultimately then, if the multiscale simulation work from these four chapters could be integrated with other computational and experimental biophysical methods, this could perhaps lay the foundations of a framework designed to characterise the complete cycle of membrane association, binding (potentially including catalysis), penetration and eventual dissociation. This would represent one possible means to link structural and biophysical studies through to systems level descriptions of plasma membrane signalling events [419]. An exciting and very recent development has been the publication of a whole-cell computational model of a bacterium, *Mycoplasma genitalium* [420]. This represents the first time an entire organism has been modelled in terms of its molecular components, and the model has already provided novel insights into *in vivo* processes such as protein-DNA association. If an analogous model of a human cell could be developed, this may yield similar insights into protein-membrane association and signalling events at the plasma membrane.

Appendix A

PTEN gene mutations

THE publicly available version of the HGMD [399] contained nearly 90000 mutations for almost 3500 genes at the time of writing. Tabulated below are all of the sequence variants for PTEN listed in the database, along with their phenotype and the reference for the publication in which they were first reported.

Codon change	Amino acid change	Codon number	Phenotype
TTA-TAA	Leu-Term	23	Cowden's disease [421]
TTA-TGA	Leu-Term	23	Cowden's disease [374]
aGAC-TAC	Asp-Tyr	24	Bannayan-Riley-Ruvalcaba syndrome [422]
GCT-GAT	Ala-Asp	34	Bannayan-Riley-Ruvalcaba syndrome [423]
ATG-ACG	Met-Thr	35	Proteus-like syndrome [424]
ATG-AGG	Met-Arg	35	Juvenile polyposis coli [425]
gGGA-AGA	Gly-Arg	36	Cowden's disease [426]

Codon change	Amino acid change	Codon number	Phenotype
tGCA-CCA	Ala-Pro	39	Cowden's disease [427]
GGC-GAC	Gly-Asp	44	Developmental delay and macrocephaly [428]
AACa-AAA	Asn-Lys	48	Cowden's disease [400]
TCA-TGA	Ser-Term	59	Cowden's disease [429]
gCAT-GAT	His-Asp	61	Macrocephaly, ventriculomegaly, VATER association [430]
ATA-AGA	Ile-Arg	67	Cowden's disease [431]
aTAC-CAC	Tyr-His	68	Bannayan-Zonana syndrome [431]
aTAC-GAC	Tyr-Asp	68	Proteus syndrome [432]
ACCg-ACT	Thr-Thr	78	Cowden's disease? [429]
cGCC-ACC	Ala-Thr	79	Breast cancer [433]
AAT-ACT	Asn-Thr	82	Breast cancer [433]
aCAA-TAA	Gln-Term	87	Cowden's disease [431]
aTAT-CAT	Tyr-His	88	Lhermitte-Duclos disease [434]
GACc-GAG	Asp-Glu	92	Cowden's disease [374]
CAT-CGT	His-Arg	93	Autism spectrum disorder and macrocephaly [435]
cCAT-TAT	His-Tyr	93	Cowden's disease [436]
CCA-CAA	Pro-Gln	96	Cowden's disease [437]
aCAG-TAG	Gln-Term	97	Cowden's disease [438]
TGT-TAT	Cys-Tyr	105	Bannayan-Riley-Ruvalcaba syndrome [423]
CTT-CCT	Leu-Pro	108	PTEN hamartoma-tumour syndrome [439]
cCAA-TAA	Gln-Term	110	Cowden's disease [421]

Codon change	Amino acid change	Codon number	Phenotype
aTGG-CGG	Trp-Arg	111	Proteus syndrome [424]
CTA-CCA	Leu-Pro	112	Cowden's disease [440]
CAT-CCT	His-Pro	118	Neurodevelopmental disordersneurodevelopmental disorders and macrocephaly without autism [441]
tGTT-ATT	Val-Ile	119	Multiple cancers [442]
CAC-CGC	His-Arg	123	Cowden's disease [443]
tCAC-GAC	His-Asp	123	Cowden's disease [437]
cTGT-CGT	Cys-Arg	124	Cowden's disease [443]
TGT-TAT	Cys-Tyr	124	Cowden's disease [444]
GGA-GAA	Gly-Glu	129	Cowden's disease [403]
aCGA-TGA	Arg-Term	130	Cowden's disease [443]
CGA-CAA	Arg-Gln	130	Cowden's disease [445]
CGA-CTA	Arg-Leu	130	Cowden's disease [431]
GGT-GAT	Gly-Asp	132	Cowden's disease [446]
GGT-GCT	Gly-Ala	132	PTEN hamartoma-tumour syndrome [439]
GGT-GTT	Gly-Val	132	PTEN hamartoma-tumour syndrome [447]
ATG-AGG	Met-Arg	134	Bannayan-Riley-Ruvalcaba syndrome [433]
ATA-AGA	Ile-Arg	135	Bannayan-Riley-Ruvalcaba syndrome [448]
gATA-GTA	Ile-Val	135	Bannayan-Riley-Ruvalcaba syndrome [423]
aTGT-CGT	Cys-Arg	136	Cowden's disease [449]
TGT-TAT	Cys-Tyr	136	Cowden's disease [450]
TTA-TAA	Leu-Term	139	Cowden's disease [374]

Codon change	Amino acid change	Codon number	Phenotype
TTA-TGA	Leu-Term	139	Cowden's disease, Sjögren's syndrome [451]
TAT-TGT	Tyr-Cys	155	Cowden's disease [452]
GAA-GGA	Glu-Gly	157	Autism spectrum disorders, developmental delay and macrocephaly [428]
gGAA-TAA	Glu-Term	157	Cowden's disease [403]
aGTA-ATA	Val-Ile	158	Multiple cancers [442]
GGA-GAA	Gly-Glu	165	Cowden's disease [438]
GGA-GTA	Gly-Val	165	Cowden's disease [431]
AGTc-AGA	Ser-Arg	170	Cowden's disease [453]
tCAG-TAG	Gln-Term	171	Cowden's disease [454]
CGC-CAC	Arg-His	173	Bannayan-Riley-Ruvalcaba syndrome [455]
gCGC-TGC	Arg-Cys	173	Bannayan-Riley-Ruvalcaba syndrome [455]
TAT-TGT	Tyr-Cys	176	Neurodevelopmental disorders and macrocephaly with autism [441]
TATt-TAA	Tyr-Term	176	Cowden's disease [374]
TATt-TAG	Tyr-Term	176	Cowden's disease [437]
TATa-TAA	Tyr-Term	178	Bannayan-Riley-Ruvalcaba syndrome [374]
TATa-TAG	Tyr-Term	178	Cowden's disease [456]
TACc-TAG	Tyr-Term	180	PTEN hamartoma-tumour syndrome [439]
CTG-CCG	Leu-Pro	181	Cowden's disease [429]
tAGA-TGA	Arg-Term	189	Cowden's disease [421]
cAAG-TAG	Lys-Term	197	Cowden's's disease [444]
ACT-ATT	Thr-Ile	202	Developmental delay and macrocephaly [428]

Codon change	Amino acid change	Codon number	Phenotype
TGCa-TGA	Cys-Term	211	Proteus syndrome [424]
tCAG-TAG	Gln-Term	214	Bannayan-Zonana syndrome [457]
GTC-GAC	Val-Asp	217	Cowden's disease [458]
cCAG-TAG	Gln-Term	219	Cowden's disease [459]
aCGA-TGA	Arg-Term	233	Cowden's disease [403]
CGG-CAG	Arg-Gln	234	Multiple cancers [401]
TTT-TCT	Phe-Ser	241	Autism spectrum disorder and macrocephaly [435]
tCAG-TAG	Gln-Term	245	Cowden's disease [431]
CCG-CTG	Pro-Leu	246	Bannayan-Riley-Ruvalcaba syndrome [423]
GAT-GGT	Asp-Gly	252	Autism spectrum disorder and macrocephaly [435]
aGAG-TAG	Glu-Term	256	Bannayan-Zonana syndrome [457]
AAT-AGT	Asn-Ser	276	Neurodevelopmental disorders and macrocephaly with autism [441]
aAAA-GAA	Lys-Glu	289	Cowden's disease [460]
TATc-TAA	Tyr-Term	315	Cowden's disease [374]
TATc-TAG	Tyr-Term	315	Cowden's disease [431]
TTA-TAA	Leu-Term	320	Cowden's disease [438]
TTA-TGA	Leu-Term	320	Cowden's's disease [374]
tGAC-AAC	Asp-Asn	326	Autism spectrum disorder and macrocephaly [461]
cCGA-TGA	Arg-Term	335	Cowden's disease [421]
CGA-CTA	Arg-Leu	335	Cowden's disease [402]

Codon change	Amino acid change	Codon number	Phenotype
TTT-TCT	Phe-Ser	337	Cowden's disease [455]
GTG-GAG	Val-Glu	343	Cowden's disease [421]
gCTG-GTG	Leu-Val	345	Cowden's disease [429]
cTTC-CTC	Phe-Leu	347	Cowden's disease [421]

Table A.1: PTEN missense/nonsense point mutations extracted from the HGMD. At the time of writing there were 94 sequence variants listed in the database.

References

[1] Singer S.J. and Nicolson G.L. The fluid mosaic model of the structure of cell membranes. *Science*, 175, 720–731, 1972.

[2] Jacobson K., Sheets E.D., and Simson R. Revisiting the fluid mosaic model of membranes. *Science*, 268, 1441–1442, 1995.

[3] Cherezov V., Rosenbaum D.M., Hanson M.A., Rasmussen S.G., Thian F.S., Kobilka T.S., Choi H.-J., Kuhn P., Weis W.I., Kobilka B.K., and Stevens R.C. High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor. *Science*, 23, 1258–1265, 2007.

[4] Lietzke S.E., Bose S., Cronin T., Klarlund J., Chawla A., Czech M.P., and Lambright D.G. Structural basis of 3-phosphoinositide recognition by pleckstrin homology domains. *Mol. Cell*, 6, 385–394, 2000.

[5] Wenk M.R. The emerging field of lipidomics. *Nat. Rev. Drug Discov.*, 4, 594–610, 2005.

[6] Simons K. and Toomre D. Lipid rafts and signal transduction. *Nat. Rev. Mol. Cell Biol.*, 1, 31–41, 2000.

[7] Munro S. Lipid rafts: Elusive or illusive? *Cell*, 115, 377–388, 2003.

[8] McLaughlin S., Wang J., Gambhir A., and Murray D. PIP_2 and proteins: Interactions, organization and information flow. *Annu. Rev. Biophys. Biomol. Struct.*, 31, 151–175, 2002.

[9] McLaughlin S. and Murray D. Plasma membrane phosphoinositide organization by protein electrostatics. *Nature*, 438, 605–611, 2005.

[10] Vance J.E. and Steenbergen R. Metabolism and functions of phosphatidylserine. *Prog. Lipid Res.*, 44, 207–234, 2005.

[11] Sampaio J.L., Gerl M.J., Klose C., Ejsing C.S., Beug H., Simons K., and Shevchenko A. Membrane lipidome of an epithelial cell line. *Proc. Natl. Acad. Sci. U.S.A.*, 108, 1903–1907, 2011.

[12] Lemmon M.A. Membrane recognition by phospholipid-binding domains. *Nat. Rev. Mol. Cell Biol.*, 9, 99–111, 2008.

[13] Yeung T., Gilbert G.E., Shi J., Silvius J., Kapus A., and Grinstein S. Membrane phosphatidylserine regulates surface charge and protein localization. *Science*, 319, 210–213, 2008.

[14] McLaughlin S. The electrostatic properties of membranes. *Annu. Rev. Biophys. Biophys. Chem.*, 18, 113–136, 1989.

[15] Mulgrew-Nesbitt A., Diraviyam K., Wang J., Singh S., Murray P., Li Z., Rogers L., Mirkovic N., and Murray D. The role of electrostatics in protein-membrane interactions. *Biochim. Biophys. Acta*, 1761, 812–826, 2006.

[16] Michell R.H. Inositol derivatives: Evolution and functions. *Nat. Rev. Mol. Cell Biol.*, 9, 151–161, 2008.

[17] Kooijman E.E., King K.E., Gangoda M., and Gericke A. Ionization properties of phosphatidylinositol polyphosphates in mixed model membranes. *Biochemistry*, 48, 9360–9371, 2009.

[18] Vanhaesebroeck B., Leevers S.J., Ahmadi K., Timms J., Katso R., Driscoll P.C., Woscholski R., Parker P.J., and Waterfield M.D. Synthesis and function of 3-phosphorylated inositol lipids. *Annu. Rev. Biochem.*, 70, 535–602, 2001.

[19] Folch J. Brain diphosphoinositide, a new phosphoinositide having inositol metadiphosphate as a constituent. *J. Biol. Chem.*, 177, 505–519, 1949.

[20] Michell R.H. Inositol phospholipids and cell surface receptor function. *Biochim. Biophys. Acta*, 415, 81–147, 1975.

[21] Berridge M.J. and Irvine R.F. Inositol trisphosphate, a novel second messenger in cellular signal transduction. *Nature*, 312, 315–321, 1984.

[22] Whitman M., Downes C.P., Keeler M., Keller T., and Cantley L. Type I phosphatidylinositol kinase makes a novel inositol phospholipid, phosphatidylinositol-3-phosphate. *Nature*, 332, 644–646, 1988.

[23] Toker A. and Cantley L.C. Signalling through the lipid products of phosphoinositide-3-OH kinase. *Nature*, 387, 673–676, 1997.

[24] Harlan J.E., Hajduk P.J., Yoon H.S., and Fesik S.W. Pleckstrin homology domains bind to phosphatidylinositol-4,5-bisphosphate. *Nature*, 371, 168–170, 1994.

[25] Stahelin R.V. Lipid binding domains: More than simple lipid effectors. *J. Lipid Res.*, 50, S299–S304, 2009.

[26] Hurley J.H. and Misra S. Signaling and subcellular targeting by membrane-binding domains. *Annu. Rev. Biophys. Biomol. Struct.*, 29, 49–79, 2000.

[27] Cho W. and Stahelin R.V. Membrane-protein interactions in cell signaling and membrane trafficking. *Annu. Rev. Biophys. Biomol. Struct.*, 34, 119–151, 2005.

[28] Kutateladze T.G. Translation of the phosphoinositide code by PI effectors. *Nature Chem. Biol.*, 6, 507–513, 2010.

[29] Peter B.J., Kent H.M., Mills I.G., Vallis Y., Butler P.J.G., Evans P.R., and McMahon H.T. BAR domains as sensors of membrane curvature: The amphiphysin BAR structure. *Science*, 303, 495–499, 2004.

[30] Essen L.-O., Perisic O., Lynch D.E., Katan M., and Williams R.L. A ternary metal binding site in the C2 domain of phosphoinositide-specific phospholipase C-δ1. *Biochemistry*, 36, 2753–2762, 1997.

[31] Banner D.W., D'Arcy A., Chène C., Winkler F.K., Guha A., Konigsberg W.H., Nemerson Y., and Kirchhofer D. The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature*, 7, 41–46, 1996.

