Interactions of peripheral membrane proteins with phosphatidylinositol lipids: insights from molecular dynamics simulations



Fiona Naughton

Lincoln College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2017

Abstract

Interactions of peripheral membrane proteins with phosphatidylinositol lipids: insights from molecular dynamics simulations

Fiona Naughton

Lincoln College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Michaelmas 2017

Interactions between proteins and membranes are central to many signalling pathways and other cellular processes. Phosphatidylinositol phosphates (PIPs) are a family of lipids often acting as second messengers and targeted by peripheral proteins in these processes. A pipeline was developed combining the molecular dynamics (MD) approaches of umbrella sampling and coarse-grain modelling, and used to quantify and compare the interactions with PIP-containing model membranes of 13 pleckstrin homology (PH) domains, a common lipid-binding domain found in many proteins showing varied affinities and specificities for PIPs. Lipid selectivity generally agreed with previous observations. Several membrane-binding modes were identified, revealing PIP interactions through a secondary site are more common than suggested experimentally and appear to be related to overall affinity. Results suggest that simultaneous binding of multiple PIP lipids is required to achieve the high affinities characteristic of PH domains. Multiscale MD, combining coarse-grain binding simulations and atomistic refinement, was used to investigate PTEN, a tumour suppressor catalysing interconversion of PIPs and associated with many cancers and other disorders. Regions often ignored in previous studies were revealed to favour productive binding, largely via electrostatics. PIP clustering by bound PTEN and membrane insertion in the productive mode were demonstrated. Existence of an N-terminal PIP-binding site was supported, with this region appearing disordered, rather than helical as previously suggested. Changes in interdomain orientation when bound and with the clinically-relevant R173C mutation further suggest the importance of the interdomain interface for productive binding. Together, this work demonstrates the important contributions MD can make towards understanding protein/membrane interactions, particularly in the context of managing the diseases caused by their disruption.

Acknowledgements

First thanks go to my supervisor, Mark Sansom, for giving me this opportunity and for always being ready to help, even when I spring last-minute deadlines on him. Thanks also to my co-supervisor Antreas Kalli for all his help and guidance, particularly back when I was still learning the ropes.

SBCBers past, present and honorary alike: to Laura and Tom, for many wonderful PhTea Breaks; to Tom and Tonya, for (among other things) all those chicken nuggets; to Maria for 'not leaving the lab before me' and helping out through these final stressful stages; and to everyone else - thank you for making the lab such a great place to work, and for putting up with my (slight) obsession with cats.

On the other side of the world: to my lovely cats Mica and Tara, for all the distance de-stressing. You've probably already forgotten me since I left Australia, but I love you both anyway.

To James, for so many hours of great discussions. Thank you for listening to all my rants, and for always giving your best input - even when I disagree with your advice.

And finally: my parents, to whom this thesis is dedicated. You might not know quite exactly what it is I do, but you're always willing to support me in every way you can - even when I disagree with your advice, too. Thank you both for everything.

Fiona Naughton

November 2017, University of Oxford

List of publications arising from this thesis

Naughton, F. B., Kalli, A. C. & Sansom, M. S. Modes of Interaction of Pleckstrin Homology Domains with Membranes: Toward a Computational Biochemistry of Membrane Recognition. *Journal of Molecular Biology* **430,** 372–388 (2018)

Naughton, F. B., Kalli, A. C. & Sansom, M. S. P. Association of Peripheral Membrane Proteins with Membranes: Free Energy of Binding of GRP1 PH Domain with Phosphatidylinositol Phosphate-Containing Model Bilayers. *The Journal of Physical Chemistry Letters* **7,** 1219–1224 (2016)

Other publications

Ni, T., Kalli, A. C., **Naughton, F. B.**, Yates, L. A., Naneh, O., Kozorog, M., Anderluh, G., Sansom, M. S. & Gilbert, R. J. Structure and lipid-binding properties of the kindlin-3 pleckstrin homology domain. *Biochemical Journal* **474,** 539–556 (2017)

Code

Code written as part of this thesis has been included as the 'auxiliary' module in the open-source software package MDAnalysis, available online from *https://www.mdanalysis.org*.

Contents

List of figures **v**

List of tables **vii**

List of acronyms and abbreviations **viii**

If proteins were cats: a graphical representation of this thesis **1**

Chapter 1 - Introduction **2**

1.1 **The role of protein/membrane interactions** . . . 2

1.1.1 Structure and composition of the membrane . . . 2

1.1.2 Proteins and the membrane . . . 3

1.1.3 Targeting membranes: phosphatidylinositol phosphate lipids . . . 5

1.1.3.1 *PIP-binding and the PH domain* . . . 5

1.1.3.2 *PIP distribution and PTEN* . . . 9

1.2 **Studying protein-PIP interactions** . . . 11

1.2.1 Experimental methods . . . 12

1.2.2 Computational methods: the case for MD . . . 13

1.3 **Work presented in this thesis** . . . 14

Chapter 2 - Theory and methods **16**

2.1 **Molecular Dynamics simulations** . . . 16

2.1.1 Force fields . . . 16

2.1.1.1 *Atomistic force fields* . . . 18

2.1.1.2 *Coarse grain force fields* . . . 19

2.1.2 Running MD simulations . . . 20

2.2 **Calculation of binding free energies** . . . 24

2.2.1 Binding free energy from experiment . . . 25

2.2.2 Binding free energy from simulation . . . 25

2.2.2.1 *Umbrella sampling and WHAM* . . . 26

2.2.2.2 *Potential of mean force profiles* . . . 28

2.2.2.3 *Extension to other coordinates* . . . 29

2.2.2.4 *Umbrella Sampling in practice* . . . 30

Chapter 3 - Structures and energetics of peripheral protein/membrane interactions through simulation: developing a US-CG approach with GRP1-PH **31**

3.1 **Background** . . . 31
3.1.1 Why combine Umbrella Sampling and Coarse Grain modelling? . . . 31
3.1.2 Why GRP1-PH as an initial case? . . . 32
3.2 **Methods** . . . 33
3.2.1 Details of the protein-membrane system . . . 34
3.2.2 Simulation details . . . 38
3.2.3 Analysis . . . 39
3.3 **Developing a protocol for CG/US simulations of protein-membrane interactions** . . . 42
3.3.1 Determining suitable values for US parameters . . . 42
3.3.1.1 *Window spacing, restraint and number: optimising coverage of the reaction coordinate* . . . 42
3.3.1.2 *Simulation time and xy protein restraint: ensuring adequate sampling of orthogonal space* . . . 44
3.3.1.3 *The steered MD simulation: efficient generation of windows* 46
3.3.2 Combining structural and energetic analysis . . . 47
3.3.2.1 *A Python wrapper for the Grossfield WHAM* . . . 48
3.3.2.2 *Free energy simulations and MDAnalysis* . . . 48
3.4 **Results and discussion: membrane binding of the GRP1-PH domain** 49
3.4.1 Checking the initial structure . . . 49
3.4.2 Results from US/CG . . . 52
3.4.2.1 *Energetic analysis: reproduction of PIP_3 selectivity and negative effect of C-site mutations* . . . 54
3.4.2.2 *Structural analysis: observation of both A- and C-site binding modes* . . . 55
3.4.2.3 *Determinants of binding mode and affinity* . . . 59
3.4.2.4 *Binding of multiple PIPs?* . . . 60
3.5 **Further considerations of the US/CG method** . . . 64
3.5.1 Additional tests . . . 64
3.5.1.1 *Charge of the PIP lipids* . . . 65
3.5.1.2 *Polarisable water* . . . 66
3.5.1.3 *Removal of the lipid restraint* . . . 67
3.5.2 Caveats of the US/CG method . . . 68
3.6 **Conclusions** . . . 68

Chapter 4 - Interactions of members of the PH domain family with membranes containing PIP lipids **70**

4.1 **Background** . . . 70
4.1.1 Unresolved questions surrounding the PH domain family . . . 70
4.1.1.1 *What determines selectivity and affinity?* . . . 70
4.1.1.2 *PIP binding site: canonical or alternate?* . . . 74
4.1.2 US/CG and the PH family . . . 76
4.2 **Methods** . . . 76
4.2.1 Details of the protein-membrane system . . . 76
4.2.2 Simulation details . . . 80
4.2.3 Analysis . . . 80
4.3 **Results and discussion** . . . 82
4.3.1 Sensitivity to initial structure . . . 82
4.3.2 Energetics: selectivity and affinity . . . 85
4.3.2.1 *PMF profiles* . . . 85
4.3.2.2 *Comparison with experiment* . . . 87
4.3.3 Structures: membrane binding modes . . . 89
4.3.3.1 *The canonical vs. atypical site* . . . 89
4.3.3.2 *Modes of interaction on the membrane* . . . 93
4.3.4 From sequence to structure to function: determining PIP binding of PH domains . . . 98
4.3.4.1 *What determines binding site?* . . . 98
4.3.4.2 *What determines binding mode?* . . . 102
4.3.4.3 *What determines selectivity/affinity?* . . . 104
4.3.5 Additional PH/PIP systems . . . 105
4.3.5.1 *The E17K-AKT1 mutant* . . . 105
4.3.5.2 *An additional test for initial system configuration bias: C-site ArhGAP9* . . . 106
4.4 **Further comments of dual binding sites: role of the A site?** . . . 107
4.5 **Caveats of the US/CG method: an updated discussion** . . . 108
4.6 **Conclusions** . . . 109

Chapter 5 - Interactions of PTEN with membranes containing PIP_3 lipids **111**

5.1 **Background** . . . 111
5.2 **Methods** . . . 113
5.2.1 Structural details . . . 113
5.2.2 Simulation details . . . 115

5.2.3 Analysis 116
5.3 **Results and discussion** 118
5.3.1 Modelling full-length PTEN 118
5.3.2 Binding modes and the binding process: CG binding simulations . 119
5.3.2.1 *Binding landscapes* 122
5.3.2.2 *PIP interactions* 124
5.3.2.3 *Conformation of the helix, loop and tail regions* 127
5.3.3 The productive binding mode: atomistic refinement 127
5.3.3.1 *Changes to the binding mode: penetration and PIP contacts* 127
5.3.3.2 *The helix, loop and tail regions* 129
5.3.4 PTEN mutations and the role of the PD/C2 interface 131
5.3.4.1 *Relative domain orientation* 131
5.3.4.2 *The R173C mutation* 135
5.4 **Further discussion and conclusions** 136

Chapter 6 - Conclusions **138**
6.1 **Summary** 138
6.2 **Future directions** 139

Appendix A: Overview of selected code written for this thesis **142**

Appendix B: Clinically relevant PTEN missense mutations **145**

References **163**

List of figures

1.1 Components of a typical membrane 3
1.2 Structures of common phospholipids 4
1.3 Interconversion of PIPs 6
1.4 Structures of PIP-binding domains/proteins 7
1.5 Overview of the AKT1 signalling pathway 10

2.1 Bonded and nonbonded interactions in forcefields 17
2.2 Coarse-grain and atomistic representations of lipids and water 19

3.1 Structure of GRP1-PH 33
3.2 Overview of the US/CG pipeline 39
3.3 Assessing coverage of the reaction coordinate 43
3.4 Assessing orthogonal sampling 45
3.5 Testing of steered MD simulations 46
3.6 Initial structure of GRP1-PH bound to the membrane 50
3.7 Demonstration of convergence and sufficient sampling 51
3.8 Structural and energetic analysis of GRP1-PH/membrane binding . . 52
3.9 GRP1-PH/PIP interaction modes 56
3.10 A-site interaction of GRP1-PH 58
3.11 Multiple PIP binding by GRP1-PH 61
3.12 Convergence and sampling of additional US simulations 63
3.13 Additional profiles to check the US/CG method 65

4.1 Structures of PH domains bound to PIP analogues 75
4.2 Overview of the US/CG simulation pipeline 79
4.3 Convergence analysis for PH/PIP US simulations 83

4.4 Distribution of initial configurations 84
4.5 PMF profiles of PH domains interaction with PIP-containing membranes 85
4.6 Comparison of profile well depths 87
4.7 Comparison of with experiment 88
4.8 Energy landscapes of binding angle vs. separation 90
4.9 Site of PIP interactions 91
4.10 Sequence alignment and PIP-contacting residues 93
4.11 Common membrane binding modes. 98
4.12 Binding mode landscapes for each PH/PIP system 100
4.13 Representative structures of PIP-binding modes for each PH domain 102
4.14 Determining binding site from sequence 103
4.15 Relation between binding site and affinity 105
4.16 Effect of the AKT1-PH E17K mutation 106
4.17 Effect of initial PIP binding site for ArhGAP9 107

5.1 Structure of PTEN 112
5.2 Secondary structure prediction of missing PTEN regions 119
5.3 Dynamics of the helix, loop and tail regions 120
5.4 Electrostatic surface profile of PTEN 121
5.5 Modes of interaction of PTEN with the membrane 123
5.6 Clustering of PIP_3 around PTEN 125
5.7 PIP_3-PTEN interactions 126
5.8 Atomistic refinement of the productive binding pose 128
5.9 Membrane penetration and orientation of the productive binding mode 129
5.10 Clinical missense mutations of PTEN 132
5.11 Relative orientation of the PD and C2 domains 133
5.12 Coarse-grain investigations of the R173C mutant 134

List of tables

1.1 Roles of selected PH domains 8

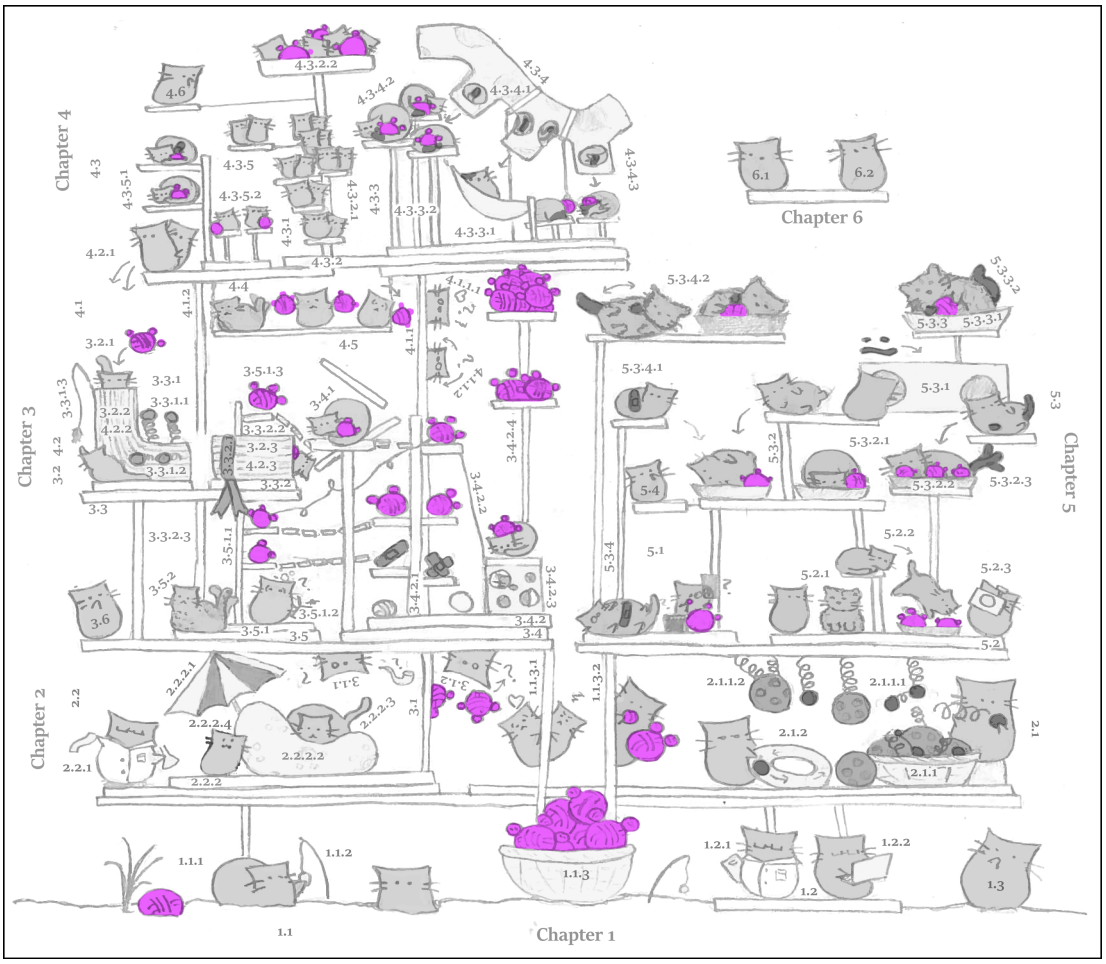
3.1 Summary of structural studies of GRP1-PH/PIP interactions 34
3.2 Summary of biophysical/biochemical studies of GRP1-PH/PIP interactions 35
3.3 Summary of simulations performed for Chapter 3 37

4.1 Structural studies identifying PIP-binding site of PH domains 71
4.2 Summary of biophysical studies PH/PIP interactions 72
4.3 Summary of US simulations performed in Chapter 4 77
4.4 PIP contacts in the A and C binding sites 95
4.5 Favourable PIP binding sites in the tightly and loosely bound regions 97

5.1 Summary of simulations performed for Chapter 5 114

Acronyms and abbreviations

CG	coarse-grained
COM	center of mass
IP_4	inositol (1,3,4,5) phosphate
MD	molecular dynamics
PC	phosphatidylcholine
PD	phosphatase domain
PDB	protein data bank
PH	pleckstrin homology
PI(3,4)P_2	phosphatidylinositol-(3,4)-bisphosphate
PI(4,5)P_2, PIP_2	phosphatidylinositol-(4,5)-bisphosphate
PIP	phosphatidylinositol phosphate
PI(3,4,5)P_3, PIP_3	phosphatidylinositol-(3,4,5)-triphosphate
PMF	potential of mean force
POPC	1-palmitoyl-2-oleoyl-sn-glycerol-3-phosphatidylcholine
POPS	1-palmitoyl-2-oleoyl-sn-glycerol-3-phosphatidylserine
PS	phosphatidylserine
PTEN	phosphatase and tensin homologue
US	Umbrella Sampling
VDW	van der Waals
WHAM	weighted histogram analyis method



If proteins were cats: a graphical representation of this thesis

Chapter 1

Introduction

1.1 - The role of protein/membrane interactions

1.1.1 - *Structure and composition of the membrane*

Biological membranes are bilayers comprised of lipids, into which are inserted membrane glycoproteins. They surround the cell and its organelles, primarily serving the vital function of compartmentalisation. These barriers prevent the free diffusion of ions and other molecules between different regions, helping to ensure maintenance of a regulated environment within the cell and individual organelles. In certain situations, it is necessary for a membrane to be bypassed, allowing, for example, the uptake of nutrients and the transmission of signals between the cell and its environment. To facilitate the range of different transmembrane processes, membranes possess a complex composition that is often asymmetrical between the two layers (Figure 1.1).

The key component of each layer, or 'leaflet', of a bilayer membrane are phospholipids, which align such that the acyl tails form a hydrophobic core to the bilayer while the polar or charged headgroups form the external surface. Common phospholipids include phosphatidylcholine (PC) and phosphatidylethanolamine (PE), both zwitterionic at neutral pH, and phosphatidylserine (PS), phosphatidic acid (PA) and the phosphatidylinositides (PI and PIPs), all negatively charged (Figure 1.2) [1]. Several

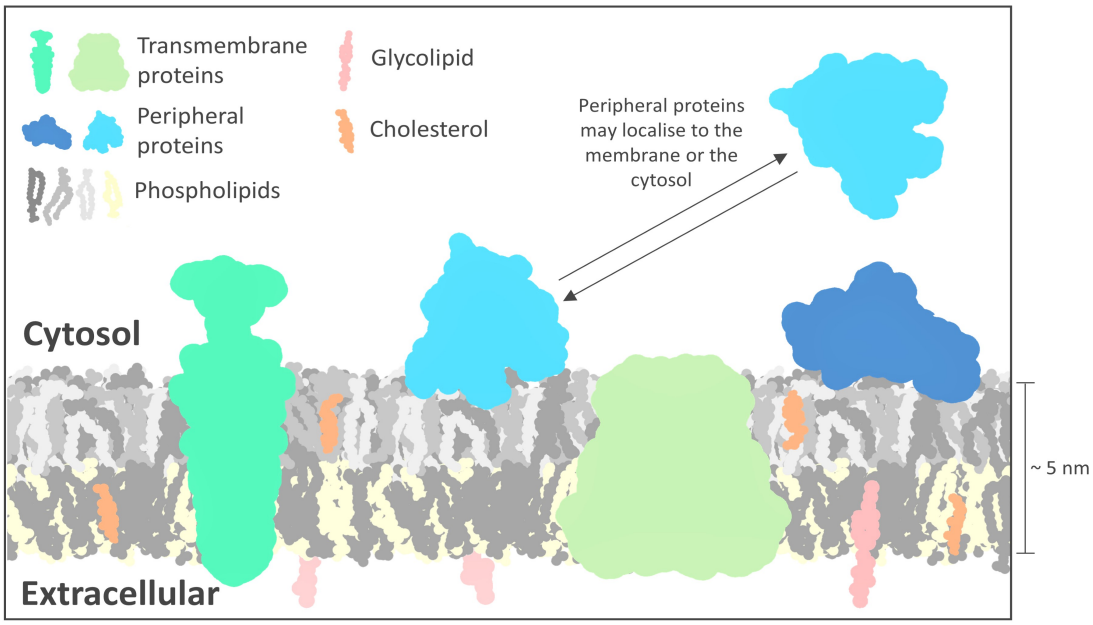


Figure 1.1: Components of a typical membrane. Each leaflet may contain a number of different lipid species (indicated here by different colours) which need not be the same, leading to an asymmetric lipid distribution. Lipids may be modified by the addition of carbohydrate, forming glycolipids. Some integral membrane proteins insert only partially to the membrane, rather than being transmembrane (not shown). While peripheral proteins often have a soluble and a membrane-bound form, in a given set of conditions it may be that only one form is seen.

different version of each exist, depending on the form of the fatty acid tails. These lipids, along with other membrane components such as proteins, sphingolipids and sterols (e.g. cholesterol), are found in different compositions in different membranes; the plasma membrane, for example, contains largely PC, sphingolipids and cholesterol in the outer leaflet, and PE, PS, PIs and cholesterol in the inner leaflet [2].

1.1.2 - *Proteins and the membrane*

Proteins that interact with the membrane may be embedded within the bilayer (integral) or associated with the surface (peripheral). Transmembrane integral proteins may allow molecules to pass through the bilayer, for example as an ion channel or a transporter, or transmit signals across the membrane through structural changes (e.g. Receptor Tyrosine Kinases (RKTs) [3]). Peripheral proteins may exist both bound

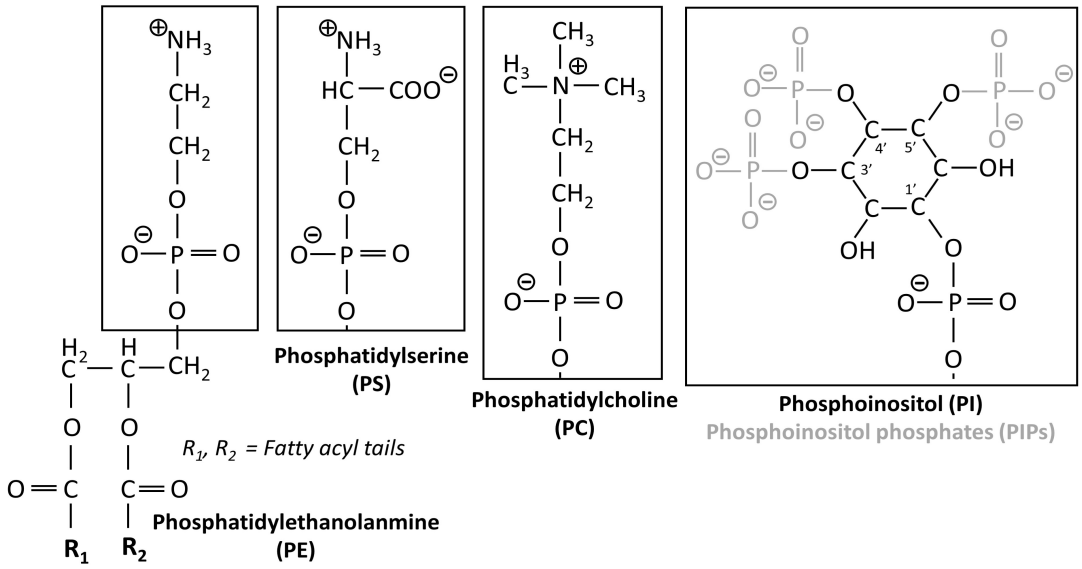


Figure 1.2: Structures of common phospholipids. Only the headgroups are shown for PS, PC and PI/PIP. In each case, the two fatty acid tails may be any of several types. The potential phosphorylation of PI at the 3', 4' and/or 5' positions of the inositol ring is shown in grey.

to the membrane and free in the cytosol (Figure 1.1), providing a convenient 'on/off' switch that makes these proteins well suited to roles in signal transduction [4]. It is peripheral proteins that form the focus of this thesis.

In addition to the propagation of signals, protein interactions at the membrane may also serve a structural role, for example enabling interactions with the membrane-proximal cytoskeleton. Further, the membrane surface provides an ideal platform for protein interactions in general, providing a scaffold that may promote productive associations by reducing degrees of freedom, both translational (from 3D to 2D) and rotational (by favouring certain orientations on the membrane).

Protein-membrane interactions, and particularly peripheral protein-membrane interactions, are thus vital to a large number of cellular processes. These interactions are mediated by 'lipid binding domains'. The ability to target membranes on both a spatial and temporal scale is essential to ensure protein function is carried out correctly, especially in the context of signalling. Many lipid-binding domains therefore recognise specific (controllable) features of membranes, allowing targeted recruitment

[5]. These may involve larger-scale characteristics such as membrane curvature or general electrostatic profile, or binding of specific lipids [5]. Often, two or more requirements are combined, allowing more control over the targeting process (known as 'coincidence detection' [6]).

To this end, many membrane lipids act as second messengers. After being generated in response to an external signal, they may then attract downstream effector proteins, triggering the rest of the signalling cascade [5]. The phosphatidylinositol phosphate (PIP) family are notable examples of this.

1.1.3 - *Targeting membranes: phosphatidylinositol phosphate lipids*

PIPs are a family of anionic lipids, sharing a common tail and inositol headgroup structure. The inositol headgroup may be variably phosphorylated; *in vivo*, phosphorylation of the 3', 4' and 5' groups is observed, leading to three monophosphates (PI(3)P, PI(4)P and PI(5)P), three bisphosphates (PI(3,4)P_2, PI(3,5)P_2 and PI(4,5)P_2) and one triphosphate (PI(3,4,5)P_3), or seven different PIP species in total (Figure 1.3) (see reviews, see e.g. [7–10]). These lipids possess a larger, more negatively charged headgroup compared to other common phospholipids (Figure 1.2) and so may be more recognisable to proteins. They are well suited for recognition by lipid binding domains in the targeting of peripheral proteins.

1.1.3.1 - *PIP-binding and the PH domain*

Recognition of PIPs may be achieved through a number of PIP-associating domains, such as the PH, C2, FYVE, PX, ENTH, ANTH, BAR and FERM families [7, 9]. These domains show differing ranges of selectivities and affinities for the PIP family. The FYVE domain [11] (Figure 1.4 A), for example, is an ∼80 residue zinc-finger domain usually observed to preferentially bind PI(3)P, with roles in membrane trafficking. The C2 domain [12], on the other hand, is distinguished by a lack of clear 'binding site' and so is often observed to indiscriminately bind anionic lipids in a calcium-dependant manner. This ∼130 residue domain has a beta-barrel structure (Figure 1.4 B), with three loops responsible for coordination of calcium and membrane binding, though not all C2 domains require Ca^{2+} to bind. Additional interactions in

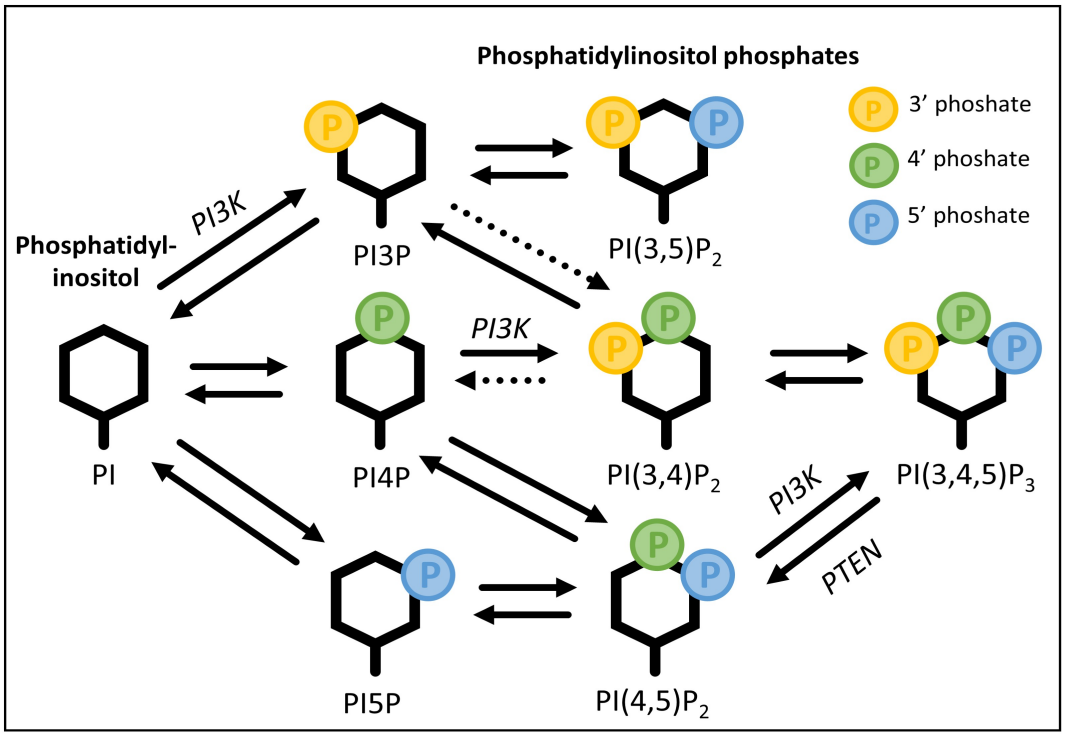


Figure 1.3: Interconversion of PIPs. Representation of the various conversions between PIP lipids that occur with in the cell. Steps carried out by PTEN and PI3K are labelled. Adapted from [9].

a nearby 'cationic groove' may provide some degree of PIP selectivity for some C2 domains [12]. Many C2 domains also appear to assist membrane anchoring through partial insertion into the membrane.

The PH domain

The most common lipid binding domain found in the human genome is the pleckstrin homology (PH) domain [9, 10, 13]. This highly structurally conserved domain (∼120 residues) possesses an open seven-stranded beta-barrel structure followed by a C-terminal α-helix (Figure 1.4 C) [13]. The canonical (C) binding site is found in between the β1-β2, β3-β4 and β6-β7 loops, with a second 'atypical' or 'alternate' (A) binging site located adjacent, between the β1-β2 and β5-β6 loops [13]. Both sites allow a high degree of contacts with the PIP headgroup, and are flanked by loops showing a high degree of sequence variation. This leads to a wide range of selectivities and affinities [13].

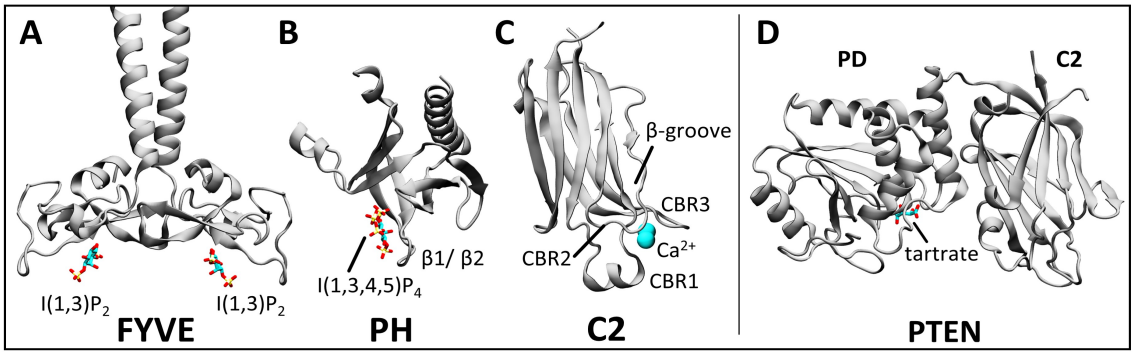


Figure 1.4: Structures of PIP-binding domains/proteins. Representative structures of **(A)** FYVE (PDB ID 1JOC), **(B)** PH (PDB ID 1UNQ) and **(C)** C2 (PDB ID 1RWL) domains. The FYVE and PH domain structures show the positions of bound PIP analogues. The C2 structure shows the location of bound calcium ions in blue, with the three calcium-binding region loops labelled. The 'β-groove' region is also identified. **(D)** Structure of PTEN (PDB ID 1D5R), which also contains a C2 domain. A tartrate molecules, bound at the active site, is also shown.