[32] Bravo J., Karathanassis D., Pacold C.M., Pacold M.E., Ellson C.D., Anderson K.E., Butler P.J., Lavenir I., Perisic O., Hawkins P.T., Stephens L., and Williams R.L. The crystal structure of the PX domain from $p40^{phox}$ bound to phosphatidylinositol 3-phosphate. *Mol. Cell*, 8, 829–839, 2001.

[33] Dumas J.J., Merithew E, Sudharshan E., Rajamani D., Hayes S., Lawe D., Corvera S., and Lambright D.G. Multivalent endosome targeting by homodimeric EEA1. *Mol. Cell*, 8, 947–958, 2001.

[34] Punta M., Coggill P.C., Eberhardt R.Y., Mistry J., Tate J., Boursnell C., Pang N., Forslund K., Ceric G., Clements J., Heger A., Holm L., Sonnhammer E.L.L., Eddy

S.R., Bateman A., and Finn R.D. The Pfam protein families database. *Nucleic Acids Res.*, 40, D290–D301, 2012.

[35] Haslam R.J., Koide H.B., and Hemmings B.A. Pleckstrin domain homology. *Nature*, 363, 309–310, 1993.

[36] Mayer B.J., Ren R., Clark K.L., and Baltimore D. A putative modular domain present in diverse signalling proteins. *Cell*, 73, 629–630, 1993.

[37] Lemmon M.A., Ferguson K.M., and Schlessinger J. PH domains: Diverse sequences with a common fold recruit signaling molecules to the cell surface. *Cell*, 85, 621–624, 1996.

[38] Kavran J.M., Klein D.E., Lee A., Falasca M., Isakoff S.J., Skolnik E.Y., and Lemmon M.A. Specificity and promiscuity in phosphoinositide binding by pleckstrin homology domains. *J. Biol. Chem.*, 273, 30497–30508, 1998.

[39] Yu J.W., Mendrola J.M., Audhya A., Singh S., Keleti D., DeWald D.B., Murray D., Emr S.D., and Lemmon M.A. Genome-wide analysis of membrane targeting by *S. cerevisiae* pleckstrin homology domains. *Mol. Cell*, 13, 677–688, 2004.

[40] Nalefski E.A. and Falke J.J. The C2 domain calcium-binding motif: Structural and functional diversity. *Protein Sci.*, 5, 2375–2390, 1996.

[41] Stahelin R.V., Rafter J.D., Das S., and Cho W. The molecular basis of differential subcellular localization of C2 domains of protein kinase C-alpha and group IVa cytosolic phospholipase A2. *J. Biol. Chem.*, 278, 12452–12460, 2003.

[42] Nalefski E.A. and Falke J.J. Location of the membrane-docking face on the Ca^{2+}-activated C2 domain of cytosolic phospholipase A_2. *Biochemistry*, 37, 17642–17650, 1998.

[43] Ponting C.P. Novel domains in NADPH oxidase subunits, sorting nexins, and PtdIns 3-kinases: Binding partners of SH3 domains? *Protein Sci.*, 5, 2353–2357, 1996.

[44] Song X., Xu W., Zhang A., Huang G., Liang X., Virbasius J.V., and Czech M.P. Phox homology domains specifically bind phosphatidylinositol phosphates. *Biochemistry*, 40, 8940–8944, 2001.

[45] Stahelin R.V., Karathanassis D., Bruzik K.S., Waterfield M.D., Bravo J., Williams R.L., and Cho W. Structural and membrane binding analysis of the Phox homology domain of phosphoinositide 3-kinase-C2α. *J. Biol. Chem.*, 281, 39396–39406, 2006.

[46] Stenmark H., Aasland R., Toh B.-H., and D'Arrigo A. Endosomal localization of the autoantigen EEA1 is mediated by a zinc-binding FYVE finger. *Nature*, 271, 24048–24054, 1996.

[47] Burd C.G. and Emr S.D. Phosphatidylinositol(3)-phosphate signaling mediated by specific binding to RING FYVE domains. *Mol. Cell*, 2, 157–162, 1998.

[48] Gaullier J.-M., Simonsen A., D'Arrigo A., Bremnes B., Stenmark H., and Aasland R. FYVE fingers bind PtdIns(3)P. *Nature*, 394, 432–433, 1998.

[49] Patki V., Lawe D.C., Corvera S., Virbasius J.V., and Chawla A. A functional PtdIns(3)P-binding motif. *Nature*, 394, 433–434, 1998.

[50] Misra S. and Hurley J.H. Crystal structure of a phosphatidylinositol 3-phosphate-specific membrane-targeting motif the FYVE domain of Vps27p. *Cell*, 97, 657–666, 1999.

[51] Dowler S., Currie R.A., Campbell D.G., Deak M., Kular G., Downes C.P., and Alessi D.R. Identification of pleckstrin-homology-domain-containing proteins with novel phosphoinositide-binding specificities. *Biochem. J.*, 351, 19–31, 2000.

[52] He J., Scott J.L., Heroux A., Roy S., Lenoir M., Overduin M., Stahelin R.V., and Kutateladze T.G. Molecular basis of phosphatidylinositol 4-phosphate and ARF1 GTPase recognition by the FAPP1 pleckstrin homology (PH) domain. *J. Biol. Chem.*, 286, 18650–18657, 2011.

[53] Whiteford C.C., Brearley C.A., and Ulug E.T. Phosphatidylinositol 3,5-bisphosphate defines a novel PI 3-kinase pathway in resting mouse fibroblasts. *Biochem. J.*, 323, 597–601, 1997.

[54] Dove S.K., Cooke F.T., Douglas M.R., Sayers L.G., Parker P.J., and Michell R.H. Osmotic stress activates phosphatidylinositol 3,5-bisphospate synthesis. *Biochem. J.*, 390, 187–192, 1997.

[55] Rameh L.E., Tolias K.F., Duckworth B.C., and Cantley L.C. A new pathway for synthesis of phosphatidylinositol-4,5-bisphosphate. *Nature*, 390, 192–196, 1997.

[56] Dove S.K., Piper R.C., McEwen R.K., Yu J.W., King M.C., Hughes D.C., Thuring J., Holmes A.B., Cooke F.T., Michell R.H., Parker P.J., and Lemmon M.A. Svp1p defines a family of phosphatidylinositol 3,5-bisphosphate effectors. *EMBO J.*, 23, 1922–1933, 2004.

[57] Gozani O., Karuman P., Jones D.R., Ivanov D., Cha J., Lugovskoy A.A., Baird C.L., Zhu H., Field S.J., Lessnick S.L., Villasenor J., Mehrotra B., Chen J., Rao V.R., Brugge J.S., Ferguson C.G., Payrastre B., Myszka D.G., Cantley L.C., Wagner G., Divecha N., Prestwich G.D., and Yuan J. The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. *Cell*, 114, 99–111, 2003.

[58] Stauffer T.P., Ahn S., and Meyer T. Receptor-induced transient reduction in plasma membrane PtdIns(4,5)P_2 concentration monitored in living cells. *Curr. Biol.*, 8, 343–346, 1998.

[59] Venkateswarlu K., Oatey P.B., Tavaré J.M., and Cullen P.J. Insulin-dependent translocation of ARNO to the plasma membrane of adipocytes requires phosphatidylinositol 3-kinase. *Curr. Biol.*, 8, 463–466, 1998.

[60] Watt S.A., Kular G., Fleming I.N., Downes C.P., and Lucocq J.M. Subcellular localization of phosphatidylinositol 4,5-bisphosphate using the pleckstrin homology domain of phospholipase Cδ1. *Biochem. J.*, 363, 657–666, 2002.

[61] Irvine R.F. Inositol lipids: To PHix or not to PHix? *Curr. Biol.*, 14, R308–R310, 2004.

[62] Yoon Y., Lee P.J., Kurilova S., and Cho W. *In situ* quantitative imaging of cellular lipids using molecular sensors. *Nat. Chem.*, 3, 868–874, 2011.

[63] van den Bogaart G., Meyenberg K., Risselada H.J., Amin H., Willig K.I., Hubrich B.E., Dier M., Hell S.W., Grubmüller H., Diederichsen U., and Jahn R. Membrane protein sequestering by ionic protein-lipid interactions. *Nature*, 479, 552–555, 2011.

[64] Tabin C.J., Bradley S.M., Bargmann C.I., Weinberg R.A., Papageorge A.G., Scolnick E.M., Dhar R., Lowy D.R., and Chang E.H. Mechanism of activation of a human oncogene. *Nature*, 300, 143–149, 1982.

[65] Chan T.O., Rodeck U., Chan A.M., Kimmelman A.C., Rittenhouse S.E., Panayotou G., and Tsichlis P.N. Small GTPases and tyrosine kinases coregulate a molecular switch in the phosphoinositide 3-kinase regulatory subunit. *Cancer Cell*, 1, 181–191, 2002.

[66] Cantley L.C. The phosphoinositide 3-kinase pathway. *Science*, 296, 1655–1657, 2002.

[67] Liu P., Cheng H., Roberts T.M., and Zhao J.J. Targeting the phosphoinositide 3-kinase pathway in cancer. *Nat. Rev. Drug Discov.*, 8, 627–644, 2009.

[68] Vivanco I. and Sawyers C.L. The phosphatidylinositol 3-kinase-AKT pathway in human cancer. *Nat. Rev. Cancer*, 2, 489–501, 2002.

[69] Wymann M.P. and Schneiter R. Lipid signaling in disease. *Nat. Rev. Mol. Cell Biol.*, 9, 162–176, 2008.

[70] Stambolic V., Suzuki A., de la Pompa J.L., Brothers G.M., Mirtsos C., Sasaki T., Ruland J., Penninger J.M., Siderovski D.P., and Mak T.W. Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell*, 95, 29–39, 1998.

[71] Bader A.G., Kang S., Zhao J.J., and Vogt P.K. Oncogenic PI3k deregulates transcription and translation. *Nat. Rev. Cancer*, 5, 921–929, 2005.

[72] Samuels Y., Wang Z., Bardelli A., Silliman N., Ptak J., Szabo S., Yan H., Gazdar A., Powell S.M., Riggins G.J., Willson J.K.V., Markowitz S., Kinzler K.W., Vogelstein B., and Velculescu V.E. High frequency of mutations of the *PIK3CA* gene in human cancers. *Science*, 304, 554, 2004.

[73] Scott J.L., Musselman C.A., Adu-Gyamfi E., Kutateladze T.G., and Stahelin R.V. Emerging methodologies to investigate lipid-protein interactions. *Integr. Biol.*, 4, 247–258, 2012.

[74] Mittermaier A. and Kay L.E. New tools to provide new insights in NMR studies of protein dynamics. *Science*, 312, 224–228, 2006.

[75] Dyson H.J. and Wright P.E. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.*, 104, 3607–3622, 2004.

[76] Pickford A.R. and Campbell I.D. NMR studies of modular protein structures and their interactions. *Chem. Rev.*, 104, 3557–3565, 2004.

[77] Clarkson J. and Campbell I.D. Studies of protein-ligand interactions by NMR. *Biochem. Soc. Trans.*, 31, 1006–1009, 2003.

[78] Lumb C.N., He J., Xue Y., Stansfeld P.J., Stahelin R.V., Kutateladze T.G., and Sansom M.S.P. Biophysical and computational studies of membrane penetration by the GRP1 pleckstrin homology domain. *Structure*, 19, 1338–1346, 2011.

[79] Lee S.A., Kovacs J., Stahelin R.V., Cheever M.L., Overduin M., Gangi Setty T., Burd C.G., Cho W., and Kutateladze T.G. Molecular mechanism of membrane docking by the Vam7p PX domain. *J. Biol. Chem.*, 281, 37091–37101, 2006.

[80] Franks W.T., Linden A.H., Kunert B., van Rossum B.-J., and Oschkinat H. Solid-state magic-angle spinning NMR of membrane proteins and protein-ligand interactions. *Eur. J. Cell Biol.*, 91, 340–348, 2012.

[81] Judge P.J. and Watts A. Recent contributions from solid-state NMR to the understanding of membrane protein structure and function. *Curr. Opin. Chem. Biol.*, 15, 690–695, 2011.

[82] Cross T.A., Sharma M., Yi M., and Zhou H.-X. Influence of solubilizing environments on membrane protein structures. *Trends Biochem. Sci.*, 36, 117–125, 2011.

[83] Park S.H., Berkamp S., Cook G.A., Chan M.K., Viadiu H., and Opella S.J. Nanodiscs versus macrodiscs for NMR of membrane proteins. *Biochemistry*, 50, 8983–8985, 2011.

[84] Pinheiro J.T. and Watts A. Resolution of individual lipids in mixed phospholipid membranes and specific lipid-cytochrome *c* interactions by magic-angle spinning solid-state phosphorus-31 NMR. *Biochemistry*, 33, 2459–2467, 1994.

[85] Tuzi S., Uekama N., Okada M., Yamaguchi S., Saitô H., and Yagisawa H. Structure and dynamics of the phospholipase C-δ1 pleckstrin homology domain located at the lipid bilayer surface. *J. Biol. Chem.*, 278, 28019–28025, 2003.

[86] Malmberg N.J., van Buskirk D.R., and Falke J.J. Membrane-docking loops of the $cPLA_2$ C2 domain: Detailed structural analysis of the protein-membrane interface via site-directed spin-labeling. *Biochemistry*, 42, 13227–13240, 2003.

[87] Lai C.-L., Landgraf K.E., Voth G.A., and Falke J.J. Membrane docking geometry and target lipid stoichiometry of membrane-bound PKCα C2 domain: A combined molecular dynamics and experimental study. *J. Mol. Biol.*, 402, 301–310, 2010.

[88] Chen H.-C., Ziemba B.P., Landgraf K.E., Corbin J.E., and Falke J.J. Membrane docking geometry of GRP1 PH domain bound to a target lipid bilayer: An EPR site-directed spin-labeling and relaxation study. *PLoS ONE*, 7, e33640, 2012.

[89] Hamman B.D., Pollok B.A., Bennett T., Allen J., and Heim R. Binding of a pleckstrin homology domain protein to phosphoinositide in membranes: A miniaturized FRET-based assay for drug screening. *J. Biomol. Screen.*, 7, 45–55, 2002.

[90] Sato M., Ueda Y., Takagi T., and Umezawa Y. Production of $PtdInsP_3$ at endomembranes is triggered by receptor endocytosis. *Nat. Cell Biol.*, 5, 1016–1022, 2003.

[91] Beseničar M., Maček P., Lakey J.H., and Anderluh G. Surface plasmon resonance in protein-membrane interactions. *Chem. Phys. Lipids*, 141, 169–178, 2006.

[92] Stahelin R.V. and Cho W. Differential roles of ionic, aliphatic, and aromatic residues in membrane-protein interactions: A surface plasmon resonance study on phospholipases A_2. *Biochemistry*, 40, 4672–4678, 2001.

[93] Ananthanarayanan B., Stahelin R.V., Digman M.A., and Cho W. Activation mechanisms of conventional protein kinase C isoforms are determined by the ligand affinity and conformational flexibility of their C1 domains. *J. Biol. Chem.*, 278, 46886–46894, 2003.

[94] Stahelin R.V., Digman M.A., Medkova M., Ananthanarayanan B., Melowic H.R., Rafter J.D., and Cho W. Diacylglycerol-induced membrane targeting and activation of protein kinase Cϵ: Mechanistic differences between protein kinases Cδ and Cϵ. *J. Biol. Chem.*, 280, 19784–19793, 2005.

[95] Maynard J.A., Lindquist N.C., Sutherland J.N., Lesuffleur A., Warrington A.E., Rodriguez M., and Sang-Hyun O. Surface plasmon resonance for high-throughput ligand screening of membrane-bound proteins. *Biotechnol. J.*, 4, 1542–1558, 2009.

[96] Knight J.D. and Falke J.J. Single molecule fluorescence studies of a PH domain: New insights into the membrane docking reaction. *Biophys. J.*, 96, 566–582, 2009.

[97] Fox C.B., Wayment J.R., Myers G.A., Endicott S.K., and Harris J.M. Single-molecule fluorescence imaging of peptide binding to supported lipid bilayers. *Anal. Chem.*, 81, 5130–5138, 2009.

[98] He J., Haney R.M., Vora M.S., Verkhusha V.V., Stahelin R.V., and Kutateladze T.G. Molecular mechanism of membrane targeting by the Grp1 PH domain. *J. Lipid Res.*, 49, 1807–1815, 2008.

[99] McCammon J.A., Gelin B.R., and Karplus M. Dynamics of folded proteins. *Nature*, 267, 585–590, 1977.

[100] van der Ploeg P. and Berendsen H.J.C. Molecular dynamics simulation of a bilayer membrane. *J. Chem. Phys.*, 76, 3271–3276, 1982.

[101] Heller H., Schaefer M., and Schulten K. Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid-crystal phases. *J. Phys. Chem.*, 97, 8343–8360, 1993.

[102] Venable R.M., Zhang Y., and Pastor R.W. Molecular dynamics simulations of a lipid bilayer and of hexadecane: An investigation of membrane fluidity. *Science*, 262, 223–226, 1993.

[103] Robinson A.J., Richards W.G., Thomas P.J., and Hann M.M. Behavior of cholesterol and its effect on head group and chain conformations in lipid bilayers: A molecular dynamics study. *Biophys. J.*, 68, 164–170, 1995.

[104] Woolf T.B. and Roux B. Molecular dynamics simulation of the gramicidin channel in a phospholipid bilayer. *Proc. Natl. Acad. Sci. U.S.A.*, 91, 11631–11635, 1994.

[105] Woolf T.B. and Roux B. Structure, energetics, and dynamics of lipid-protein interactions: A molecular dynamics study of the gramicidin a channel in a DMPC bilayer. *Proteins: Struct., Funct., Genet.*, 24, 92–114, 1996.

[106] Edholm O., Berger O., and Jähnig F. Structure and fluctuations of bacteriorhodopsin in the purple membrane: A molecular dynamics study. *J. Mol. Biol.*, 250, 94–111, 1995.

[107] Zhou F. and Schulten K. Molecular dynamics study of phospholipase A_2 on a membrane surface. *Proteins: Struct., Funct., Genet.*, 25, 12–27, 1996.

[108] Lindahl E. and Sansom M.S.P. Membrane proteins: Molecular dynamics simulations. *Curr. Opin. Struct. Biol.*, 18, 425–431, 2008.

[109] Blood P.D. and Voth G.A. Direct observation of Bin/amphiphysin/Rvs (BAR) domain-induced membrane curvature by means of molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.*, 103, 15068–15072, 2006.

[110] Arkhipov A., Yin Y., and Schulten K. Four-scale description of membrane sculpting by BAR domains. *Biophys. J.*, 95, 2806–2821, 2008.

[111] Jaud S., Tobias D.J., Falke J.J., and White S.H. Self-induced docking site of a deeply embedded peripheral membrane protein. *Biophys. J.*, 92, 517–524, 2007.

[112] Manna D., Bhardwaj N., Vora M.S., Stahelin R.V., Lu H., and Cho W. Differential roles of phosphatidylserine PtdIns(4,5)P_2 and PtdIns(3,4,5)P_3 in plasma membrane targeting of C2 domains. *J. Biol. Chem.*, 283 (38), 26047–26058, 2008.

[113] Ohkubo Y.Z. and Tajkhorshid E. Distinct structural and adhesive roles of Ca^{2+} in membrane binding of blood coagulation factors. *Structure*, 16, 72–81, 2008.

[114] Psachoulia E. and Sansom M.S.P. Interactions of the pleckstrin homology domain with phosphatidylinositol phosphate and membranes: Characterization via molecular dynamics simulations. *Biochemistry*, 47, 4211–4220, 2008.

[115] Psachoulia E. and Sansom M.S.P. PX- and FYVE-mediated interactions with membranes: Simulation studies. *Biochemistry*, 48, 5090–5095, 2009.

[116] Bhardwaj N., Stahelin R.V., Langlois R.E., Cho W., and Lu H. Structural bioinformatics prediction of membrane-binding proteins. *J. Mol. Biol.*, 359, 486–495, 2006.

[117] Bhardwaj N., Stahelin R.V., Zhao G., Cho W., and Lu H. MeTaDoR: A comprehensive resource for membrane targeting domains and their host proteins. *Bioinformatics*, 23, 3110–3112, 2007.

[118] Lomize M.A., Lomize A.L., Pogozheva I.D., and Mosberg H.I. OPM: Orientations of Proteins in Membranes database. *Bioinformatics*, 22, 623–625, 2006.

[119] Lomize A.L., Pogozheva I.D., Lomize M.A., and Mosberg H.I. Positioning of proteins in membranes: A computational approach. *Protein Sci.*, 15, 1318–1333, 2006.

[120] Lomize A.L., Pogozheva I.D., Lomize M.A., and Mosberg H.I. The role of hydrophobic interactions in positioning of peripheral proteins in membranes. *BMC Struct. Biol.*, 7, 44, 2007.

[121] Diraviyam K., Stahelin R.V., Cho W., and Murray D. Computer modeling of the membrane interaction of FYVE domains. *J. Mol. Biol.*, 328, 721–736, 2003.

[122] Singh S.M. and Murray D. Molecular modeling of the membrane targeting of phospholipase C pleckstrin homology domains. *Protein Sci.*, 12, 1934–1953, 2003.

[123] Medkova M. and Cho W. Interplay of C1 and C2 domains of protein kinase C-α in its membrane binding and activation. *J. Biol. Chem.*, 274, 19852–19861, 1999.

[124] Rizo J. and Südhof T.C. C2 domains, structure and function of a universal Ca^{2+}-binding domain. *J. Biol. Chem.*, 273, 15879–15882, 1998.

[125] Murray D. and Honig B. Electrostatic control of the membrane targeting of C2 domains. *Mol. Cell*, 9, 145–154, 2002.

[126] Hammes-Schiffer S. and Benkovic S.J. Relating protein motion to catalysis. *Annu. Rev. Biochem.*, 75, 519–541, 2006.

[127] Smock R.G. and Gierasch L.M. Sending signals dynamically. *Science*, 324, 198–203, 2009.

[128] Rong S.-B., Hu Y., Enyedy I., Powis G., Meuillet E.J., Wu X., Wang R., Wang S., and Kozikowski A.P. Molecular modeling studies of the Akt PH domain and its interaction with phosphoinositides. *J. Med. Chem.*, 44, 898–908, 2001.

[129] Mills S.J., Komander D., Trusselle M.N., Safrany S.T., van Aalten D.M.F., and Potter B.V.L. Novel inositol phospholipid headgroup surrogate crystallized in the pleckstrin homology domain of protein kinase Bα. *ACS Chem. Biol.*, 2, 242–246, 2007.

[130] Dancea F., Kami K., and Overduin M. Lipid interaction networks of peripheral membrane proteins revealed by data-driven micelle docking. *Biophys. J.*, 94, 515–524, 2008.

[131] Rosen S.A.J., Gaffney P.R.J., Spiess B., and Gould I.R. Understanding the relative affinity and specificity of the pleckstrin homology domain of protein kinase B for inositol phosphates. *Phys. Chem. Chem. Phys.*, 14, 929–936, 2012.

[132] Manna D., Albanese A., Park W.S., and Cho W. Mechanistic basis of differential cellular responses of phosphatidylinositol 3,4-bisphosphate- and phosphatidylinositol 3,4,5-trisphosphate-binding pleckstrin homology domains. *J. Biol. Chem.*, 282, 32093–32105, 2007.

[133] Ohkubo Y.Z., Pogorelov T.V., Arcario M.J., Christiansen G.A., and Tajkhorshid E. Accelerating membrane insertion of peripheral proteins with a novel membrane mimetic model. *Biophys. J.*, 102, 2130–2139, 2012.

[134] Karplus M. and McCammon J.A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.*, 9, 646–652, 2002.

[135] Verlet L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159, 98–103, 1967.

[136] Hansen J.-P. and McDonald I.R. *Theory of Simple Liquids*. Academic Press, 3rd edition, 2006.

[137] Hockney R.W. and Eastwood J.W. *Computer Simulations Using Particles*. McGraw-Hill, New York, 1981.

[138] Frenkel D. and Smit B. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 2nd edition, 2002.

[139] Hess B., Bekker H., Berendsen H.J.C., and Fraaije J.G.E.M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18, 1463–1472, 1997.

[140] Brooks B.R., Brooks III C.L., Mackerell A.D., Nilsson L., Petrella R.J., Roux B., Won Y., Archontis G., Bartels C., Boresch S., Caflisch A., Caves L., Cui Q., Dinner A.R., Feig M., Fischer S., Gao J., Hodoscek M., Im W., Kuczera K., Lazaridis T., Ma J., Ovchinnikov V., Paci E., Pastor R.W., Post C.B., Pu J.Z., Schaefer M., Tidor B., Venable R.M., Woodcock H.L., Wu X., Yang W., York D.M., and Karplus M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.*, 30, 1545–1615, 2009.

[141] Case D.A., Cheatham III D.E., Darden T., Gohlke H., Luo R., Merz Jr K.M., Onufriev A., Simmerling C., Wang B., and Woods R. The Amber biomolecular simulation program. *J. Comput. Chem.*, 26, 1668–1688, 2005.

[142] van Gunsteren W.F., Billeter S.R., Eising A.A., Hünenberger P.H., Krüger P., Mark A.E., Scott W., and Tironi I.G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide.* Biomos & Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996.

[143] Jorgensen W.L., Madura J.D., and Swenson C.J. Optimized intermolecular potential functions for liquid hydrocarbons. *J. Am. Chem. Soc.*, 106, 6638–6646, 1984.

[144] Berger O., Edholm O., and Jähnig F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.*, 72, 2002–2013, 1997.

[145] Oostenbrink C., Villa A., Mark A.E., and van Gunsteren W.F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, 25, 1656–1676, 2004.

[146] Hermans J., Berendsen H.J.C., van Gunsteren W.F., and Postma J.P.M. A consistent empirical potential for water-protein interactions. *Biopolymers*, 23, 1513–1518, 1984.

[147] Schuettelkopf A.W. and van Aalten D.M.F. PRODRG - A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr.*, D60, 1355–1363, 2004.

[148] Lemkul J.A., Allen W.J., and Bevan D.R. Practical considerations for building GROMOS-compatible small-molecule topologies. *J. Chem. Inf. Model.*, 50, 2221–2235, 2010.

[149] Frisch M. J., Trucks G. W., Schlegel H. B., Scuseria G. E., Robb M. A., Cheeseman J. R., Montgomery Jr J. A., Vreven T., Kudin K. N., Burant J. C., Millam J. M., Iyengar S. S., Tomasi J., Barone V., Mennucci B., Cossi M., Scalmani G., Rega N., Petersson G. A., Nakatsuji H., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Klene M., Li X., Knox J. E., Hratchian H. P., Cross J. B., Bakken V., Adamo C., Jaramillo J., Gomperts R., Stratmann R. E., Yazyev O., Austin A. J., Cammi R., Pomelli C., Ochterski J. W., Ayala P. Y., Morokuma K., Voth G. A., Salvador P., Dannenberg J. J., Zakrzewski V. G., Dapprich S., Daniels A. D., Strain M. C., Farkas O., Malick D. K., Rabuck A. D., Raghavachari K., Foresman J. B., Ortiz J. V., Cui Q., Baboul A. G., Clifford S., Cioslowski J., Stefanov B. B., Liu G., Liashenko A., Piskorz P., Komaromi I., Martin R. L., Fox D. J., Keith T., Al-Laham M. A., Peng C. Y., Nanayakkara A., Challacombe M., Gill P. M. W., Johnson B., Chen W., Wong M. W., Gonzalez C., and Pople J. A. Gaussian 03 Revision E.01. Gaussian Inc. Wallingford CT 2004.

[150] Becke A.D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98, 5648–5652, 1993.

[151] Lee C., Yang W., and Parr R.G. Development of the Colle-Salvetti correlation energy formula into a functional of the electron density. *Phys. Rev. B*, 37, 785–789, 1988.

[152] Blood P.D., Swenson R.D., and Voth G.A. Factors influencing local membrane curvature induction by N-BAR domains as revealed by molecular dynamics simulations. *Biophys. J.*, 95, 1866–1876, 2008.

[153] Lupyan D., Mezel M., Logothetis D.E., and Osman R. A molecular dynamics investigation of lipid bilayer perturbation by PIP2. *Biophys. J.*, 98, 240–247, 2010.

[154] Kaminski G.A., Friesner R.A., Tirado-Rives J., and Jorgensen W.L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B*, 105, 6474–6487, 2001.

[155] Jorgensen W.L., Chandrasekhar J., Madura J.D., Impey R.W., and Klein M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79, 926–934, 1983.

[156] Leach A. R. *Molecular Modelling: Principles and Applications*. Prentice Hall, 2nd edition, 2001.

[157] Steinbach P.J. and Brooks B.R. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.*, 15, 667–683, 1994.

[158] Deserno M. and Holm C. How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines. *J. Chem. Phys.*, 109, 7678–7693, 1998.

[159] Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A., and Haak J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81, 3684–3690, 1984.

[160] Morishita T. Fluctuation formulas in molecular dynamics simulations with the weak coupling heat bath. *J. Chem. Phys.*, 113, 2976–2982, 2000.

[161] Andersen H.C. Molecular dynamics at constant pressure and/or temperature. *J. Chem. Phys.*, 72, 2384–2393, 1980.

[162] Parrinello M. and Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52, 7182–7190, 1981.

[163] Nosé S. and Klein M.L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.*, 50, 1055–1076, 1983.

[164] Hammond G.R.V., Sim Y., Lagnado L., and Irvine R.F. Reversible binding and rapid diffusion of proteins in complex with inositol lipids serves to coordinate free movement with spatial information. *J. Cell Biol.*, 184, 297–308, 2009.

[165] Grubmüller H. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E*, 52, 2893–2906, 1995.

[166] Dellago C., Bolhuis P.G., and Chandler D. Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements. *J. Chem. Phys.*, 108, 9236–9245, 1998.