Originally identified in pleckstrin [14], which contains two PH domains, this domain has since been identified in over 200 other human proteins [15]. This covers a diverse range of roles (see examples in Table 1.1). Within these proteins, the PH domain may serve only as membrane anchor (e.g. β-spectrin), may direct membrane binding as part of a signalling cascade (e.g. PLCδ1), or may through its membrane binding lead to conformational changes that activate the protein (e.g. AKT1). Mutations within the PH domain have been linked with disease, including for a cancer-causing mutation in AKT1 [16] and X-linked agammaglobulinemia mutations in BTK [17].

It is noted that while many PH domains do show strong PIP interactions, many do not appear to associate with PIP-containing membranes [18]. This may be due to the requirement for additional binding partners to allow favourable binding ('coincidence detection'), or may indicate more diverse roles of the PH domain [18]. A $KX_n(K/R)XR$ motif in the β1/2 loop has been proposed to predict PIP binding [19, 20], but does identify interactions in the alternate site, and many questions remain regarding the determinants of affinity and selectivity between PH domains.

Table 1.1: Roles of selected PH domains. The function of the 12 proteins whose PH domains will be investigated in Chapter 4, if known, and the likely contribution of the PH domain to carrying out this role.

Protein	Function	Role of PH domain	Ref.
ArhGAP9	GAP activity at inner plasma membrane; involved in adhesion	May recruit/aid binding to plasma membrane	[21]
AKT1	Kinase activity as part many signalling pathways including apoptosis, proliferation, migration, glucose metabolism	Recruit to the inner plasma membrane in response to upstream signalling (changing PIP concentration) PIP binding causes conformational change that enables activation Localise to membrane, alongside activator and targets	[22]
ARNO	GEF activity at inner plasma membrane; involved in cytoskeletal remodelling, adhesion, migration	Assist in recruiting protein to membrane PIP binding activates enzyme	[23]
ASAP1	GAP activity at inner plasma membrane; involved in cytoskeletal remodelling	May assist in anchoring protein to membrane PIP binding increases enzyme activity	[23]
β-spectrin	Structural protein (cytoskeleton)	Assists in anchoring to membrane	[24]
BTK	Kinase activity as part of B-lymphocyte development and signalling pathways	Recruit to the inner plasma membrane in response to upstream signalling (changing PIP concentration) Localise to membrane, alongside activator	[22]
DAPP1	Adaptor protein assists the assembly of protein complexes; involved in adhesion	Recruit to the inner plasma membrane in response to upstream signalling (changing PIP concentration) Localise to membrane, alongside targets	[22]
KIN2	Activation of integrins; involved in cell adhesion	Assist in recruiting to plasma membrane	[25]
PDK1	Kinase activity in signalling pathways including cell survival, proliferation, adhesion, glucose metabolism	Recruit to the inner plasma membrane in response to upstream signalling (changing PIP concentration) Localise to membrane, alongside targets	[26]
PEPP1	*Unknown*		
PLCδ1	Phospholipase activity, producing second messengers IP_3 and DAG	Recruit to membrane where substrate is located	[27]
Pleckstrin	Actin organisation	Localise to membrane, where it is activated	[28]

1.1.3.2 - *PIP distribution and PTEN*

Different PIP species show unique distributions throughout membranes in the cell [7, 8], allowing targeting to different organelles via appropriate lipid binding domains. PI(4)P, for example, is found in the Golgi, while PI(4,5)P_2 is predominantly located at the inner leaflet of the plasma membrane, as is PI(3,4,5)P_3, though at much lower concentrations [7, 8]. Furthermore, PIP species may be readily interconverted by the addition or removal of phosphates. An extensive network of lipid phosphatases and kinases exists to catalyse these conversions, forming part of many signalling transduction pathways (Figure 1.3) [7–9].

PIP_3 in particularly plays an important role in the AKT1/PKB signalling pathway (Figure 1.5) [29]. PIP_3 is usually present in low quantities at the inner leaflet of the plasma membrane, but upon binding of a growth factor to protein receptors in the membrane and subsequent activation of phosphatidylinositol 3-kinase (PI3K), this kinase catalyses conversion of the more abundant PI(4,5)P_2 to PIP_3, greatly raising PIP_3 levels. AKT1 possesses a PIP_3-selective PH domain (see Table 1.1) and so is recruited to the plasma membrane, where it is then activated and may begin to propagate the signal, eventually leading to promotion of proliferation and cell survival [29]. Many interaction partners of AKT1 also possess PH (or other) PIP_3-binding domains, such as PDK1 which contributes to activation of AKT1, and are similarly recruited to the membrane, aiding this progression. Figure 1.5 shows an overview of this pathway.

PTEN

Understandably, dysregulation of PIP-regulated pathways such as the AKT1/PKB signalling pathway is often associated with disease [30]. Of particular note is the 3' phosphatase PTEN (phosphatase and tensin homologue), which catalyses the conversion of PIP_3 to PI(4,5)P_2 (reviewed recently in [31, 32]). Loss-of-function of PTEN thus leads to a build-up of PIP_3 at the plasma membrane and excessive signalling through the cell-growth promoting ATK1/PKB pathway, leading to tumour formation. PTEN was initially identified as a tumour suppressor [33, 34] and has since been observed to be one of the most frequently mutated genes in cancers [31].

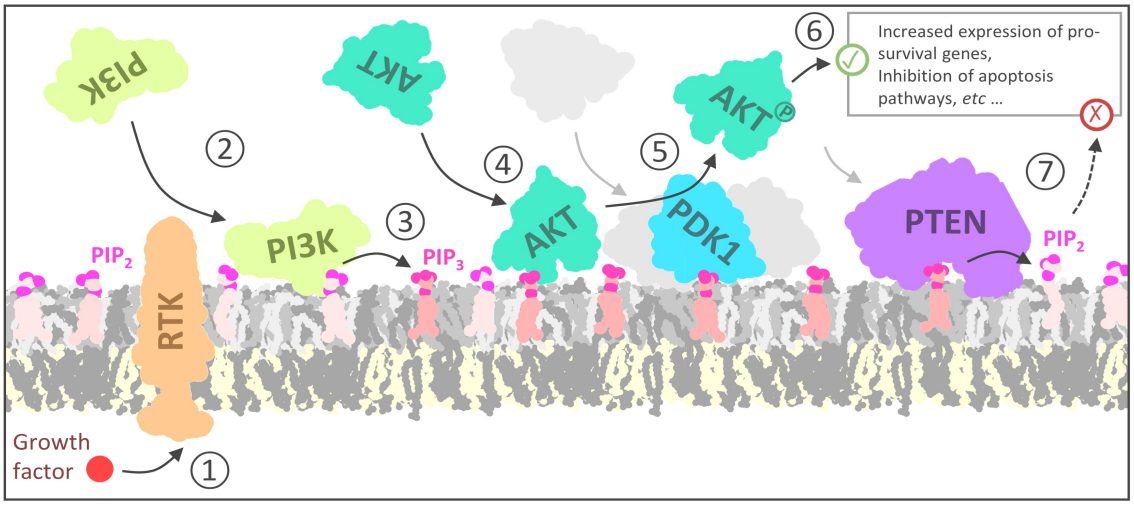


Figure 1.5: Overview of the AKT1 signalling pathway. *1.* Outside the cell, a growth factor binds and stimulates dimerisation of a receptor tyrosine kinase (RTK). *2.* Inside the cell, PI3K is recruited to and activated by the RKT, whereupon it *3.* catalyses conversion of the existing pool of $PI(4,5)P_2$ lipids to PIP_3. *4.* The increased level of PIP_3 recruits AKT1 to the membrane through its PH domain (along with several other domains, such as PDK1). *5.* Through interactions with the membrane and other proteins present there (particularly PDK1, which phosphorylates AKT1), AKT1 is activated and released from the membrane. *6.* Activated AKT1 now travels elsewhere in the cell, phosphorylating multiple targets, ultimately leading to suppression of apoptosis and promotion of cell survival. *7.* Meanwhile, PTEN is also recruited to the membrane by PIP_3, and begins to catalyse conversion of PIP_3 back to $PI(4,5)P_2$. Eventually, the initial lipid composition is regained, and AKT1-propagated signalling halts.

Representative structures of relevant proteins were obtained from the PDB: RTK transmembrane and juxtamembrane regions (2M20), PI3K (1XO5), ATK1 PH domain (1UNQ), PDK1 PH domain (1W1D), PTEN core region (1D5R).

Germline mutations are associated with several syndromes characterised by increased hamartoma formation [35]. PTEN mutations have also been associated with Autism Spectrum Disorder [36]. Significantly, even a slight downregulation of PTEN can have a large impact on tumorigenesis [37].

While the primary function of PTEN does appear to be its lipid phosphatase activity, a number of roles have been identified for PTEN that do not require membrane binding, though all ultimately contribute to its role as a tumour suppressor [31, 32]. In some cases, these roles appear to be linked with the protein phosphatase activity of

PTEN; while in others, including localisation to the nucleus and effecting chromatin remodelling and centrosome stability, they appear to be phosphatase independent [31, 32]. A secreted variant of PTEN is also observed, containing an N-terminal extension including a secretion sequence [38].

PTEN contains a Ca^{2+} independent [39] PIP-binding C2 domain and a phosphatase domain (PD) containing the HCXXGXXR phosphatase catalytic motif, which also contributes to and is required for membrane binding [40] (Figure 1.4 D). The large area of interaction between the two domains has led to suggestion that the PD/C2 interface is vital for optimal orientation of the active site on the membrane [39]. The membrane binding interface had been well mapped out across these two domains by several studies (e.g. [39–46]. It has been proposed that PTEN binds the membrane as a dimer, with each subunit interacting through this interface [47, 48].

Additional domains at the N-terminal and C-terminal are suggested to provide a $PI(4,5)P_2$ interaction site [49] and to regulate binding when phosphorylated [32], respectively. These regions, along with a loop in the C2 region, are missing from the crystal structure [39] and so are often not included in structural and experimental studies, limiting knowledge of their function. While a number of studies have looked at the role of C-terminal tail phosphorylation, these provide conflicting results about the exact nature of this regulation (e.g. [40, 41, 46, 47, 50–52]).

1.2 - Studying protein-PIP interactions

As outlined above, protein-membrane interactions, such as those involving peripheral proteins an PIP lipids, play vital roles in many cellular processes and may be associated with disease when disrupted. Gaining a thorough understanding of the details of these interactions, in both health and disease, is important particularly when considering the potential management of these diseases.

Understanding an interaction between biomolecules may be considered to have two key components: the structural ('*how* do they interact?') and the energetic ('how *strongly* do they interact?'). Many tools are available for structural and energetic investigations both experimentally and computationally, with varied success in application to a peripheral protein/membrane systems [53, 54]. These are outlined below.

1.2.1 - *Experimental methods*

Determining structural details

High-resolution structural data is frequently obtained experimentally using NMR or X-ray crystallography. However, these are complicated in the context of a membrane environment [53]. For example, most crystal structures do not provide details of interactions with lipid molecules. At best, soluble analogues of PIP lipids are used. The resulting structures do reveal information about the likely PIP binding site and interactions, but lack the broader context of the membrane, and further show only static snapshots of the interaction. NMR also faces a limitation in regards to protein size, though observation of NMR shifts may reveal residues that interact with PIP analogues. Cryo-EM [55] may provide an alternative structural determination method, but is not yet widely used for membrane-binding structures. Structural membrane-binding data may be pieced together from other studies, such as mutational analysis to identify likely membrane-binding residues, or neutron reflectivity to obtain density profiles of the bound protein along the membrane normal; if a structural model is available, it may then be fit to this density profile to identify the likely binding membrane-bound orientation.

Determining energetic details

The presence of a membrane environment can also complicate energetic investigations. The simplest biophysical investigations of protein/membrane interactions assess the *relative* binding of proteins to PIPs, though, for example, centrifugation to separate bound and unbound fraction [56], and protein-lipid overlays or 'lipid blots' [57]. While (without modification to the method) these return only qualitative data, lipid blots in particular provide a simple way by which to quickly assess PIP selectivity of a protein. Quantified data, in the form of binding constants K_D, may be obtained using isothermal titration calorimetry (ITC) [58], though this again often utilises soluble analogues of PIP and so may not reflect the *in vivo* binding. Quantified binding data in a membrane environment may instead be obtained through surface plasmon resonance (SPR) [59] and fluorescence resonance energy transfer (FRET). However, the increased complexity of these methods means the simpler soluble or qualitative-based methods are often preferably used.

1.2.2 - *Computational methods: the case for MD*

Molecular dynamics (MD) simulations allow the dynamics of biological systems to be investigated on atomistic or near atomistic scales. The theory and practice of MD simulations will be outlined in Section 2.2. The computational expense involved currently limits the typical simulation below ∼100 nm length scales and µs timescales, and simulations usually rely on experimental structural data as a starting point. Simulation timescales may be improved through use of, for example, a coarse-grained (CG) model [60], in which groups of atoms (typically ∼4 heavy atoms) are replaced by a single representative bead, providing ∼3-fold improvement in accessible timescale [61]. Multiscale simulations may be performed to obtain the advantages of increased timescales of CG simulations and increased resolution of atomistic simulations.

MD simulations provide dynamic structural data at resolutions difficult to obtain experimentally and have widely been used to study membrane interactions [42]. Both coarse-grained and atomistic MD simulations have previously been used to investigate the membrane binding behaviour of many proteins, including PTEN [44, 62, 63] and PH [64–66] and C2 [67–69] domains, with results consistent with experimental observations.

Energetic analysis may be carried out within MD simulations through use of one of a number of free energy calculation methods. Structural and energetic data may then be directly combined, allowing favourable structures to be readily identified and understood. 'Alchemical' free energy methods such as thermodynamic integration (TI) and free energy perturbation (FEP) [70] allow the direct calculation of the difference in energy of two similar structures, e.g. two potential drug molecules in the same binding site, by switching atoms 'on' or 'off' to move between them.

'Physical' free energy methods typically rely on use of a set of restraints to move the system around a set of physical coordinates until enough data has been sampled to obtain a reasonable energy estimate. Methods under this scope include metadynamics [71] adaptive bias force [72] and umbrella sampling (US) [73]. US in particular allows the generation of a potential of mean force (PMF) profile, a one-dimensional energy profile along a 'reaction coordinate' of choice; from this 2D energy landscapes along other dimensions may also be calculated.

1.3 - Work presented in this thesis

It is of interest to investigate the interactions of peripheral proteins with membranes in order to better understand the role of these associations, and, in view of the many diseases associated with perturbed protein/membrane interactions, to identify the molecular basis of the effect of mutations on structure and role. MD simulations provide an ideal way to investigate these interactions, with creation and use of model membrane systems relatively straightforward, without the difficulties that may be faced experimentally. Additionally, the option to combine structural and energetic data allowed by MD will aid the identification of binding modes, their relative favourability, and the factors that distinguish them.

PIPs represent a key target for peripheral proteins. In this thesis, the behaviours of two key peripheral proteins/PIP systems were investigated using MD simulations: first, interactions with different PIPs of domains in the PH domain family, as a widespread domain for which the exact determinants of e.g. different selectivity and affinity is currently often unclear; and secondly, interaction of PTEN with PIP_3-containing membranes, representing an important clinically-relevant system where MD may aid in understanding the molecular function of wild-type and mutant and contribute to management of PTEN-related diseases.

Chapter 2 introduces the background and theory to the MD simulation method and the calculation of binding energies.

In *Chapter* 3, a pipeline is developed for the setup, running and analysis of a combined US/CG approach for studying membrane interactions, using GRP1-PH as an initial case. A set of parameters was chosen that allow for adequate sampling and convergence of the tested systems, and resultant trends in lipid selectivity and mutational effect were in agreement with previous experimental results. Favourable interactions at the secondary 'A' site, not previously reported for this protein, were observed, and the presence of multiple PIP molecules was demonstrated to have a large impact on the binding energy.

Chapter 4 extends this US/CG method to a further 12 PH systems. For most domains, no particular preference for PIP_3 or PIP_2 was observed, but those that did were

in agreement with experimental results. A number of different membrane-binding modes were observed, including a larger degree of A-site interactions that suggested previously. Domains could be grouped according to their relative preference for the A and C sites while close to or slightly displaced from the membrane (showing A-site interactions at any separation, only while loosely bound, or not at all). These groups could be related to short sequence motifs or charge distribution on binding loops. The importance of simultaneous binding to multiple PIP lipids to achieve the high affinities and membrane targeting characteristic of many PH domains is suggested.

Chapter 5 focusses on PTEN, using modelling to generate a full-length PTEN structure and performing multiscale simulations to investigate its membrane binding. Simulations suggest the presence of the completed N- and C-terminal regions is required for directing productive binding of PTEN, largely through changes in the electrostatic surface. The productive binding pose obtained was consistent with previous experimental and computational results. Further evidence for the potential role of the PD/C2 interface was seen, with a change in relative orientation of the two subunits on productive binding, and differences in relative orientations observed with the clinically-relevant R173C mutation.

An overall conclusion for this work is provided in *Chapter* 6, along with future directions outlining how use of MD simulation techniques such as those demonstrated here may enable us to continue to better understand the intricacies of the protein/membrane interactions.

Chapter 2

Theory and methods

2.1 - Molecular Dynamics simulations

The overall aim of MD simulations is to simplify the complex quantum-mechanical nature of atomic interactions so that the behaviour of molecules on biologically relevant timescales may be easily modelled with current computational infrastructure. To this end, atoms (or groups of atoms) are represented as a single point, allowable through the Born-Oppenheimer approximation in which electrons adjust instantly to changes in the nucleus. Given an approximation to the net force on each atom due to the presence of all others, their motions may then be calculated through straightforward application of Newtons equations of motion.

In performing MD simulations, there are two aspects that must be considered: first, the representation of each atom (or group of atoms) and the functions describing the interactions between them (known as a 'force field'), and second, the actual calculation of progression of movement. An overview is provided below; more detailed treatment can be found in e.g. [74].

2.1.1 - *Force fields*

A force field is a set of potential energy functions describing all possible interactions between particles (usually atoms) in a system. For a given atom i, the potential

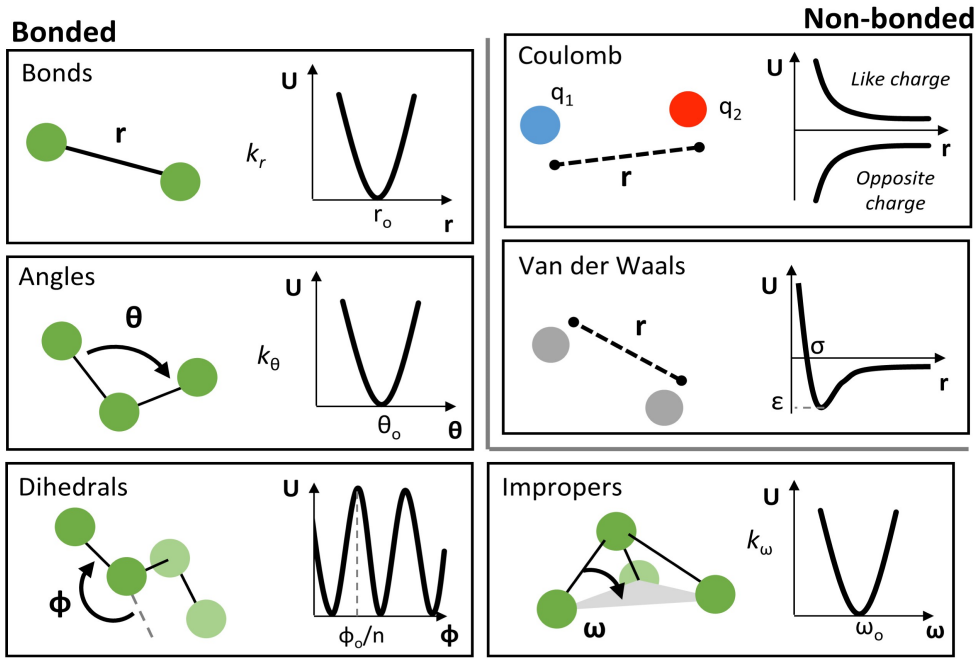


Figure 2.1: Bonded and nonbonded interactions in forcefields. Overview of the different bonded and non-bonded interaction types usually considered by MD force fields, along with the forms of the potentials commonly used to represent them.

energy is written as a sum of bonded and non-bonded interactions:

$$U_i(x^N) = U_{i,bonded}(x^N) + U_{i,nonbonded}(x^N) \tag{2.1}$$

where x^N gives the coordinates of all particles in the system.

The bonded term represents the effective restrictions on bond lengths and angles arising from the quantum mechanical interactions. These are usually well approximated by harmonic potentials (for bonds, angles and improper dihedral angles) or a simple oscillating potential (for dihedral angles) (Figure 2.1):

$$\begin{aligned} U_{i,bonded}(x^N) = &\sum_{bonds} \frac{1}{2} k_r (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \\ &+ \sum_{dihedrals} \frac{1}{2} k_\phi \left(1 + \cos(n\phi - \phi_0)\right) \\ &+ \sum_{impropers} \frac{1}{2} k_\omega (\omega - \omega_0)^2 \end{aligned} \tag{2.2}$$

r, θ, ϕ and ω are the current values of the bond length or angle; the k terms represent

the force constants and r_0, θ_0, ϕ_0 and ω_0 are the respective equilibrium values, and n sets the periodicity of the dihedral energy function.

The nonbonded term includes the charged and Van der Waals interactions, described using Coulombs law and a Lennard-Jones (LJ) potential, respectively, and computed over all other atoms j:

$$U_{i,nonbonded}(x^N) = \sum_j \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_j 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{-12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{-6}\right] \tag{2.3}$$

where q_i and q_j are the charges of the particles, ϵ_0 is the permittivity of free space, r_{ij} is the distance between the pair, and ϵ_{ij} and σ_{ij} give the depth and finite-distance-at-zero-potential of the LJ potential for the pair, respectively. Within the LJ potential, the r^{-12} term describes the short-range repulsion due to the Pauli exclusion principle, while the r^{-6} term represents the longer-range attraction due to the dispersion force.

The force field then provides the collection of charges (and masses) of each particle type, and the set of force constants, equilibrium values and LJ parameters for all possible bonds, angles and interaction pairs between them. Values are carefully chosen such that simulations best match experimentally and quantum-mechanically derived observations; this process is referred to as 'parametrisation'. Many different force fields exist, each parametrised separately and with their own versions of different biological molecules, including proteins (as individual amino acid building blocks) and lipids. In some cases multiple versions of a given molecule are available.

The choice of force field depends on the presence of a parameter set for the molecules of interest, compatibility with software, and the desired resolution, with forcefields available to carry out simulations at both an atomistic or reduced representation (coarse-grain) level.

2.1.1.1 - *Atomistic force fields*

'All-atom' force fields consider every atom in a system, while 'united atom' force fields exclude non-polar hydrogens. Common examples include CHARMM [75] and GROMOS (e.g. [76]), respectively. Atomistic simulations performed in this thesis used the GROMOS 43a1 [76] force field. The 'Berger lipids', derived from initial

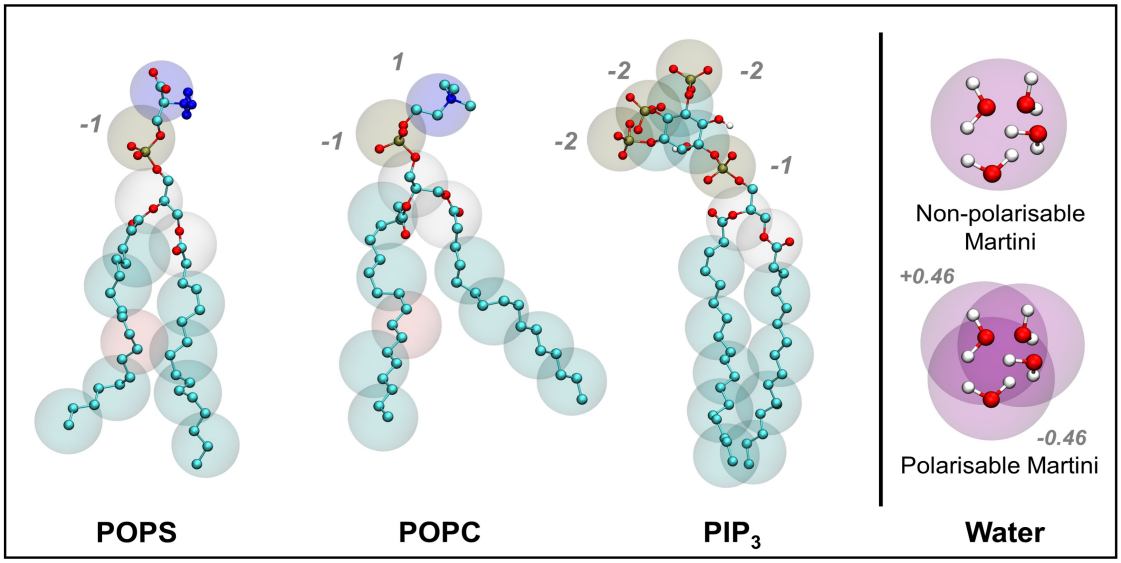


Figure 2.2: Coarse-grain and atomistic representations of lipids and water. Atomistic (from GROMOS 43a [76]) and coarse-grain (from Martini 2.1 [61, 78]) representations of POPS, POPC, PIP_3 and water, shown with CPK representation and as transparent VdW spheres, respectively. Both polarisable and non-polarisable water models are shown for Martini. Numbers indicate the charges on Martini beads, beads not annotated are neutral.

parametrisation by Berger *et al.* [77], are a commonly used parameter sets for lipids under the GROMOS force field, including POPC, POPS and PIP_3, the lipids of interest in this study (Figure 2.2).

Representation of water is particularly difficult due to its unusual properties, and the desire for a relatively simple model due to the large presence of water in any given simulation; many water models, including the SPC, TI3P and TI4P (among the first models developed and still in widespread use) don't include the polarisability of water [79]. Atomistic simulations presented in this thesis use the SPC water model.

2.1.1.2 - *Coarse grain force fields*

In coarse-grain force fields, such as Martini [61, 78], the simulated particles are representative beads rather than individual atoms; interactions between these particles, both bonded and non-bonded, are modelled and parametrised in a similar way to atoms in atomistic force fields. Despite the loss in resolution, coarse-grain force fields

have the advantage of allowing much larger timescales to be simulated Martini, for example, offers a 2-3 order of magnitude improvement in timescale [61]. This is both due to the reduced total number of particles in a given system, greatly reducing the number of calculations to be made at each step, the increase in particle size/mass leading to slower degrees of motion and allowing larger simulation timestep (see Section 2.1.2), and use of shorter-range potentials. Dynamics may also be faster as 'friction' is reduced due to smoother movement of particles.

The Martini force field (version 2.1 [61, 78]) was used for all CG simulations in this thesis. In this force field, particles (or 'beads') represent groups of approximately 4 heavy atoms. In this representation, amino acids consist of a backbone bead and 0 to 4 sidechain beads; alanine and glycine are both represented as a single backbone bead [61]. Parameter sets for POPC, POPS and PIP are again available (Figure 2.2). Initial parametrisation of the PIP headgroup was obtained for PI(3,4)P_2 [80]. Models for PIP_3 and PI(4,5)P_2 were then generated by mirroring the appropriate phosphate beads, but have not been properly tested. It is also noted that in the simplified model used, PIP becomes symmetric and PI(3,4)P_2 and PI(4,5)P_2 are essentially identical.

The four-heavy-atoms approximation of the Martini model also extends to water, with one Martini water bead representing approximately four water molecules. As for many atomistic water models, polarisability is not modelled. A more complicated Martini water model has been developed to mimic polarisable behaviour [81] (Figure 2.2), though with the introduction of two new particles for every water bead this model greatly increases the computational time required.

2.1.2 - *Running MD simulations*

Having obtained a forcefield with which to represent a system and the interactions within it, the next step is to calculate the resultant particle movements. From Newtons second law ($F = ma$), evolution of the position r over time t of a particle of mass m can be related to the net force F acting on it via:

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} \tag{2.4}$$

The net force can be obtained from the potential energy according to the forcefield,

which in turn depends on the position of the particle (and of all other particles in the system):

$$F = -\frac{dU(r^N)}{dx} \tag{2.5}$$

This set of differential equations cannot be solved analytically; instead, a numerical approach is used. There are a number of simulation engines available, including GROMACS [82], NAMD [83] and CHARMM [75], which, given a force field and an initial set of coordinates (and velocities), utilise different methods to estimate new particle positions after a small timestep; from these, a new set of forces are calculated and the cycle repeated to progress the simulation.

All simulation in this thesis used GROMACS versions 4.5 or 4.6, which by default uses a 'leap-frog' integrator. Once every step, at the half-step interval, velocities are updated using the previous set of forces following Equation 2.4:

$$v\left(t + \frac{1}{2}\delta t\right) = v\left(t - \frac{1}{2}\delta t\right) + \frac{F(t)}{m}\delta t \tag{2.6}$$

where v is the velocity of a particle, t is the current time δt is the timestep, m is the particle mass and $F(t)$ is the net force on that particle at time t.

Every step, positions are updated according to the previous positions and this set of velocities:

$$r(t + \delta t) = r(t) + v\left(t + \frac{1}{2}\delta t\right)\delta t \tag{2.7}$$

where r is the position of a particle.

Ideally, the timestep used should be faster than the fastest degree of motion in the system in order to accurately capture particle motions. For atomistic systems, this is vibration of C-H bonds; however, given that bond stretching is not of particular interest, bond lengths are often fixed using, for example, LINCS [84], allowing larger timesteps of the order 1-2 fs. Slower variation amongst the 'bigger' particles of coarse grain simulations allow greater timesteps, in the order of 20 fs.

Several other factors important aspects of running MD simulations are outlined below. Further details may be found in the GROMACS manual [85].

Periodic boundary conditions - to prevent edge effects arising from discontinuities at the boundaries, periodic boundary conditions are often used, indefinitely projecting identical images of the central box to give a continuous system. A particle that leaves one side of the simulation box will re-enter through the other. Different box shapes may be employed, the most common being a rectangular prism. Care must be taken when choosing the box dimensions to prevent unwanted interactions between a molecule and its projected image.

Nearest neighbour list and cut-offs - given the large number of atoms in a system the calculation of non-bonded interactions is a large bottleneck in the MD process. The LJ potential decays with r^{-6}, and it is therefore not unreasonable to specify a distance cut-off (around 1 nm) above which LJ interactions are not calculated. Contribution of interactions may be gradually shifted to zero starting from an earlier distance to avoid introducing a discontinuity. To avoid having to search through all particles to find those within the cut-off every timestep, a 'nearest neighbour' list of particles within or near this cut-off may be maintained and updated only every 10 or so steps. LJ interactions between close-bonded neighbours is assumed to be described by the bonded potentials and so particles within 2 bonds are excluded from the LJ calculation. Coulomb interactions decay with r^{-1}; a similar cut-off scheme may be sufficient for CG simulations, but with an atomistic resolution a more accurate treatment is needed to avoid errors.

Particle Mesh Ewald (PME) - PME [86] is used as an alternative to cut-off schemes to speed calculation of Coulomb interactions. In the Ewald method, electrostatic interactions are split into a short-range and a long-range component. The short-range component will converge quickly with r, while the long-range component will not. However, given the periodicity of the system under periodic boundary conditions, the long-range interactions will converge quickly in Fourier space. The two components may then be truncated separately in their respective spaces, without significant error. In the PME method, charges are assigned to a grid, allowing use of fast Fourier transform for better scaling with system size ($Nlog(N)$ ranger than N^2, for number of particles N).

Temperature and pressure coupling - Many simulations, including those here, are performed under NPT (isothermal, isobaric) conditions, reflecting experimental conditions. In order to maintain temperature and pressure of a simulation system, a thermostat and barostat must be employed. Simple thermostats, such as the Berendsen thermostat, take advantage of the relation between kinetic energy and temperature, rescaling the velocities of all particles to implement necessary changes in temperature.

In the Berendsen thermostat [87], temperature is rescaled according to:

$$\frac{dT(t)}{dt} = \frac{1}{\tau_T}(T_{ref} - T(t)) \tag{2.8}$$

where T_{ref} is the desired reference temperature and τ_T is a coupling constant. The correction is applied after velocities are calculated. Groups of molecules within the system (lipids, protein, etc) are often treated separately to ensure even distribution of kinetic energy. In this thesis, CG simulations use the Berendsen thermostat, while atomistic simulations use the more accurate v-rescale method, which adds an additional term to Equation 2.7 to ensure a correct kinetic energy distribution.

Pressure coupling may be achieved by rescaling of the box dimensions and corresponding particle coordinates, performed when coordinates are updated. A Berendsen barostat [87] works in a similar way to the Berendsen thermostat, with pressure changed according to:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_P}(P_{ref} - P(t)) \tag{2.9}$$

Changes are passed on to the box dimensions according to a scaling matrix that depends on a specified compressibility. In membrane simulations, such as those performed here, it is useful to for pressure coupling to be semi-isotropic, with the xy dimensions (membrane plane) scaled together and the z dimension scaled separately; this helps limit distortions in the membrane. CG simulations in this study use the Berendsen barostat, while atomistic simulations use the more complicated and accurate Parrinello-Rahman coupling [88], which also adds an additional term to Equation 2.4.