[167] Laio A. and Parrinello M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.*, 99, 12562–12566, 2002.

[168] Perilla J.R., Beckstein O., Denning E.J., and Woolf T.B. Computing ensembles of transitions from stable states: Dynamic importance sampling. *J. Comput. Chem.*, 32, 196–209, 2011.

[169] Lu H., Isralewitz B., Krammer A., Vogel V., and Schulten K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.*, 75, 662–671, 1998.

[170] Isralewitz B., Gao M., and Schulten K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.*, 11, 224–230, 2001.

[171] Zhang D., Gullingsrud J., and McCammon J.A. Potentials of mean force for acetylcholine unbinding from the alpha7 nicotinic acetylcholine receptor ligand-binding domain. *J. Am. Chem. Soc.*, 128, 3019–3026, 2006.

[172] Gräter F., de Groot B.L., Jiang H., and Grubmüller H. Ligand-release pathways in the pheromone-binding protein of *Bombyx mori*. *Structure*, 14, 1567–1576, 2006.

[173] Jarzynski C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78, 2690–2693, 1997.

[174] Park S., Khalili-Araghi F., Tajkhorshid E., and Schulten K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys.*, 119, 3559–3566, 2003.

[175] Marrink S.-J., Risselada H.J., Yefimov S., Tieleman D.P., and de Vries A.H. The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, 111, 7812–7824, 2007.

[176] Monticelli L., Kandasamy S.K., Periole X., Larson R.G., Tieleman D.P., and Marrink S-.J. The MARTINI coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput.*, 4, 819–834, 2008.

[177] Ayton G.S., Blood P.D., and Voth G.A. Membrane remodeling from N-BAR domain interactions: Insights from multi-scale simulation. *Biophys. J.*, 92, 3595–3602, 2007.

[178] Ayton G.S., Lyman E., Krishna V., Swenson R.D., Mim C., Unger V.M., and Voth G.A. New insights into BAR domain-induced membrane remodeling. *Biophys. J.*, 97, 1616–1625, 2009.

[179] Helfrich W. Elastic properties of lipid bilayers: Theory and possible experiments. *Z. Naturforsch. [C]*, 28, 693–703, 1973.

[180] Arkhipov A., Yin Y., and Schulten K. Four-scale description of membrane sculpting by BAR domains. *Biophys. J.*, 95, 2806–2821, 2008.

[181] Bond P.J. and Sansom M.S.P. Insertion and assembly of membrane proteins via simulation. *J. Am. Chem. Soc.*, 128, 2697–2704, 2006.

[182] Shelley J.C., Shelley M.Y., Reeder R.C., Bandyopadhyay S., and Klein M.L. A coarse grain model for phospholipid simulations. *J. Phys. Chem. B*, 105, 4464–4470, 2001.

[183] Stevens M.J. Coarse-grained simulations of lipid bilayers. *J. Chem. Phys.*, 121, 11942–11948, 2004.

[184] Arkhipov A., Freddolino P.L., and Schulten K. Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure*, 14, 1767–1777, 2006.

[185] Lindahl E., Hess B., and van der Spoel D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.*, 7, 306–317, 2001.

[186] Hess B., Kutzner C., van der Spoel D., and Lindahl E. GROMACS 4: Algorithms for highly efficient load-balanced and scalable molecular simulation. *J. Chem. Theor. Comput.*, 4, 435–447, 2008.

[187] Lu Y., Wang R., Yang C.-Y., and Wang S. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J. Chem. Inf. Model.*, 47, 668–675, 2007.

[188] Langevin P. Sur la théorie du mouvement Brownien. *C. R. Acad. Sci.*, 146, 530–533, 1908.

[189] Einstein A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.*, 322, 549–560, 1905.

[190] Gabdoulline R.R. and Wade R.C. Simulation of the diffusional association of barnase and barstar. *Biophys. J.*, 72, 1917–1929, 1997.

[191] Davis M.E., Madura J.D, Luty B.A., and McCammon J.A. Electrostatics and diffusion of molecules in solution: Simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun.*, 62, 187–197, 1991.

[192] Madura J.D., Briggs J.M., Wade R.C., Davis M.E., Luty B.A., Ilin A., Antosiewicz J., Gilson M.K., Bagheri B., Scott L.R., and McCammon J.A. Electrostatics and diffusion of molecules in solution: Simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun.*, 91, 57–95, 1995.

[193] Northrup S., Laughner T., and Stevenson G. Macrodox macromolecular simulation program, 1997.

[194] Ermak D.L. and McCammon J.A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.*, 69, 1352–1360, 1978.

[195] Rotne J. and Prager S. Variational treatment of hydrodynamic interaction in polymers. *J. Chem. Phys.*, 50, 4831–4837, 1969.

[196] Yamakawa H. Transport properties of polymer chains in dilute solution: Hydrodynamic interaction. *J. Chem. Phys.*, 53, 436–443, 1970.

[197] Mazo R.M. *Brownian Motion*. Oxford University Press, 1st edition, 2002.

[198] Frembgen-Kesner T. and Elcock A.H. Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins. *J. Chem. Theory Comput.*, 5, 242–256, 2009.

[199] Frembgen-Kesner T. and Elcock A.H. Absolute protein-protein association rate constants from flexible coarse-grained Brownian dynamics simulations: The role of intermolecular hydrodynamic interactions in barnase-barstar association. *Biophys. J.*, 99, L75–L77, 2010.

[200] Ando T. and Skolnick J. Crowding and hydrodynamic interactions likely dominate *in vivo* macromolecular motion. *Proc. Natl. Acad. Sci. U.S.A.*, 43, 18457–18462, 2010.

[201] Geyer T. and Winter U. An $\mathcal{O}(N^2)$ approximation for hydrodynamic interactions in Brownian dynamics simulations. *J. Chem. Phys.*, 130, 114905, 2009.

[202] Northrup S.H., Allison S.A., and McCammon J.A. Brownian dynamics simulation of diffusion-induced biomolecular reactions. *J. Chem. Phys.*, 80, 1517–1524, 1984.

[203] Smoluchowski M. Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation der Kolloidteilchen. *Phys. Z.*, 17, 557–585, 1916.

[204] Hansen J.-P. and Löwen H. Effective interactions between electric double layers. *Annu. Rev. Phys. Chem.*, 51, 209–242, 2000.

[205] Gouy L.G. Sur la constitution de la charge électrique à la surface d'un électrolyte. *C. R. Hebd. Acad. Sci.*, 149, 654–657, 1909.

[206] Chapman D.L. A contribution to the theory of electrocapillarity. *Philos. Mag.*, 25, 475–481, 1913.

[207] Dill K.A. and Bromberg S. *Molecular Driving Forces*. Garland Science, 1st edition, 2003.

[208] Debye P. and Hückel E. Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen. *Phys. Z.*, 24, 185–206, 1923.

[209] Baker N.A. Poisson-Boltzmann methods for biomolecular electrostatics. *Methods Enzymol.*, 383, 94–118, 2004.

[210] Baker N.A., Sept D., Joseph S., Holst M.J., and McCammon J.A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, 98, 10037–10041, 2001.

[211] Rocchia W., Alexov E., and Honig B. Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions. *J. Phys. Chem. B*, 105, 6507–6514, 2001.

[212] Rocchia W., Sridharan S., Nicholls A., Alexov E., Chiabrera A., and Honig B. Rapid grid-based construction of the molecular surface for both molecules and geometric objects: Applications to the finite difference Poisson-Boltzmann method. *J. Comput. Chem.*, 23, 128–137, 2002.

[213] Holst M. and Saied F. Numerical solution of the nonlinear Poisson-Boltzmann equation: Developing more robust and efficient methods. *J. Comput. Chem.*, 16, 337–364, 1995.

[214] Guldbrand L., Jönnson B., Wennerström H., and Linse P. Electrical double layer forces. a Monte Carlo study. *J. Chem. Phys.*, 80, 2221–2228, 1984.

[215] Dolinsky T.J., Nielsen J.E., McCammon J.A., and Baker N.A. PDB2PQR: An automated pipeline for the setup execution and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.*, 32, W665–W667, 2004.

[216] Elcock A.H., Gabdoulline R.R., Wade R.C., and McCammon J.A. Computer simulation of protein-protein assocation kinetics: Acetylcholinesterase-fasciculin. *J. Mol. Biol.*, 291, 149–162, 1999.

[217] de la Torre J.G., Huertas M.L., and Carrasco B. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.*, 78, 719–730, 2000.

[218] Gabdoulline R.R. and Wade R.C. Effective charges for macromolecules in solvent. *J. Phys. Chem.*, 100, 3868–3878, 1996.

[219] Marshall G.R. Computer-aided drug design. *Ann. Rev. Pharmacol. Toxicol.*, 27, 193–213, 1987.

[220] Morris G.M., Huey R., Hart W.E., Belew R.K., and Olson A.J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, 19, 1639–1662, 1998.

[221] Huey R., Morris G.M., Olson A.J., and Goodsell D.S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.*, 28, 1145–1152, 2007.

[222] Verdonk M.L., Cole J.C., Hartshorn M.J., Murray C.W., and Taylor R.D. Improved protein-ligand docking using GOLD. *Proteins*, 52, 609–623, 2003.

[223] Friesner R.A., Banks J.L., Murphy R.B., Halgren T.A., Klicic J.J., Mainz D.T., Repasky M.P., Knoll E.H., Shelley M., Perry J.K., Shaw D.E., Francis P., and Shenkin P.S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, 47, 1739–1749, 2004.

[224] Venkatachalam C.M., Jiang X., Oldfield T., and Waldman M. LigandFit: A novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graph. Model.*, 21, 289–307, 2003.

[225] Klarlund J.K., Guilherme A., Holik J.J., Virbasius J.V., Chawla A., and Czech M.P. Signaling by phosphoinositide-3,4,5-trisphosphate through proteins containing pleckstrin and Sec7 homology domains. *Science*, 275, 1927–1930, 1997.

[226] Tannert A., Voigt P., Burgold S., Tannert S., and Schaefer M. Signal amplification between G$\beta\gamma$ release and PI3Kγ-mediated PI(3,4,5)P_3 formation monitored by a fluorescent G$\beta\gamma$ biosensor protein and repetitive two component total internal reflection/fluorescence redistribution after photobleaching analysis. *Biochemistry*, 47, 11239–11250, 2008.

[227] Sakaguchi R., Tainaka K., Shimada N., Nakano S., Inoue M., Kiyonaka S., Mori Y., and Morii T. An *in vivo* fluorescent sensor reveals intracellular Ins(1,3,4,5)P4 dynamics in single cells. *Angew. Chem.*, 49, 2150–2153, 2010.

[228] D'Souza-Schorey C. and Chavrier P. ARF proteins: Roles in membrane traffic and beyond. *Nat. Rev. Mol. Cell Biol.*, 7, 347–358, 2006.

[229] Tague S.E., Muralidharan V., and D'Souza-Schorey C. ADP-ribosylation factor 6 regulates tumor cell invasion through the activation of the MEK/ERK signaling pathway. *Proc. Natl. Acad. Sci. U.S.A.*, 101, 9671–9676, 2004.

[230] DiNitto J.P., Delprato A., Lee M.-T., Cronin T., Huang S., Guilherme A., Czech M.P., and Lambright D.G. Structural basis and mechanism of autoregulation in 3-phosphoinositide-dependent Grp1 family Arf GTPase exchange factors. *Mol. Cell*, 28, 569–583, 2007.

[231] Klarlund J.K., Tsiaras W., Holik J.J., Chawla A., and Czech M.P. Distinct polyphosphoinositide binding selectivities for pleckstrin homology domains of GRP1-like proteins based on diglycine *versus* triglycine motifs. *J. Biol. Chem.*, 275, 32816–32821, 2000.

[232] Cronin T., DiNitto J.P., Czech M.P., and Lambright D.G. Structural determinants of phosphoinositide selectivity in splice variants of Grp1 family PH domains. *EMBO J.*, 23, 3711–3720, 2004.

[233] Ferguson K.M., Kavran J.M., Sankaran V.G., Fournier E., Isakoff S.J., Skolnik E.Y., and Lemmon M.A. Structural basis for discrimination of 3-phosphoinositides by pleckstrin homology domains. *Mol. Cell*, 6, 373–384, 2000.

[234] Li Z., Venable R.M., Rogers L.A., Murray D., and Pastor R.W. Molecular dynamics simulations of PIP_2 and PIP_3 in lipid bilayers: Determination of ring orientation and the effects of surface roughness on a Poisson-Boltzmann description. *Biophys. J.*, 97, 155–163, 2009.

[235] Darden T., York D., and Pedersen L. Particle mesh Ewald: An N log N method for Ewald sums in large systems. *J. Chem. Phys.*, 98, 10089–10092, 1993.

[236] Hermans J., Berendsen H.J.C., van Gunsteren W.F., and Postma J.P.M. A consistent empirical potential for water-protein interactions. *Biopolymers*, 23, 1513–1518, 1984.

[237] Knight J.D., Lerner M.G., Marcano-Velázquez, Pastor R.W., and Falke J.J. Single molecule diffusion of membrane-bound proteins: Window into lipid contacts and bilayer dynamics. *Biophys. J.*, 99, 2879–2887, 2010.

[238] Corbin J.A., Dirkx R.A., and Falke J.J. GRP1 pleckstrin homology domain: Activation parameters and novel search mechanism for rare target lipid. *Biochemistry*, 43, 16161–16173, 2004.

[239] Contera S.A., Lemaître V., de Planque M.R.R., Watts A., and Ryan J.F. Unfolding and extraction of a transmembrane α-helical peptide: Dynamic force spectroscopy and molecular dynamics simulations. *Biophys. J.*, 89, 3129–3140, 2005.

[240] Marrink S.-J., Berger O., Tieleman D.P., and Jähnig F. Adhesion forces of lipids in a phospholipid membrane studied by molecular dynamics simulations. *Biophys. J.*, 74, 931–943, 1998.

[241] Vemparala S., Saiz L., Eckenhoff R.G., and Klein M.L. Partitioning of anesthetics into a lipid bilayer and their interaction with membrane-bound peptide bundles. *Biophys. J.*, 91, 2815–2825, 2006.

[242] Flesch F.M., Yu J.W., Lemmon M.A., and Burger K.N.J. Membrane activity of the phospholipase C-δ1 pleckstrin homology (PH) domain. *Biochem. J.*, 389, 435–441, 2005.

[243] Lumb C.N. and Sansom M.S.P. Finding a needle in a haystack: The role of electrostatics in target lipid recognition by PH domains. *PLoS Comput. Biol.*, 8, e1002617, 2012.

[244] Macia E., Paris S., and Chabre M. Binding of the PH and polybasic C-terminal domains of ARNO to phosphoinositides and to acidic lipids. *Biochem.*, 39, 5893–5901, 2000.

[245] Golding I. and Cox E.C. Physical nature of bacterial cytoplasm. *Phys. Rev. Lett.*, 96, 098102, 2006.