Additional restraints - it is often useful to specify additional restraints within an MD simulation, for example to prevent unwanted behaviours or encourage others, like in the umbrella sampling approach discussed below. These additional restraints are often of harmonic potential, and are applied along with the force field.

Energy minimisation - before starting an MD simulation, it is often necessary to perform a short 'energy minimisation' simulation, in order to ensure there are no steric clashes or other instabilities in the initial system. In this thesis, all energy minimisations are carried out using 'steepest decent': each step, the force on each particle is calculated, and they are displaced proportional to this force. If the potential at the new location is higher, the particle is returned to the previous location, and proportionately lower movements will be trialled in future steps. This is repeated for a specified number of steps or until the minimum force falls below a specified value.

2.2 - Calculation of binding free energies

In an NPT ensemble, free energy of associated with a particular state is given by the Gibbs free energy G:

$$G = H - TS \tag{2.10}$$

where H and S are the enthalpy and entropy of the state, respectively, and T is the temperature. Knowing the free energy of a system tells us how likely we are to observe particular states and how likely transitions between them are to occur: from the Boltzmann distribution, the probability of observing a given state i with energy E_i is:

$$p_i = \frac{\exp(-\beta E_i)}{\sum_j \exp(-\beta E_j)} \tag{2.11}$$

$\beta = k_B T$, where k_B is the Boltzmann constant.

Free energies are particularly useful to quantify the relative strengths of interactions between molecules. In this case, the 'binding free energy' ΔG_{bind} is given by the difference in free energy of a 'bound' state in which the molecules are interacting favourably, and an 'unbound' state in which they are far apart and do not 'see' each other. Binding free energies can be calculated in several ways.

2.2.1 - *Binding free energy from experiment*

Strength of binding, including between proteins and membranes (or lipid analogues) is usually quantified experimentally by a dissociation constant, K_D, calculated from the kinetics of the interaction. The dissociation constant is related to the binding energy via:

$$\Delta G_{bind} = RT \ln \frac{K_D}{c_o} \tag{2.12}$$

R is the ideal gas constant and c_o is the standard reference concentration, 1 mol L^{-1}.

The difference in binding strength between two molecules for the same partner is often reported as a 'fold selectivity', or the ratio of their respective binding constants. From Equation 2.12, this is equivalent to:

$$selectivity = \frac{K_{D1}}{K_{D2}} = \exp(\Delta G_{bind1} - \Delta G_{bind2}) \tag{2.13}$$

2.2.2 - *Binding free energy from simulation*

A key assumption of MD simulations is the 'ergodic hypothesis': that a time-average calculated from a suitably long simulation is equivalent to the ensemble average over the underlying statistical ensemble. Thus, the probability distribution of states along a particular coordinate of a system can be obtained from the frequency the state is observed in simulation (assuming the simulation is sufficiently sampled):

$$p(\xi) = \frac{n(\xi)}{N} \tag{2.14}$$

$n(\xi)$ is the number of simulation frames in which the system in state ξ and N is the total number of frames. Rearranging Equation 2.11, the free energy landscape may then be calculated as:

$$E(\xi) = -\beta \ln p(\xi) + C \tag{2.15}$$

where C is an arbitrary constant: energies are always relative to some reference point, and so have no unique value; the whole landscape may be freely shifted as one unit. Commonly, the landscape is shifted such that some convenient reference is at 0.

It is often difficult to obtain the level of sampling required for Equation 2.14 to be accurate, especially when in the presence of energy barriers. For example, given a

favourable interaction, it is likely that, once bound, the molecules will not separate for the remainder of the simulation and as such the 'unbound' states will be under-sampled.

Free energy calculation methods may therefore focus on biasing the system to force exploration of these less-favoured states. There are several ways to achieve this; umbrella sampling (US) [73] is among the more common and is the method chosen for use in this thesis. It is outlined below.

2.2.2.1 - *Umbrella sampling and WHAM*

If a restraint of form $U(\xi)$ is applied to a system, the resulting 'biased' probability distribution is given by:

$$p_{biased}(\xi) = p_{unbiased}(\xi) \exp(-\beta U(\xi)) \left\langle \exp(-\beta U(\xi)) \right\rangle^{-1} \tag{2.16}$$

The last term may be written as,

$$\langle \exp(-\beta U(\xi) \rangle = \exp(-\beta F) \tag{2.17}$$

where F gives the free energy associated with introduction of the restraining potential $U(\xi)$ to the system. It is difficult to calculate F directly; however, in the case of a single restraint, substitution into Equation 2.15 it contributes to the calculated energy as a constant offset, and so can be ignored.

Rearranging Equation 2.16, given a biased probability, an unbiased distribution may be calculated by:

$$p_{unbiased}(\xi) = p_biased(\xi) \exp(-\beta U(\xi) + \beta F)) \tag{2.18}$$

Note the calculated $p_{unbiasaed}(\xi)$ will be accurate only in the region sampled by $p_{biased}(\xi)$.

In order to obtain full sampling along a chosen coordinate - the 'reaction coordinate' - in the US method, a number of simulations (or 'windows') are performed, each restrained to cover small, overlapping sections of the reaction coordinate. These restraints are usually harmonic in form:

$$U_i(\xi) = \frac{1}{2} k_i (\xi - \xi_{o,i})^2 \tag{2.19}$$

$\xi_{o,i}$ is the value of the minima in window i. The force constants k_i may be different for each window, but are often kept the same.

Reaction coordinate data are collected from each window and the local biased ($p_{biased,i}$) and unbiased ($p_{unbiased,i}$) probability distributions calculated using Equations 2.14 and 2.18, respectively. These unbiased distributions must then be combined to form the full unbiased distribution p_o. At the simplest level, this could be done by simply rescaling the distributions (or, equivalently, shifting the corresponding partial energy profiles) to best match between adjacent windows, then deleting overlapping data. However, this involves the loss usable data and it is unclear exactly how adjacent windows should be aligned. Several methods for the more accurate method for combining windows, the most common of these being WHAM [89].

To derive the WHAM equations, the expected statistical uncertainty associated with each window's probability distribution is calculated in terms of p_o. An estimation of p_o is then made by adding the individual $p_{i,unbiased}$, assigning appropriate weights to minimise propagation of the statistical error. This yields:

$$p_o(\xi) = \frac{\sum_i n_i(\xi) g_i^{-1}}{\sum_i N_i g_i^{-1} \exp(\beta F_i - \beta U(\xi))} \tag{2.20}$$

g_i^{-1} are the statistical inefficiencies of each window, given by $g_i = 1 + 2\tau_i$ where τ_i are the integrated autocorrelation times. These are often similar between windows and will cancel out; variations within an order of magnitude do not significantly affect the PMF [90]. Otherwise, they can be calculated from the autocorrelation function $R_i(\Delta t)$ by:

$$\tau_i = \sum_{\Delta t=1}^{\infty} R_i(\Delta t) \tag{2.21}$$

To allow feasible calculation, the sum may be carried out only until $R_i(\Delta t)$ drops below a defined threshold (e.g. 0.05 in [91]).

The F_i values appearing in Equation 2.20 no longer evaluate to a constant when calculating energies, and so must be considered. Rather than directly calculating the

set of F_i values, they can be obtained in terms of $p_o(\xi)$ using Equation 2.16 and the condition that $\sum_\xi p_{biased}(\xi) = 1$:

$$\exp(-\beta F_i) = \sum_\xi p_o(\xi) \exp(-\beta U_i(\xi)) \tag{2.22}$$

Equations 2.20 and 2.22 form the WHAM equations. Given an initial guess of the F values, the two equations are iterated until the changes observed fall within a predefined threshold, giving a final probability distribution over ξ and a set of F values associated with each window.

2.2.2.2 - *Potential of mean force profiles*

The (one-dimensional) energy landscape obtained along the reaction coordinate by US (or another free energy method) is known as a potential of mean force (PMF) profile $W(\xi)$. From Equation 2.15:

$$W(\xi) = -RT \ln p_o(\xi) + C \tag{2.23}$$

Here the ideal gas constant R is used in place of Boltzmann's constant to get energy in units of kJ mol^{-1}. This profile assumes that all coordinates orthogonal to the reaction coordinate were adequately sampled. Error in $W(\xi)$ is often calculated via bootstrapping, in which multiple sets of 'fake' data are generated from the final probability distributions of each window; profiles are calculated for each fake set and the error determined from the variation in these. However, being generated from existing data, bootstrap analysis does not identify errors due to poor sampling; the degree of sampling must be assessed in other ways. Usually, converge of the profile with time - that is, when the profile shape is not significantly altered on addition of further data, is usually taken to indicate adequate sampling.

Sampling along orthogonal coordinates may be improved through use of restraints. US may be extended to multiple dimensions, but this is often prohibitively expensive. Instead, simple lateral restraints, present in all windows, may be employed, but must be accounted for in the calculation of, for example, binding free energies.

In the investigation binding by US, the reaction coordinate is usually chosen as separation between the interaction partners, and the arbitrary constant C in Equation

2.23 set so the PMF is zero in the 'unbound' state. Assuming binding is favourable, an energy well will be observed at the 'bound' state; the value of the minimum is referred to as the 'well depth' ΔW, and can loosely be taken to represent the binding energy. More correctly, the absolute binding free energy is calculated from a one-dimensional PMF taking in to account the reduced dimensionality and the effect of any orthogonal restraint. Following Doudou *et al.* [92], assuming the PMF is calculated along the z axis, perpendicular to the membrane, the absolute binding free energy is given by:

$$\Delta G^0_{bind} = \Delta W - RT \ln \left(-\frac{2\pi RT}{V_o k_{xy}} \int_{bound} \exp\left(-\frac{W(z)}{RT} \right) dz \right) + \Delta G_R \tag{2.24}$$

k_{xy} is the force constant of the lateral restraint. The integral is performed over the 'bound' region until the bulk-region is reached; here, the PMF profile is assumed to be shifted such that the minimum is at zero. $V_o = 1661 \mathring{A}^3$ is the standard-state volume. ΔG_R accounts for the effects of the lateral restraints, calculated as:

$$\Delta G_R = RT \ln \left\langle \exp\left(-\frac{k_{xy}(\Delta x^2 + \Delta y^2)}{2RT} \right) \right\rangle_{k_{xy}=0} \tag{2.25}$$

Δx and Δy are centre-of-mass displacements from the location of the orthogonal restraint minimum, averaged over a simulation in which no restraints were applied.

2.2.2.3 - *Extension to other coordinates*

The PMFs described above show the energy profile along only one coordinate. Taking a second coordinate η independent of the reaction coordinate, Equation 2.20 may be extended to obtain a two-dimensional energy landscape:

$$p_o(\xi, \eta) = \frac{\sum_i n_i(\xi, \eta)}{\sum_i N_i \exp(\beta F_i - \beta U_i(\xi))} \tag{2.26}$$

or equivalently,

$$p_o(\xi, \eta) = \frac{\sum_i \sum_f \delta(\xi_{if} - \xi)\delta(\eta_{if} - \eta)}{\sum_i N_i \exp(\beta F_i - \beta U_i(\xi_{if}))} \tag{2.27}$$

where δ is the Dirac delta function, $\sum_f$ is the sum over each frame in simulation i and ξ_{if} and η_{if} are the values of η and ξ in frame f of simulation i. In general, once the F values for each window are known, an energy profile across any number of coordinates η^X may be obtained from the appropriate timeseries of η^X and corresponding ξ:

$$p_o(\eta^N) = \frac{\sum_i \sum_f \delta(\eta^N_{if} - \eta^N)}{\sum_i N_i \exp(\beta F_i - \beta U_i(\xi_{if}))} \tag{2.28}$$

Once F-values are known, they can also be used to compute averages along the reaction coordinate of other properties:

$$\langle \eta(\xi) \rangle = \frac{\sum_i \sum_f \eta_{if} \exp(\beta U_{if}(\xi_{if}) - \beta F_i) \delta(\xi_{if} - \xi)}{\sum_i \sum_f \exp(\beta U_{if}(\xi_{if}) - \beta F_i) \delta(\xi_{if} - \xi)} \tag{2.29}$$

2.2.2.4 - *Umbrella Sampling in practice*

The umbrella sampling method has been demonstrated to obtain accurate energetic and structural data in a variety of different systems. Interactions involving membrane proteins have been studied using a coarse-grained representation in order to achieve sufficient sampling on reasonable timescales: for example, investigation of the binding of lipids to transmembrane protein [93] and the dimerisation of transmembrane helices [94, 95]. These studies were able to identify preferred binding partners and the effects of mutations, demonstrating that the umbrella sampling methodology may indeed be applied towards the study of the interaction of biological models, even with use of a coarse-grained representation.

Chapter 3

Structures and energetics of peripheral protein/membrane interactions through simulation: developing a US-CG approach with GRP1-PH

3.1 - Background

3.1.1 - *Why combine Umbrella Sampling and Coarse Grain modelling?*

Molecular Dynamics (MD) simulations are widely used to investigate dynamic atomistic or near-atomistic structural information of biomolecular systems, including those involving membranes [42]. Through free energy calculation methods, quantified energetic information may also be obtained. The Umbrella Sampling (US) method [73], outlined in Section 2.2.2.1, seems particularly well suited for the investigation of peripheral protein/membrane interactions on both a structural and energetic level, with separation along the membrane normal providing a natural reaction coordinate. However, while the US method is well established in the study of small ligands, including, for example, membrane permeation, the long timescales required to achieve adequate sampling and accurate results have limited the application of US to larger systems such as those involving the interactions of membrane proteins. Recent studies have thus taken advantage of the increase in accessible timescales offered by use of a coarse-grained model [60] (see also Section 2.1.1.2) to allow free energy calculations through US of systems involving proteins in a membrane environment, such as the

dimerisation of transmembrane helices [94–96] and the interaction of lipids with transmembrane proteins [97].

This work aimed to extend this method to the case of peripheral proteins associating with a membrane. This represents a novel application of the US methodology: to my knowledge, protein-membrane interactions had not previously been investigated in this way on a large scale (comparing multiple proteins and lipids).

Before working to address wider questions on the nature of protein-membrane binding, it was thus chosen to establish a protocol for combined US/CG simulations, using the pleckstrin homology (PH) domain of GRP1 binding to membranes containing phosphatidylinositol phosphate (PIP) lipids. This allowed the feasibility of the method to be assessed, suitable values for simulation parameters to be identified, and a pipeline to be developed for the setup, running and analysis of US/GG simulations between peripheral proteins and membranes, ready for application to further systems.

3.1.2 - *Why GRP1-PH as an initial case?*

A well characterised PH domain was chosen for this initial investigation due to its small and compact size (at ∼120 residues), and to provide a gateway to the study of members of the PH domain family and their interactions with PIP lipids. Despite the widespread nature and importance of both PH domains and PIPs (see e.g. refs. [7], [13]), the varying selectivity and affinities of PH domains for different PIPs are often poorly understood. Furthermore, structural studies indicate that while most PH domains, including GRP1 [20], interact with PIP through the 'canonical' (or C) binding site, some bind through an adjacent 'alternate' (or A) site [98] (labelled in Figure 3.1 A). Additional binding sites for other lipids have also been proposed for some domains, including GRP1 [98]. However, the importance of the presence of multiple binding sites remains again unclear.

The GRP1 (General Receptor for Phosphoinositides 1) protein is involved in cytoskeletal dynamics; its PH domain, being highly selective for PIP_3, allows GRP1 to be targeted to the plasma membrane in response to insulin-stimulated accumulation of PIP_3, whereupon it propagates this signal through guanine exchange factor activity [99]. Many previous studies, both computational and experimental, have investigated

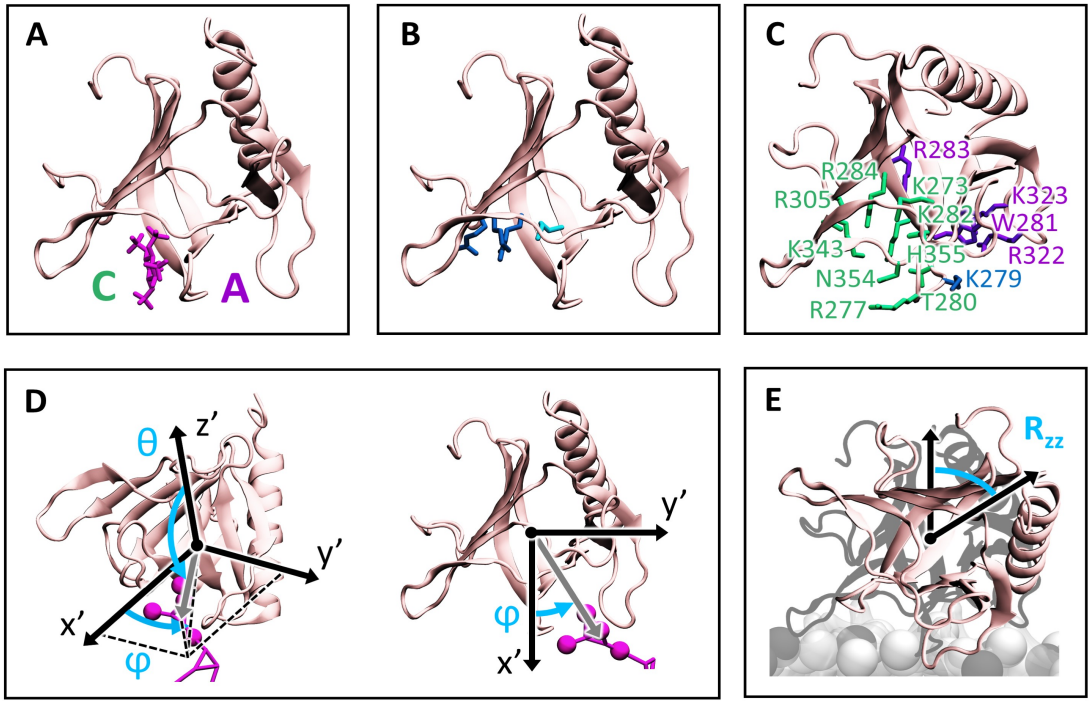


Figure 3.1: Structure of GRP1-PH. (A) Crystal structure with IP_4 (magenta, liquorice representation) bound at the canonical (C) binding site (PDB ID: 1FGY); the location of the alternate (A) PIP binding site is also indicated. **(B)** Location of four key positive residues in the C binding site (K273 in light blue; R288, R305 and K343 are shown in dark blue). **(C)** Location of residues in the A (purple) or C (green) sites. **(D)**Measurement of the 'PIP binding angle'. PIP is shown with a ball-and-stick representation in magenta. **(E)** Measurement of membrane orientation (R_{ZZ}). The reference structure (the initial binding pose, based on the crystal structure) is shown in grey.

interactions of GRP1-PH with PIPs (see Tables 3.1 and 3.2), providing data both to direct the setup and with which to compare results from this study. GRP1-PH therefore provides an ideal initial case for developing and assessing the US-CG approach for peripheral protein/membrane interactions, leading in to an investigation of PIP binding across PH domains.

3.2 - Methods

A table listing all production simulations performed for this chapter is provided as Table 3.3. See also Figure 3.2 for an overview of the setup, running and analysis process. Several additional simulations not listed were also performed in order to test different simulation parameters; these are outlined in Section 3.3.1.

Table 3.1: Summary of structural studies of GRP1-PH/PIP interactions. Studies were performed in either a membrane environment or in solution with a soluble PIP analogue: the lipid composition or analogue used, respectively, is listed. Residues identified to be in contact with or near the bound PIP or PIP analogue are listed. Where possible, only those forming sidechain contacts are listed.

Method	Environment	Residues	Ref.
X-ray crystallography	IP_4	*sidechain contacts:* K273, T280, K282, R284, Y295, R305, K343, N354, H355	[20]
X-ray crystallography	IP_4	*sidechain contacts:* K273, T280, K282, R284, Y295, R305, K343, N354, H355	[100]
Spin labelling	PC:PS:PIP_3 (74:24:2)	*orientation only*	[101]
NMR shift	diC_4PIP_3	K273, L274, G275, G276, V278, K279, R284, W285, I308, G336, K343, T344, A346, N347, E352	[102]
MD	PC (100) + 1 PIP_3	*sidechain contacts:* K273, R277, T280, K282, R284, Y295, R305, K343, H355	[66]
MD	PC(100) + 1 PIP_3	K273, G274, T280, Y295, F296, K343, T344	[103]
MD	PC:PS (75:25) + 1 PIP_3	*sidechain contacts:* K273, R277, R284, Y295, R305, K343, H355	[65]
MD	PC:PS:PIP_2:PIP_3 (73:20:5:2)	*sidechain contacts:* K273, R277, V278, K279, T280, K282, R284, K343, N354	[104]

3.2.1 - *Details of the protein-membrane system*

PH domain - An atomistic structure of the GRP1-PH domain, with I(1,3,4,5)P_4 (the PIP_3 headgroup) bound at the C-site, was obtained from the PDB (ID 1FGY). Mutant GRP1-PH was created with the Mutator plug-in of VMD [105]. Atomistic structures were converted to a coarse-grained Martini-forcefield [61, 78] representation using an in-house script.

Membranes - Martini coarse-grain membranes were generated from a preformed bilayer of xy dimensions 11.3×11.3 nm^2 consisting of 192 POPC molecules in each leaflet. An in-house script was used to replace POPC lipids with other lipid species to achieve the compositions listed in Table 3.3. For single-bound-lipid US simulations, one POPC lipid in the 'upper' leaflet was initially replaced with a PIP_3 lipid for alignment, as outlined below, and then converted to the appropriate 'bound' lipid.

Table 3.2: Summary of biophysical/biochemical studies of GRP1-PH/PIP interactions. Studies were performed in either a membrane environment or in solution with a soluble PIP analogue: the lipid composition or analogue used, respectively, is listed. Temperature and pH are listed, if reported. Absolute binding energies are calculated from reported K_D values using Equation 2.12 and listed in square brackets, using a temperature of 298 K if not otherwise reported.

For qualitative studies: 'Y' - binding seen, '<PIP_3' - binding lower than for PIP_3, 'N' - no binding seen.

Abbreviations: PC - phosphatidylcholine, PE - phosphatidylethanolamine, PS - phosphatidylserine, PI - phosphatidylinositol, SM - sphingomyelin, Ch - cholesterol

Method	**Environment**			**PIP binding K_D (nM)** [Binding energy (kJ mol^{-1})]				**Ref.**
	PIP	T (°C)	pH	PIP_3	PI(4,5) P_2	PI(3,4) P_2	No PIP	
Overlay	PIP (spotted)			Y	N	N		[106]
Overlay	PIP (spotted)			Y	<PIP_3	<PIP_3		[107]
Overlay	PIP (spotted)			Y	<PIP_3	<PIP_3		[108]
Overlay	PIP (spotted)			Y		N		[109]
Radiolabel	diC_8PIP (soluble)			50 [-41]	<PIP_3			[110]
ITC	IP (soluble)			27.3 [-43]	>4500	>2300		[106]
ITC	IP (soluble)	25		35 [-42]	6360 [-29]			[111]
ITC	diC_8PIP (soluble)	25		22 [-43]	7260 [-29]			[111]
ITC	IP (soluble)			42 [-42]	5890 [-30]			[112]
SPR	PC:PE:PIP 77:20:3	23	7.4	170 [-38]	N	730 [-35]	N	[113]
SPR	PC:PE:PIP 78:20:2	25	6	28 [-43]				[102]
SPR	PC:PE:PIP 78:20:2	25	7.4	150 [-38]				[102]
SPR	PC:PE:PIP 78:20:2	25	8	620 [-35]				[102]
SPR	PC:PE:PS:PI:Ch:PIP 9:35:22:9:22:3	25	6	4.8 [-47]				[102]

Continued on following page

Continued from previous page

SPR	PC:PE:PS:PI:Ch:PIP 9:35:22:9:22:3	25	7.4	28 [-43]				[102]
SPR	PC:PE:PS:PI:Ch:PIP 9:35:22: 9:22:3	25	8	110 [-40				[102]
FRET	PC:PE:PIP 92:5:3	25	7.4	110 [-40]				[114]
FRET	PC:PS:dPE:PIP 69:23:5:3	25	7.4	700 [-35]			N	[114]
FRET	PE:PC:PS:PI:dPE:SM:PIP 41.4:18.4:18.4:9.2:5:4.6:3	25	7.4	50 [-42]	N	8000 [-29]		[114]
FRET	PE:PC:SM:dPE:PIP 59.1:26.3:6.6:5:3	25	7.4	600 [-35]				[114]
FRET	PC:PE:PIP 93:5:2	21.5	7.5	60 [-41]			N	[98]
FRET	PC:PS:PE:PIP 69.8:23.2:5:2	21.5	7.5	5 [-47]				[98]
FRET	PC:PS:PE:PI:SM:Ch:dPE:PIP 10:21:27.5:4.5:4.5:25:5:2	25	7.4	14 [-45]	3000 [-32]			[115]
FRET	PC:dPE:PIP 92:5:3	25		481.7 [-36]			N	[65]
FRET	PC:PS:dPE:PIP 69:23:5:3	25		72.7 [-41]				[65]

Initial structures - The initial structure for the 'bound' simulation and single-bound-lipid US simulations was generated by aligning the position of the IP_4 in the crystal structure with the headgroup of the membrane PIP lipid, minimising the phosphate-bead RMSD. Initial structures for binding simulations were generated by centring the protein 7 nm above (along the $+z$ axis) the bilayer in a random orientation. An initial structure for US simulations of GRP1-PH bound to multiple PIP lipids was taken from binding simulations, selecting a snapshot by visual inspection in which there were PIP_3 molecules bound in each of the C and A sites. For the $2 \times PIP_3$ case, the remaining PIP_3 lipids were then converted to POPC.

Solvation - Using built-in tools of the GROMACS 4.5.5 [82] simulation engine, the protein-membrane system was solvated with standard or polarisable Martini water [81] (as indicated in Table 3.3) with box size $11.3 \times 11.3 \times 16$ nm^3 (binding and US-GRP1-PH/PIP_3 simulations) or $11.3 \times 11.3 \times 15$ nm^3 (all other simulations), and neutralising Na^+/Cl^- ions added to a concentration of 0.15 M. Each system was

Table 3.3: Summary of simulations performed for Chapter 3. Summary of the details of all production simulations performed in this chapter. 'Bound' lipid refers to the lipid initially located in the canonical binding site of the PH domain (and, for multi-PIP US simulations, also in the alternate site). For single-bound-lipid US simulations, the listed membrane composition indicates the composition before a single POPC lipid in the upper leaflet was converged to the bound lipid species. The net charge of the bound lipid is listed, with the charge distribution across the 3' (if applicable), 4' and 5' phosphate beads shown in brackets. If not specified, each had a charge of $-2e$. In all cases, the 1' phosphate had a charge of $-1e$.

Type	**Lipid composition** (PC:PS[:PIP])	**'Bound' lipid**	**'Bound' lipid charge**	**GRP1-PH**	***xy* restraint?**	**Polarisable water?**	**Sim. time (ns)**	**No. windows**	**Convergence time (ns)**
US	80:20	PIP_3	-7	WT	Y	N	1000	31	300
US	80:20	$PI(4,5)P_2$	-5	WT	Y	N	1000	24	300
US	80:20	$PI(3,4)P_2$	-5	WT	Y	N	1000	24	300
US	80:20	POPC	0	WT	Y	N	1000	24	500
US	80:20	POPS	-1	WT	Y	N	1000	24	300
US	80:20	PIP_3	-7	4A	Y	N	1000	24	500
US	80:20	PIP_3	-7	K273A	Y	N	1000	24	300
US	80:19.5:0.5[a]	PIP_3	-7	WT	Y	N	2000	23	500
US	76:19:5[b]	PIP_3	-7	WT	Y	N	3000	24	900
US	80:20	PIP_3	-5 (-1, -2, -1)	WT	Y	N	2000	24	700
US	80:20	PIP_3	-5 (-1.3, -1.4, -1.3)	WT	Y	N	2000	24	100
US	80:20	$PI(4,5)P_2$	-4 (-2, -1)	WT	Y	N	2000	23	200
US	80:20	PIP_3	-7	WT	Y	Y	2 x 2000[c]	24	700, 200
US	80:20	$PI(4,5)P_2$	-5	WT	Y	Y	1000	24	-
US	80:20	PIP_3	-7	WT	N	N	2000	24	200
Bound	80:20	PIP_3	-7	WT	N	N	1000	-	-
Bind	76:19:5[b]	-	-	WT	N	N	5 x 1000[d]	-	-

[a] 2 PIP_3 lipids per leaflet
[b] 10 PIP_3 lipids per leaflet
[c] Two repeats were performed, using different starting velocities
[d] Five repeats were performed, using different starting orientations of GRP1-PH

subject to energy minimisation by steepest decent for 300 steps and 5 ns equilibration in which protein atoms were restrained by force constant 1000 kJ mol^{-1} nm^{-2}.

3.2.2 - *Simulation details*

Simulations - All simulations were run using GROMACS 4.5.5 [82] with the Martini 2.1 force field [61, 78]. The timestep was 20 fs, with particle coordinates written out every 400 ps. The nearest neighbour list was updated every 10 steps. Coulombic and Lennard-Jones interactions were shifted to zero between 0-1.2 nm and 0.9-1.2 nm, respectively. Relative electrostatic screening was 2.5 for polarisable water simulations and 20 for all other simulations. Temperature and pressure were maintained at 323 K and 1 bar, respectively, using a Berendsen thermostat [87] (coupling constant 1 ps) and barostat (coupling constant 1 ps, compressibility 5 x 10^{-6} bar^{-1}; semi-isotropic). Bond lengths were constrained using the LINCS algorithm [84]. Periodic boundary conditions were applied in xyz.

Additional setup for US simulations - The reaction coordinate was chosen as the component of the protein centre-of-mass to lipid phosphate bead (for PIPs, the 1' phosphate was used) separation along the membrane normal (see 3.6 A); the systems were set up such that this was aligned with the z-axis. A steered MD simulation was performed, restraining the COM of the PH domain with force constant 1000 kJ mol^{-1} nm^{-2} to a reference point moving in the $+z$ direction at rate of 0.01 nm ps^{-1} relative to the lipid phosphate, which was restrained to its initial position with force constant 1000 kJ mol^{-1} nm^{-2}, for a total distance of 2.6 nm. Coordinates were saved every 50 steps. Frames with protein-lipid separations in ∼0.1 nm intervals for the first 2 nm and ∼0.2 nm thereafter were extracted. These frames were used as the initial structures for the umbrella sampling windows. Values for US setup were initially based on those used in other CG-US studies (e.g. [63, 94–96]) and were subsequently checked and refined (see Section 3.3.1).

US window simulations - In each window simulation, the value of the reaction coordinate was restrained to the initial value by applying a harmonic potential with force constant 1000 kJ $moll^{-1}$ nm^{-2} to the z-coordinate of the protein COM relative to bound lipid (1') phosphate. The value of the reaction coordinate was recorded every 1000 steps. The phosphate of the 'bound' lipid was restrained with a force constant of 1000 kJ mol^{-1} nm^{-2}, and the x and y components of the protein COM restrained

relative to this bead with force constant 100 kJ mol^{-1} nm^{-2}. Windows were run initially run for 1000 ns and extended if necessary, as recorded in Table 3.3.

3.2.3 - *Analysis*

PMF profiles - 1D Potential of Mean Force (PMF) profiles were generated from US simulations using the Grossfield WHAM implementation [90]. 200 bins were used and tolerance was set to 1 x $10^{-}6$. Errors were estimated from 200 bootstraps. Correlation times for bootstrap analysis were estimated following the method used by the *g_wham* WHAM implementation [91]: autocorrelation functions were calculated using GROMACS, and correlation times estimated using Equation 2.21, carrying out the sum to a threshold of 0.05. The value of the PMF was set to 0 at bulk.

Assessment of convergence - Convergence was assessed by calculating profiles from data recorded over 200 ns segments, starting at 100 ns intervals; the 'convergence time' was judged by eye as the time after which profiles begin to fluctuate about a common value. In subsequent analysis, data before this time were discarded.

Window overlap - The % overlap between adjacent windows was calculated as the

Figure 3.2 *(following page)*: Overview of the US/CG pipeline.

Setup: *1a.* The atomistic protein structure is obtained from the PDB and converted to a coarse-grained representation, while *1b.* a coarse-grained membrane of the desired composition is created from a pre-assembed coarse-grain bilayer. *2.* The protein is positioned on the membrane by aligning the bound analogue from the PDB structure and PIP in the membrane. *3.* Water and ions are added and the system equilibrated. *4.* A steered MD simulation, pulling the protein away from the membrane, is performed and snapshots selected as initial structures for Umbrella Sampling windows.

Run: *5.* Each window is simulated and reaction coordinate data recorded. *6a.* Convergence and *7a.* overlap of the data are checked; if insufficient, *6b.* simulation time is increased or *7b.* additional windows are added, respectively. Steps *5-7* are repeated until both sufficient convergence and sampling are observed.