[246] Konopka M.C., Shkel I.A., Cayley S., Record M.T., and Weisshaar J.C. Crowding and confinement effects on protein diffusion *in vivo*. *J. Bacteriol.*, 188, 6115–6123, 2006.

[247] Adam G. and Delbrück M. *Structural Chemistry and Molecular Biology*. W.H. Freeman & Company San Francisco CA., 1968.

[248] Axelrod D. New dimensions in two dimensions. *Biophys. J.*, 67, 1799–1800, 1994.

[249] McCloskey M.A. and Poo M.-M. Rates of membrane-associated reactions: Reduction of dimensionality revisited. *J. Cell Biol.*, 102, 88–96, 1986.

[250] Kholodenko B.N., Hoek J.B., and Westerhoff H.V. Why cytoplasmic signalling proteins should be recruited to cell membranes. *Trends Cell Biol.*, 10, 173–178, 2000.

[251] Felder C.E., Prilusky J., Silman I., and Sussman J.L. A server and database for dipole moments of proteins. *Nucleic Acids Res.*, 35, W512–W521, 2007.

[252] Zhang X., Rizo J., and Südhof T.C. Mechanism of phospholipid binding by the C_2A-domain of synaptotagmin I. *Biochemistry*, 37, 12395–12403, 1998.

[253] Schreiber G. and Fersht A.R. Rapid electrostatically assisted association of proteins. *Nat. Struct. Biol.*, 3, 427–431, 1996.

[254] Spaar A., Dammer C., Gabdoulline R.R., Wade R.C., and Helms V. Diffusional encounter of barnase and barstar. *Biophys. J.*, 90, 1913–1924, 2006.

[255] Madura J.D. and McCammon J.A. Brownian dynamics simulation of diffusional encounters between triose phosphate isomerase and D-glyceraldehyde phosphate. *J. Phys. Chem.*, 93, 7285–7287, 1989.

[256] McGuffee S.R. and Elcock A.H. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol.*, 6, e1000694, 2010.

[257] Gabdoulline R.R. and Wade R.C. On the contributions of diffusion and thermal activation to electron transfer between *Phormidium laminosum* plastocyanin and cytochrome *f*: Brownian dynamics simulations with explicit modeling of nonpolar desolvation interactions and electron transfer events. *J. Am. Chem. Soc.*, 131, 9230–9238, 2009.

[258] Cui M., Shen J., Briggs J.M., Fu W., Wu J., Zhang Y., Luo X., Chi Z., Ji R., Jiang H., and Chen K. Brownian dynamics simulations of the recognition of the scorpion toxin P05 with the small-conductance calcium-activated potassium channels. *J. Mol. Biol.*, 318, 417–428, 2002.

[259] Flöck D. and Helms V. A Brownian dynamics study: The effect of a membrane environment on an electron transfer system. *Biophys. J.*, 87, 65–74, 2004.

[260] Spaar A., Flöck D., and Helms V. Association of cytochrome *c* with membrane-bound cytochrome *c* oxidase proceeds parallel to the membrane rather than in bulk solution. *Biophys. J.*, 96, 1721–1732, 2009.

[261] Gorba C., Geyer T., and Helms V. Brownian dynamics simulations of simplified cytochrome *c* molecules in the presence of a charged surface. *J. Chem. Phys.*, 121, 457–464, 2004.

[262] Carlsson F., Hyltner E., Arnebrant T., Malmsten M., and Linse P. Lysozyme adsorption to charged surfaces. A Monte Carlo study. *J. Phys. Chem. B*, 108, 9871–9881, 2004.

[263] Spaar A. and Helms V. Free energy landscape of protein-protein encounter resulting from Brownian dynamics simulations of barnase:barstar. *J. Chem. Theor. Comput.*, 1, 723–736, 2005.

[264] van den Broek B., Lomholt M.A., Kalisch S.M.J., Metzler R., and Wuite G.J.L. How DNA coiling enhances target localization by proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 105, 15738–15742, 2008.

[265] Kim Y.C., Best R.B., and Mittal J. Macromolecular crowding effects on protein-protein binding affinity and specificity. *J. Chem. Phys.*, 133, 205101, 2010.

[266] Rincon V., Bocanegra R., Rodriguez-Huete A., Rivas G., and Mateu M.G. Effects of macromolecular crowding on the inhibition of virus assembly and virus-cell receptor recognition. *Biophys. J.*, 100, 738–746, 2011.

[267] Wang Q., Liang K.C., Czader A., Waxham M.N., and Cheung M.S. The effect of macromolecular crowding, ionic strength and calcium binding on calmodulin dynamics. *PLoS Comput. Biol.*, 7, e1002114, 2011.

[268] Böckmann R.A., Hac A., Heimburg T., and Grubmüller H. Effect of sodium chloride on a lipid bilayer. *Biophys. J.*, 85, 1647–1655, 2003.

[269] Saffman P.G. and Delbrück M. Brownian motion in biological membranes. *Proc. Natl. Acad. Sci. U.S.A.*, 72, 3111–3113, 1975.

[270] Gambin Y., Lopez-Esparza R., Reffay M., Sierecki E., Gov N.S., Genest M., Hodges R.S., and Urbach W. Lateral mobility of proteins in liquid membranes revisited. *Proc. Natl. Acad. Sci. U.S.A.*, 103, 2098–2102, 2006.

[271] Guigas G. and Weiss M. Size-dependent diffusion of membrane inclusions. *Biophys. J.*, 91, 2393–2398, 2006.

[272] Naji A., Levine A.J., and Pincus P.A. Corrections to the Saffman-Delbrück mobility for membrane bound proteins. *Biophys. J.*, 93, L49–L51, 2007.

[273] Wu Y., Vendome J., Shapiro L., Ben-Shaul A., and Honig B. Transforming binding affinities from three dimensions to two with application to cadherin clustering. *Nature*, 475, 510–513, 2011.

[274] Matsumoto M. and Nishimura T. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM T. Model. Comput. S.*, 8, 3–30, 1998.

[275] Rasper D.M. and Merrill A.R. Evidence for the modulation of *Pseudomonas aeruginosa* exotoxin A-induced pore formation by membrane surface charge density. *Biochemistry*, 33, 12981–12989, 1994.

[276] Narayanan V. and Scarlata S. Membrane binding and self-association of α-synucleins. *Biochemistry*, 40, 9927–9934, 2001.

[277] Harel M., Spaar A., and Schreiber G. Fruitful and futile encounters along the association reaction between proteins. *Biophys. J.*, 96, 4237–4248, 2009.

[278] van Zon R. and Schofield J. Constructing smooth potentials of mean force, radial distribution functions and probability densities from sampled data. *J. Chem. Phys.*, 132, 154110, 2010.

[279] Stansfeld P.J. and Sansom M.S.P. From coarse grained to atomistic: A serial multiscale approach to membrane protein simulations. *J. Chem. Theor. Comput.*, 7, 1157–1166, 2011.

[280] Chen C. and Pettitt B.M. The binding process of a nonspecific enzyme with DNA. *Biophys. J.*, 101, 1139–1147, 2011.

[281] Marcos E., Mestres P., and Crehuet R. Crowding induces differences in the diffusion of thermophilic and mesophilic proteins: A new look at neutron scattering results. *Biophys. J.*, 101, 2782–2789, 2011.

[282] Bunney T.D. and Katan M. Phosphoinoside signalling in cancer: Beyond PI3K and PTEN. *Nat. Rev. Cancer*, 10, 342–352, 2010.

[283] Baraldi E., Djinovic Carugo K., Hyvonen M., Surdo P.L., Riley A.M., Potter B.V., O'Brien R., Ladbury J.E., and Saraste M. Structure of the PH domain from Bruton's tyrosine kinase in complex with inositol 1,3,4,5-tetrakisphosphate. *Structure Fold. Des.*, 7, 449–460, 1999.

[284] Thomas C.C., Deak M., Alessi D.R., and van Aalten D.M.F. High-resolution structure of the pleckstrin homology domain of protein kinase B/Akt bound to phosphatidylinositol (3,4,5)-trisphosphate. *Curr. Biol.*, 12, 1256–1262, 2002.

[285] Ferguson K.M., Lemmon M.A., Schlessinger J., and Sigler P.B. Structure of the high affinity complex of inositol trisphosphate with a phospholipase C pleckstrin homology domain. *Cell*, 83, 1037–1046, 1995.

[286] Ripoll D. R., Faerman C. H., Axelsen P. H., Silman I., and Sussman J. L. An electrostatic mechanism for substrate guidance down the aromatic gorge of acetylcholinesterase. *Proc. Natl. Acad. Sci. U.S.A.*, 90, 5128–5132, 1993.

[287] Servant G., Weiner O.D., Herzmark P., Balla T., Sedat J.W., and Bourne H.R. Polarization of chemoattractant receptor signaling during neutrophil chemotaxis. *Science*, 287, 1037–1040, 2000.

[288] Shewan A., Eastburn D.J., and Mostov K. Phosphoinositides in cell architecture. *Cold Spring Harb. Perspect. Biol.*, 3, a004796, 2011.

[289] Khandelia H., Ipsen J.H., and Mouritsen O.G. The impact of peptides on lipid membranes. *Biochim. Biophys. Acta*, 1778, 1528–1536, 2008.

[290] May S., Harries D., and Ben-Shaul A. Lipid demixing and protein-protein interactions in the adsorption of charged proteins on mixed membranes. *Biophys. J.*, 79, 1747–1760, 2000.

[291] Mbamala E.C., Ben-Shaul A., and May S. Domain formation induced by the adsorption of charged proteins on mixed lipid membranes. *Biophys. J.*, 88, 1702–1714, 2005.

[292] Gambhir A., Hangyás-Mihályné G., Zaitseva I., Cafiso D.S., Wang J., Murray D., Pentyala S.N., Smith S.O., and McLaughlin S. Electrostatic sequestration of PIP_2 on phospholipid membranes by basic/aromatic regions of proteins. *Biophys. J.*, 86, 2188–2207, 2004.

[293] Wang J., Gambhir A., McLaughlin S., and Murray D. A computational model for the electrostatic sequestration of $PI(4,5)P_2$ by membrane-adsorbed basic peptides. *Biophys. J.*, 86, 1969–1986, 2004.

[294] Hinderliter A., Almeida P.F.F., Creutz C.E., and Biltonen R.L. Domain formation in a fluid mixed lipid bilayer modulated through binding of the C2 protein motif. *Biochemistry*, 40, 4181–4191, 2001.

[295] Hinderliter A., Biltonen R.L., and Almeida P.F.F. Lipid modulation of protein-induced membrane domains as a mechanism for controlling signal transduction. *Biochemistry*, 43, 7102–7110, 2004.

[296] Dias R.S., Pais A.A.C.C., Linse P., Miguel M.G., and Lindman B. Polyion adsorption onto catanionic surfaces: A Monte Carlo study. *J. Phys. Chem. B*, 109, 11781–11788, 2005.

[297] Tzlil S. and Ben-Shaul A. Flexible charged macromolecules on mixed fluid lipid membranes: Theory and Monte Carlo simulations. *Biophys. J.*, 89, 2972–2987, 2005.

[298] Tzlil S., Murray D., and Ben-Shaul A. The "electrostatic-switch" mechanism: Monte Carlo study of MARCKS-membrane interaction. *Biophys. J.*, 95, 1745–1757, 2008.

[299] Fleck C., Netz R.R., and von Grünberg H.H. Poisson-Boltzmann theory for membranes with mobile charged lipids and the pH-dependent interaction of a DNA molecule with a membrane. *Biophys. J.*, 82, 76–92, 2002.

[300] Khelashvili G., Weinstein H., and Harries D. Protein diffusion on charged membranes: A dynamic mean-field model describes time evolution and lipid reorganization. *Biophys. J.*, 94, 2580–2597, 2008.

[301] Harries D., Ben-Shaul A., and Szleifer I. Enveloping of charged proteins by lipid bilayers. *J. Phys. Chem. B*, 108, 1491–1496, 2004.

[302] Dias R.S. and Linse P. Colloid adsorption onto responsive membranes. *Biophys. J.*, 94, 3760–3768, 2008.

[303] Fleck C.C. and Netz R.R. Electrostatic colloid-membrane binding. *Europhys. Lett.*, 67, 314–320, 2004.

[304] Deserno M. and Bickel T. Wrapping of a spherical colloid by a fluid membrane. *Europhys. Lett.*, 62, 767–773, 2003.

[305] Deserno M. Elastic deformation of a fluid membrane upon colloid binding. *Phys. Rev. E*, 69, 031903, 2004.

[306] Deserno M. When do fluid membranes engulf sticky colloids? *J. Phys.: Condens. Matter*, 16, S2061–S2070, 2004.

[307] Nowak S.A. and Chou T. Membrane lipid segregation in endocytosis. *Phys. Rev. E*, 78, 021908, 2008.

[308] Smith K.A., Jasnow D., and Balazs A.C. Designing synthetic vesicles that engulf nanoscopic particles. *J. Chem. Phys.*, 127, 084703, 2007.

[309] Fošnarič M., Iglič A., Kroll D.M., and May S. Monte Carlo simulations of complex formation between a mixed fluid vesicle and a charged colloid. *J. Chem. Phys.*, 131, 105103, 2009.

[310] Larsen A.E. and Grier D.G. Like-charge attractions in metastable colloidal crystallites. *Nature*, 385, 230–233, 1997.

[311] Squires T.M. and Brenner M.P. Like-charge attraction and hydrodynamic interaction. *Phys. Rev. Lett.*, 85, 4976–4979, 2000.

[312] Cichocki B., Ekiel-Jeżewska M.L., Nägele G., and Wajnryb E. Hydrodynamic interactions between widely-separated particles at a free surface. *Europhys. Lett.*, 67, 383–389, 2004.

[313] Cichocki B., Ekiel-Jeżewska M.L., Nägele G., and Wajnryb E. Motion of spheres along a fluid-gas interface. *J. Chem. Phys.*, 121, 2305–2316, 2004.

[314] Cichocki B., Ekiel-Jeżewska M.L., and Wajnryb E. Hydrodynamic interactions between spheres in a viscous fluid with a flat free surface or hard wall. *J. Chem. Phys.*, 126, 184704, 2007.

[315] Ben-Tal N., Honig B., Peitzsch R.M., Denisov G., and McLaughlin S. Binding of small basic peptides to membranes containing acidic lipids: Theoretical models and experimental results. *Biophys. J.*, 71, 561–575, 1996.

[316] Ben-Tal N., Honig B., Bagdassarian C.K., and Ben-Shaul A. Association entropy in adsorption processes. *Biophys. J.*, 79, 1180–1187, 2000.

[317] Record Jr M.T., Anderson C.F., and Lohman T.M. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: The roles of ion association or release, screening, and ion effects on water activity. *Q. Rev. Biophys.*, 11, 103–178, 1978.

[318] Roth C.M., Sader J.E., and Lenhoff A.M. Electrostatic contribution to the energy and entropy of protein adsorption. *J. Colloid Interface Sci.*, 203, 218–221, 1998.

[319] Wagner K., Harries D., May S., Kahl V., Rädler J.O., and Ben-Shaul A. Direct evidence for counterion release upon cationic lipid-DNA condensation. *Langmuir*, 16, 303–306, 2000.