Analysis: *8.* a PMF profile is generated from the reaction coordinate data using WHAM. *9.* Structural data, e.g. binding angles, may also be obtained from the window trajectories, and *10.* combined with energetic information from the PMF to obtain e.g. two-dimensional energy landscapes.

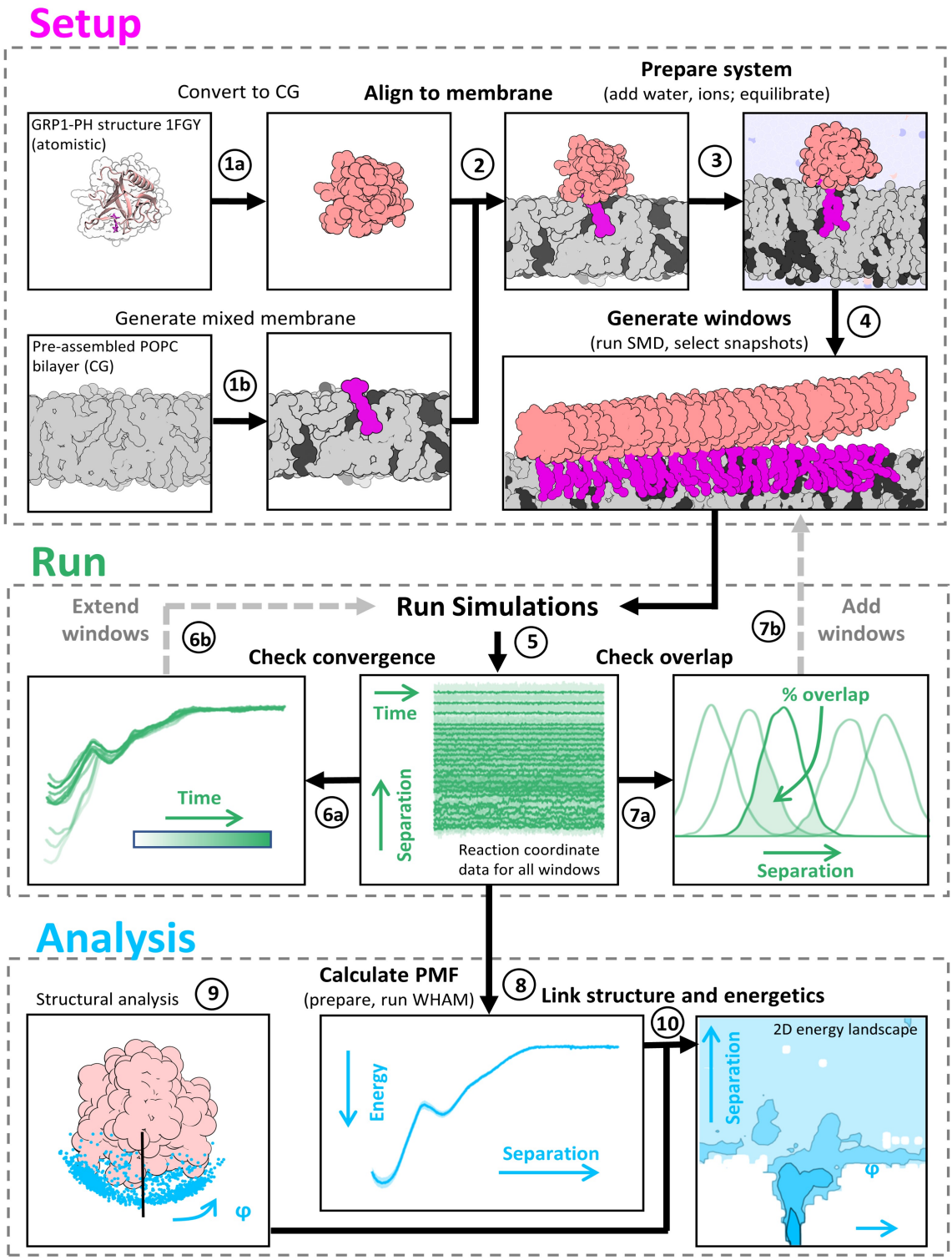


Figure 3.2: Overview of the US/CG pipeline. *Caption on previous page.*

fractional area overlap of the histograms of reaction coordinate values explored in each window.

Calculation of binding free energies and fold selectivity - Fold-selectivity was calculated from profile well depths using Equation 2.13. Absolute binding free energy for the GRP1-PH-WT/PIP_3 system was calculated from the PMF profile via Equation 2.24, using data from the 'bound' simulation to estimate the effects of xy restrains. Binding free energies were calculated for K_D values obtained in experimental studies using Equation2.12.

PIP binding angle - A reference structure was generated with the PH domain COM at the origin and the A and C sites clearly distinguished in the xy plane (see Figure 3.1 D). Backbone alignment to the PH domain was performed, and the θ and φ spherical coordinates of the COM of the lipid phosphates measured, as indicated in Figure 3.1 D. 2D energy landscapes of θ and φ vs. protein-lipid separation were then calculated using Equation 2.28, with window F-values obtained from WHAM analysis.

Membrane orientation - A backbone alignment to the initial membrane-bound structure, rotating only about the z axis, was performed. The rotation matrix of the aligned PH domain to the initial structure was calculated, and the R_{ZZ} component selected to represent the 'membrane orientation'. 2D energy landscapes of PIP binding angle vs. protein-lipid separation were then calculated using Equation 2.28, with window F-values obtained from WHAM analysis.

Contact analysis - The minimum distance between the sidechain particles of each residue and any lipid phosphate was measured for each frame; a residue was considered to form a contact with the lipid if this distance was less than 0.6 nm. The unbiased fraction of frames in contact for each residue was calculated using Equation 2.29 (assigning '1' to a frame with a given contact, and '0' without). Note that alanine and glycine do not have sidechain beads in the Martini representation, and so will always return a degree of contacts of 0 with this analysis.

Sequence alignment - Structure-based sequence alignment of PH domains was performed using the MultiSeq [116] plugin of VMD [105].

3.3 - Developing a protocol for CG/US simulations of protein-membrane interactions

An overview of the US/CG process, including set up, running and analysis is shown in Figure 3.2.

3.3.1 - *Determining suitable values for US parameters*

Accurate US calculations require thorough sampling along both the reaction coordinate and the orthogonal space. Various parameters of US, relating largely to the restraints used and location of windows, control the degree of sampling and should be chosen to ensure adequate sampling whilst also aiming to reduced computational expense. These parameters should be chosen to best fit the system in questions, and have been explored for other systems (e.g. [117]), though the parameters used across different systems are often similar (see e.g.[63, 94–96]). Values were first developed for the GRP1-PH-WT/PIP_3 system as presented here; these parameters were also found to be suitable for other GRP1-PH systems (see Section 3.4.2).

3.3.1.1 - *Window spacing, restraint and number: optimising coverage of the reaction coordinate*

To be meaningful, a PMF profile should cover a region of the reaction coordinate between the state of interest and a reference state, allowing comparison with other systems and studies. In this case, these are chosen respectively as the 'bound' state, with the protein interacting at the membrane, and the 'bulk' state, with the protein in solution and no longer able to 'see' the membrane. Windows should be chosen to fully cover this region, and with sufficient overlap - controlled by the spacing between windows and the force constant of the reaction coordinate restraint - to allow them to be accurately combined to the final PMF. There is no standard for determining an 'acceptable' level of overlap; following a previous study [117], a threshold of 5% is deemed sufficient here.

A window spacing of 0.1 nm and a restraint constant of 1000 kJ mol^{-1} nm^{-2} were chosen, based on values used in previous CG/US studies (see e.g. [63, 94–96]. Initially, 31 windows were generated with the 0.1 nm spacing (to a protein-lipid separation of

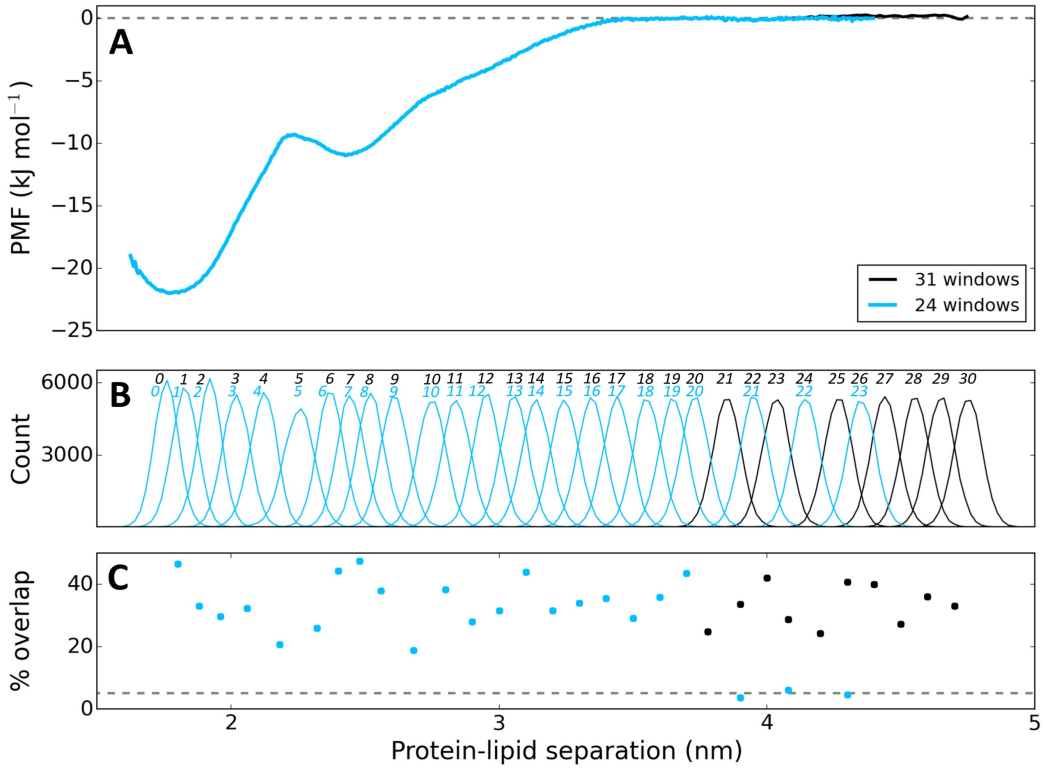


Figure 3.3: Assessing coverage of the reaction coordinate. (A) PMF profiles, **(B)** window histograms and **(C)** % overlap between windows along the reaction coordinate for the GRP1-PH-WT/PIP_3 system, using 31 windows at 0.1nm spacing (black) or 24 windows, the first 21 with 0.1nm spacing and last 2 with 0.2nm spacing (blue). Window numbering is indicated in **B**. The dotted line in **C** shows the 5% 'threshold' for window overlap.

∼4.8 nm). The resulting PMF profile clearly shows a plateau has been reached at a separation of around 3.4 nm (Figure 3.3 A), indicating the 'unbound state' has been reached. Overlap between neighbouring windows is well above the 5% threshold in all cases (Figure 3.3 C, black).

Taking every window up to a protein-lipid separation of 3.7 nm (after the plateau is reached) and a further 3 windows at 0.2 nm spacing allows the total number of windows to be reduced from 31 to 24 without affecting the profile shape (Figure 3.3 A). While the window overlap for later windows now falls near or below that of the 5% 'threshold' (Figure 3.3 C, blue), the lack of change in the profile suggests this is not an issue: given the lack of interactions, the error in the plateau region is expected to be lower and so the overlap requirement is less stringent in this region. These larger-

spaced plateau windows were run so that the plateau can be clearly seen, without too much simulation time spent on 'uninteresting' windows. Additional windows may be added for future systems if these larger spacings do not fall in the plateau region.

Based on the above observations, this set of window spacings and a 1000 kJ mol^{-1} nm^{-2} restraint were thus deemed suitable for US/CG simulations performed here.

3.3.1.2 - *Simulation time and xy protein restraint: ensuring adequate sampling of orthogonal space*

The construction of a 1D PMF from US simulations assumes thorough exploration of states orthogonal to the reaction coordinate. There will also be an initial period as the system equilibrates from the 'artificial' initial state that must be discarded to avoid biasing results. Orthogonal sampling may be improved by increasing the simulation time or reducing the sampling space using additional restraints, though these must later be accounted for. Without knowledge of the expected distribution of states, orthogonal sampling is difficult to measure directly; instead, 'convergence' of the PMF profile with time - showing no changes as additional simulation data is added - is often used. This may be assessed by calculating profiles over increasing time intervals (e.g. 0-100 ns, 0-200 ns, 0-300 ns,), though with this method it is difficult to distinguish changes in subsequent profiles due to equilibration vs. state transitions on comparable timescales to the interval length. Instead, profiles may be calculated over time slices of the same length but increasing start time (e.g. 0-200 ns, 100-300 ns, 200-400 ns); in this case, the fluctuation of profiles about a common shape will indicated convergence, while profiles generated from pre-equilibrated data will be readily distinguished.

Simulation time for each window was initially set to 1000 ns. At first, profiles were not observed to converge within this time (Figure 3.4 A). This lack of convergence was isolated to a handful of windows as the PH domain begins to dissociate from the lipid and translate in xy, parallel to the membrane (Figure 3.4 C); in these windows, there exist both bound and unbound states, with insufficient time to explore both. To avoid increasing simulation time and the associated computational expense, a lateral restraint was introduced to the PH domain center-of-mass (COM) relative to the PIP_3 lipid. The effects of such a restraint may be accounted for later using Equation 2.25.

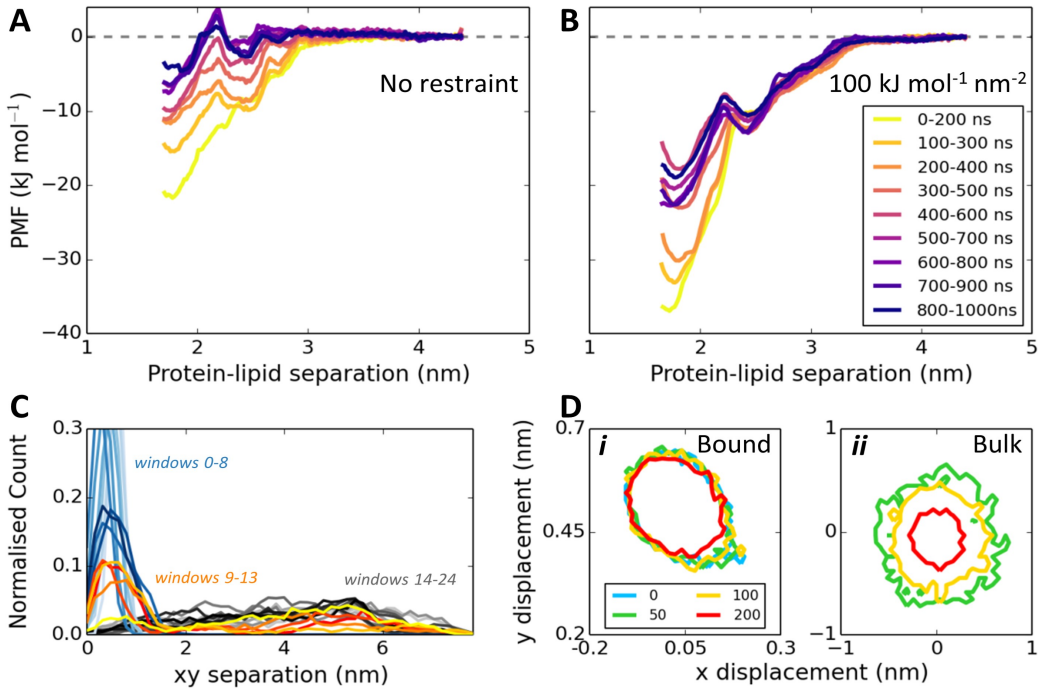


Figure 3.4: Assessing orthogonal sampling. PMF profiles calculated over adjacent time intervals with **(A)** no xy protein restraint and **(B)** a 100 kJ mol^{-1} nm-2 restraint. **(C)**Separation of the GRP1-PH domain from PIP_3 in the xy plane for each US window with no lateral restraint. **(D)**Contours showing exploration in xy space of the GRP1-PH center-of-mass with a 500 kJ mol^{-1} nm-2 (red), 100 kJ mol^{-1} nm-2 (orange), 50 kJ mol^{-1} nm-2 (green) or no (blue) lateral restraint in a *(i)* bound state window (window 0) and a *(ii)* bulk state window (window 15). The contour level is set to encapsulate every xy bin visited at least once in the simulation. No contour is seen for no restraint in bulk as all bins were visited at least once.

Ideally, the lateral restraint used should reduce the unbound xy space to be sampled but not affect the exploration of bound states. Several values for the force constant were tested on windows representing the bound (in the well) and bulk (plateau region) states. The reduction in sampling space with increasing force constant is clearly seen in the bulk state (Figure 3.4 D *ii*). In the bound state, restraining by up to 100 kJ mol^{-1} nm^{-2} showed negligible effect on xy distribution, while a slight perturbation is observed with 500 kJ mol^{-1} nm^{-2} (Figure 3.4 D *i*).

Based on these observations, it was decided to use a 100 kJ mol^{-1} nm^{-2} lateral restraint on the PH domain relative to the bound lipid in subsequent simulations.

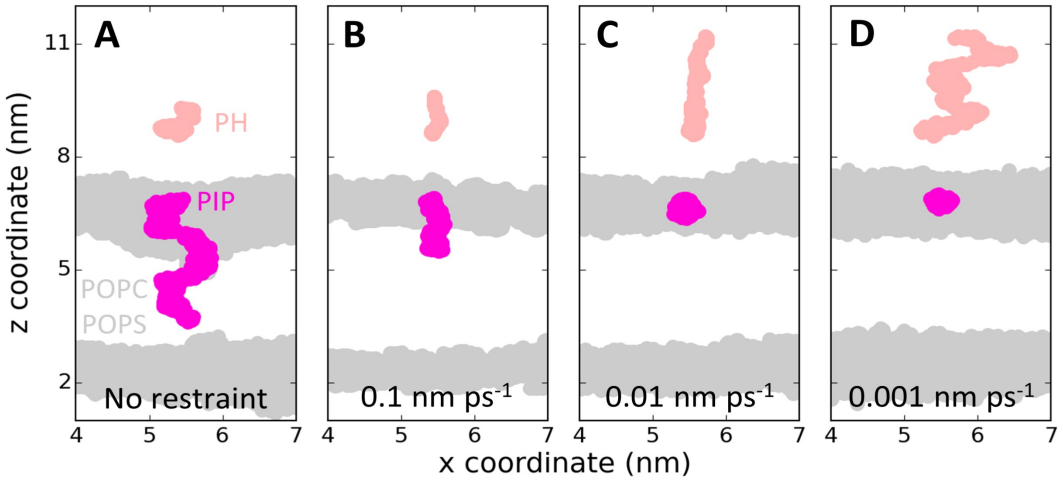


Figure 3.5: Testing of steered MD simulations. xz positions of the PIP_3 1' phosphate (magenta), GRP1-PH-WT center-of-mass (light pink) and POPC and POPS phosphates (grey) throughout steered MD simulations with **(A)** no z restraint on PIP_3 and pulling speed 0.01 nm ps^{-1}, or a 1000 kJ mol^{-1} nm^{-2} restraint on the 1' phosphate of PIP_3 and a pulling speed of **(B)** 0.1, **(C)** 0.01 or **(D)** 0.001 nm ps^{-1}. Note the use of the x coordinate was an arbitrary choice; any projection about the z axis is expected to show similar results.

For the GRP1-PH-WT/PIP_3 system, US performed with this restraint indeed shows convergence within the original 1000 ns (Figure 3.4 B). This simulation time may later be increased for other systems if satisfactory convergence is not demonstrated.

3.3.1.3 - *The steered MD simulation: efficient generation of windows*

When choosing initial structures for each window simulation, the use of a near-equilibrated structure at the desired reaction coordinate value will reduce the time spent generating data that will be discarded as part of the equilibration time. Steered MD (SMD) simulations are often used to obtain initial snapshots near chosen reaction coordinate values. A faster pulling speed allows snapshots to be obtained quicker but may result structures that are further from equilibrium.

SMD simulations were used to pull the GRP1-PH domain away from the membrane. Pulling speeds of 0.001, 0.01 and 0.1 nm ps^{-1} were tested, saving coordinates every 500, 50, and 5 steps, respectively - approximately one frame per 0.01 nm (assuming movement of GRP1-PH is constant), to allow frames close to the desired 0.1 nm

spacing to be identified. Initially, the PIP_3 molecule was observed to displace instead of GRP1-PH (Figure 3.5 A). To prevent this in all further SMD simulations, a 1000 kJ mol^{-1} nm^{-2} restraint was applied to the 1' phosphate bead. Despite this restraint, a pulling speed of 0.1 nm ps^{-1} still caused displacement of the PIP_3 lipid (Figure 3.5 B), while pulling speeds of 0.01 and 0.001 nm ps^{-1} both enabled selection of snapshots within ~0.01 nm of the target 0.1 nm spacing, without significant movement of PIP_3 relative to the membrane (Figure 3.4 C, D). In the interest of reducing computational time, the faster 0.01 nm ps^{-1} was used for subsequent systems.

3.5 demonstrates the fluctuations in the z position of membrane lipids on simulation timescales. PIP_3 is expected to have the greatest effect of GRP1-PH binding. The reaction coordinate was thus chosen using PIP_3 as a reference, rather than the wider membrane, to avoid possible complications as the relative movement of PIP_3 changes without change in average membrane position. Furthermore, longer simulation times may be required if PIP_3 is free to explore z; to reduce this possible effect, the above restraint on PIP_3 1' was maintained in window simulations. This was later shown to have minimal effect (see Section 3.5.1.3).

3.3.2 - *Combining structural and energetic analysis*

Reaction coordinate data from US simulations is typically unbiased and combined to calculate the final PMF profile using WHAM [73]. Several implementations of WHAM exist; among the more widely available are those that come with the simulation packages GROMACS (known as *g_wham* [82]) and CHARMM [75], and a standalone version written by Alan Grossfield [90]. These are, however, run from the command line and/or first require the creation of specific input files, making their integration into flexible simulation/analysis pipelines difficult. Further, it may be useful to perform WHAM energetic analysis directly alongside structural analysis, for example to project the free energy landscape along other coordinates or calculate ensemble averages as the reaction coordinate changes as discussed in Section 2.2.2.3.

Writing code for structural analysis, particularly when pipelining, may be achieved with MDAnalysis [118, 119], a Python library providing tools for loading MD trajectories and performing common analysis tasks. MDAnalysis initially lacked capabilities

for dealing with free energy calculations in US simulations. Integration of WHAM-based analysis required a way to run WHAM from Python, and being able to gather and use both the input and output data for WHAM from within the MDAnalysis framework. This was achieved through a combination of direct additions to the MDAnalysis library and construction of separate modules/scripts, as outlined below.

3.3.2.1 - *A Python wrapper for the Grossfield WHAM*

Rather than creating a new WHAM implementation, I decided to write a 'wrapper' to launch an existing version from Python. The Grossfield WHAM was chosen as it is not associated with a specific simulation package and reports the final 'F-values', the energy associated with each window due to the reaction coordinate restraint, that may be directly used to calculate ensemble averages and projections of the profile along other coordinates through Equations 2.28 and 2.29.

A Python module was written providing functions to, given the reaction coordinate time series and required metadata (namely, the temperature, location and force constant) for each window, generate input files of the appropriate format and run the Grossfield WHAM. The input data may be provided directly, or as a list of the relevant GROMACS file names, from which the information will then be extracted. Other required input - such as the start and end values of the reaction coordinate, whether to run bootstrap error analysis and how many bootstraps to run - may optionally be specified, otherwise default values are used. In the case of start/end reaction coordinate values, these can be automatically determined from the timeseries data. Additional options, such as selecting a time range, have also been added. For more details on use and availability, see Appendix A.

3.3.2.2 - *Free energy simulations and MDAnalysis*

Launching WHAM with MDAnalysis is not a straightforward task as much of the data required for WHAM is not currently read when a trajectory is loaded by MDAnalysis. This includes the timeseries of reaction coordinate values: while these could be calculated for all frames in the trajectory using MDAnalysis, it is common to save reaction coordinate values to a separate file at a higher frequency while running a US

simulation, allowing more data for running WHAM without the prohibitive file sizes that would result from saving the full set of coordinates. The ability to read external timeseries data such as these reaction coordinate values alongside a trajectory could in general be very useful. To this end, I developed a new 'auxiliary' module for MDAnalysis. External (or 'auxiliary') timeseries data may be provided directly or as a filename, along with the corresponding simulation times (specifying the appropriate column of the input data or the timestep and start time); this module will then allow the auxiliary data to be read and updated as trajectory frames are updated, with the value corresponding to the current time stored alongside atom coordinates and other frame data. Options exist to store the value from the closest timestep if there is no exact match, or an average value if the auxiliary data is more frequent. This module is now included with MDAnalysis from version 0.16.0 [119].

Simulation metadata is also not read by MDAnalysis, and currently must be passed to the WHAM wrapper outlined in Section 3.3.2.1 either directly or via the appropriate GROMACS *tpr* files. To simplify the task of associating each MDAnalysis window trajectory with the corresponding metadata, a module was written making use of MDSynthesis [120], a Python library providing an interface to MDAnalysis and allowing persistent storage of associated data, to group the trajectories and store metadata in named fields. Once calculated via WHAM, the F-value of each window is also stored; additional functions may then be used to compute averages and projections along other coordinates via Equations 2.28 and 2.29, using any frame-based MDAnalysis structural analysis. See Appendix A for more details.

3.4 - Results and discussion: membrane binding of the GRP1-PH domain

3.4.1 - *Checking the initial structure*

A good initial structure of GRP1-PH binding PIP_3 will reduce the MD simulation time that must be discarded as equilibration, and minimise potential bias, particularly given the presence of the protein xy restraints. The PDB structure of GRP1-PH with IP_4 bound allowed an initial guess to be made (Figure 3.6 A). Given this structure was determined independent of a membrane environment [20], the initial mode was first checked by performing simulations in which GRP1-PH was unrestrained and

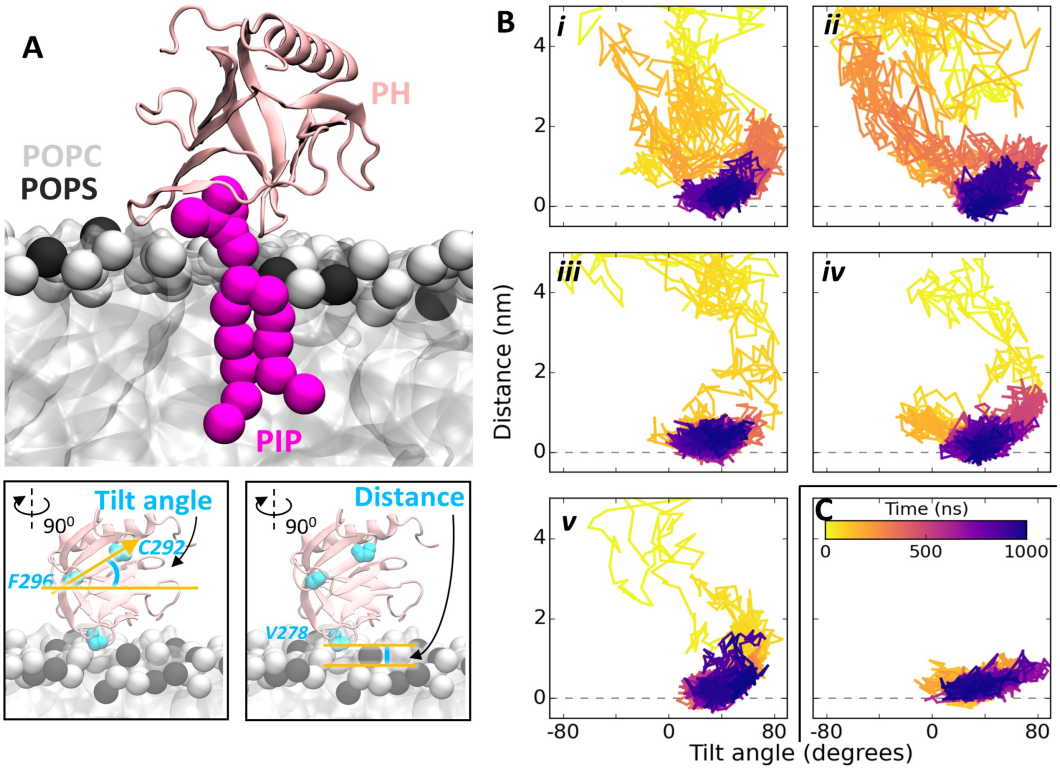


Figure 3.6: Initial structure of GRP1-PH bound to the membrane. **(A)** The initial coarse-grain GRP1-PH/PIP_3 binding structure generated from alignment of the crystal structure (Figure 3.1 A) with PIP_3 in the membrane. **(B)** Evolution of orientation and separation of GRP1-PH relative to the membrane in each of five binding simulations with different starting orientations. **(C)** Evolution of orientation and separation of GRP1-PH relative to the membrane in the unrestrained bound simulation.
Orientation is measured as the angle of the vector defined by residues C292 and F296 to the membrane surface, and separation as the perpendicular distance of V278 from the average lipid phosphate plane (see inset).
GRP1-PH is shown by the cartoon representation of the aligned crystal structure (light pink). PIP_3 is shown with van der Waals representation (magenta). The membrane is shown by a surface representation (white), with POPC and POPS phosphates shown explicitly in can der Waals representation (light and dark grey, respectively). Water and ions are omitted for clarity.

initially displaced from the membrane, allowing 'natural' membrane binding to be observed. A higher PIP_3 concentration (10 PIP_3 molecules per leaflet) was used for these simulations to ensure membrane binding.

This initial structure was in good agreement with previous experimentally and computationally derived models [65, 66, 101]. Following definitions used in an experimentally

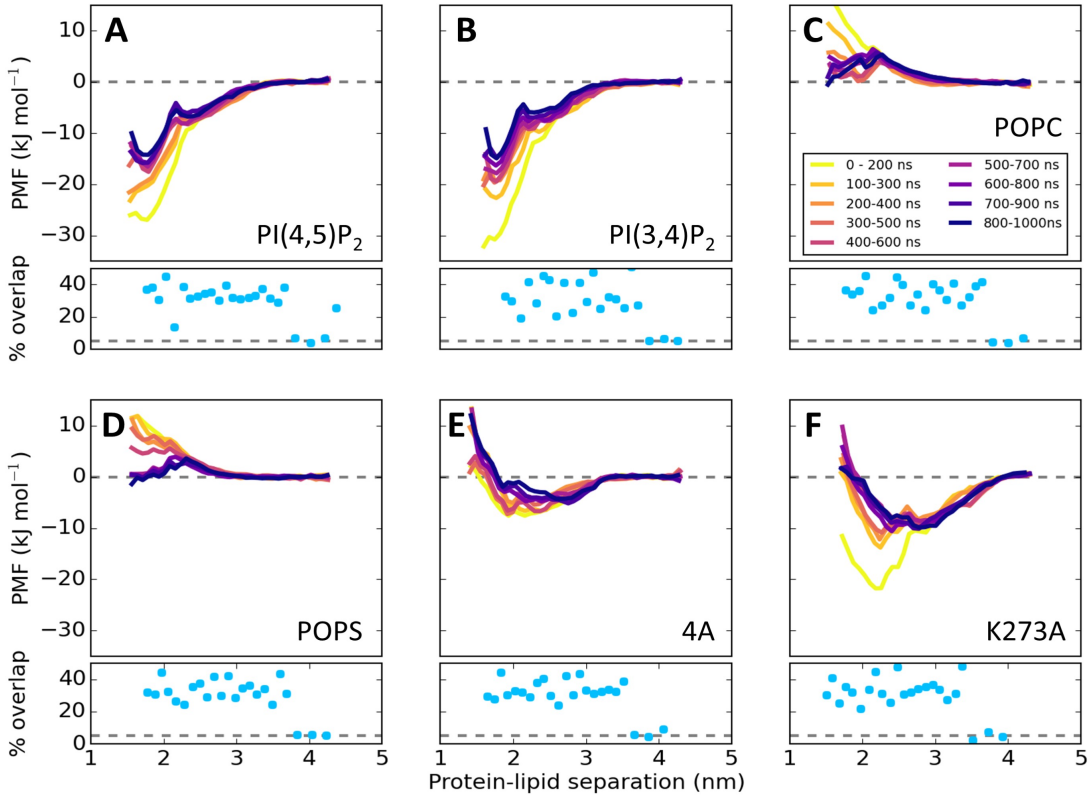


Figure 3.7: Demonstration of convergence and sufficient sampling. PMF profiles calculated at increasing time intervals and % reaction coordinate overlap between adjacent windows for **(A)** GRP1-PH-WT/PI(4,5)P_2, **(B)** GRP1-PH-WT/PI(3,4)P_2, **(C)** GRP1-PH-WT/POPC, **(D)** GRP1-PH-WT/POPS, **(E)** GRP1-PH-4A/PIP_3 and **(F)** GRP1-PH-K273A/PIP_3 US systems. Similar calculations for the GRP1-PH-WT/PIP_3 system are shown in Figures 3.3 C and 3.4 B.

driven structural study [101], the orientation (angle relative to the bilayer, measured from the vector joining residues C292 and F296 to the membrane surface) and distance (penetration into the bilayer (the distance of V278 below the average lipid phosphate plane) of the PH domain was measured. These values for the initial structure were 47° and 0.18 nm, respectively, within the ranges 46 ± 7° and 0.24 ± 0.2 nm determined by site-directed spin labelling [101]. Tracing the distance and orientation relative to the membrane throughout binding simulations indicates a common, stable binding pose matching this structure is reached in each of five repeats (Figure 3.6 B). Tracing distance and orientation throughout the unrestrained 'bound' simulation explored further suggests the stability of this interaction mode (Figure 3.6 C).

3.4.2 - *Results from US/CG*

Experimental studies in membrane environments have shown that GRP1-PH binds preferably to PIP_3, with a lower affinity for other PIPs including $PI(3,4)P_2$ and $PI(4,5)P_2$, and no binding seen in the absence of PIPs (see Table 3.2 and references therein). To explore the membrane binding of GRP1-PH and test if the US/CG approach can distinguish binding to different PIPs, US/CG simulations were performed as outlined in Section 3.2 for GRP1-PH binding PIP_3, $PI(3,4)P_2$, and $PI(4,5)P_2$, as well as the non-PIP lipids POPC and POPS. Note that in the coarse-grained model used, $PI(3,4)P_2$ and $PI(4,5)P_2$ are identical; only the initial orientation will differ between these systems. The effects of modifications to the GRP1-PH domain were also investigated by performing additional US/CG simulations with a '4A' mutant in which four positive residues key to PIP_3 binding (K273, R284, R305 and K343; see Figure 3.1 B) were mutated to alanine, and with a single K273A mutation known experimentally to eliminate binding [110]. Convergence and sufficient overlap was observed for all systems using the parameters initially developed for GRP1-PH/PIP_3 (Figure 3.7). All simulations were analysed structurally and energetically as outlined in Section 3.2.3 (Figure 3.8).

Figure 3.8 *(following page)*: Structural and energetic analysis of GRP1-PH/membrane binding. Results from US simulations of GRP1-PH-WT binding to **(A)** PIP_3, **(B)** $PI(4,5)P_2$, **(C)** $PI(3,4)P_2$, **(D)** POPC **(E)** POPS, and **(F)** GRP1-PH-4A and **(G)** GRP1-PH-K273A mutants binding PIP_2. (i) 1D PMF profiles along protein-lipid separation, showing bootstrap errors. *(ii-iii)* 2D energy landscapes of protein-lipid separation against PIP-binding angle. *(iv)* 2D energy landscape of protein-lipid separation against membrane orientation. *(v)* Average degree of contacts for each residue along protein-lipid separation. The inset shows calculation of binding angle, orientation (R_{ZZ}) and the locations of residues showing a high degree of contacts, coloured according to position at the C site (green) or A site (purple). These figures are reproduced larger in Figure 3.1. Vertical dotted lines indicated the location of wells in each profile. Binding modes identified from PIP angle and orientation are labelled in orange; representative snapshots are shown in Figure 3.9.

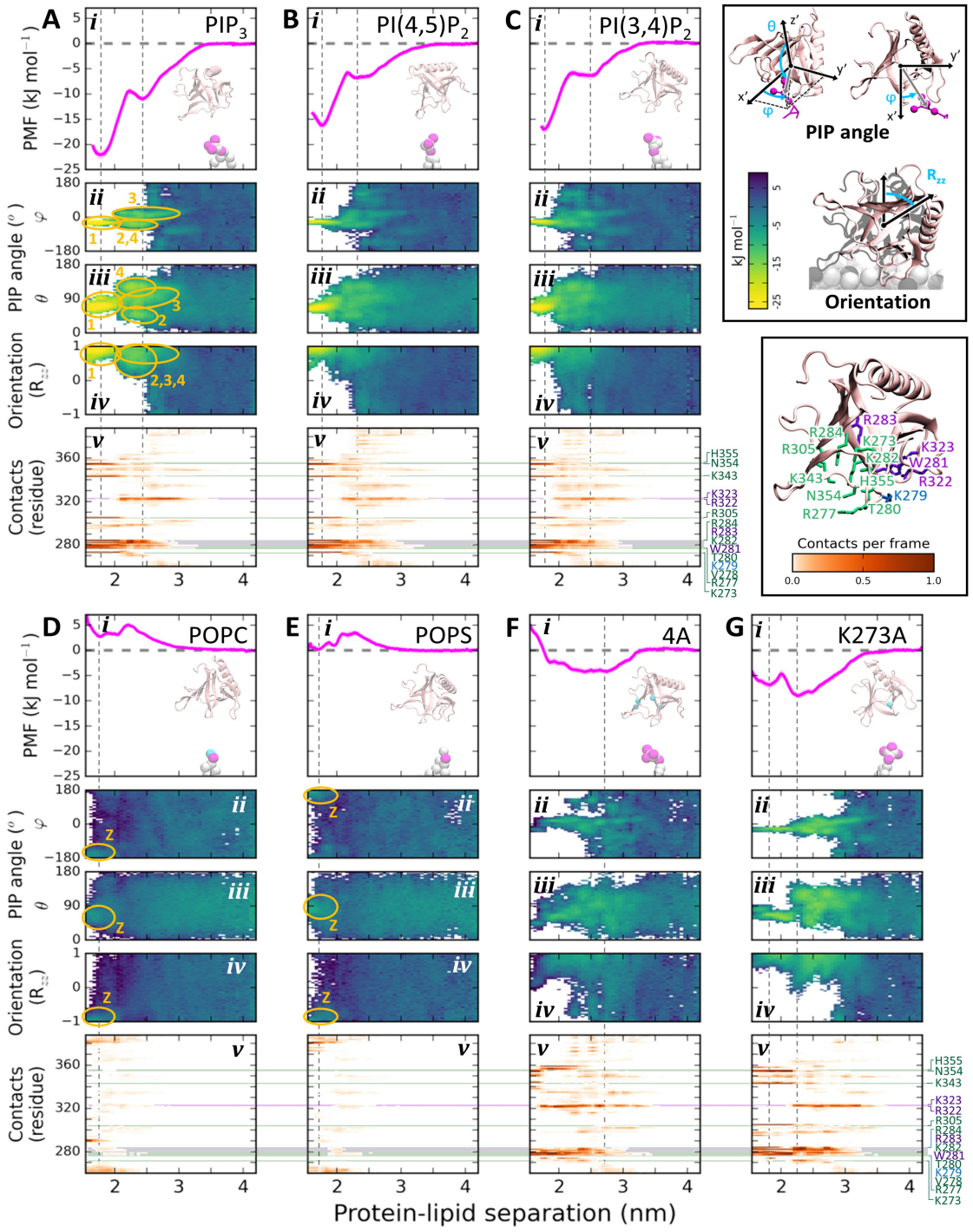


Figure 3.8: Structural and energetic analysis of GRP-PH/Membrane binding. *Caption on previous page.*

3.4.2.1 - *Energetic analysis: reproduction of PIP_3 selectivity and negative effect of C-site mutations*

Comparison of PMF profiles reveals that GRP1-PH-WT binds most favourably to PIP_3, with a well depth of -22.3 ± 0.6 kJ mol^{-1} (Figure 3.8 A *i*), followed by $PI(3,4)P_2$/$PI(4,5)P_2$ (well depths -16.4 ± 0.5 and -16.8 ± 0.5 kJ mol^{-1}, respectively; Figure 3.8 B-C *i*). Neither POPC nor POPS show favourable binding, with the membrane-associated state showing no change or a higher PMF compared to bulk, respectively (Figure 3.8 D-E *i*). The GRP1-PH K283A mutant demonstrated reduced affinity for PIP_3 (well depth -9.1 ± 0.6 kJ mol^{-1}) (Figure 3.8 G *i*). These results are qualitatively consistent with those seen experimentally (Table 3.2). The GRP1-4A mutant also showed a significantly reduced energy minima (well depth -4.3 ± 0.6 kJ mol^{-1}; Figure 3.8 F *i*), as expected.

To allow qualitative comparison with previous results, the absolute binding free energy for the GRP1-PH-WT/PIP_3 system was calculated from the PMF via Equation 2.24, yielding a final standard binding free energy of 13.4 kJ mol^{-1}. Unrestrained *xy* data for determining the effect of the *xy* protein restraint on the bound state were obtained from an additional simulation performed as outlined in Section 3.2. The *xy* restraint was chosen to have minimal effect on the bound state (see Section 3.3.1.2. Calculating the individual terms in Equation 2.24, the bound-state correction term is indeed low (-1.3 kJ mol^{-1}); the largest correction to the binding energy comes from the effective volume sampled given the *xy* restraint compared to standard volume (10.0 kJ mol^{-1}). Assuming a similar, flat, correction will apply to the PIP_2 systems, the fold-selectivity of GRP1-PH for PIP_3 over PIP_2 can be estimated using Equation 2.13 and the respective well depths, yielding ∼10 fold selectivity for PIP_3. This is similar to the range seen experimentally (∼50-200 and ∼5-200 for PIP_3 over $PI(4,5)P_2$ and $PI(3,4)P_2$ respectively; see Table 3.2 and references therein).

The binding free energy for GRP1-PH binding PIP_3 is somewhat below the experimental range of ∼35-46 kJ mol^{-1} (Table 3.2); however, it is worth noting the high variation within experimental results, with dependence on system conditions including as pH, temperature and background lipid composition demonstrated - there is no one 'true' answer. Particularly, there is only one PIP lipid available for GRP1-PH to bind

in the simulations performed here, whereas a PH domain in an experimental system is likely to encounter more than one PIP molecule. Binding to multiple PIPs is expected to significantly increase the binding energy, towards the range seen experimentally. This is examined further in Section 3.4.2.4.

Examining the profile for binding of GRP1-PH-WT to PIP_3, a second, shallower well (depth -11.6 $\pm$ 0.6 kJ mol^{-1}) is observed at a separation of 2.4 nm, in addition to the initial well at 1.8 nm (Figure 3.8 A *i*). This suggests the presence of at least two binding modes, with the one at the lower separation being more energetically favourable. Evidence of this second mode is also seen for PIP_2, though here it forms an even shallower well (Figure 3.8 B-C *i*), suggesting the mode is relatively less favoured than for PIP_3. Interestingly, it is primarily the initial well that is perturbed by the 4A and K273A mutations, having vanished for the former, and, while retained for the later, is significantly reduced to -6.7 kJ mol^{-1}. In both cases, the second well is now the more favourable (Figure 3.8 F-G *i*). This suggests that while the primary binding mode is associated with the canonical binding site and thus affected by the mutations studied here, a secondary mode exists involving association of PIP with a different region of the protein. Further details of this secondary mode may be obtained through structural analysis.

3.4.2.2 - *Structural analysis: observation of both A- and C-site binding modes*

Structural details of the interaction of the GRP1-PH domain were investigated using three measures outlined in Section 3.2.3 (Figure 3.8 *ii-v*): to identify the site of PIP-interaction, the position of the PIP (or POPC/POPS) on the PH domain surface was measured in spherical coordinates, defined such that the A and C site interactions may be readily distinguished by azimuthal angle φ, corresponding to positive ($\sim$0-60°) and negative (-60-0°) angles, respectively (Figure 3.1 D). Binding modes were further characterised by measuring the orientation of the domain on the membrane relative to the initial structure (Figure 3.1 E); the initial structure will have an R_{ZZ} of 1, while an 'upside down' orientation will have R_{ZZ} -1. Lastly, to assist identification of binding sites and to begin considering how sequence determines the affinity and selectivity of a PH domain, contacts between each residue sidechain and the PIP (or POPC/POPS) phosphates were counted and averaged over all simulation frames.

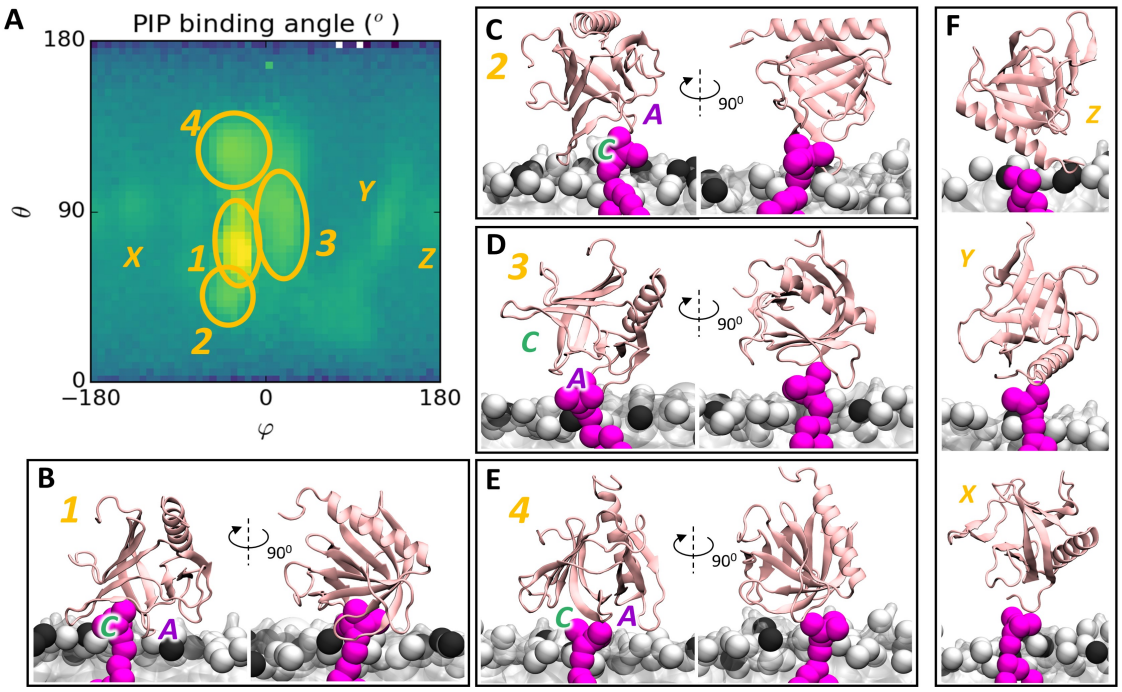


Figure 3.9: GRP1-PH/PIP interaction modes. (A) Energy landscape showing the locations of PIP (or POPC/POPS) interactions with the GRP1-PH in spherical (θ and φ) coordinates (see Figure 3.1 D), averaged over all systems show in Figure 3.8. The four main binding modes are circled and representative snapshots shown in (**B-E**). The locations of the A and C sites are labelled in each snapshot. Additional upside down modes are also indicated and representative snapshots shown in **(F)**.

The PH domain is show in light pink as the cartoon representation of the atomistic crystal structure aligned to the coarse-grained domain; PIP (or for mode Z, POPC) is show in van der Waals representation in magenta. The membrane is shown be surface representation in grey, with phosphate beads of POPC and POPS are shown explicitly in light and dark grey, respectively. Water and ions are omitted for clarity.

For the GRP1-PH-WT/PIP_3 system, 2D landscapes of protein-lipid separation vs. PIP binding angle and membrane orientation indicate the presence of one primary binding mode (mode '1') at protein-lipid separation ~1.5-2 nm, corresponding to the first, deeper well (Figure 3.8 A *ii-iv*). As expected, this mode corresponds to the initial structure, with a PIP-binding angle of around -45° to -20° (i.e. in the C-site) and membrane orientation of 0.8-1; a representative snapshot is shown in Figure 3.9 B. The residues showing the most PIP-contacts are K273, R277, K279, T280, K282, R284, R305, K343, N354 and H355, located in the C-site of the GRP1-PH domain

(green in Figure 3.8 inset; K279, shown in blue, appears to extend towards the A-site in the crystal structure, but given its location on a flexible loop is likely to be able to orient to either site). These contacts are consistent with those observed with IP_4 in the crystal structure [20] and suggested by other experimental and computational studies (see Table 3.1 and references therein).

Three secondary binding modes are observed from a protein-lipid separation of ∼2.1 to 2.5 nm or greater, corresponding to the second, shallower well. These modes are clearly distinguished by θ binding angle (angles of between ∼40 and 60° correspond to mode 2, between ∼75 and 115° to mode 3, and ∼110 to 140° to mode 4; Figure 3.8 A *iii*), but collapse along φ and R_{ZZ} to show only two or one distinct modes, respectively (Figure 3.8 A *iii-iv*). This demonstrates the importance of considering multiple coordinates to avoid 'missing' interaction modes when projecting a complicated, multi-dimensional interaction along a single coordinate.

The correspondence between binding angles for each mode is more clearly shown by plotting the 2D energy landscape of θ vs. φ (Figure 3.9 A; averaged over all 7 systems). Mode '2' shows a similar range of φ angles (-50 to -25°) and slightly lower θ angle (40 to 60°) to mode '1', also corresponding to binding at the C-site. Several of the C-site residues seen to interact with PIP_3 in mode '1' again display a high degree of contacts here (Figure 3.8 A *v*). These can be separated in to those towards the 'front' of the C-site (R277, K279, K343 and N354) and those towards the back (K282 and R284); fewer contacts are seen with those residues in the 'middle' (Figure 3.1 C). The lower θ suggests PIP_3 binds further 'forward' in the C-site in mode '2' compared to mode '1', contributing the 'forward' set of contacts. A representative structure is shown in Figure 3.9 C.

In addition to the expected C-site residues, contacts are also seen with the residues W281, R283, R322 and R323. These are located within the A-site (purple in Figure 3.8 inset), along with K279 (mentioned above to appear free to orient toward either binding site). Indeed, an interaction mode (mode '3') is identified with φ in the range 0-30°, corresponding to the A site (Figure 3.8 A *ii*; representative structure in Figure 3.9 D). This mode matches A-site binding modes observed in PDB structures for

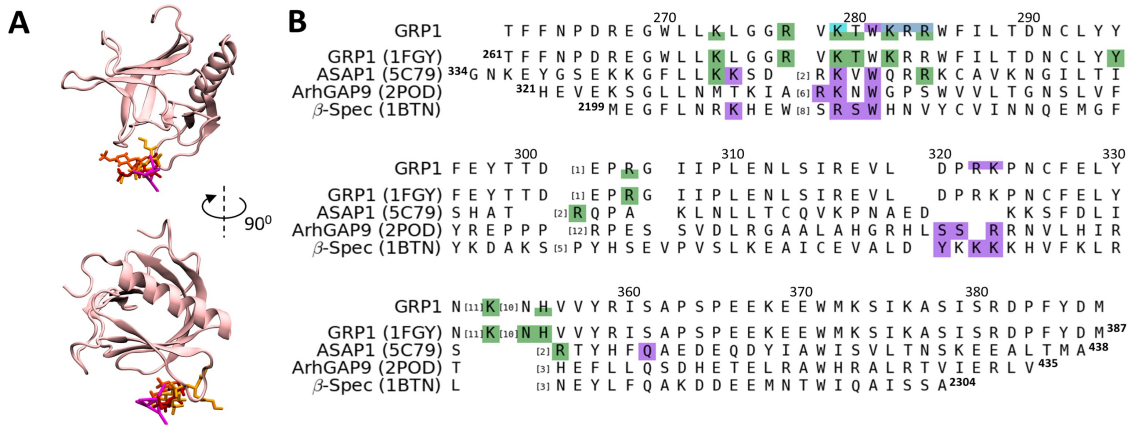


Figure 3.10: A-site interaction of GRP1-PH. (A) Relative location in the A site of PIP binding to GRP1-PH, from a representative snapshot of mode 3 (magenta), and PIP analogues from PDB structures of the ArhGAP9 (red; PDB ID 2POD), β-spectrin (orange; 1BTN) and ASAP1 (yellow; 5C79). Backbone alignment of the PH domains was performed; only the GRP1-PH domain is shown (light pink). **(B)**Structure-based sequence alignment of the PH domains of GRP1, ASAP1, ArhGAP9 and β-spectrin. Residues found to contact PIP analogues bound in the C (green) or A (purple) sites in the indicated PDB structures are indicated in the lower four rows; the top row indicates contacts observed in US simulations at close protein lipid separation (i.e. mode 1, green; lower half) or larger protein lipid separations (i.e. mode 2, green; mode 3, purple; both mode 2 and 3, light blue; and mode 4, dark blue; upper half).

three other PH domains [121–123], with equivalent residues of all but R283 observed to make contacts with the PIP analogues in these structures (3.10).

The final interaction mode, mode '4', initially appears to correspond to the C-site, with a φ angle of -50 to -10°; however, as indicated by the much higher θ (110-140°), PIP_3 in this mode is located further 'back' on the protein, and in fact binds behind the C site, in an extension of the A site (Figure 3.9 E). This explains the contacts observed with residues at the back of both the C site (K282 and R284) and the A site (R283).

While the presence of a secondary interaction site for anionic lipids has been suggested for GRP1-PH [98], PIP binding outside of the canonical site has not previously been demonstrated. The existence of these A site modes confirms the suggestion of a non-

canonical binding mode at larger protein-lipid separation based on the PMF profiles, as discussed in Section 3.4.2.1.

The same binding modes are observed for the interaction of GRP1-PH-WT with PI(3,4)P_2/PI(4,5)P_2 (Figure 3.8 B-C *ii-v*). The initial mode '1' is noticeably reduced for the K273A mutant, showing a lower energy in 2D binding landscapes, while the secondary modes 2-4 are remain at a similar energy (Figure 3.8 G *ii-iii*). For the 4A mutant, mode '1' is absent, both from orientation and angle landscapes and the pattern of contacts, instead, the A-site interaction mode appears from a protein-lipid separation of around 1.6 nm (Figure 3.8 F *ii-v*). These observations are consistent with relative depths of the two wells in the PMF profiles, compared to that of the WT GRP1 with PIP_3, as discussed at the end of Section 3.4.2.1.

When PIP is absent, GRP1-PH appears to interact with POPC and POPS preferably in an 'upside down' orientation, with R_{ZZ} and φ binding angle of -1 and 180°), respectively (Figure 3.8 D-E; representative snapshot shown in Figure 3.9 F), though this mode is not favourable compared to bulk. This may not represent a physiological membrane-interaction mode when considering the full-length GRP1 protein, but nonetheless demonstrates the lack of favourable binding in the absence of PIP.

3.4.2.3 - *Determinants of binding mode and affinity*

As suggested by previous experimental studies (e.g. [20, 111]), the high affinity of GRP1-PH for PIP_3 is likely due to the nature of the C site: closed in on three sides by the β1-β2, β3-β4 and β1'-β2' loops (the latter being an insert not found in most other PH domains). Each loop contains multiple positive or polar residues, creating an ideal cavity for PIP binding (see Figure 3.1). PIP_2, possessing fewer phosphate groups and a lower overall charge, is not able to maximise contacts with these residues and so displays a lower affinity. This is seen in Figure 3.8, with the primary binding mode showing lower energetic favourability for PI(3,4)P_2/PI(4,5)P_2 despite the degree of PIP binding for each residue being similar. If multiple contacts between each residue and PIP were considered, PIP_3 would be expected to show a higher level of contacts. The larger number of contacts formed within the C-site also

allows the primary interaction mode '1' to be retained with the mutation of a single key residue (K273A), though with a lower binding energy.

Meanwhile, the A site is shallower and offers less positively charged or polar residues (see Figure 3.1). Interactions with the A site (mode '3' and, to a lesser extent, mode '4') are therefore less favourable and unable to compete with 'mode 1' at lower protein-lipid separations, though the 4A mutant demonstrates that transitions to the A site are possible at these separations. As protein-lipid separation is increased, PIP is no longer able to bind completely in the C site cavity, instead sitting towards the front of the site and interacting with fewer residues (mode '2'); modes 3 and 4, now comparable in favourability, are also seen. These secondary modes appear more strong for PIP_3 compared to $PI(3,4)P_2$/$PI(4,5)P_2$; this may again be due to the increased charge of PIP_3.

None of the three secondary modes involve direct contacts with K273 and are thus minimally perturbed by the K273A mutation. However, all become energetically less favourable with the '4A' mutation; modes 2 and 4 due to direct loss of contacts (K343 and R284, respectively), and the A site likely indirectly through its proximity to the mutations, which are likely to substantially shift the electrostatic surface of the PH domain.

POPC and POPS, having smaller, lower charged headgroups, are less able to access the C site cavity or the longer-range C/A interactions, instead binding 'upside down' and interacting with the more accessible C and N termini (though, as mentioned above, this may not represent a physiological binding mode).

3.4.2.4 - *Binding of multiple PIPs?*

It is noted that the A and C sites are located on the same face of the PH domain, and interaction modes involving each have similar orientation on the membrane (Figure 3.8). This provides further weight to the idea that the GRP1-PH domain (and other PH domains) may simultaneously bind multiple PIP lipids. Indeed, further analysis of the unrestrained binding simulations shows that on average 3-5 PIP_3 molecules (of the ten per leaflet) are in contact with (within 0.6 nm of) the PH domain while bound (Figure 3.11 C). The density of PIP_3 relative to GRP1-PH shows several

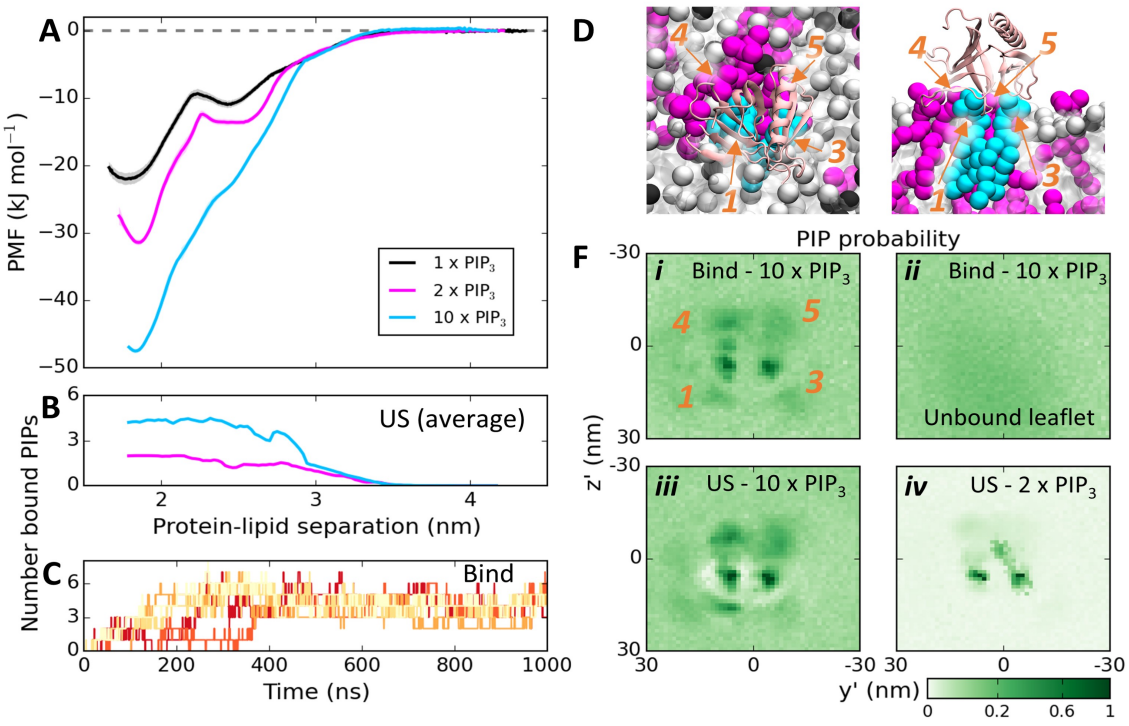


Figure 3.11: Multiple PIP binding by GRP1-PH. (A) PMF profiles for GRP1-PH binding PIP_3 in membranes containing a total of one (black, reproduced from Figure 3.8 A *i*), 2 (pink) or 10 (blue) PIP_3 molecules in the bound leaflet. **(B)** Average number of PIP interacting with (within 0.6 nm of) GRP1-PH along protein-lipid separation, unbiased for reaction coordinate, from 2 × PIP_3 (pink) and 10 × PIP_3 (blue) US simulations. **(C)** Number of PIPs interacting with (within 0.6 nm of) GRP1-PH with simulation time for each of the five binding simulation repeats. **(D)** Initial structures used for multi-PIP US-simulations. PIPs binding in sites equivalent to those shown in Figure 3.9 are labelled with the corresponding mode. **(E)** Density of PIP_3 along in the $y'z'$ plane following alignment to the reference structure shown in Figure 3.1 D, for the *(i)* bound or *(ii)* unbound leaflets, averaged over all 10 x PIP_3 binding simulations, and from US simulations with *(iii)* 10 × PIP_3 and *(iv)* 2 × PIP_3, unbiased for reaction coordinate restraints. Hotspots are labelled with the equivalent binding mode (see Figure 3.9).

The PH domain is shown in light pink cartoon representation of the aligned atomistic structure; PIPs are shown in van der Waal representation, with the two PIPs common to both systems shown in blue, and those present only in the 10 x PIP_3 system in pink. POPC and POPS phosphates are shown in light and dark grey, respectively, and the membrane is represented by a transparent white surface.

hotspots indicating the positions of interaction sites (Figure 3.11 E *i*). The y' and z' coordinates of PIPs were used in place of the θ and φ angles used previously, in order

to separate unbound lipid noise from actual bound lipids. The axes are presented such that modes around (0°, 90°) in (φ, θ), or equivalently (0, 0) in (y', z') will occupy similar positions in the respective landscapes. This covers all main binding modes identified in Figure 3.9. Good agreement with single-bound PIP US simulations is observed, with two main interaction sites corresponding to mode 1 (C site) and mode 3 (A site); a third site corresponds to mode 4 (Figure 3.11 E *i*, Figure 3.9 A). Several additional, lower probability interaction sites are also present. One of these - in which a lipid is bound in the initial structure - is labelled as '5' (Figure 3.11 D, E).

To quantify the effects of multiple-lipid binding, additional US simulations were performed with multiple PIP_3 molecules in the bound leaflet. An initial structure was chosen from visual inspection of binding simulations, with a PIP_3 molecule bound in each of the three main sites identified in Figure 3.11 E *i*. US simulations were performed for a 10 × PIP_3 system, using this initial structure directly, and a 2 × PIP_3 system, in which all PIPs but those in the A and C sites were replaced with POPC. In each case, the PIP initially located at the C site was restrained and used as the reference for the reaction coordinate and GRP1-PH restraints, while other PIP_3 lipids were free to diffuse. This increased the sampling space, such that longer simulation times were required to ensure convergence was reached (Figure 3.12 AB). The presence of multiple PIP lipids resulted in a significant increase in well depth to -31.4 ± 0.5 kJ mol^{-1} for 2 × PIP_3 (an increase of 9.8 kJ mol^{-1}, approximately the depth of the secondary well in the GRP1-PH-WT/PIP_3 profile) and -47.5 ± 0.3 kJ mol^{-1} for 10 × PIP_3. Assuming the correction due to the xy protein restraints will be similar to that calculated for GRP1-PH-WT/PIP_3 above (an increase of around 10 kJ mol^{-1}), this brings the estimated binding free energy from simulation into the range seen experimentally (∼35-46 kJ mol^{-1}; Table 3.2).