[320] Chandler D. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437, 640–647, 2005.

[321] Moser M., Legate K.R., Zent R., and Fässler R. The tail of integrins, talin, and kindlins. *Science*, 324, 895–899, 2009.

[322] Chishti A.H., Kim A.C., Marfatia S.M., Lutchman M., Hanspal M., Jindal H., Liu S.-C., Low P.S., Rouleau G.A., Mohandas N., Chasis J.A., Conboy J.G., Gascard P., Takakuwa Y., Huang S.-C., Benz Jr E.J., Bretscher A., Fehon R.G., Gusella J.F.,

Ramesh V., Solomon F., Marchesi V.T., Tsukita S., Tsukita S., Arpin M., Louvard D., Tonks N.K., Anderson J.M., Fanning A.S., Bryant P.J., Woods D.F., and Hoover K.B. The FERM domain: A unique module involved in the linkage of cytoplasmic proteins to the membrane. *Trends Biochem. Sci.*, 23, 281–282, 1998.

[323] Pearson M.A., Reczek D., Bretscher A., and Karplus P.A. Structure of the ERM protein moesin reveals the FERM domain fold masked by an extended actin binding tail domain. *Cell*, 101, 259–270, 2000.

[324] Hamada K., Shimizu T., Matsui T., Tsukita S., Tsukita S., and Hakoshima T. Structural basis of the membrane-targeting and unmasking mechanisms of the radixin FERM domain. *EMBO J.*, 19, 4449–4462, 2000.

[325] Goult B.T., Bouaouina M., Harburger D.S., Bate N., Patel B., Anthis N.J., Campbell I.D., Calderwood D.A., Barsukov I.L., Roberts G.C., and Critchley D.R. The structure of the N-terminus of kindlin-1: A domain important for αIIbβ3 integrin activation. *J. Mol. Biol.*, 394, 944 – 956, 2009.

[326] P.R. Elliott, Goult B.T., Kopp P.M., Bate N., Grossman J.G., Roberts G.C.K., Critchley D.R., and Barsukov I.L. The structure of the talin head reveals a novel extended conformation of the FERM domain. *Structure*, 18, 1289–1299, 2010.

[327] Campbell I.D. The talin FERM domain is not so FERM. *Structure*, 18, 1222–1223, 2010.

[328] Ussar S., Moser M., Widmaier M., Rognoni E., Harrer C., Genzel-Boroviczeny O., and Fässler R. Loss of kindlin-1 causes skin atrophy and lethal neonatal intestinal epithelial dysfunction. *PLoS Genet*, 4, e1000289, 2008.

[329] White S.J. and McLean W.H.I. Kindler surprise: Mutations in a novel actin-associated protein cause Kindler syndrome. *J. Dermatol. Sci.*, 38, 169–175, 2005.

[330] Svensson L., Howarth K., McDowall A., Patzak I., Evans R., Ussar S., Moser M., Metin A., Fried M., Tomlinson I., and Hogg N. Leukocyte adhesion deficiency-III is caused by mutations in *KINDLIN3* affecting integrin activation. *Nat. Med.*, 15, 306–312, 2009.

[331] Malinin N.L., Zhang L., Choi J., Ciocea A., Razorenova O., Ma Y.-Q., Podrez E.A., Tosi M., Lennon D.P., Caplan A.I., Shurin S.B., Plow E.F., and Byzova T.V. A point mutation in *KINDLIN3* ablates activation of three integrin subfamilies in humans. *Nat. Med.*, 15, 313–318, 2009.

[332] Malinin N.L., Plow E.F., and Byzova T.V. Kindlins in FERM adhesion. *Blood*, 115, 4011–4017, 2010.

[333] Qu H., Tu Y., Shi X., Larjava H., Saleem M. A., Shattil S.J., Fukuda K., Qin J., Kretzler M., and Wu C. Kindlin-2 regulates podocyte adhesion and fibronectin matrix deposition through interactions with phosphoinositides and integrins. *J. Cell Sci.*, 124, 879–891, 2011.

[334] Liu J., Fukuda K., Xu Z., Ma Y.-Q., Hirbawi J., Mao X., Wu C., Plow E.F., and Qin J. Structural basis of phosphoinositide binding to kindlin-2 protein pleckstrin homology domain in regulating integrin activation. *J. Biol. Chem.*, 286, 43334–43342, 2011.

[335] Bandyopadhyay A., Rothschild G., Kim S., Calderwood D.A., and Raghavan S. Functional differences between kindlin-1 and kindlin-2 in keratinocytes. *J. Cell. Sci.*, 125, 2172–2184, 2012.

[336] Tsujishita Y., Guo S., Stolz L.E., York J.D., and Hurley J.H. Specificity determinants in phosphoinositide dephosphorylation: Crystal structure of an archetypal inositol polyphosphate 5-phosphatase. *Cell*, 105, 379–389, 2001.

[337] Miller G.J., Wilson M.P., Majerus P.W., and Hurley J.H. Specificity determinants in inositol polyphosphate synthesis: Crystal structure of inositol 1,3,4-trisphosphate 5/6-kinase. *Mol. Cell*, 18, 201–212, 2005.

[338] Begley M.J., Taylor G.S., Brock M.A., Ghosh P., Woods V.L., and Dixon J.E. Molecular basis for substrate recognition by MTMR2, a myotubularin family phosphoinositide phosphatase. *Proc. Natl. Acad. Sci. U.S.A.*, 103, 927–932, 2006.

[339] Milburn C.C., Deak M., Kelly S.M., Price N.C., Alessi D.R., and van Aalten D.M.F. Binding of phosphatidylinositol 3,4,5-trisphosphate to the pleckstrin homology domain of protein kinase B induces a conformational change. *Biochem. J.*, 375, 531–538, 2003.

[340] Gasteiger J. and Marsili M. A new model for calculating atomic charges in molecules. *Tet. Lett.*, 19, 3181–3184, 1978.

[341] Li C., Xu L., Wolan D.W., Wilson I.A., and Olson A.J. Virtual screening of human 5-aminoimidazole-4-carboxamide ribonucleotide transformylase against the NCI diversity set by use of AutoDock to identify novel nonfolate inhibitors. *J. Med. Chem.*, 47, 6681–6690, 2004.

[342] Grosdidier A., Zoete V., and Michielin O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.*, 39, W270–W277, 2011.

[343] Zoete V., Cuendet M.A., Grosdidier A., and Michielin O. SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.*, 32, 2359–2368, 2011.

[344] Durrant J.D, de Oliveira C.A.F., and McCammon J.A. POVME: An algorithm for measuring binding-pocket volumes. *J. Mol. Graph. Model.*, 29, 773–776, 2011.

[345] Sali A. and Blundell T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, 234, 779–815, 1993.

[346] Feig M., Karanicolas J., and Brooks C.L. 3rd. MMTSB Tool Set: Enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph Model.*, 22, 377–395, 2004.

[347] Kumar S. and Nussinov R. Close-range electrostatic interactions in proteins. *ChemBioChem*, 3, 604–617, 2002.

[348] Carpten J.D., Faber A.L., Horn C., Donoho G.P., Briggs S.L., Robbins C.M., Hostetter G., Boguslawski S., Moses T.Y., Savage S., Uhlik M., Lin A., Du J., Qian Y.-W., Zeckner D.J., Tucker-Kellogg G., Touchman J., Patel K., Mousses S., Bittner M., Schevitz R., Lai M.-H. T., Blanchard K.L., and Thomas J.E. A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature*, 448, 439–444, 2007.

[349] Landgraf K.E., Pilling C., and Falke J.J. Molecular mechanism of an oncogenic mutation that alters membrane targeting: Glu17Lys modifies the PIP lipid specificity of the AKT1 PH domain. *Biochemistry*, 47, 12260–12269, 2008.

[350] Pilling C., Landgraf K.E., and Falke J.J. The GRP1 PH domain, like the AKT1 PH domain, possesses a sentry glutamate residue essential for specific targeting to plasma membrane PI(3,4,5)P_3. *Biochemistry*, 50, 9845–9856, 2011.

[351] Warren G.L., Andrews C.W., Capelli A.-M., Clarke B., LaLonde J., Lambert M.H., Lindvall M., Nevins N., Semus S.F., Senger S., Tedesco G., Wall I.D., Woolven J.M., Peishoff C.E., and Head M.S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.*, 49, 5912–5931, 2006.

[352] Gilson M.K. and Zhou H.-X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.*, 36, 21–42, 2007.

[353] Liu Y., Zhu Y., and Ye S. Zhang R. Crystal structure of kindlin-2 PH domain reveals a conformational transition for its membrane anchoring and regulation of integrin activation. *Protein Cell*, 3, 434–440, 2012.

[354] Lemmon M.A., Ferguson K.M., and Abrams C.S. Pleckstrin homology domains and the cytoskeleton. *FEBS Lett.*, 513, 71–76, 2002.

[355] Yates L.A., Lumb C.N., Brahme N.N., Bird L.E., De Colibus L., Owens R.J., Calderwood D.A., Sansom M.S.P., and Gilbert R.J.C. Structural and functional characterisation of the kindlin-1 pleckstrin homology domain. *J. Biol. Chem., in press*, 2012.

[356] Edgar R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, 32, 1792–1797, 2004.

[357] Komander D., Fairservice A., Deak M., Kular G.S., Prescott A.R., Downes C.P., Safrany S.T., Alessi D.R., and van Aalten D.M.F. Structural insights into the regulation of PDK1 by phosphoinositides and inositol phosphates. *EMBO J.*, 23, 3918–3928, 2004.

[358] Hyvonen M., Macias M.J., Nilges M., Oschkinat H., Saraste M., and Wilmanns M. Structure of the binding site for inositol phosphates in a PH domain. *EMBO J.*, 14, 4676–4685, 1995.

[359] Worthylake D.K., Rossman K.L., and Sondek J. Crystal structure of Rac1 in complex with the guanine nucleotide exchange region of Tiam1. *Nature*, 408, 682–688, 2000.

[360] Thomas C.C., Dowler S., Deak M., Alessi D.R., and van Aalten D.M.F. Crystal structure of the phosphatidylinositol 3,4-bisphosphate-binding pleckstrin homology (PH) domain of tandem PH-domain-containing protein 1 (TAPP1): Molecular basis of lipid specificity. *Biochem. J.*, 358, 287–294, 2001.

[361] Jackson S.G., Zhang Y., Bao X., Zhang K., Summerfield R., Haslam R.J., and Junop M.S. Structure of the carboxy-terminal PH domain of pleckstrin at 2.1 Angstroms. *Acta Crystallogr. Sect. D*, 62, 324–330, 2006.

[362] Ceccarelli D.F.J., Blasutig I.M., Goudreault M., Li Z., Ruston J., Pawson T., and Sicheri F. Non-canonical interaction of phosphoinositides with pleckstrin homology domains of Tiam1 and ArhGAP9. *J. Biol. Chem.*, 282, 13864–13874, 2007.

[363] Waterhouse A.M., Procter J.B., Martin D.M.A., Clamp M., and Barton G.J. Jalview version 2: A multiple sequence alignment and analysis workbench. *Bioinformatics*, 25, 1189–1191, 2009.

[364] Clamp M., Cuff J., Searle S.M., and Barton G.J. The Jalview Java alignment editor. *Bioinformatics*, 20, 426–427, 2004.

[365] Felsenstein J. Phylogeny Inference Package (Version 3.2). *Cladistics*, 5, 164–166, 1989.

[366] Felsenstein J. An alternating least squares approach to inferring phylogenies from pairwise distances. *Syst. Biol.*, 46, 101–111, 1997.

[367] Anand K., Maeda K., and Gavin A.-C. Structural analyses of the Slm1-PH domain demonstrate ligand binding in the non-canonical site. *PLoS ONE*, 7, e36526, 2012.

[368] Ali I.U., Schriml L.M., and Dean M. Mutational spectra of PTEN/MMAC1 gene: A tumor suppressor with lipid phosphatase activity. *J. Natl. Cancer Inst.*, 91, 1922–1932, 1999.

[369] Hanahan D. and Weinberg R.A. The hallmarks of cancer. *Cell*, 100, 57–70, 2000.

[370] Hanahan D. and Weinberg R.A. Hallmarks of cancer: The next generation. *Cell*, 144, 646–674, 2011.

[371] Lee J.-O., Yang H., Georgescu M.-M., Di Cristofano A., Maehama T., Shi Y., Dixon J.E., Pandolfi P., and Pavletich N.P. Crystal structure of the PTEN tumor suppressor: Implications for its phosphoinositide phosphatase activity and membrane association. *Cell*, 99, 323–334, 1999.

[372] Steck P.A., Pershouse M.A., Jasser S.A., Yung W.K.A., Lin H., Ligon A.H., Langford L.A., Baumgard M.L., Hattier T., Davis T., Frye C., Hu R., Swedlund B., Teng D.H.R., and Tavtigian S.V. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nat. Genet.*, 15, 356–362, 1997.

[373] Li J., Yen C., Liaw D., Podsypanina K., Bose S., Wang S.I., Puc J., Miliaresis C., Rodgers L., McCombie R., Bigner S.H., Giovanella B.C., Ittmann M., Tycko B., Hibshoosh H., Wigler M.H., and Parsons R. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science*, 275, 1943–1947, 1997.

[374] Eng C. *PTEN*: One gene, many syndromes. *Hum. Mutat.*, 22, 183–198, 2003.

[375] Maehama T. and Dixon J.E. The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *J. Biol. Chem.*, 273, 13375–13378, 1998.

[376] Myers M.P., Pass I., Batty I.H., J. Van der Kaay, Stolarov J.P., Hemmings B.A., Wigler M.H., Downes C.P., and Tonks N.K. The lipid phosphatase activity of PTEN is critical for its tumor suppressor function. *Proc. Natl. Acad. Sci. U.S.A.*, 95, 13513–13518, 1998.

[377] Iida S., Ono A., Sayama K., Hamaguchi T., Fujii H., Nakajima H., Namba M., Hanafusa T., Matsuzawa Y., and Moriwaki K. Accelerated decline of blood glucose after intravenous glucose injection in a patient with Cowden disease having a

heterozygous germline mutation of the PTEN/MMAC1 gene. *Anticancer Res.*, 20, 1901–1904, 2000.

[378] Stiles B., Wang Y., Stahl A., Bassilian S., Lee W.P., Kim Y.J., Sherwin R., Devaskar S., Lesche R., Magnuson M.A., and Wu H. Liver-specific deletion of negative regulator Pten results in fatty liver and insulin hypersensitivity. *Proc. Natl. Acad. Sci. U.S.A.*, 101, 2082–2087, 2004.

[379] Stiles B.L., Kuralwalla-Martinez C., Guo W., Gregorian C., Wang Y., Tian J., Magnuson M.A., and Wu H. Selective deletion of Pten in pancreatic beta cells leads to increased islet mass and resistance to STZ-induced diabetes. *Mol. Cell Biol.*, 26, 2772–2781, 2006.