Calculation of the number of PIP molecules interacting with GRP1-PH along protein-lipid separation shows on average ∼4 PIP lipids are bound while GRP1-PH is close to the membrane, in agreement with the figure calculated from binding simulations, though note that some bias will be present as the GRP1-PH domain is restrained relative to the C-site PIP. PIP density (Figure 3.11 E *iii*) was again in good agreement with results from binding simulation (Figure 3.11 E *i*) and the single-bound-lipid US

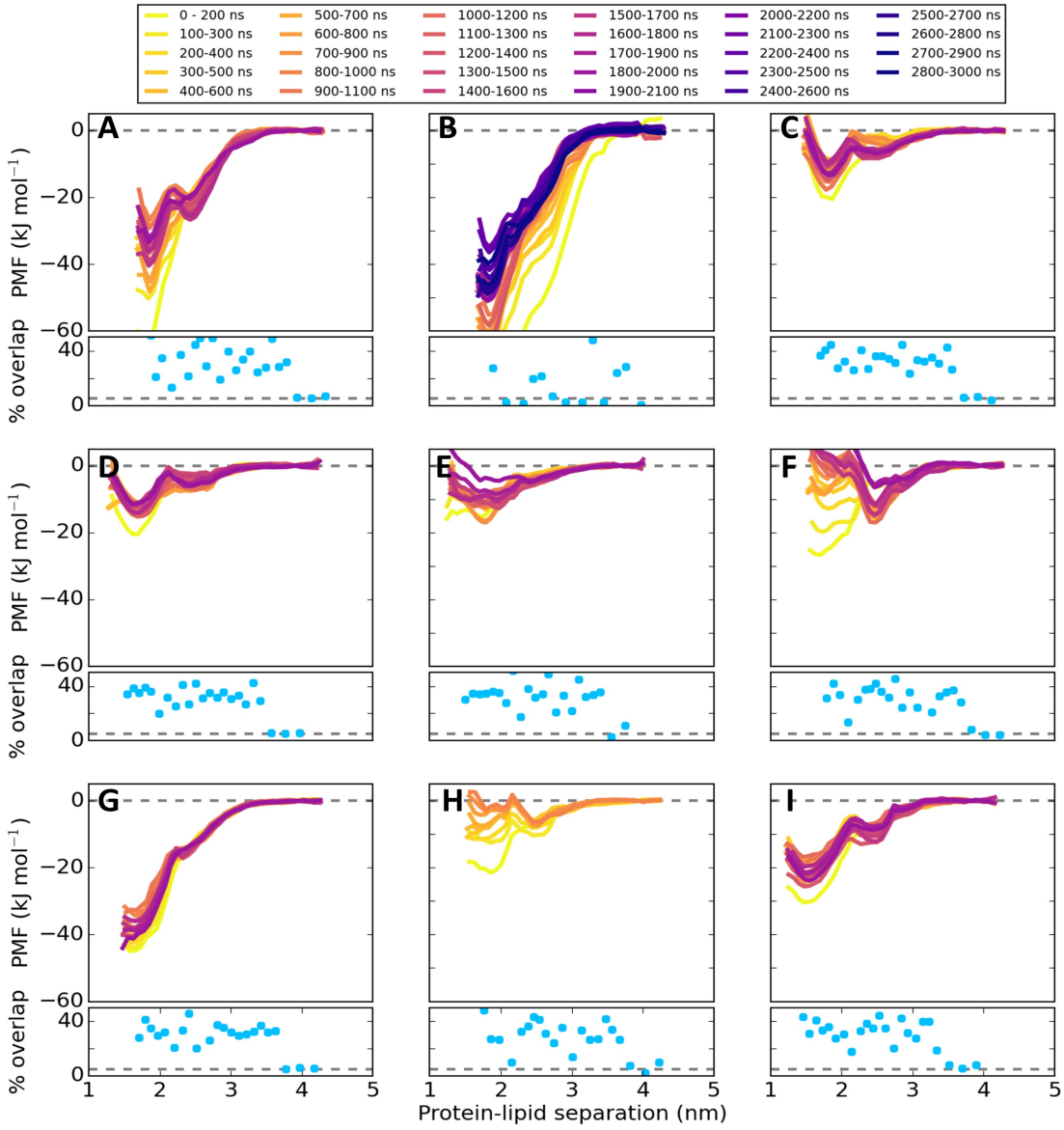


Figure 3.12: Convergence and sampling of additional US simulations. PMF profiles calculated at increasing time intervals and % reaction coordinate overlap between adjacent windows for GRP1-PH-WT binding to: **(A)** PIP_3, with two PIP_3 molecules in total in the bound leaflet; **(B)** PIP_3 with ten PIP_3 molecules in total in the bound leaflet; **(C)** PIP_3 -5e, with charge split 1/2/1 over the 3'/4'/5' phosphates; **(D)** PIP_3 -5e, with charge split 1.3/1.4/1.3 over the 3'/4'/5' phosphates; **(E)** PI(4,5)P2 -4e; **(F-G)** PIP_3, with a polarisable water model (two repeats); **(H)** PI(4,5)P_2, with a polarisable water model; and **(I)** PIP_3, with no restraint on the lipid.

simulations (Figure 3.9 A). The less dense regions around the binding modes indicate regions inaccessible while GRP1-PH is tightly bound, due to steric clashes with the protein, and are likely more visible here compared to the binding simulations due to the more thorough sampling enabled by US simulations. The density profile from the 2 × PIP_3 system indicates that the C and A sites are preferred over other interaction sites, with only a slight density observed at other locations. The background PIP density is much lower (effectively zero), indicating binding of both PIPs are able to stably bind GRP1-PH (Figure 3.11 E *iv*). A non-zero density is seen for 10 x PIP_3 systems as the interaction sites on GRP1-PH become 'filled', leaving 'spare', unbound PIPs that contribute to the background.

The number of PIPs bound and density profiles from both binding and US simulations indicate that GRP1-PH facilitates the clustering of PIP lipids, as has also been suggested by previous studies (e.g. [104]). The overall concentration of PIP_3 in the plasma membrane appears to be quite low, estimated to be below 1% [7], compared to the 5% composition used for 10 × PIP_3 simulations here. While clustering of PIPs, by GRP1-PH or by other mechanisms will increase the local concentration, it is unknown how well the 10 PIP_3 per leaflet system described above will reflect physiological conditions. These results nonetheless demonstrates the large effect that binding of multi PIP lipids can have on PH binding, and confirms the difference in number of PIP molecules is likely to explain the initial underestimation of binding free energy calculated of GRP1-PH to PIP_3 compared to experimental results, many of which use non-physiological PIP concentrations at a similar level to that used in the 10-PIP system (∼5% PIP_3; see Table 3.2 and references therein). Some broader implications of multiple PIP binding will be discussed further in Chapter 4.

3.5 - Further considerations of the US/CG method

3.5.1 - *Additional tests*

Following initial US investigations, PMF binding profiles for several additional systems were obtained for additional checks on aspects of the method and setup (Figure 3.13). Most of these simulations were extended to 2000 ns, and appeared to be

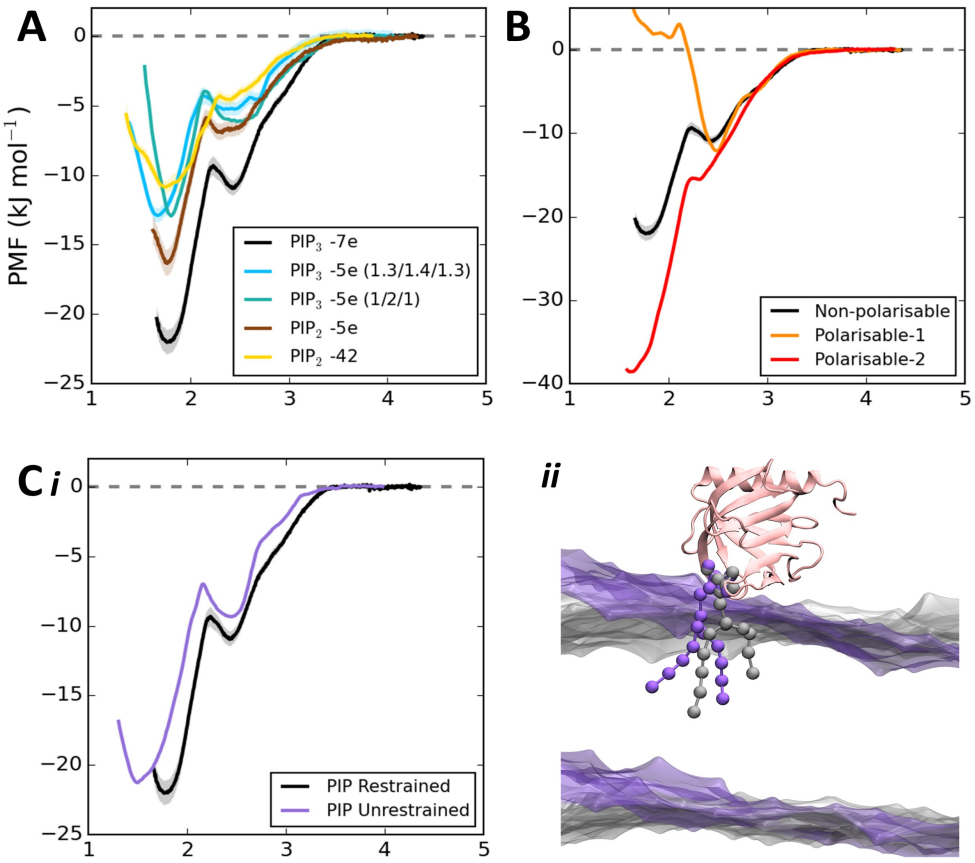


Figure 3.13: Additional profiles to check the US/CG method. **(A)** Binding to PIPs with altered charge patterns: PIP_2 -4e (yellow) and PIP_3 -5e with an partial (blue) or integer (green) split of charge across phosphates, $PI(4,5)P_2$ -5e and PIP_3 -7e, reproduced from Figure 3.8, are also shown in brown and black, respectively, for reference. **(B)** Binding of GRP1-PH to PIP_3 in the presence of polarisable water from two repeats (red, orange). Binding in the presence of 'normal' Martini water, reproduced from Figure 3.8 A *i*, is shown in black for reference. **(C)** *(i)* Profile for GRP1-PH-WT/PIP_3 where PIP_3 was unrestrained (purple) or restrained (black; reproduced from Figure 3.8 A *i*). *(ii)* Structure of GRP1-PH showing relative position of PIP_3 and the membrane at the well minima for the unrestrained (purple) or restrained (grey) system.

converged and well sampled (Figure 3.12 C-I); the exception, involving polarisable water, is discussed in Section 3.5.1.2.

3.5.1.1 - *Charge of the PIP lipids*

In the above simulations, the charge on the PIP lipids was set to the default values provided by the Martini [61, 78] forcefield: a -1e charge on the 1' phosphate and

-2e on for 3', 4' and 5' phosphates, giving a net charge of -7e for PIP_3 and -5e for $PI(3,4)P_2$/$PI(4,5)P_2$. Several studies suggest that the charge state of PIPs varies over physiological pH ranges (see e.g. [124, 125]). Additional US simulations were therefore performed for GRP1-PH-WT binding to PIP_3 and $PI(4,5)P_2$ with a reduced charge of -5e and -4e, respectively - the net charge at neutral pH determined by [125], rounded to integer charge to facilitate neutralisation of the system. For $PI(4,5)P_2$, this charge was split with -2e on the 4' phosphate and -1e on each the 1' and 5' phosphates. For PIP_3, two simulations were performed: one with an integer charge split of -2e for the 4' phosphate and -1e for all others, and a second with partial charges of -1.3e, -1.4e, -1.3e respectively for the 3', 4' and 5' phosphates, following values calculated in [125].

The resulting PMF profiles show a reduced well depth compared to the initial profiles (Figure 3.13 A). PIP charge is likely to vary with pH; this further demonstrates the high dependence of calculated binding energies on the exact system conditions. Interestingly, while the overall depth of the PIP_3 -5e profiles is closer to that of -5e PIP_2, the shape of the profile more closely resembles that of the -7e PIP_3 profile, with a clearer second well; meanwhile, the -4e PIP_2 profile is perturbed in this region but shows no clear well. This indicates that, despite the coarse-grained representation, the US approach may distinguish between distribution of charges among the phosphate groups, rather than simply overall charge. While the exact value of binding energy may change with environment conditions, insights drawn regarding structural details and favoured binding sites likely remain valid.

3.5.1.2 - *Polarisable water*

A limitation of the standard Martini model, used in the above simulations, is that the polarisability of water is ignored. A polarisable version of Martini water is available [81], and has in some case shown to improve free energy estimations (e.g. [126]). Additional US simulations were run using polarisable water to assess the possible improvement offered. However, the polarisable water model appeared to significantly perturb the underlying energy landscape compared to that seen elsewhere in this chapter. Two repeats with the GRP1-PH-WT/PIP_3 system were performed. Both required extension to 2000 ns to ensure convergence had been reached (Figure

3.12 F-G), and converged to values very different from both the original GRP1-PH-WT/PIP_3 profile and from each other (Figure 3.13 B). Further simulation would be required to obtain a truly converged profile representative of the GRP1-PH-WT/PIP_3 system, and it is not clear if the result would indeed be more accurate than without polarisable water. A polarisable water US simulation was also started for the GRP1-PH-WT/PI(4,5)P_2 system; however, after 1000 ns this system was showing similar behaviour to the PIP_3 systems and so not continued to convergence (Figure 3.12 H).

Even without considering the extension of each window from 1000 ns to 2000 ns, simulations with polarisable Martini water carry a greater computational expense due to the much higher total number of interactions to be considered. This does against the original purpose of choosing a CG representation - sacrificing accuracy of the model to increase accessible timescales. Given that the results achieved with the normal Martini water appear to be reasonable, and the difficulty obtaining results with polarisable water, it was decided that use of a polarisable water model was, at this stage, unnecessary.

3.5.1.3 - *Removal of the lipid restraint*

A restraint on the xyz motion of the 'bound lipid' phosphate was initially employed to simplify the system. To check that this restraint was not biasing the system, an additional US simulation was performed for the GRP1-PH-WT/PIP_3 system in which this restraint had been removed. The resulting PMF profile is a good match for that obtained with the restraint, differing only in the location along protein-lipid separation of the deepest binding well (Figure 3.13 C *i*). This difference can be explained as PIP_3, now free to move, rises slightly from the membrane plane with a corresponding slight rotation of the PH domain; this reduces the z component of protein-lipid separation (i.e. the reaction coordinate) while leaving the overall PIP-protein interaction unchanged (Figure 3.13 C *ii*). Profiles calculated on 200 ns intervals show greater variation when the lipid is unrestrained, which required simulation time to be extended to 2000 ns to confirm convergence (Figure 3.12 I), indicating the restraint may indeed have improved convergence as intended. Given this and the lack of a large effect on the resulting profile, maintaining the lipid xyz restraint for future investigations was determined to be reasonable.

3.5.2 - *Caveats of the US/CG method*

It is noted that there are several limitations of the CG/US methodology; these will be discussed further in Chapter 4, but are presented first here. Firstly, the use of a coarse-grained model results in loss of accuracy. As discussed above, this was deemed a necessary compromise to allow simulations to be run on feasible timescales, particularly as future directions look to larger-scale application of the method. However, even despite the lower resolution, PIP lipids and residue mutations could be distinguished.

One of the largest issues of MD simulations and particularly free energy calculations is the question of adequate sampling; this is also discussed above. Several measures were used which appear to show adequate sampling in the systems investigated, however sampling may be difficult to assess especially if structural changes occur on timescales on or over the order of the simulation length. Bootstrap analysis, as used here, is commonly used to calculate errors associated with PMF profiles from WHAM calculations, but does not take into account the degree of sampling. A better measure of possible error due to inadequate sampling may be obtained by comparing profiles obtained for systems with differences not expected to have major effects on the profile. This includes, for example the PI(3,4)P_2 vs PI(4,5)P_2 system (due to the equivalence of these lipids in the coarse-grained force field), GRP1-PH-WT/PIP_3 system with and without the lipid restraint and the two GRP1-PH-WT/PIP_3 (-5e) systems; all demonstrate good agreement. However, it is noted that all systems investigated through US/CG here start with very similar initial structures, and the potential bias of this initial structure on the interaction space sampled is not investigated.

3.6 - Conclusions

In this chapter, a coarse-grained model was successfully combined with umbrella sampling simulations to allow investigation of the interactions of GRP1-PH with membranes containing PIP lipids on both an energetic and structural front. Good qualitative agreement with previous studies, both computational and experimental, was seen for the primary interaction mode, lipid selectivity, and effect of a single amino acid mutation. Although the initial quantitative values estimated from single-bound-PIP US simulations appear to underestimate existing experimental results,

adding additional PIP lipids to the membrane brings energies within the range seen experimentally.

To our knowledge, a combined energetic/structural analysis approach such as this had not previously been applied to large scale studies of peripheral protein/membrane interactions. By carefully testing simulation parameters and developing an automated pipeline through investigation of the GRP1-PH/PIP system, expansion to larger scale investigations - particularly looking at the wider PH family - may be readily achieved. Already, new insights have been made in relation to the presence of multiple binding sites in PH domains, with identification of PIP-interactions at the A-site of GRP1-PH, not previously observed for this protein. Now, application of US/CG to further PH domains may allow the broader roles of these multiple binding sites and the determinants of PIP-selectivity, intrinsic to the function of PH domains, to be understood.

Chapter 4

Interactions of members of the PH domain family with membranes containing PIP lipids

4.1 - Background

4.1.1 - *Unresolved questions surrounding the PH domain family*

The PH domain is among the most prevalent domains in the human genome, essential for the localisation and/or activation of many proteins; understanding the membrane association of PH domains and how it is driven by different PIP species is therefore of interest. As such, numerous studies have investigated the interactions of various PH domains with PIPs or their analogues both structurally (see Table 4.1) and biophysically (see Table 4.2). However, several important questions remain unclear or unanswered.

4.1.1.1 - *What determines selectivity and affinity?*

Through a variety of qualitative and/or quantitative methods, biophysical and biochemical studies have reported PH domains selective towards PIP_3, PIP_2, other PIPs, or showing no clear preference, with binding constants ranging over several orders of magnitude (see Table 4.2). In conjunction with structural studies, explanations for the individual PIP preferences of several domains have been proposed (e.g. [100, 111, 137], but overall predictors of affinity for each PIP remain largely unclear. A KX_n(K/R)XR motif was initially proposed to identify 3' specific binding domains (e.g. [19, 20]), but there are many exceptions (see e.g. [108]).

Table 4.1: Structural studies identifying PIP-binding site of PH domains. Structural studies reporting evidence for binding of PIP or PIP analogues in the C, A or both binding sites for different PH domains, including the 12 domains studied in this chapter. For studies associated with a PDB entry, the PDB ID is listed.

Protein	Binding site		
	C-site	A and C site	A-site
ACAP1		MD [127]	
ArhGAP9		MD [104]	X-ray (2POD, 2POH), Mutation [122]
AKT1	X-ray (1UNQ, 1H10, 2UVM), Mutation [128]	MD [104]	
ARNO	X-ray (1U27, 1U29), MD [104]		
ASAP1		X-ray (5C79), Mutation [123], MD [104]	
β-spectrin		MD [104]	X-ray (1BTN), NMR shift [121]
BTK	X-ray (1B55, 2Z0P), MD [104]		
CERT	NMR shift [129]		
DAPP1	X-ray (1FAO), MD [104]		
Dok7		MD [130]	
Dynamin	NMR shift [131, 132], Mutation [131]		
Gab1	Mutation [133]		
GRP1	*see Table*3.1		
Kindlin-2	NMR (2LKO), MD [104]		
Ldp		Mutation [134]	
PDK1	X-ray (1W1D, 1W1G), MD [104]		
PEPP1	X-ray (1UPR), MD [104]		
PLCδ	X-ray (1MAI), MD [104]		
Pleckstrin-C	X-ray (215C, 215F), NMR shift [135], MD [104]		
P-Rex	X-ray (5D3X, 5D3Y)		
Slm1		Mutation[136]	X-ray (4A6H, 4A6K)
Tiam1			Mutation [122]

Table 4.2: Summary of biophysical studies PH/PIP interactions. Affinities and selectivities reported for the 12 PH domains studied in this chapter binding to either $PI(4,5)P_2$ (or its analogues) and/or PIP_3 (or its analogues). Studies performed in a membrane environment are indicated by grey backgrounds. Values obtained in the same study but with different conditions are listed in the same row, comma separated. Results of qualitative studies are indicated by 'fold selectivities' of >1 (PIP_3 selective), <1 (PIP_2 selective) or ∼1 (no clear preference).

Protein	Method	PIP_2/analogue affinity (μM)	PIP_3/analogue affinity (μM)	Fold selectivity PIP_3/PIP_2	Ref.
ArhGAP9	Fluorescence	0.1	0.3	0.3	[122]
AKT	SPR	>0.1	0.035	>3	[144]
	SPR		0.59		[113]
	FRET	>10	0.023	>430	[139]
	SDS-Page			>1	[145]
	Fluorescence	1.2, 2.5	1.5, 0.5	0.8, 5	[138]
	Fluorescence	9.1	0.13	70	[122]
	Radiolabel			>1	[146]
	Radiolabel			>1	[147]
	Overlay			>1	[106]
	Overlay			>1	[148]
	Overlay			≈	[108]
	Microarray			>1	[149]
ARNO	SPR		1.7		[113]
	SDS-Page			>1	[150]
	ITC	4.15, 2.66	1.02, 1.64	4, 1.6	[111]
	Radiolabel	>10	0.0693	>1	[151]
	Radiolabel		1.6	≈	[110]
	Overlay			≈	[108]
ASAP1	Co-sedimentation	7			[123]
	ITC	>100			[123]
	ITC	75			[152]
	Overlay			≈	[108]
	Overlay			<1	[153]
BTK	SPR			>1	[131]
	SPR		0.08		[113]

Continued on following page

Continued from previous page

	ITC		0.04		[154]
	Radiolabel		0.8	>1	[155]
	Radiolabel		0.04	>1	[156]
	Overlay			≈	[108]
β-spectrin	SPR	0.125		<1	[157]
	SDS-Page			<1	[25]
	CD shift	40		≈	[121]
	Microarray	0.006			[158]
	Overlay			≈	[106]
	Overlay			≈	[25]
DAPP1	SPR		0.003		[148]
	SPR		0.67		[113]
	ITC	>10	0.043	>130	[106]
	ITC	0.049			[100]
	Radiolabel			>1	[147]
	Overlay			>1	[148]
	Overlay			>1	[107]
	Overlay			>1	[159]
	Overlay			>1	[108]
Kindlin-2	SPR		2.12	>1	[137]
	SPR		159		[141]
	Co-sedimentation			>1	[160]
	Overlay			>1	[161]
	Overlay			≈	[108]
PDK1	SPR	0.024	0.0016	15	[144]
	SPR		0.06		[102]
	SPR		0.052, 0.03	>1	[162]
	FRET		0.012	>1	[163]
	Radiolabel			>1	[147]
	Overlay			>1	[148]
	Overlay			>1	[163]
PEPP1	-				
PLCδ1	SPR	2.1			[164]
	SPR	6.6			[165]
	SPR	0.8			[142]

Continued on following page

Continued from previous page

FRET	0.19	>10	<0.019	[139]
FRET	0.15	>10	<0.015	[115]
ITC	0.2			[136]
SDS-Page	1.36			[140]
Co-sedimentation	1.7			[123]
Co-sedimentation			>1	[143]
Overlay			<1	[165]
SPR	0.104			[165]
ITC	0.21, 1.66	>13	<0.016	[166]
ITC	0.18			[123]
Radiolabel	0.3		<1	[167]
Overlay			<1	[168]
Overlay			<1	[106]
Overlay			<1	[107]
Overlay			<1	[108]

PleckstrinC -

Comparing selectivity and affinity across PH domains is complicated by the incomplete results for some domains and a lack of consistency in reported binding affinities among those that are better studied (Table 4.2). Many studies use PIPs spotted onto a paper surface or water-soluble analogues in solution, neither of which fully reflect the native membrane environment. Dissociation constants may vary by orders of magnitude for different methods- for example, AKT1-PH has been reported to bind to IP_4 with a K_D of 1.5 µM (using a fluorescence based assay [138]) and to PIP_3 itself with K_Ds from 0.59 µM (using SPR [113]) to 0.023 µM (using FRET [139]). Even considering only a membrane context, individual studies have demonstrated dependence of PH-PIP interactions on experimental conditions including pH [20], salt concentration [140], buffer [141] and lipid composition [102, 115, 140, 142, 143]. Largescale approaches investigating many domains with the same conditions may help to resolve these differences.

4.1.1.2 - *PIP binding site: canonical or alternate?*

Structural studies have utilised NMR, X-ray crystallography and mutational analysis to isolate the PIP binding site on several PH domains (see e.g. Table 4.1); this

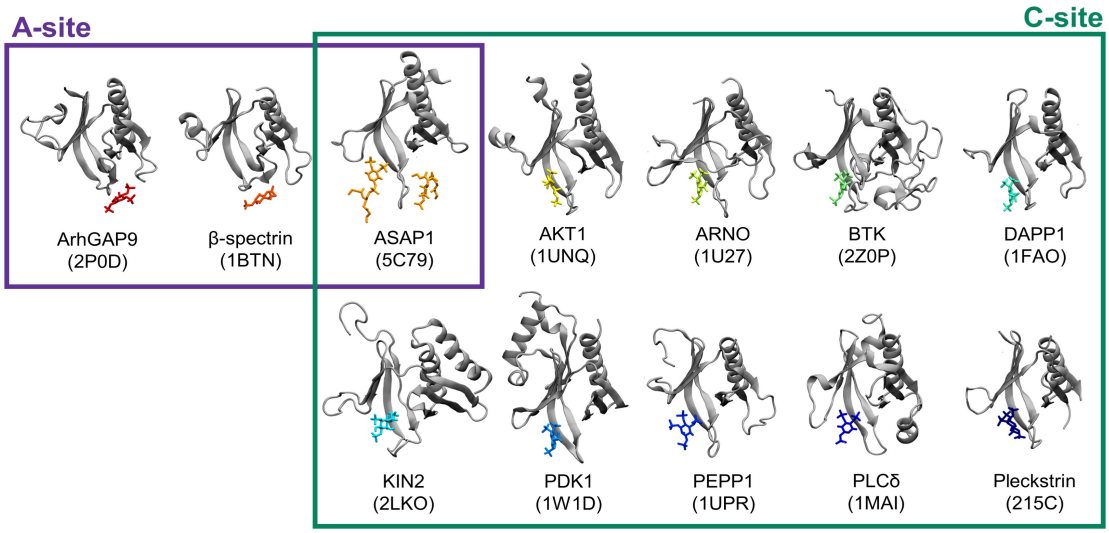


Figure 4.1: Structures of PH domains bound to PIP analogues. PDB structures of the 12 PH domains studied in this chapter bound to PIP analogues. Each domain has been aligned to the same reference structure; PIP analogues may be bound in the canonical (C-site) and/or alternate (A-site) sites, as indicated.

includes over 20 structures of PH domains bound to PIP analogues deposited in the PDB, representing 15 unique proteins. Most studies suggest PIP binding occurs in the 'canonical' (C) site formed between the β1-β2, β3-β4 and β6-β7 loops; less common is an 'alternate' or 'atypical' (A) site on the opposite side of the β1-β2 loop, between this and the β4-β5 loop [10] (Table 4.1; Figure 4.1). Only a few domains have been suggested experimentally to bind with both sites, with a structure available only for ASAP1-PH [123]. Recent computational studies, including the work presented in Chapter 3, have expanded the list of dual-binding PH domains [104, 127, 130].

As suggested in Chapter 3, the binding of multiple PIPs may be integral to achieving high membrane affinities; however, the presence of different binding sites across the PH had not yet been widely assessed. The K-X_n-(K/R)-X-R sequence motif mentioned above may better correlate with binding at the C-site, but the distinguishing features of A-site and dual binding domains are unclear. Results in Chapter 3 and [104] have shown A-site PIP interactions in domains previously seen as C-site binders, suggesting a computational approach may be required to reveal the dual-binding nature of some PH domains.

Further, most of the experimental structural studies investigating PH-PIP interactions, including all PDB structures, use soluble analogues of PIP (Table 4.1) and as such there is limited information regarding the exact modes of interaction of PH domains on the membrane surface and how these differ between domains. Computational studies will be integral to determining these details.

4.1.2 - *US/CG and the PH family*

Chapter 3 demonstrated the success of a combined US/CG approach for describing the binding of the well-studied GRP1-PH domain to a model membrane containing PIP molecules, including capturing PIP selectivity and observing both A- and C-site interaction modes. In this chapter, this methodology was expanded to examine the protein-membrane interactions of a further 12 PH domains with model membranes containing $PI(4,5)P_2$ or $PI(3,4,5)P_3$ lipid molecules. This study represents the first large-scale quantified investigation of protein-membrane interactions through simulation, and is an important step towards answering the questions presented above, enabling a systematic comparison of the strength and specificity of PIP interactions within the PH family and providing a comparative, global definition of the different modes of binding of PH domains to PIP-containing bilayers.

4.2 - Methods

The details of all US simulations performed in this chapter are summarised in Table 4.3. A brief overview of the US/CG pipeline used is shown in Figure 4.2; a more detailed pipeline can be found in Figure 3.2.

4.2.1 - *Details of the protein-membrane system*

PH domains - Atomistic structures for 12 PH domains bound to PIP analogues, ranging in resolution from 0.98 to 2.58 Å, were obtained from the PDB (see Figure 4.1). A mutant version of the AKT1-PH domain was generated using the Mutator plug-in of VMD [105]. Atomistic structures were converted to a coarse-grained Martini-forcefield [61, 78] representation using an in-house script.

Membranes - Martini coarse-grain membranes of xy size 11.3 × 11.3 nm and composi-

Table 4.3: Summary of US simulations performed in Chapter 4. Details of all Umbrella Sampling simulations performed in this chapter. The initial configuration used is indicated under 'Initial struc.' by 'c' (crystal) or 's' (simulation). 'Separation covered' indicates the protein-lipid distance range over which the PMF profile was calculated.

Protein	PIP	Initial struc.	No. windows	Sim. time (ns)	Converge time (ns)	Separation covered (nm)	Location of minimum (nm)
ArhGAP9	PIP_2	c	25	3500	1000	1.35 - 3.95	1.44
		s	26	3000	600	1.68 - 4.28	1.62
	PIP_3	c	26	3000	400	1.42 - 4.21	1.48
		s	24	3000	600	1.60 - 4.20	1.73
AKT1	PIP_2	c	25	4000	600	1.49 - 4.09	1.56
		s	24	4000	600	1.56 - 4.05	1.61
	PIP_3	c	24	3500	200	1.45 - 4.07	1.53
		s	24	4000	200	1.66 - 4.27	1.62
ARNO	PIP_2	c	29	2500	400	1.11 - 3.92	1.5
		s	25	2500	600	1.30 - 3.91	1.68
	PIP_3	c	24	2500	400	1.54 - 4.13	1.57
		s	24	2500	400	1.47 - 4.07	1.54
ASAP1	PIP_2	c	24	3500	800	1.44 - 4.05	1.77
C-site		s	24	3500	400	1.72 - 4.31	2.05
	PIP_3	c	24	3500	600	1.82 - 4.42	2.12
		s	23	3500	600	1.51 - 3.90	2.12
ASAP1	PIP_2	c	26	4000	600	1.55 - 4.36	1.63
A-site		s	24	4000	200	1.54 - 4.15	1.82
	PIP_3	c	26	4000	400	1.77 - 4.57	1.79
		s	24	4000	200	1.73 - 4.33	1.97
β-spectrin	PIP_2	c	24	4000	1800	1.12 - 3.73	1.25
		s	24	4000	2400	1.02 - 3.62	1.31
	PIP_3	c	24	4000	600	0.95 - 3.54	2.36
		s	23	4000	1000	1.66 - 4.07	2.48
BTK	PIP_2	c	24	4000	1000	1.35 - 3.96	1.85
		s	24	4000	1400	1.85 - 4.35	2.74
	PIP_3	c	27	4000	200	1.17 - 3.97	1.23
		s	26	4000	1200	1.61 - 4.39	2.64
DAPP1	PIP_2	c	24	2500	600	1.63 - 4.23	1.64
		s	26	2500	600	1.19 - 3.99	1.19
	PIP_3	c	23	3000	800	0.88 - 3.28	1.11
		s	24	3000	200	1.25 - 3.84	1.29
Kindlin-2	PIP_2	c	23	4000	400	1.32 - 3.72	1.6
		s	23	4000	600	1.36 - 3.76	1.42

Continued on following page

Continued from previous page

	PIP_3	c	24	4000	2400	1.29 - 3.79	1.45
		s	23	4000	400	1.61 - 4.02	1.55
PDK1	PIP_2	c	24	2500	400	1.72 - 4.36	1.85
		s	24	2500	1000	1.52 - 4.01	1.48
	PIP_3	c	24	3000	200	1.65 - 4.26	1.8
		s	23	3000	600	1.39 - 3.78	1.58
PEPP1	PIP_2	c	24	3500	200	1.47 - 4.07	1.53
		s	24	3500	200	1.26 - 3.85	1.42
	PIP_3	c	26	3500	200	1.21 - 4.01	1.32
		s	23	3500	1200	1.12 - 3.53	1.22
PLCδ	PIP_2	c	24	2500	200	1.58 - 4.18	1.68
		s	24	2500	200	1.67 - 4.24	1.61
	PIP_3	c	24	2500	600	1.47 - 3.87	1.79
		s	24	2500	600	1.54 - 4.13	1.54
pleckstrin-C	PIP_2	c	25	3000	1200	1.14 - 3.83	1.1
		s	24	3000	600	1.20 - 3.81	1.26
	PIP_3	c	23	3000	200	1.47 - 3.86	1.74
		s	23	4000	600	1.63 - 4.02	1.91
AKT1-E17K	PIP_2	-	26	1500	200	1.30 - 4.20	1.42
	PIP_3	-	26	1500	600	1.39 - 4.17	1.5
AG9	PIP_3	C start	24	3000	1000	1.71 - 4.23	1.72

tion POPC:POPS 80:20 with one molecule of either PIP_3 or $PI(4,5)P_2$ in the 'upper' leaflet (192 lipids in total per membrane) were generated as in Section 3.2.1.