[380] Tanaka M. and Grossman H.B. *In vivo* gene therapy of human bladder cancer with PTEN suppresses tumor growth, downregulates phosphorylated Akt, and increases sensitivity to doxorubicin. *Gene Ther.*, 10, 1636–1642, 2003.

[381] Tanaka M., Rosser C.J., and Grossman H.B. PTEN gene therapy induces growth inhibition and increases efficacy of chemotherapy in prostate cancer. *Cancer Detect. Prev.*, 29, 170–174, 2005.

[382] Anai S., Goodison S., Shiverick K., Iczkowski K., Tanaka M., and Rosser C.J. Combination of *PTEN* gene therapy and radiation inhibits the growth of human prostate cancer xenografts. *Hum. Gene Ther.*, 17, 975–984, 2006.

[383] Saito Y., Swanson X., Mhashilkar A.M., Oida1 Y., Schrock R., Branch C.D., Chada S., Zumstein L., and Ramesh R. Adenovirus-mediated transfer of the PTEN gene inhibits human colorectal cancer growth *in vitro* and *in vivo*. *Gene Ther.*, 10, 1961–1969, 2003.

[384] Courtney K.D., Corcoran R.B., and Engelman J.A. The PI3K pathway as a drug target in human cancer. *J. Clin. Oncol.*, 28, 1075–1083, 2010.

[385] Ihle N.T., Williams R., Chow S., Chew W., Berggren M.I., Paine-Murrieta G., Minion D.J., Halter R.J., Wipf P., Abraham R., Kirkpatrick L., and Powis G. Molecular pharmacology and antitumor activity of PX-866, a novel inhibitor of phosphoinositide-3-kinase signaling. *Mol. Cancer Ther.*, 3, 763–772, 2004.

[386] Kong D. and Yamori T. Phosphatidylinositol 3-kinase inhibitors: Promising drug candidates for cancer therapy. *Cancer Sci.*, 99, 1734–1740, 2008.

[387] Willems L., Tamburini J., Chapuis N., Lacombe C., Mayeux P., and Bouscary D. PI3K and mTOR signaling pathways in cancer: New data on targeted therapies. *Curr. Oncol. Rep.*, 14, 129–138, 2012.

[388] Hers I., Vincent E.E., and Tavaré J.M. Akt signalling in health and disease. *Cell. Signal.*, 23, 1515–1527, 2011.

[389] Lindsley C.W., Zhao Z., Leister W.H., Robinson R.G., Barnett S.F., Defeo-Jones D., Jones R.E., Hartman G.D., Huff J.R., Huber H.E., and Duggan M.E. Allosteric Akt (PKB) inhibitors: Discovery and SAR of isozyme selective inhibitors. *Bioorg. Med. Chem. Lett.*, 15, 761–764, 2005.

[390] Barnett S.F., Defeo-Jones D., Fu S., Hancock P.J., Haskell K.M., Jones R.E., Kahana J.A., Kral A.M., Leander K., Lee L.L., Malinowski J., McAvoy E.M., Nahas D.D., Robinson R.G., and Huber H.E. Identification and characterization of pleckstrin-homology-domain-dependent and isoenzyme-specific Akt inhibitors. *Biochem. J.*, 385, 399–408, 2005.

[391] Gills J.J. and Dennis P.A. Perifosine: Update on a novel Akt inhibitor. *Curr. Oncol. Rep.*, 11, 102–110, 2009.

[392] Garrett C.R., Coppola D., Wenham R.M., Cubitt C.L., Neuger A.M., Frost T.J., Lush R.M., Sullivan D.M., Cheng J.Q., and Sebti S.M. Phase I pharmacokinetic and pharmacodynamic study of triciribine phosphate monohydrate, a small-molecule

inhibitor of AKT phosphorylation, in adult subjects with solid tumors containing activated AKT. *Invest. New Drug.*, 29, 1381–1389, 2011.

[393] Ahad A.M., Zuohe S., Du-Cuny L., Moses S.A., Zhou L.L., Zhang S., Powis G., Meuillet E.J., and Mash E.A. Development of sulfonamide AKT PH domain inhibitors. *Bioorg. Med. Chem.*, 19, 2046–2054, 2011.

[394] Sutton R.B., Ernst J.A., and Brunger A.T. Crystal structure of the cytosolic C2A-C2B domains of synaptotagmin III: Implications for ca^{2+}-independent SNARE complex interaction. *J. Cell Biol.*, 147, 589–598, 1999.

[395] Wang Q., Wei Y., Mottamal M. Roberts M.F., and Krilov G. Understanding the stereospecific interactions of 3-deoxyphosphatidylinositol derivatives with the PTEN phosphatase domain. *J. Mol. Graph. Model.*, 29, 102–114, 2010.

[396] Das S., Dixon J.E., and Cho W. Membrane-binding and activation mechanism of PTEN. *Proc. Natl. Acad. Sci. U.S.A.*, 100, 7491–7496, 2003.

[397] Rahdar M., Inoue T., Meyer T., Zhang J., Vazquez F., and Devreotes P.N. A phosphorylation-dependent intramolecular interaction regulates the membrane association and activity of the tumor suppressor PTEN. *Proc. Natl. Acad. Sci. U.S.A.*, 106, 480–485, 2009.

[398] Georgescu M.-M., Kirsch K.H., Kaloudis P., Yang H., Pavletich N.P., and Hanafusa H. Stabilization and productive positioning roles of the C2 domain of PTEN tumor suppressor. *Cancer Res.*, 60, 7033–7038, 2000.

[399] Stenson P.D., Mort M., Ball E.V., Howells K., Phillips A.D., Thomas N.S.T., and Cooper D.N. The Human Gene Mutation Database: 2008 update. *Genome Med.*, 1, 13, 2009.

[400] Vega A., Torres J., Torres M., Cameselle-Teijeiro J., Macia M., Carracedo A., and Pulido R. A novel loss-of-function mutation (N48K) in the *PTEN* gene in a Spanish patient with Cowden disease. *J. Invest. Dermatol.*, 121, 1356–1359, 2003.

[401] Staal F.J.T, van der Luijt R.B., Baert M.R.M., van Drunen J., van Bakel H., Peters E., de Valk I., van Amstel H.K.P., Taphoorn M.J.B., Jansen G.H., van Veelen C.W.M., Burgering B., and Staal G.E.J. A novel germline mutation of PTEN associated with brain tumours of multiple lineages. *Br. J. Cancer*, 86, 1586–1591, 2002.

[402] Sawada T., Hamano N., Satoh H., Okada T., Takeda Y., and Mabuchi H. Mutation analysis of the PTEN/MMAC1 gene in Japanese patients with Cowden disease. *Cancer Sci.*, 91, 700–705, 2000.

[403] Liaw D., Marsh D.J., Li J., Dahia P.L.M., Wang S.I., Zheng Z., Bose S., Call K.M., Tsou H.C., Peacoke M., Eng C., and Parsons R. Germline mutations of the *PTEN* gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nat. Genet.*, 16, 64–67, 1997.

[404] Shenoy S., Shekhar P., Heinrich F., Daou M.-C., Gericke A., Ross A.H., and Lösche M. Membrane association of the PTEN tumor suppressor: Molecular details of the protein-membrane complex from SPR binding studies and neutron reflection. *PLoS ONE*, 7, e32591, 2012.

[405] Vazquez F. and Sellers W.R. The PTEN tumor suppressor protein: An antagonist of phosphoinositide 3-kinase signaling. *Biochim. Biophys. Acta*, 1470, M21–M35, 2000.

[406] Guan R., Han D., Harrison S.C., and Kirchhausen T. Structure of the PTEN-like region auxilin, a detector of clathrin-coated vesicle budding. *Structure*, 18, 1191–1198, 2010.

[407] Matsuda M., Takeshita K., Kurokawa T., Sakata S., Suzuki M., Yamashita E., Okamura Y., and Nakagawa A. Crystal structure of the cytoplasmic phosphatase and tensin homolog (PTEN)-like region of *Ciona intestinalis* voltage-sensing phosphatase provides insight into substrate specificity and redox regulation of the phosphoinositide activity. *J. Biol. Chem.*, 286, 23368–23377, 2011.

[408] Liu L., Kohout S.C., Xu Q., Müller S., Kimberlin C.R., Isacoff E.Y., and Minor Jr D.L. A glutamate switch controls voltage-sensitive phosphatase function. *Nat. Struct. Mol. Biol.*, 19, 633–641, 2012.

[409] Stansfeld P.J., Hopkinson R., Ashcroft F.M., and Sansom M.S.P. PIP_2-binding site in Kir channels: Definition by multiscale biomolecular simulation. *Biochemistry*, 48, 10926–10933, 2009.

[410] Hansen S.B., Tao X., and MacKinnon R. Structural basis of PIP_2 activation of the classical inward rectifier K^+ channel Kir2.2. *Nature*, 477, 495–498, 2011.

[411] Shan Y., Kim E.T., Eastwood M.P., Dror R.O., Seeliger M.A., and Shaw D.E. How does a drug molecule finds its target binding site? *J. Am. Chem. Soc.*, 133, 9181–9183, 2011.

[412] López-Canut V., Roca M., Bertrán J., Moliner V., and Tũnón I. Promiscuity in alkaline phosphatase superfamily. Unraveling evolution through molecular simulations. *J. Am. Chem. Soc.*, 133, 12050–12062, 2011.

[413] Hou G. and Cui Q. QM/MM analysis suggests that alkaline phosphatase (AP) and nucleotide pyrophosphatase/phosphodiesterase slightly tighten the transition state for phosphate diester hydrolysis relative to solution: Implication for catalytic promiscuity in the AP superfamily. *J. Am. Chem. Soc.*, 134, 229–246, 2011.

[414] Šali D., Bycroft M., and Fersht A.R. Stabilization of protein structure by interaction of α-helix dipole with a charged side chain. *Nature*, 335, 740–743, 1988.

[415] Ripoll D.R., Vila J.A., and Scheraga H.A. On the orientation of the backbone dipoles in native folds. *Proc. Natl. Acad. Sci. U.S.A.*, 102, 7559–7564, 2005.

[416] Tartaglia G.G., Cavalli A., Pellarin R., and Caflisch A. The role of aromaticity, exposed surface, and dipole moment in determining protein aggregation rates. *Protein Sci.*, 13, 1939–1941, 2004.

[417] Rottman A., Gilboa H., Schechter Y., and Silver B.L. The determination of intracellular sodium concentration in human red blood cells: Nuclear magnetic resonance measurements. *Anal. Biochem.*, 201, 48–51, 1992.

[418] Lo C.-J., Leake M.C., and Berry R.M. Fluorescence measurement of intracellular sodium concentration in single *Escherichia coli* cells. *Biophys. J.*, 90, 357–365, 2006.

[419] Bandara S., Schlöder J.P., Eils R., Bock H.G., and Meyer T. Optimal experimental design for parameter estimation of a cell signaling model. *PLoS Comput. Biol.*, 5, e1000558, 2009.

[420] Karr J.R., Sanghvi J.C., Macklin D.N., Gutschow M.V., Jacobs J.M., Bolival Jr B., Assad-Garcia N., Glass J.I., and Covert M.W. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150, 389–401, 2012.

[421] Lynch E.D., Ostermeyer E.D., Lee M.K., Arena J.F., Ji H., Dann J., Swisshelm K., Suchard D., MacLeod P.M., Kvinnsland S., Gjertsen B.T., Heimdal K., Lubs H., Møller P., and King M.-C. Inherited mutations in PTEN that are associated with breast cancer, Cowden disease, and juvenile polyposis. *Am. J. Hum. Genet.*, 61, 1254–1260, 1997.

[422] Çelebi J.T., Chen F.F., Zhang H., Ping X.L., Tsou H.C., and Peacocke M. Identification of *PTEN* mutations in five families with Bannayan-Zonana syndrome. *Exp. Dermatol.*, 8, 134–139, 1999.

[423] Marsh D.J., Kum J.B., Lunetta K.L., Bennett M.J., Gorlin R.J., Ahmed S.F., Bodurtha J., Crowe C., Curtis M.A., Dasouki M., Dunn T., Feit H., Geraghty M.T, Graham J.M., Hodgson S.V., Hunter A., Korf B.R., Manchester D., Miesfeldt S., Murday V.A., Nathanson K.L., Parisi M., Pober B., Romano C., Tolmie J.L., Trembath R., Winter R.M., Zackai E.H., Zori R.T., Weng L.-P., Dahia P.L.M., and Eng C. *PTEN* mutation spectrum and genotype-phenotype correlations in Bannayan-Riley-Ruvalcaba syndrome suggest a single entity with Cowden syndrome. *Hum. Mol. Genet.*, 8, 1461–1472, 1999.

[424] Zhou X.-P., Hampel H., Thiele H., Gorlin R.J., Hennekam R.C.M., Parisi M., Winter R.M., and Eng C. Association of germline mutation in the *PTEN* tumour suppressor gene and Proteus and Proteus-like syndromes. *Lancet*, 358, 210–211, 2001.

[425] Olschwang S., Serova-Sinilnikova O.M., and Lenoir G.M. PTEN germ-line mutations in juvenile polyposis coli. *Nat. Genet.*, 18, 12–14, 1998.

[426] Çelebi J.T., Ping X.L., Zhang H., Remington T., Sulica V.I., Tsou H.C., and Peacocke M. Germline *PTEN* mutations in three families with Cowden syndrome. *Exp. Dermatol.*, 9, 152–156, 2000.

[427] Tate G., Suzuki T., Endo Y., and Mitsuya T. A novel mutation of the *PTEN* gene in a Japanese patient with Cowden syndrome and bilateral breast cancer. *Cancer Genet. Cytogenet.*, 184, 67–71, 2008.

[428] Varga E.A., Pastore M., Prior T., Herman G.E., and McBride K.L. The prevalence of *PTEN* mutations in a clinical pediatric cohort with autism spectrum disorders, developmental delay, and macrocephaly. *Genet. Med.*, 11, 111–117, 2009.

[429] Thiffault I., Schwartz C.E., Der Kaloustian V., and Foulkes W.D. Mutation analysis of the tumor suppressor *PTEN* and the glypican 3 (*GPC3*) gene in patients diagnosed with Proteus syndrome. *Am. J. Med. Genet. A*, 130A(2), 123–127, 2004.

[430] Reardon W., Zhou X.-P., and Eng C. A novel germline mutation of the *PTEN* gene in a patient with macrocephaly, ventricular dilatation, and features of VATER association. *J. Med. Genet.*, 38, 820–823, 2001.

[431] Marsh D.J., Coulon V., Lunetta K.L., Rocca-Serra P., Dahia P.L.M., Zheng Z., Liaw D., Caron S., Duboué B., Lin A.Y., Richardson A.-L., Bonnetblanc J.-M., Bressieux J.-M., Cabarrot-Moreau A., Chompret A., Demange L., Eeles R.A., Yahanda A.M., Fearon E.R., Fricker J.-P., Gorlin R.J., Hodgson S.V., Huson S., Lacombe D., LePrat F., Odent S., Toulouse C., Olopade O.I., Sobol H., Tishler S., Woods C.G., Robinson B.G., Weber H.C., Parsons R., Peacocke M., Longy M., and Eng C. Mutation spectrum and genotype-phenotype analyses in Cowden disease and Bannayan-Zonana syndrome, two hamartoma syndromes with germline *PTEN* mutation. *Hum. Mol. Genet.*, 7, 507–515, 1998.