Initial structures - The 'crystal' set of initial structures were generated by aligning each PH crystal structure to the bilayer, minimising the RMSD between phosphate groups of the PIP lipid and the bound PIP analogue, before conversion to a coarse-grained representation. For ASAP1, where the crystal structure contains two bound PIP analogues, two structures were generated by aligning to each bound ligand in turn. For some domains, the PIP analogue was found to be bound 'upside down' in the crystal structure (such as in the ASAP1 A-site, see also [104]); in these cases, pairs to be aligned were chosen for a more appropriate orientation. This structure was then solvated and energy minimised as described in Section 3.2.1. Following performance of unrestrained simulations starting with the 'crystal' structure, representative snapshots were chosen as the set of 'simulation' initial structures by visual inspection of

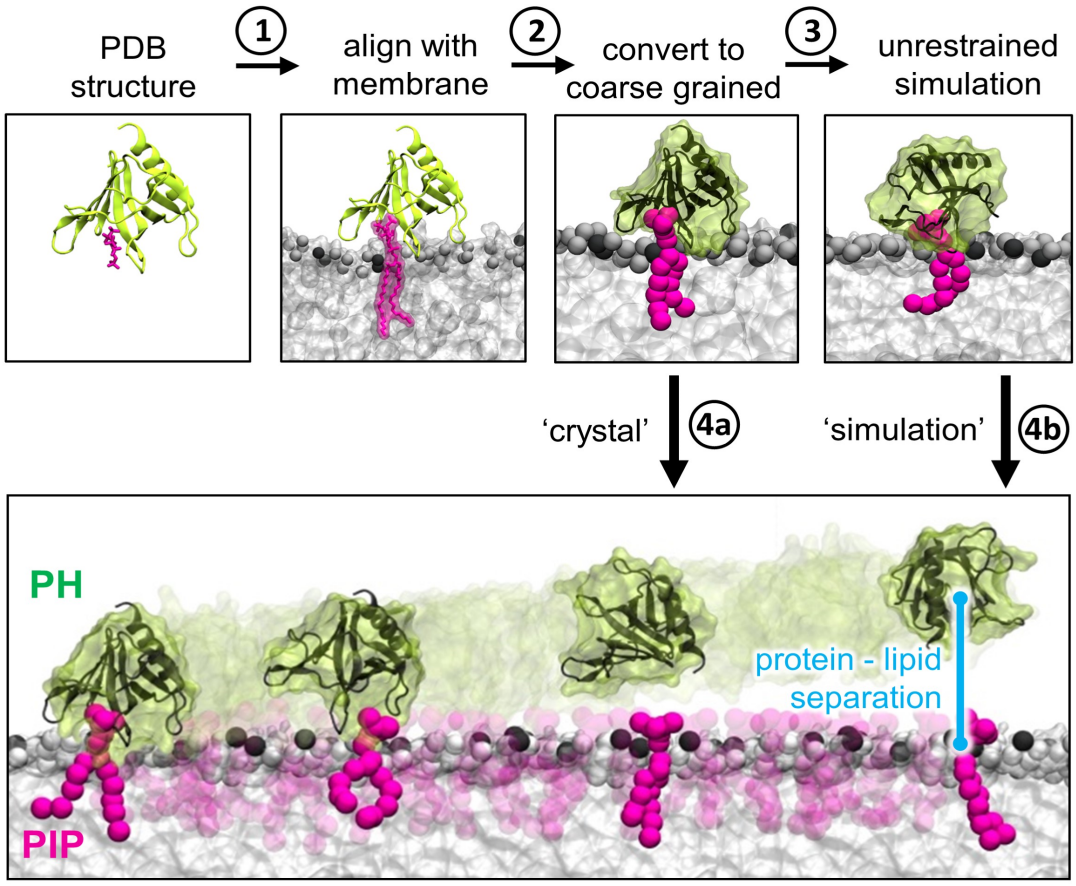


Figure 4.2: Overview of the US/CG simulation pipeline. *1.* A structure of the PH domain, with PIP analogue bound, is obtained from the PDB and aligned via the analogue with a PIP-containing membrane. *2.* Both protein and membrane are converted to coarse-grained resolution. *3.* An 1000 ns unrestrained simulation of the protein-membrane system is performed. Membrane-bound structures obtained *4a.* from the initially aligned structure in step *2* or *4b.* from unrestrained simulations in step *3*, selected by visual inspection from clustered structures, are each pulled from the membrane using steered MD simulations to generate a series initial structures for umbrella windows. Windows may then be simulated, reaction coordinate data collected, and a PMF profile calculated based on each initial structure.

Shown for the PH domain of ARNO (yellow) binding to PIP_3 (magenta). The lower panel shows a snapshot taken from each US window. The 'protein-lipid separation' reaction coordinate is indicated in blue. See Figure 3.2 for an extended pipeline, including Analysis.

clustering of the PIP phosphates relative to the PH domain. An additional 'C site' initial structure was generated for the ArhGAP9-PH domain by structural alignment to the crystal initial structure of AKT1-PH.

4.2.2 - *Simulation details*

US simulations - US simulations were set up and performed as outlined in Section 3.2.2, using GROMACS 4.6.5 [82]. Each window was simulated for upwards of 2000 ns, as indicated in Table 4.3, extending until satisfactory convergence was reached. In some cases, additional windows were added to ensure coverage adequate coverage of the reaction coordinate (see Section 3.3.1.1).

Unrestrained simulations - Additional 1 µs 'unrestrained' simulations were performed for each PH-PIP system, starting from the 'crystal' aligned structure. The 1' phosphate of the bound PIP was restrained to its initial position with force constant 1000 kJ mol^{-1} nm^{-2}, while no restraint was applied between protein and lipid.

4.2.3 - *Analysis*

PMF profiles and assessing sampling - Profiles were calculated using the Grossfield WHAM as described in Section 3.2.3. Convergence was assessed by calculating profiles over 400 ns segments starting at 200 ns intervals. In subsequent analysis, time before convergence was discarded. Window overlap was calculated from window histograms as described in Section 3.2.3.

Binding free energies and selectivity - Absolute binding free energies were calculated via Equation 2.24, using the xy displacement of the protein COM from the corresponding 'unrestrained' simulation to estimate the effect of the xy restraints. Fold selectivity was calculated from well depths using Equation 2.13, and experimental binding free energies were calculated using Equation 2.12.

Defining binding regions - 'tight' and 'loosely' bound regions were defined for each PH-PIP system, covering protein-lipid separation from -0.1 to 0.4 nm and 0.8 to 1.2 nm relative to the well minimum respectively. The minimum was taken as the deepest well common to both profiles generated for the PH-PIP system.

PIP binding angle - The φ component of the PIP binding angle and 2D landscapes of φ vs. protein-lipid separation were calculated as outlined in Section 3.2.3, using the same GRP1-PH reference. (Normalised) histograms representing the distribution of φ while tightly or loosely bound were obtained from 2D landscapes using the protein-

lipid separations defined above. An A (or C) site mode was determined to be present within a given binding region if the cumulative sum between 0 to 60° (or -50 to 10°) was greater than 0.3, or 'potentially present' if greater than 0.2.

Sequence alignment - A structure-based sequence alignment of PH domains was performed using the VMD [105] plugin MultiSeq [116].

Contact analysis - Contacts between protein residues and the PIP lipid were calculated and unbiased as described in Section 3.2.3 - a residue is considered in contact with PIP if the minimum distance between and sidechain particle of the residue and a phosphate bead of PIP is less than 0.6 nm. Known A- and C-site residues were identified as those in equivalent positions to a contact observed with a PIP in the appropriate site from any crystal structure. The average number of A, C and total residue contacts while tightly or loosely bound were obtained from unbiased contact counts using the protein-lipid separations defined above. An A (or C) site mode was determined to be present within a given binding region if the fraction of contacts belonging to A (or C) site residues was greater than 0.3, or 'potentially present' if greater than 0.2. A contact score was calculated taking the (relative) number of A contacts minus (relative) C contacts, each scaled by the respective number of A/C residues present and with sidechain particles (i.e. not glycine or alanine) in the relevant PH domain.

Degree of binding - Taking greater than two contacts as a 'bound state', the 'degree of C-site binding' was calculated as the fraction of bound-state frames in which the interaction was with the C-site (at least two C-site interactions and a φ binding angle between -50 and 10°). 'Degree of A-site binding' was calculated as the fraction of non-C binding frames in which the interaction was with the A-site (at least two A-site interactions and a φ binding angle between 0 and 60°). See Figure 4.9 B for indication of measurement of φ and A/C site residues.

Membrane orientation - Orientation of each PH domain on the membrane was calculated as outlined in Section 3.2.3, using the same GRP1-PH reference structure.

4.3 - Results and discussion

To investigate the interactions of the PH domain family with membranes containing PIP lipids, the US/CG methodology as described in Section 4.2 was used to quantify the binding of 12 PH domains to a single PIP lipid, either PIP_3 or $PI(4,5)P_2$, in a bilayer containing POPC and POPS lipids. These 12 domains were chosen based on availability of structures in the PDB with PIP analogues bound (Figure 4.1). PIP_3 and $PI(4,5)P_2$ were selected as the PIP species for investigation due to their presence at the inner leaflet of the plasma membrane, where most of the 12 PH domains to be investigated are believed to carry out their roles (Table 1.1); here, $PI(4,5)P_2$ is constitutively present, while the rarer PIP_3 is a vital second messenger and target for many PH domains [29]. Each US system showed satisfactory convergence within 4000 ns (Figure 4.3).

4.3.1 - *Sensitivity to initial structure*

Following discussion in Section 3.5, to sensitivity of the US approach to the exact starting model were assessed by performing two sets of US simulations. The first used an initial structure obtained directly from alignment of the appropriate crystal structure, as for Chapter 3 ('crystal'), while the second used representative snapshots taken from unrestrained simulations of each system ('simulation'). This unrestrained simulation was intended to allow the PH/PIP system to equilibrate and maximise optimal contacts. For some PH/PIP systems, the initial binding mode was stable and so the two starting structures were very similar; in many cases, however, shifts in binding orientation were observed and the two starting structures are readily distinguished (Figure 4.4).

Despite the difference in initial orientation, the overall shapes of the PMF profiles show good agreement between the 'crystal' and 'simulation' repeats for all PH-PIP systems except for the BTK-PH domain (Figure 4.5). This is further seen when comparing the energy minima (i.e. well depths) of the two US sets, with only BTK (binding to PIP_3) and DAPP1 (binding to both PIP_2 or PIP_3) showing variation greater than the 'chemical accuracy' of $\sim$4 kJ mol^{-1} (Figure 4.6 A). As will be shown

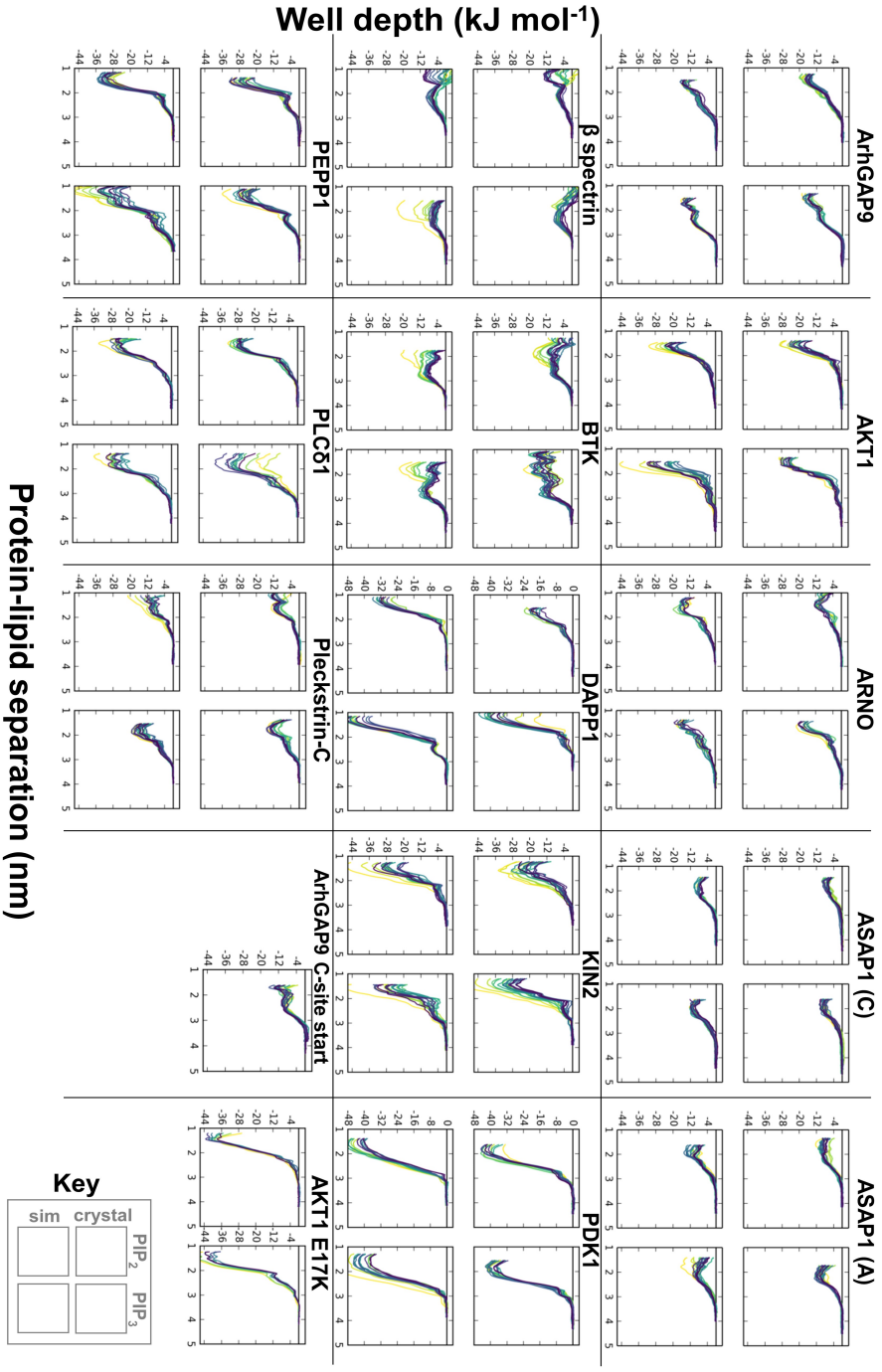


Figure 4.3: Convergence analysis for PH/PIP US simulations. PMF profiles calculated over 400 ns slices, starting in 200 ns intervals, coloured from yellow (start of simulation) to blue (end of simulation), for each US simulation performed in this chapter. Profiles are grouped by PH domain; the position of profiles for the four systems containing a given PH domain (two lipids × two starting structures) are as indicated in the key, bottom left.

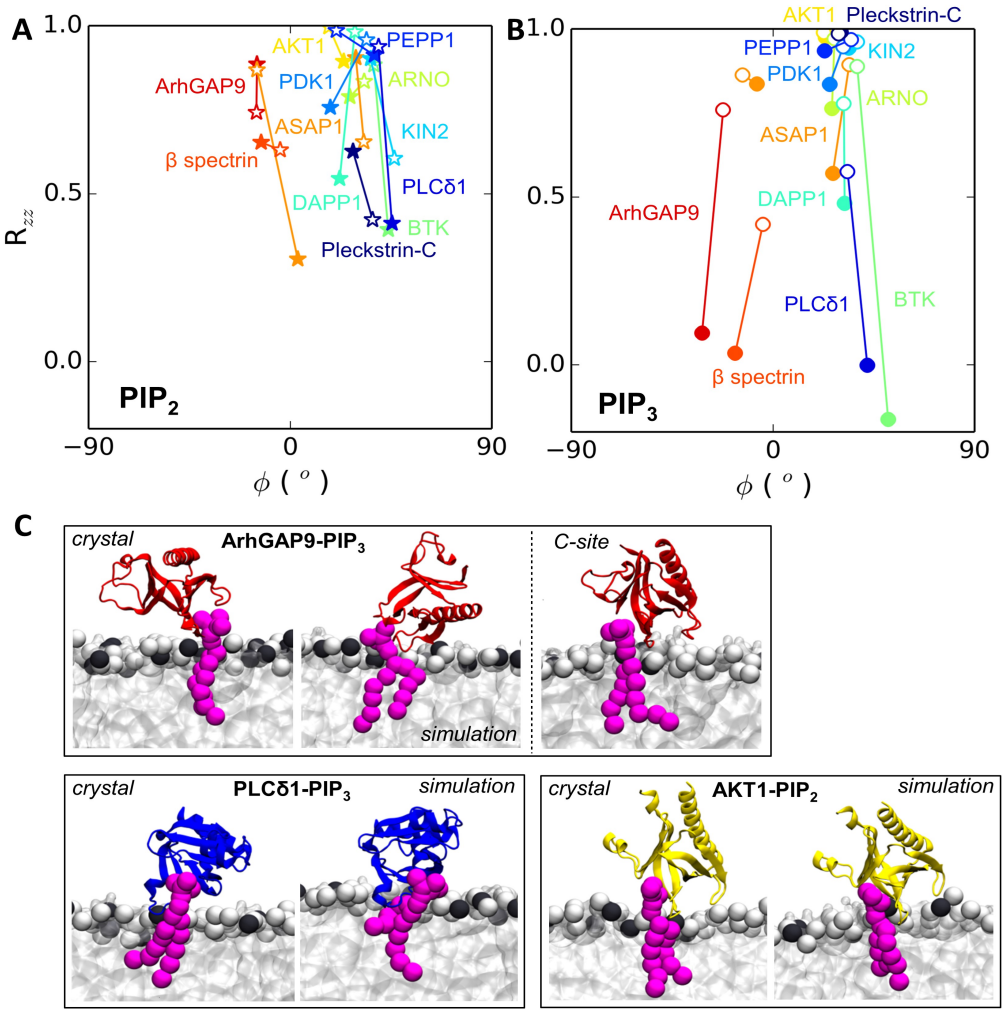


Figure 4.4: Distribution of initial configurations. PIP binding angle φ and R_{ZZ} of 'crystal' (empty markers) and 'simulation' (filled markers) initial structures for each PH domain binding to **(A)** PIP_2 and **(B)** PIP_3. The crystal/simulation structures for each system are linked by a line. **(C)** Snapshots for selected initial configurations, including the ArgGAP9 'C-site' aligned structure, generated from alignment with the AKT1 crystal structure.

in Section 4.3.3, the same patterns of binding modes are also seen between the 'crystal' and 'simulation' repeats for all domains other than BTK.

The differences observed in BTK simulations may be due to the xy restraints. In a complicated binding energy landscape, if the two starting structures are sufficiently displaced in xy these restraints may result in sampling vertical slices from the full energy landscape in which the favourability of binding modes differs. BTK indeed

shows among the highest shifts in orientation between the two initial strictures, especially for PIP_3 (Figure 4.4). This effect may also explain the relatively large error between repeats for DAPP1. However, the good agreement between the set of profiles for most cases suggests this is not a common issue, and the US/CG method is reasonably robust to initial orientation.

4.3.2 - *Energetics: selectivity and affinity*

4.3.2.1 - *PMF profiles*

The 1D PMD profiles calculated for each PH/PIP system (Figure 4.5) all show a clear global minimum, indicating favourable binding to PIP, ranging from ca. -40 (PDK1) to -10 kJ/mol (ASAP1). A previous study [169] predicting possible membrane interaction sites across PH domains, including those studied here, failed to identify potential membrane interactions for ArhGAP9, β-spectrin or BTK, in line with the lower binding energies of these domains seen here.

The PIP selectivity of each PH domain was assessed by comparing the depths of the respective minima from the PIP_3 and PIP_2 profiles (Figure 4.6 B). Most PH domains bound equally strongly to PIP_3 and PIP_2 at the current CG level of description. Only the AKT1, DAPP1 and β-spectrin PH domains were predicted to exhibit clear-cut selectivity: AKT1 showed 4.5-fold selectivity for PIP_3; DAPP1 showed 300-fold selectivity for PIP_3; whilst β-spectrin showed 3-fold selectivity for PIP_2.

For most systems, the global minimum occurs at a protein-lipid separation of around 1.5 nm; this was referred to as the 'tightly bound' state, T. In some cases (namely for β-spectrin/PIP_3 and BTK), the global minimum occurs at a larger separation of around 2.2 nm; many other profiles show a local minimum or inflection at this distance. This second region was labelled the 'loosely bound' state, L. These two

Figure 4.5 *(following page)*: PMF profiles of PH domains interaction with PIP-containing membranes. PMF profiles calculated from US simulations for each PH domain binding to **(A)** PIP_2 or **(B)** PIP_3, using an initial configuration based directly on the crystal (or NMR) structure (magenta) or from an unrestrained simulation (black). Bootstrap errors are shown. The 'tightly' (T) and 'loosely' (L) bound regions are indicated for ArhGAP9 binding PIP_2.

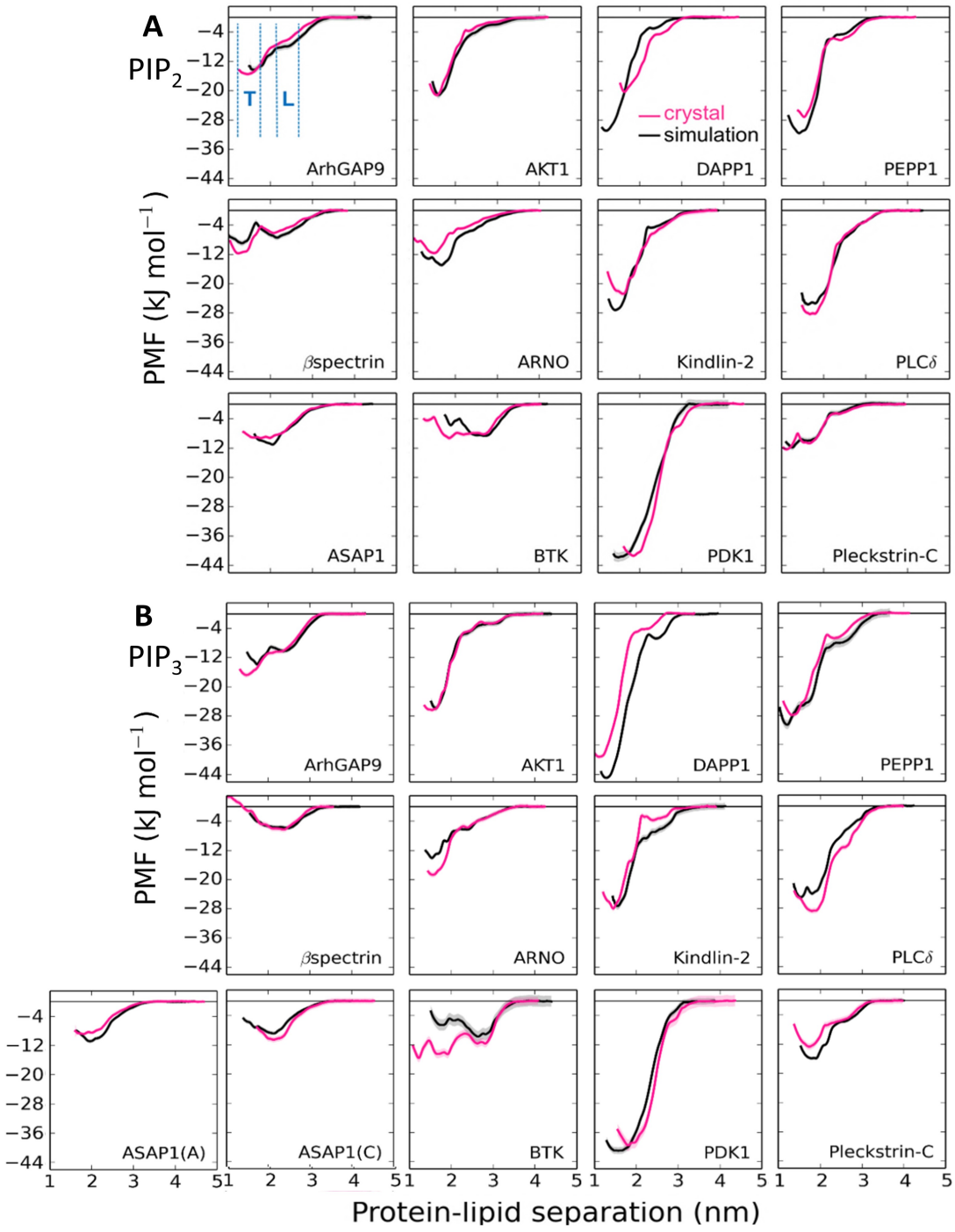


Figure 4.5: PMF profiles of PH domains interaction with PIP-containing membranes. *Caption on previous page.*

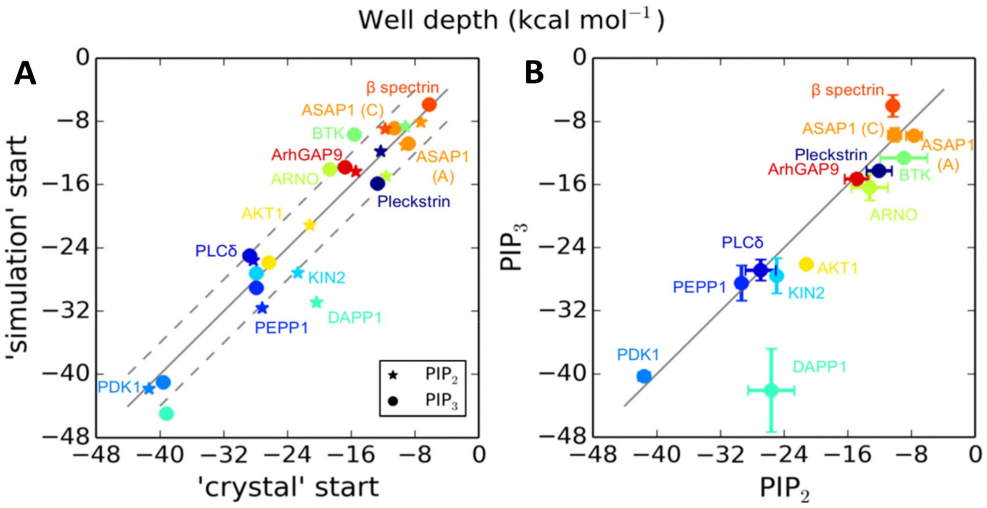


Figure 4.6: Comparison of profile well depths. **(A)** Reproducibility of results from US simulations, indicated by comparison of well depths calculated from 'crystal' vs. 'simulation' initial configurations for each PH/PIP system. The dotted lines indicate a difference of 4 kJ mol^{-1}. **(B)** Selectivity of each PH domain, indicated by the average well depth for binding to PIP_3 vs. PIP_2. Domains falling above the line indicates preference for PIP_2, below the line for PIP_3. Error indicates the range from the two US simulations.

regions are shown for ArhGAP9/PIP_2 in Figure 4.5 A, and match the approximate locations of the two wells observed for the binding of GRP1-PH to PIPs, described in Section 3.4.2.1. In Chapter 3, these wells were found to correspond to the presence of two sets of interaction modes, favourable at different protein-lipid separations. Similarly, more complex profile shapes observed for domains investigated in this chapter likely indicate the presence of multiple interaction modes. These interaction modes were investigated further through structural analysis and will be discussed in Section 4.3.3.

4.3.2.2 - *Comparison with experiment*

Standard free energies of binding were calculated from the PMF profiles for each system, averaging between the 'crystal' and 'simulation' repeats (Figure 4.7 A). Binding free energies were also calculated from experimentally reported dissociation constant values (Figure 4.7 A, Table 4.2). The binding free energies derived from simulations are generally smaller than those measured experimentally, as was the case for GRP1-PH (Section 3.4.2.1). Again, this may be in part due to limitations in the

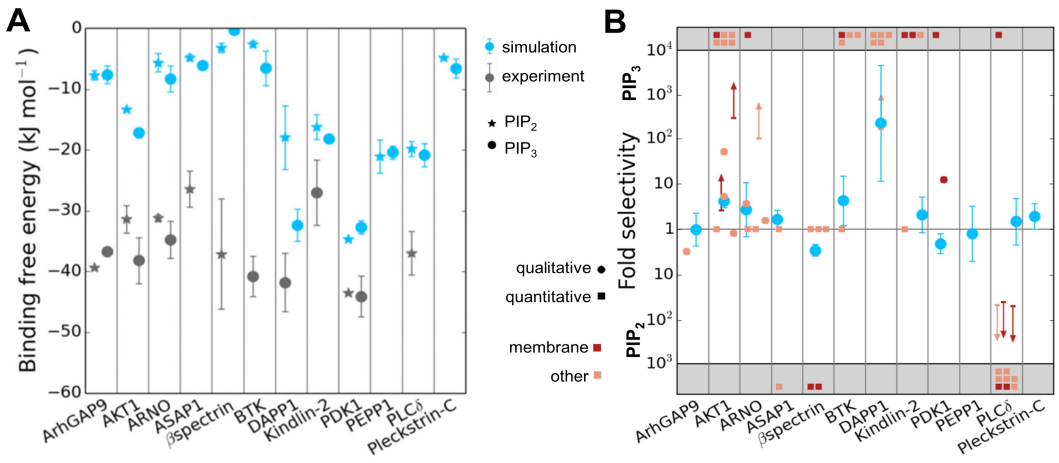


Figure 4.7: Comparison of with experiment. **(A)** Binding free energies and **(B)** fold selectivities calculated from simulations (blue) and the experimentally determined K_D values reported by studies listed in Table 4.2 (red where a membrane environment was used, pink otherwise, grey for average). Simulation binding free energy is calculated from well depths in Figure 4.5 using Equation 2.24; experimental binding free energies are calculated from Table 4.2 using Equation 2.12. In **B**, quantitative results are represented by circles; arrows indicate upper/lower bounds where binding was not observed within the detectable range of an experiment. Qualitative selectivity results are indicated by squares placed above (PIP_3 selective), below (PIP_2 selective) or along the centre line (no clear preference for either PIP_2/PIP_3) of the selectivity plot. Errors indicate standard deviation.

coarse-grained model and force field used, but more likely is due to the presence of only one PIP molecule, whereas the PH domain in experimental systems may be expected to cause local nanoscale clustering of and interact with multiple PIP molecules, as suggested by computational [104, 130] studies and demonstrated experimentally for other proteins (e.g. [170]). In Chapter 3, it was demonstrated that the presence of multiple PIP molecules could bring the computational binding energy in line with that determined from experiment. Thus, it should be emphasized that the binding free energies estimated in this chapter correspond to 1:1 PH-PIP interactions at the surface of a lipid bilayer.

Direct comparison of computationally observed selectivities with experiment is hindered by the lack of reliable quantitative or even qualitative data, with many studies using non-membrane environments and reporting widely varying results, as discussed in Section 4.1.1.1. However, the three domains suggested to exhibit PIP selectivity

from US results (DAPP1 for PIP_2, ATK1 and β-spectrin for PIP_3) are all generally reported to show the same preference experimentally (Figure 4.7 B). Three additional domains are suggested from experiment to be PIP_3-selective (ARNO, BTK and kindlin-2); the computational results are consistent with this assignment, suggesting either PIP_3 selectivity or equal binding to PIP2 and PIP_3. Discrepancy between computational and experimental results is seen only for the PH domain of PLCδ1, for which most (though not all) experimental studies report PIP_2 selectivity. This domain was observed computationally to bind equally to PIP_2 and PIP_3; it is not clear what the origin of this difference was, but may be a limitation of the coarse-grain model.

4.3.3 - *Structures: membrane binding modes*

4.3.3.1 - *The canonical vs. atypical site*

The φ PIP binding angle was defined in Chapter 3 in order to differentiate A and C site interactions within the GRP1-PH domain (Figure 3.1 D). Given high structural similarity of PH domains, and the use of an elastic network restraining the protein secondary structure within the coarse-grain model, this same angle is likely suitable for distinguishing A and C interactions for other PH domains. Thus, 2D energy landscapes indicating the favourability of different φ PIP binding angles with protein-lipid separation were calculated as outlined in Section 4.2.3 (Figure 4.8). These landscapes revealed multiple binding sites; most fall within the range of binding angles ascribed to A or C site interactions (0 to 60° and -50 to 10°, respectively) and at distances corresponding to either the tight or loose binding regions observed in Section 4.3.2.1. The exact distribution across these or other modes differs between domains, but typically shows good agreement between the 'crystal' and 'simulation' derived landscapes (Figure 4.8).