[432] Loffeld A., McLellan N.J., Cole T., Payne S.J., Fricker D., and Moss C. Epidermal naevus in Proteus syndrome showing loss of heterozygosity for an inherited *PTEN* mutation. *Br. J. Dermatol.*, 154, 1194–1198, 2006.

[433] Figer A., Kaplan A., Frydman M., Lev D., Paswell J., Papa M.Z., Goldman B., and Friedman E. Germline mutations in the *PTEN* gene in Israeli patients with Bannayan-Riley-Ruvalcaba syndrome and women with familial breast cancer. *Clin. Genet.*, 62, 298–302, 2002.

[434] Zhou X.-P., Marsh D.J., Morrison C.D., Chaudhury A.R., Maxwell M., Reifenberger G., and Eng C. Germline inactivation of *PTEN* and dysregulation of the phosphoinositol-3-kinase/Akt pathway cause human Lhermitte-Duclos disease in adults. *Am. J. Hum. Genet.*, 73(5), 1191–1198, 2003.

[435] Butler M.G., Dasouki M.J., Zhou X.-P., Talebizadeh Z., Brown M., Takahashi T.N., Miles J.H., Wang C.H., Stratton R., Pilarski R., and Eng C. Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with

germline *PTEN* tumour suppressor gene mutations. *J. Med. Genet.*, 42, 318–321, 2005.

[436] Kohno T., Takahashi M., Fukutomi T., Ushio K., and Yokota J. Germline mutations of the *PTEN/MMAC1* gene in Japanese patients with Cowden disease. *Cancer Sci.*, 89, 471–474, 1998.

[437] Bussaglia E., Pujol R.M., Gil M.J., Marti R.M., Tuneu A., Febrer M.I., Garcia-Patos V., Ruiz E.M., Barnadas M., Alegre M., Serrano S., and Matias-Guiu X. PTEN mutations in eight Spanish families and one Brazilian family with Cowden syndrome. *J. Invest. Dermatol.*, 118, 639–644, 2002.

[438] Nelen M.R., Kremer H., Konings I.B.M., Schoute F., van Essen A.J., Koch R., Woods C.G., Fryns J.-P., Hamel B., Hoefsloot L.H., Peeters E.A.J., and Padberg G.W. Novel *PTEN* mutations in patients with Cowden disease: Absence of clear genotype-phenotype correlations. *Eur. J. Hum. Genet.*, 7, 267–273, 1999.

[439] Tan W.-H., Baris H.N., Burrows P.E., Robson C.D., Alomari A.I., Mulliken J.B., Fishman S.J., and Irons M.B. The spectrum of vascular anomalies in patients with *PTEN* mutations: implications for diagnosis and management. *J. Med. Genet.*, 44, 594–602, 2007.

[440] Tsou H.C., Ping X.L., Xie X.X., Gruener A.C., Zhang H., Nini R., Swisshelm K., Sybert V., Diamond T.M., Sutphen R., and Peacocke M. The genetic basis of Cowden's syndrome: Three novel mutations in *PTEN/MMAC1/TEP1*. *Hum. Genet.*, 102, 467–473, 1998.

[441] Orrico A., Galli L., Buoni S., Orsi A., Vonella G., and Sorrentino V. Novel *PTEN* mutations in neurodevelopmental disorders and macrocephaly. *Clin. Genet.*, 75, 195–198, 2009.

[442] De Vivo I., Gertig D.M., Nagase S., Hankinson S.E., O'Brien R., Speizer F.E., Parsons R., and Hunter D.J. Novel germline mutations in the *PTEN* tumour suppressor gene found in women with multiple cancers. *J. Med. Genet.*, 37, 336–341, 2000.

[443] Nelen M.R., van Staveren W.C.G., Peeters E.A.J., Ben Hassel M., Gorlin R.J., Hamm H., Lindboe C.F., Fryns J.-P., Sijmons R.H., Woods D.G., Mariman E.C.M., Padberg G.W., and Kremer H. Germline mutations in the *PTEN/MMAC1* gene in patients with Cowden disease. *Hum. Mol. Genet.*, 6, 1383–1387, 1997.

[444] Sawada T., Okada T., Miwa K., Satoh H., Asano A., and Mabuchi H. Two novel mutations of *PTEN* gene in Japanese patients with Cowden syndrome. *Am. J. Med. Gen. A*, 128A, 12–14, 2004.

[445] Kurose K., Araki T., Matsunaka T., Takada Y., and Emi M. Variant manifestation of Cowden disease in Japan: Hamartomatous polyposis of the digestive tract with mutation of the *PTEN* gene. *Am. J. Hum. Genet.*, 64, 308–310, 1999.

[446] Derrey S., Proust F., Debono B., Langlois O., Layet A., Layet V., Longy M., Fréger P., and Laquerrière A. Association between Cowden syndrome and Lhermitte-Duclos disease: Report of two cases and review of the literature. *Surg. Neurol.*, 61, 447–454, 2004.

[447] Tekin M., Hişmi B.Ö., Fitoz S., Yalçinkaya F., Ekim M., Kansu A., Ertem M., Deda G., Tutar E., Arsan S., Zhou X.-P., Pilarski R., Eng C., and Akar N. A germline *PTEN* mutation with manifestations of prenatal onset and verrucous epidermal nevus. *Am. J. Med. Genet. A*, 140A, 1472–1475, 2006.

[448] Boccone L., Dessì V., Zappu A., Piga S., Piludu M.B., Rais M., Massidda C., De Virgiliis S., Cao A., and Loudianos G. Bannayan-Riley-Ruvalcaba syndrome with reactive nodular lymphoid hyperplasia and autism and a PTEN mutation. *Am. J. Med. Genet. A*, 140A, 1965–1969, 2006.

[449] Kubo Y., Urano Y., Hida Y., Ikeuchi T., Nomoto M., Kunitomo K., and Arase S. A novel *PTEN* mutation in a Japanese patient with Cowden disease. *Br. J. Dermatol.*, 142, 1100–1105, 2000.

[450] Scala S., Bruni P., Lo Muzio L., Mignogna M., Viglietto G., and Fusco A. Novel mutation of the PTEN gene in an Italian Cowden's disease kindred. *Int. J. Oncol.*, 13, 665–668, 1998.

[451] Raizis A.M., Ferguson M.M., Robinson B.A., Atkinson C.H., and George P.M. Identification of a novel PTEN mutation (L139X) in a patient with Cowden disease and Sjögren's syndrome. *Mol. Pathol.*, 51, 339–341, 1998.

[452] Gicquel J.-J., Vabres P., Bonneau D., Mercié M., Handiri L., and Dighiero P. Retinal angioma in a patient with Cowden disease. *Am. J. Ophthalmol.*, 135, 400–402, 2003.

[453] Marsh D.J., Dahia P.L.M., Zheng Z., Liaw D., Parsons R., Gorlin R.J., and Eng C. Germline mutations in *PTEN* are present in Bannayan-Zonana syndrome. *Nat. Genet.*, 16, 333–334, 1997.

[454] Davies M.P.A., Gibbs F.E.M., Halliwell N., Joyce K.A., Roebuck M.M., Rossi M.L., Salisbury J., Sibson D.R., Tacconi L., and Walker C. Mutation in the *PTEN/MMAC1* gene in archival low grade and high grade gliomas. *Br. J. Cancer*, 79, 1542–1548, 1999.

[455] Lachlan K.L., Lucassen A.M., Bunyan D., and Temple I.K. Cowden syndrome and Bannayan-Riley-Ruvalcaba syndrome represent one condition with variable expression and age-related penetrance: Results of a clinical study of *PTEN* mutation carriers. *J. Med. Gen.*, 44, 579–585, 2007.

[456] Goffin A., Hoefsloot L.H., Bosgoed E., Swillen A., and Fryns J.-P. PTEN mutation in a family with Cowden syndrome and autism. *Am. J. Med. Genet.*, 105, 521–524, 2001.

[457] Longy M., Coulon. V., Duboué B., David A., Larrègue M., Eng C., Amati P., Kraimps J.L., Bottani A., Lacombe D., and Bonneau D. Mutations of PTEN in patients with Bannayan-Riley-Ruvalcaba phenotype. *J. Med. Genet.*, 35, 886–889, 1998.

[458] Kim D.-K., Myung S.-J., Yang S.-K., Hong S.S., Kim K.J., Byeon J.-S., Lee G.H., Kim J.-H., Min Y.I., Lee S.M., Jeong J.-Y., Song K., and Jung S.-A. Analysis of *PTEN* gene mutations in Korean patients with Cowden syndrome and polyposis syndrome. *Dis. Colon Rectum*, 48, 1714–1722, 2005.

[459] Cho M.-Y., Kim H.S., Eng C., Kim D.S., Kang S.J., Eom M., Yi S.Y., and Bronner M.P. First report of ovarian dysgerminoma in Cowden syndrome with germline PTEN mutation and PTEN-related 10q loss of tumor heterozygosity. *Am. J. Surg. Pathol.*, 32, 1258–1264, 2008.

[460] Chi S.G., Kim H.J., Park B.J., Min H.J., Park J.H., Kim Y.W., Dong S.H., Kim B.H., Lee J.I., Chang Y.W., Chang R., Kim W.K., and Yang M.H. Mutational abrogation of the *PTEN/MMAC1* gene in gastrointestinal polyps in patients with Cowden disease. *Gastroenterology*, 115, 1084–1089, 1998.

[461] Buxbaum J.D., Cai G., Chaste P., Nygren G., Goldsmith J., Reichert J., Anckarsäter H., Rastam M., Smith C.J., Silverman J.M., Hollander E., Leboyer M., Gillberg C., Verloes A., and Betancur C. Mutation screening of the PTEN gene in patients with autism spectrum disorders and macrocephaly. *Am. J. Med. Genet. B*, 144B, 484–491, 2007.

Glossary

AChE	Acetylcholinesterase
Akt	Protein kinase B
AMBER	Assisted Model Building with Energy Refinement
APBS	Adaptive Poisson-Boltzmann Solver
Arf	ADP ribosylation factor
ARNO	Arf nucleotide binding site opener
B3LYP	Becke three parameter Lee-Yang-Parr
BAD	Bcl-2 associated death promoter
Bcl-2	B-cell lymphoma 2
BD	Brownian dynamics
C-TT	C-terminal tail
CADD	Computer aided drug design
CBR	Calcium binding region
CFP	Cyan fluorescent protein

CG	Coarse grained
CHARMM	Chemistry at Harvard Molecular Mechanics
CHO	Chinese hamster ovary
Ci-VSP	*Ciona intestinalis* voltage sensing phosphatase
CW EPR	Continuous wave electron paramagnetic resonance
DFT	Density functional theory
DPC	Dodecylphosphatidylcholine
DPPC	1,2-dipalmitoylphosphatidylcholine
ECM	Effective charge method
ENTH	Epsin N-terminal homology
EPR	Electron paramagnetic resonance
FA	Focal adhesion
FERM	4.1, Ezrin, Radixin, Moesin
FKHR	Forkhead in human rhabdomyosarcoma
FYVE	Fab1, YOTB, Vac1, EEA1
GDP	Guanosine 5′-diphosphate
GFP	Green fluorescent protein
GLA	γ-carboxyglutamic acid rich
GPCR	G protein coupled receptor
Grb2	Growth factor receptor bound protein 2
GROMACS	Groningen Machine for Chemical Simulations
GROMOS	Groningen Molecular Simulation
GRP1	General receptor for phosphoinositides 1
GSK3	Glycogen synthase kinase 3
GTP	Guanosine 5′-triphosphate

HEK293	Human embryonic kidney
HF	Hartree-Fock
HGMD	Human Gene Mutation Database
HSQC	Heteronuclear single quantum coherence
Kir	Inwardly rectifying potassium
LAD3	Leukocyte adhesion deficiency syndrome III
LINCS	Linear constraint solver
LJ	Lennard-Jones
LMP2	Local Møller-Plesset second order perturbation theory
MD	Molecular dynamics
MDM2	Murine double minute oncoprotein
MeTaDoR	Membrane Targeting Domains Resource
MMAC1	Mutated in multiple advanced cancers
MSD	Mean square displacement
mTOR	Mammalian target of rapamycin
N-PBM	N-terminal PI(4,5)P_2 binding module
NAM	Northrup-Allison-McCammon
NFκB	Nuclear factor κ B
NMR	Nuclear magnetic resonance
OPLS-AA/L	Optimised Potentials for Liquid Simulations - All Atom/LMP2
OPM	Orientation of Proteins in Membranes
PB	Poisson-Boltzmann
PDB	Protein Data Bank
PDK	Phosphoinositide dependent kinase
PE	Phosphatidylethanolamine

PG	Phosphatidylglycerol
PH	Pleckstrin homology
PHYLIP	Phylogeny Inference Package
PI3-K	Phosphatidylinositol 3-kinase
PI	Phosphatidylinositol
PIP	Phosphatidylinositol phosphate
PI(3)P	Phosphatidylinositol 3-phosphate
PI(3,4,5)P_3	Phosphatidylinositol (3,4,5)-trisphosphate
PI(4)P	Phosphatidylinositol 4-phosphate
PI(3,4)P_2	Phosphatidylinositol (3,4)-bisphosphate
PI(3,5)P_2	Phosphatidylinositol (3,5)-bisphosphate
PI(4,5)P_2	Phosphatidylinositol (4,5)-bisphosphate
PI(5)P	Phosphatidylinositol 5-phosphate
PKB	Protein kinase B
PKC	Protein kinase C
PLCδ1	Phospholipase C-δ1
PM	Plasma membrane
PME	Particle mesh Ewald
PMF	Potential of mean force
POPC	1-palmitoyl-2-oleoylphosphatidylcholine
POPS	1-palmitoyl-2-oleoylphosphatidylserine
PPI	Protein protein interaction
PTEN	Phosphatase and tensin homologue deleted on chromosome ten
PX	Phox homology
QM/MM	Quantum mechanical/molecular mechanical

RAC1	Ras-related C3 botulinum toxin substrate
Raf	Rapidly accelerated fibrosarcoma
Ras	Rat sarcoma
RDF	Radial distribution function
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RPY	Rotne-Prager-Yamakawa
RTK	Receptor tyrosine kinase
S6K	Ribosomal protein S6 kinase
SDA	Simulation of Diffusional Association
SDSL	Site directed spin labelling
SGK	Serum/glucocorticoid regulated kinase
SOS	Son of sevenless
SPC	Simple point charge
SPR	Surface plasmon resonance
ssNMR	Solid state NMR
STED	Super resolution stimulated emission depletion microscopy
TIP4P	Four point transferable intermolecular potential
TIRFM	Total internal reflection fluorescence microscopy
UHBD	University of Houston Brownian Dynamics
vdW	van der Waals
YFP	Yellow fluorescent protein