A greater number of US/CG simulations were performed in this chapter than for GRP1-PH previously. To facilitate assignment of favourable binding sites without requiring manual inspection of full angle and contact distributions for all 52 systems (12 PH domains + 1 extra for ASAP1 × 2 PIPs × 2 initial structures), protein-lipid separation ranges corresponding to the tight and loosely bound states were formally defined relative to the global minimum of each profile, and the average φ

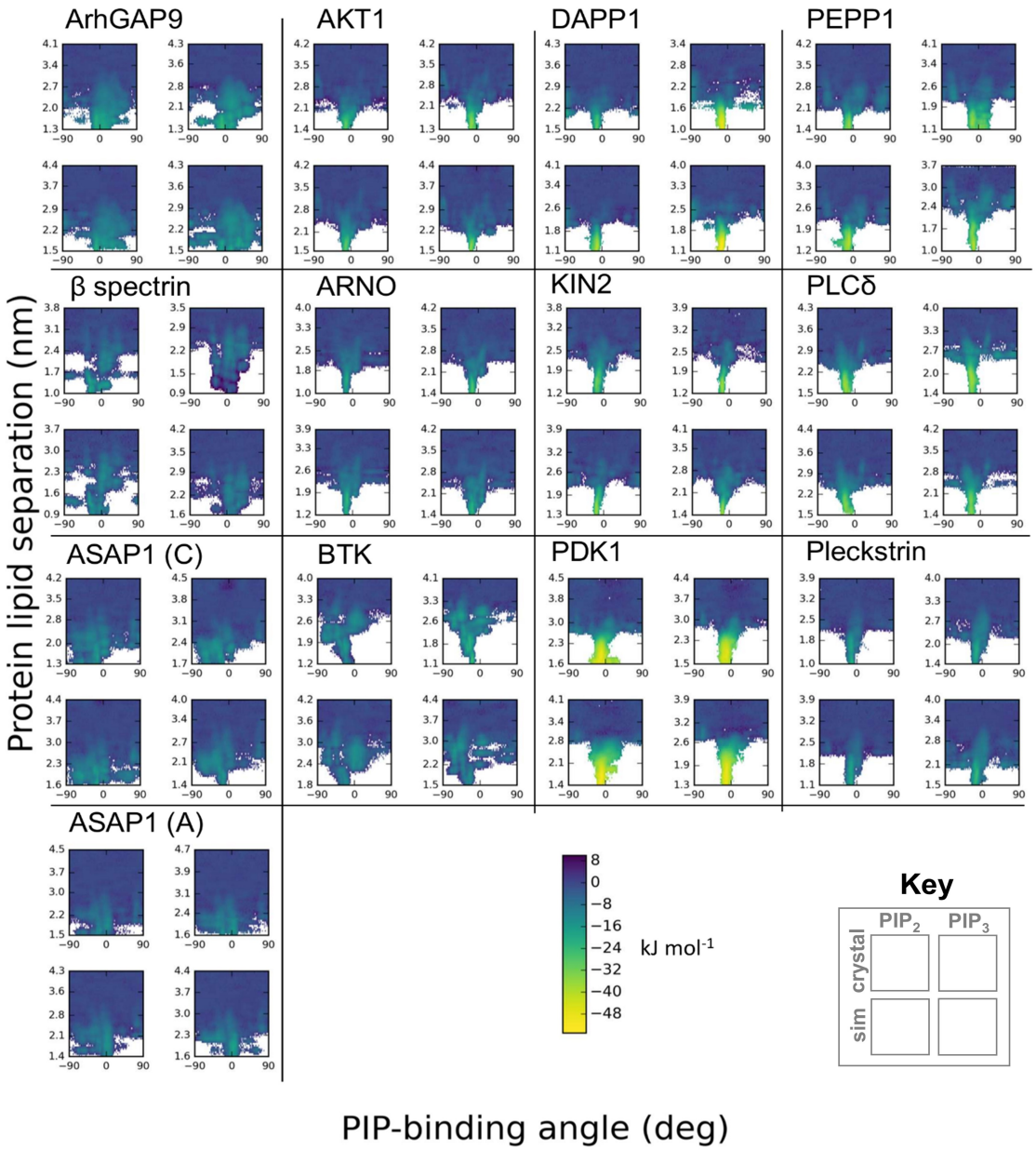


Figure 4.8: Energy landscapes of binding angle vs. separation. 2D energy landscapes of protein-lipid separation vs. φ PIP-binding angle for each system. Figure 4.9 B shows measurement of the PIP-binding angle. Landscapes are grouped by PH domain; the position of landscapes for the four systems containing a given PH domain (two lipids × two starting structures) are as indicated in the key, bottom right.

distribution over each range calculated (Figure 4.9). These regions were chosen to optimally encompass in turn the initial close-separation well observed in all PMFs and the secondary feature (e.g. a well, or plateau) observed in many profiles (Figure

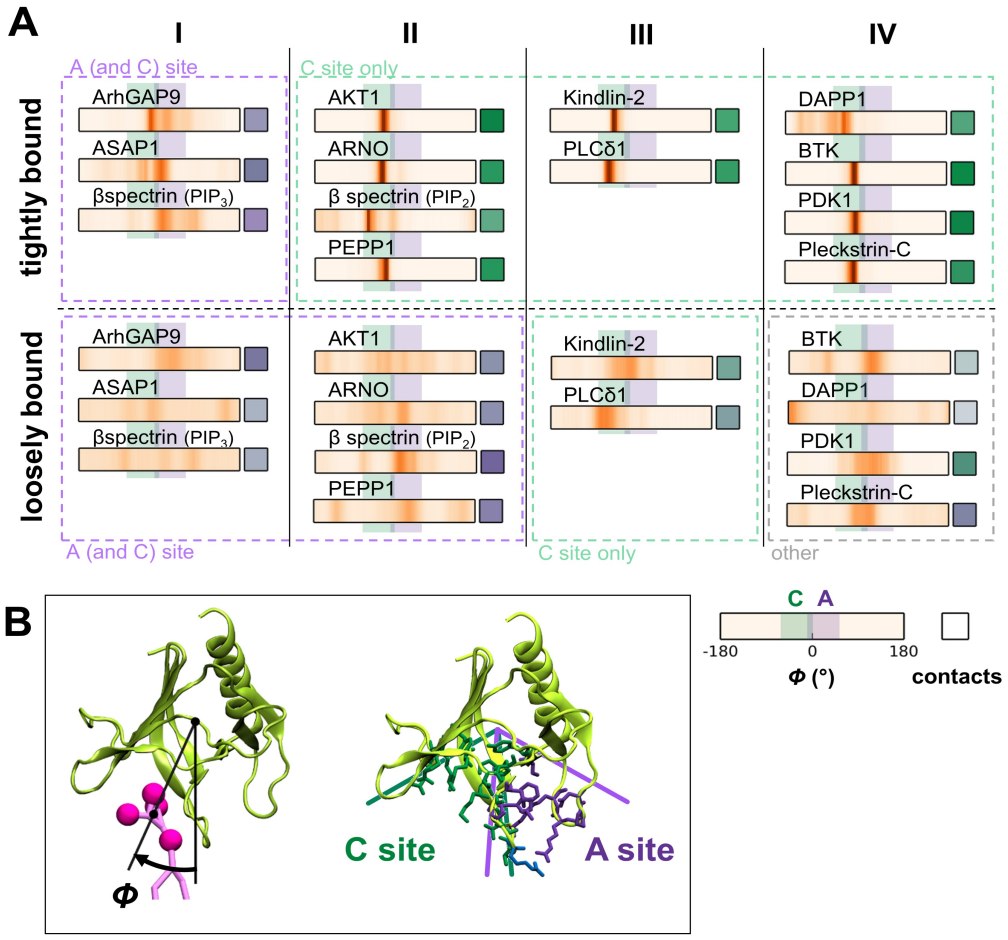


Figure 4.9: Site of PIP interactions. (A) Distribution of PIP-binding angle (orange) and fraction of contacts with A-site (purple) or C-site (green) residues for each PH domain in the 'tightly' and 'loosely' bound regions. Shown for PIP_2 (representative of PIP_3) unless otherwise indicated. Domains are grouped by the favourable binding site(s) in each region. Four main groups, showing different interaction sites between the two regions, are indicated. **(B)** Measurement of binding angle and A-site (purple) and C-site (green) residues shown for the ARNO PH domain.

4.5). The presence of A and/or C site interactions was then determined based on the fraction of states falling within the A or C angle ranges. See Section 4.2.3 for details of tight/loose range definition and assignment of A/C site binding.

The φ binging angle is a one-dimensional summary of a complicated, multi-dimensional interaction, and not necessarily representative of A/C site interactions; this was seen in Chapter 3 with the interaction 'mode 4', which showed a 'C-site'

φ angle but was located behind the C-site, as evidenced by θ binding angle and/or contact analysis. To reduce the level of data to be analysed across these 52 systems, only PIP contacts - being more intuitive and relatable to other studies - were measured here. The unbiased fraction of time each residue spends in contact with PIP vs. protein-lipid separation was calculated; good agreement was again seen between the 'crystal' and 'simulation' repeats. Figure 4.10 shows the structure-based sequence alignment of the 12 PH domains studied in this chapter with residues coloured by average degree of contacts over the tightly and loosely bound regions defined above; a noticeable shift in the pattern of interacting residues is observed for e.g. ARNO, PEPP1 and DAPP1, again indicating the changes in interaction mode.

To better identify A- and C-interactions from contact analysis, the positions of known A or C site residues were identified based on interactions observed with PIP analogues in PDB structures (boxed in Figure 4.10). The average number of A, C and total residues in contact with PIP over the tight and loose binding regions was calculated (Table 4.4), and presence of A and/or C site binding assigned according to the fraction of contacts that are 'A site' or 'C site', respectively. A 'contact score' indicating the relative level of A-site contacts vs. C-site contacts was also calculated and represented in Figure 4.9.

Comparison of the presence of A and/or C site interactions in the tight or loosely bound regions based on φ or contact analysis shows general agreement between the two measures (Table 4.5). In a small number of cases, the two measures disagree (e.g. for BTK and pleckstrin while loosely bound), suggesting that the interaction corresponds to another site near but not quite the A or C site; these modes were generally classified as 'other', and emphasise the necessity of using both φ and contact analysis to confirm binding site. Figure 4.9 summarises the binding site observations, and demonstrates that PH domains may be categorised in to four main groups based on the relative favourability of A and C site contacts while tightly or loosely bound:

Group I: PH domains which may bind PIP at the A-site at any separation, i.e. ASAP1, ArhGAP9, and β-spectrin (the latter when binding to PIP_3)

Group II: PH domains for which A-site interactions are observed only at larger

separations, i.e. ARNO, AKT1, PEPP1, and β-spectrin (the latter when binding to PIP_2);

Group III: PH domains for which C-site interactions are favourable at both short and larger separations, and which exhibit few A-site interactions, i.e. kindlin-2, PLCδ1

Group IV: PH domains for which other (i.e. non-A/C- site) interactions become predominant at larger separation, i.e. BTK, DAPP1, PDK1 and pleckstrin-C.

Excepting BTK-PIP_3, the interaction site(s) exhibited while tightly bound are consistent with the binding sites determined through experimental methods (Table 4.1), perhaps not surprising given the setup of the simulations. These groups will be discussed further in relation to their structural determinants in Section 4.3.3.

4.3.3.2 - *Modes of interaction on the membrane*

The binding modes of PH domains may be further distinguished by the orientation of the domain relative to the membrane. Following from GRP1-PH in Chapter 3, this was assessed by measuring the R_{ZZ} component of the rotation matrix of the PH domain relative to a reference structure. An R_{ZZ} of 1 corresponds to the 'canonical' binding mode while -1 indicates an 'upside down' mode. Unbiased 2D landscapes of R_{ZZ} vs. φ while tightly or loosely bound were calculated and averaged across all systems - weighting to allow equal contribution by each of the four groups identified

Figure 4.10 *(following page)*: Sequence alignment and PIP-contacting residues. Structure-based sequence alignment of PH domains used in this study. The fraction of frames in which each residue is observed to make sidechain contacts with PIP while 'tightly' (bottom) or 'loosely' (top) is indicated (orange). Specific residues and loop regions characteristic of A-site or C-site binding, determined form simulations, are indicated above or below the sequence in purple or green, respectively. Where an increase positive/polar residues in a given region corresponds to a particular binding orientation, the mode number per Figure 4.11 is indicated. Residues observed to contact PIP-analogues in the PDB structures are indicated by purple (A-site) or green (C-site) boxes. Conserved secondary structure elements are indicated below the sequence in black.
Sequences are taken from the appropriate PDB structures: 2POD (ArhGAP9), 1BTN (β-spectrin), 5C79 (ASAP1), 1UNQ (AKT1), 1U27 (ARNO), 2Z0P (BTK), 1FAO (DAPP1), 2LKO (Kindlin-2), 1W1D (PDK1), 1UPR (PEPP1), 1MAI (PLCδ1), 215C (Pleckstrin-C).

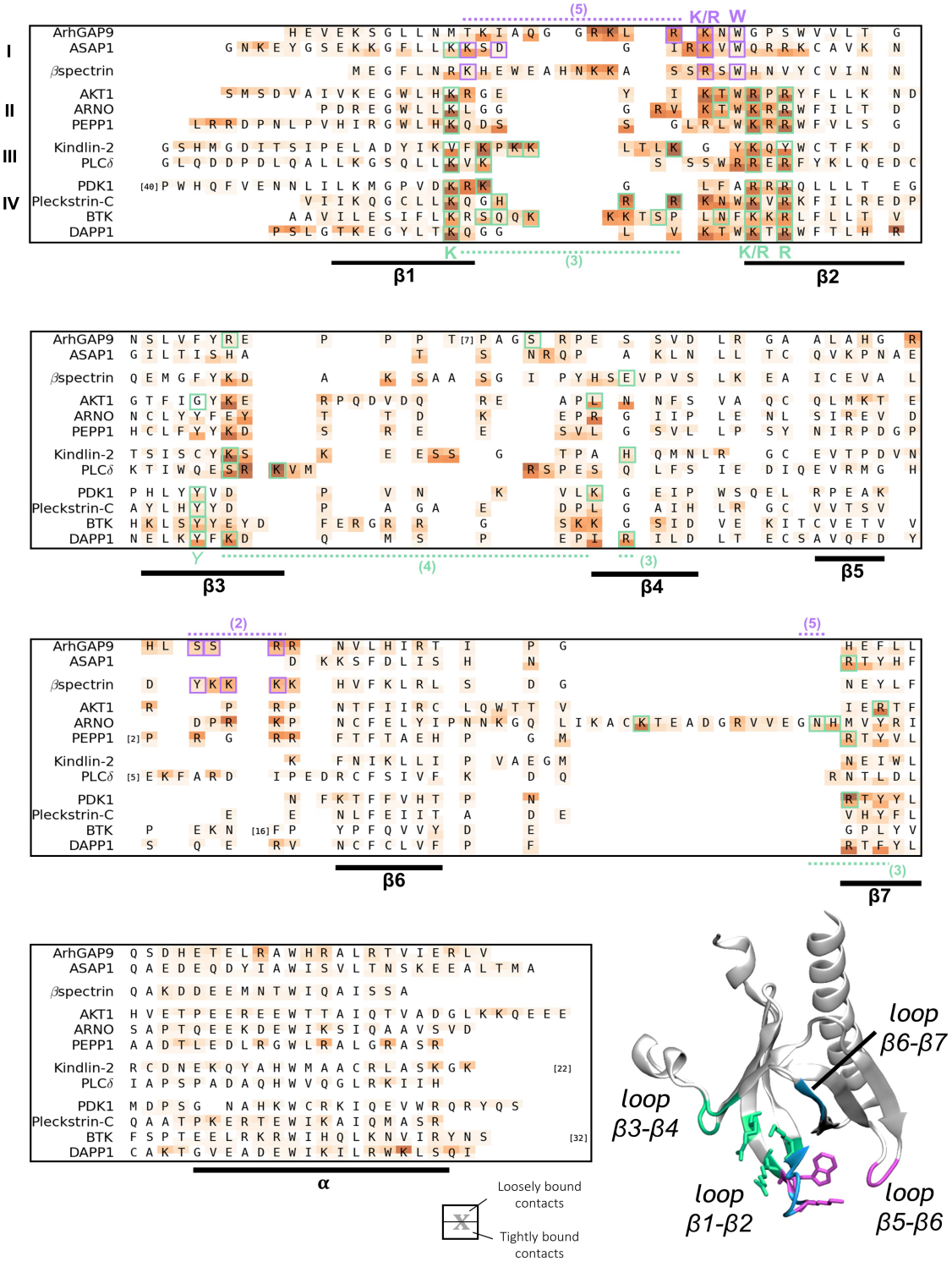


Figure 4.10: Sequence alignment and PIP-contacting residues. *Caption on previous page.*

Table 4.4: PIP contacts in the A and C binding sites. Average number of PIP sidechain contacts per frame with 'A site', 'C site' or any residue while in the 'tightly' or 'loosely' bound region for each system. Note that multiple contacts between a residue and PIP are only counted once and backbone interactions not considered; the actual number of protein-PIP contacts will be higher than indicated by 'total'. Error is based on the two repeats (crystal/simulation) for each system.

Protein	PIP	Average number of contacting residues					
		'Tightly' bound region			'Loosely' bound region		
		A site	*C site*		**A site**	**C site**	**All**
ArhGAP9	**PIP_2**	2.1 ± 0.1	1.6 ± 0.1	4.9 ± 0.3	1.0 ± 0.2	0.4 ± 0.1	2.4 ± 0.3
	PIP_3	2.0 ± 0.3	1.2 ± 0.9	5.2 ± 0.3	1.3 ± 0.1	0.4 ± 0.1	2.8 ± 0.3
AKT1	**PIP_2**	1.4 ± 0.1	5.0 ± 0.2	6.8 ± 0.3	0.6 ± 0.1	0.7 ± 0.1	2.2 ± 0.1
	PIP_3	1.1 ± 0.01	6.1 ± 0.1	7.7 ± 0.1	0.6 ± 0.2	0.8 ± 0.4	2.6 ± 0.6
ARNO	**PIP_2**	1.1 ± 0.03	5.0 ± 0.02	7.0 ± 0.01	1.0 ± 0.1	0.3 ± 0.1	2.1 ± 0.2
	PIP_3	1.5 ± 0.2	5.6 ± 0.8	8.2 ± 03	1.4 ± 0.2	0.5 ± 01	3.0 ± 0.4
ASAP1	**PIP_2**	1.4 ± 0.3	2.0 ± 0.4	5.2 ± 0.2	0.3 ± 0.1	0.3 ± 0.1	1.0 ± 0.3
	PIP_3	1.1 ± 0.4	1.5 ± 0.1	4.7 ± 0.3	0.2 ± 0.04	0.3 ± 0.1	0.8 ± 0.1
β-spectrin	**PIP_2**	0.7 ± 0.3	2.5 ± 0.1	5.8 ± 0.2	1.8 ± 0.2	0.5 ± 0.2	3.1 ± 0.2
	PIP_3	1.6 ± 0.1	0.4 ± 0.2	2.6 ± 0.1	0.0 ± 0.01	0.0 ± 0.01	0.0 ± 0.01
BTK	**PIP_2**	0.1 ± 0.2	1.8 ± 1	4.3 ± 2	0.0 ± 0.01	0.5 ± 0.7	1.5 ± 2
	PIP_3	0.1 ± 0.2	3.8 ± 2	6.6 ± 2	0.0 ± 0.04	1.3 ± 1	2.9 ± 2
DAPP1	**PIP_2**	0.8 ± 0.1	5.7 ± 0.5	7.0 ± 0.5	0.3 ± 0.04	0.4 ± 0.3	2.6 ± 0.9
	PIP_3	0.8 ± 0.01	7.3 ± 0.5	8.5 ± 0.5	0.3 ± 0.04	0.3 ± 0.04	3.8 ± 0.8
Kindlin-2	**PIP_2**	0.2 ± 0.1	5.3 ± 0.5	7.2 ± 0.5	0.1 ± 0.01	1.2 ± 0.4	2.9 ± 0.3
	PIP_3	0.4 ± 0.01	5.5 ± 0.1	7.8 ± 0.1	0.1 ± 0.01	2.4 ± 0.02	3.5 ± 0.4
PDK1	**PIP_2**	0.0 ± 0.02	4.9 ± 0.01	5.4 ± 0.01	0.1 ± 0.01	1.6 ± 0.5	2.6 ± 0.3
	PIP_3	0.0 ± 0.02	4.0 ± 0.4	4.3 ± 0.4	0.0 ± 0.01	1.5 ± 0.1	2.3 ± 0.2
PEPP1	**PIP_2**	0.1 ± 0.04	5.1 ± 0.1	6.0 ± 0.3	0.6 ± 0.01	0.3 ± 0.2	2.5 ± 0.1
	PIP_3	0.3 ± 0.3	6.1 ± 1	7.5 ± 0.7	0.7 ± 0.1	1.0 ± 1	3.1 ± 0.3
PLCδ1	**PIP_2**	0.0 ± 0.02	5.3 ± 0.7	7.3 ± 0.5	0.1 ± 0.1	1.3 ± 0.3	2.6 ± 0.3
	PIP_3	0.1 ± 0.03	5.4 ± 0.1	7.8 ± 0.1	0.1 ± 0.1	1.6 ± 0.4	2.9 ± 0.6
Pleckstrin C	**PIP_2**	0.7 ± 0.04	5.2 ± 0.04	6.1 ± 0.02	1.8 ± 0.5	0.6 ± 0.1	3.0 ± 0.4
	PIP_3	0.8 ± 0.1	4.6 ± 0.3	5.7 ± 0.2	1.0 ± 0.2	0.3 ± 0.1	1.5 ± 0.3

- revealing several common binding modes (Figure 4.11) labelled 1-5 (tightly bound) and 0, 1* and 2* (loosely bound). The individual landscapes for each US/CG system are also shown in Figure 4.12 and labelled correspondingly; again, general good

agreement between 'crystal' and 'simulation' results is seen. Additional modes were generally unique to each domain and are labelled X-Z, with Z corresponding to the collection of 'upside down' modes. Representative structures for the binding modes observed for each PH domain are shown in Figure 4.13, and summarised in Figure 4.11.

While tightly bound, 5 main modes are identified: three C-site interacting modes, and two A-site (Figure 4.11). The most favourable mode of each set corresponds to the 'canonical' orientation ($R_{ZZ} \sim 1$); the C-site 'mode 1', exhibited by AKT1, ARNO, ASAP1, DAPP1, kindlin-2, PDK1, PEPP1, pleckstrin-C and matching 'mode 1' observed for GRP1-PH in Chapter 3; and the A-site 'mode 2', exhibited by ASAP1 and β-spectrin (Figure 4.12). Interactions with each the A and C site are also observed with $R_{ZZ} \sim 0.5$ - the C-site 'mode 3' (ArhGAP9, DAPP1, pleckstrin-C (PIP_2 only) and to a lesser extent AKT1) and A-site 'mode 5' (ASAP1 and to a lesser extent ArhGAP9 and β-spectrin) (Figure 4.12). Representative structures (Figure 4.11, Figure 4.13) indicate these orientations correspond the PH domain rolling forward onto its 'front' face. A final C-site mode - 'mode 4' - with R_{ZZ} under 0.5 is exhibited by β-spectrin (PIP_2 only), BTK and PLCδ1 (Figure 4.12), and corresponds to sideways rotation (Figure 4.11). Alternate membrane-binding orientations were also identified form unrestrained MD simulations for AKT1, ASAP1, DAPP1, PLCδ1 and pleckstrin-C in ref. [104].

The combined map for the 'loosely' bound interactions of PIPs reveals both A- and C-site binding modes, although distinct orientations relative to the bilayer normal are no longer observed; these modes are collectively labelled as 1* (AKT1, β-spectrin, kindlin-2, PEPP1, PLCδ1, pleckstrin-C) and 2* (ArhGAP9, AKT1, ARNO, β-spectrin and PEPP1) (Figures 4.11 and 4.12). A third mode, 'mode 0', in which φ provides no clear distinction between A or C site while the domain is still in the canonical orientation, was identified for several domains (ASAP1, PDK1, pleckstrin-C); representative structures indicate this corresponds to interaction of PIP with the loop separating the two sites, rather than binding 'in' either side (Figure 4.11, Figure 4.13). Additional 'other' modes were also seen, but these were largely adopted by only

Table 4.5: Favourable PIP binding sites in the tightly and loosely bound regions. Presence of interactions of PIP with the 'A' and/or 'C' site, based on contact analysis (Table 4.4) or PIP binding angle (Figure 4.9), and the final assignment of favourable binding site(s). Sites listed in brackets were 'potentially' present based on the corresponding analysis (see Section 4.2.3 for determination of 'present' or 'potentially present').

Protein	PIP	PIP binding site					
		'Tightly' bound region			'Loosely' bound region		
		contact analysis	*angle analysis*	**Conclusion**	*contact analysis*	*angle analysis*	**Conclusion**
ArhGAP9	**PIP_2**	A C	A C	**A C**	A	A	**A**
	PIP_3	A (C)	(A C)	**A C**	A	A	**A**
AKT1	**PIP_2**	(A) C	C	**C**	(A C)	(A C)	**A C**
	PIP_3	C	C	**C**	(A C)	(A C)	**A C**
ARNO	**PIP_2**	C	C	**C**	A	(A C)	**A**
	PIP_3	C	C	**C**	A	(A C)	**A**
ASAP1	**PIP_2**	(A) C	(A) C	**A C**	(A C)	(C)	**C**
	PIP_3	(A C)	(A C)	**A C**	(A C)	(C)	**C**
β-spectrin	**PIP_2**	C	(A) C	**C**	A	A	**A**
	PIP_3	A	A	**A**	(A)	-	**-**
BTK	**PIP_2**	C	C	**C**	(C)	A (C)	**-**[a]
	PIP_3	C	C	**C**	C	(A)	**-**[a]
DAPP1	**PIP_2**	C	C	**C**	-	-	**-**
	PIP_3	C	C	**C**	-	-	**-**
Kindlin-2	**PIP_2**	C	C	**C**	C	(A C)	**C**
	PIP_3	C	C	**C**	C	(A) C	**C**
PDK1	**PIP_2**	C	C	**C**	C	A (C)	**-**[a]
	PIP_3	C	C	**C**	C	A (C)	**-**[a]
PEPP1	**PIP_2**	C	C	**C**	(A)	(A)	**A**
	PIP_3	C	(A) C	**C**	(A C)	(A C)	**A C**
PLCδ1	**PIP_2**	C	C	**C**	C	C	**C**
	PIP_3	C	C	**C**	C	C	**C**
pleckstrin	**PIP_2**	C	C	**C**	A	(C)	**-**[a]
C	**PIP_3**	C	C	**C**	A	(C)	**-**[a]

[a] In these cases, the binding site indicated from contact and binding angle analyses disagree; this suggests that PIP is instead interacting in some 'other' site.

one domain e.g. DAPP1 or BTK while loosely bound, and so are shown indidually in Figure 4.13, but not in Figure 4.11.

4.3.4 - *From sequence to structure to function: determining PIP binding of PH domains*

4.3.4.1 - *What determines binding site?*

Examination of the structure-based sequence alignment of the PH domains studied here (Figure 4.10) reveals that those which exhibit predominately C-site interactions while tightly bound (i.e. groups II, III and IV) possess the K-X_n-(K/R)-X-R motif in the β1 and β2 strands identified previously as a requirement for PIP binding [19, 20]. Additional C-site residues are found on the β1-β2, β3-β4 and β6-β7 loops.

Meanwhile, domains that show A-site interactions while either tightly or loosely bound possess a short (K/R)-X-W motif, also located in the β1-β2 loop region, but out of phase by one residue such that the sidechains face the opposite direction (i.e. in to the A site) to those in the C-site motif. Most also possess a long β5-β6 loop containing multiple positively charged residues. Neither of these features are exhibited by C-site only binding domains (Group III) (Figure 4.10).

Identifying Group IV domains is more difficult due to the lack of one particular interaction mode among these domains. PDK1 and pleckstrin-C both possess the β1-β2 C-site motif but, unlike Group III domains, no positive charges in the β3-β4

Figure 4.11 ***(following page)*****: Common membrane binding modes. (A)** Density map of orientation (R_{ZZ}) vs. PIP binding angle (φ) in the 'tightly' and 'loosely' bound regions based on data from all PH-PIP systems, revealing several common binding modes. Results are normalised within each Group I-IV before averaging to the final landscape. Only a section of the full 2D landscape, containing the most frequently observed binding modes, is shown in each case. Measurement of R_{ZZ} and φ, indicated for the PH domain of ARNO, is shown in the inset. R_{ZZ} is measured relative to the reference structure in grey, corresponding to the initial GRP1-PIP bound structure used in Chapter 3. This orientation corresponds to an R_{ZZ} of 1. **(B)** Representative structures of each of the binding modes identified in A. Modes adopted by only one domain e.g. for DAPP1 and BTK while loosely bound are not shown. PIP is shown in magenta; the membrane is shown in grey with phosphate groups shown explicitly in light (POPC) or dark (POPS) grey. Protein is shown by the cartoon representation of the aligned crystal structure, with the main loops that bound the A- and C- sites labelled. The binding landscapes and representative structures of each individual system are shown in Figures 4.12 and 4.13.

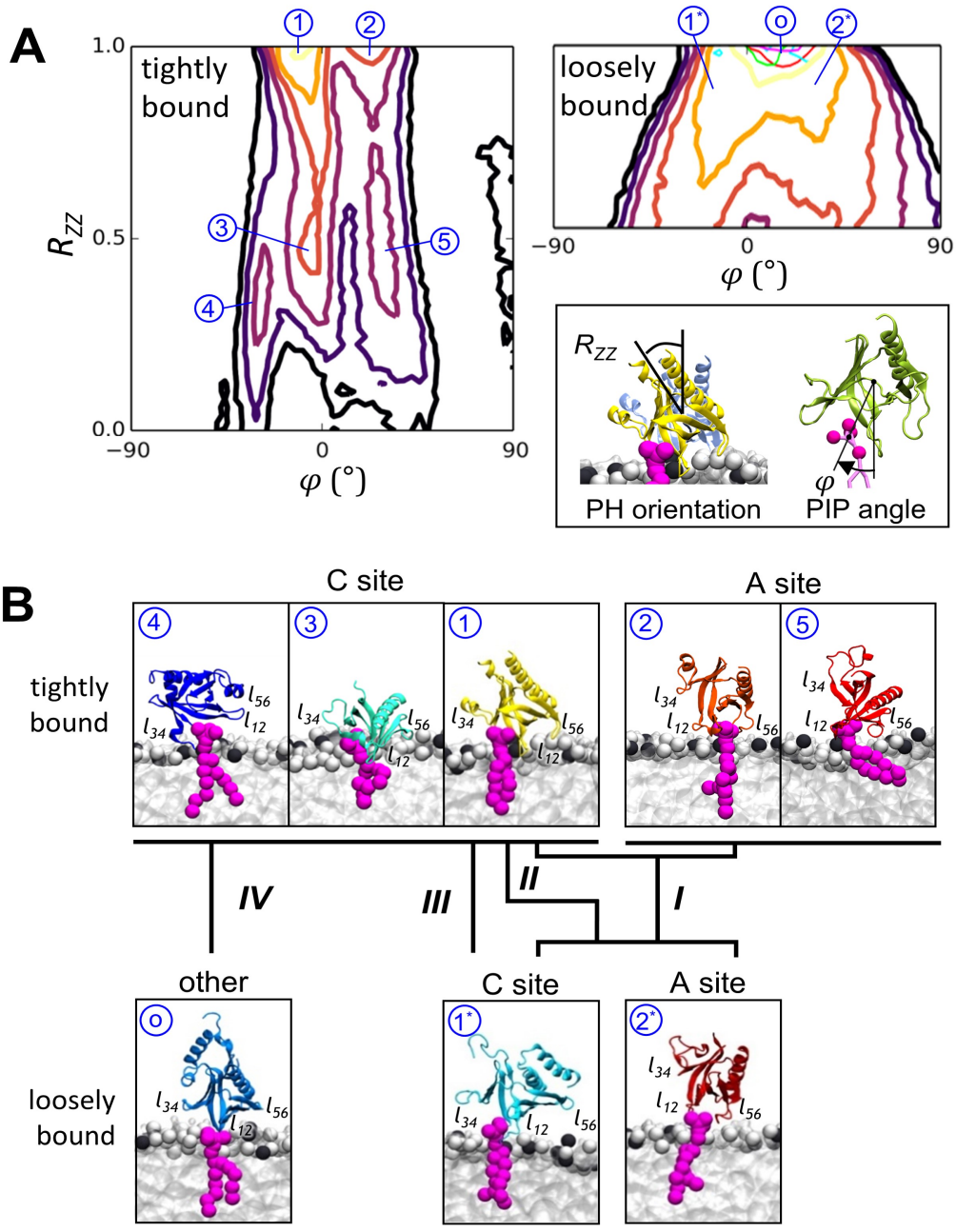


Figure 4.11: Common membrane binding modes. *Caption on previous page.*

loop. Pleckstrin-C also possesses the β1-β2 A-site motif but again lacks the other A-site feature (β5-β6), while ASAP1 lacks both this secondary A-site feature and the C-site features; that is, the PIP-interacting residues of all three domains are primarily located on the 'central' β1-β2 loop with few to none on the 'flanking' β3-β4 and β5-β6 loops. This is consistent with these domains interacting with the β1-β2 loop rather than strictly in either site while loosely bound (i.e. mode 0).

DAPP1 contains both A- and C- motifs and so would be expected to be part of Group II; however, a unique 'upside down' orientation is seen when this domain is loosely bound, leading to its Group IV classification. This alternate orientation involves PIP interactions with the end of the PH domain helix, which has previously been predicted as a potential membrane interaction site [169] but may not represent a physiological binding mode when considering the full-length parent protein. An upside-down interaction mode of DAPP1 with the membrane was also observed in unrestrained binding simulations [104]. It is not clear why this interaction is favoured above A-site binding.

Thus, in summary: Group I domains are generally those lacking the K-X_n-(K/R)-X-R motif but containing (K/R)-X-W sequence; Group III domains possess the K-X_n-(K/R)-X-R and a β3-β4 containing multiple positive/polar residues, but lacking the (K/R)-X-W sequence and multiple positive/polar residues in the β5-β6 loop; Group II domains possess both sequences and a positive/polar β5-β6 loop; while Group IV contains the domains not otherwise classified and additional domains due to as-of-yet undetermined factors. This summary is provided as a flow chart in Figure 4.14.

Figure 4.12 *(following page)*: Binding mode landscapes for each PH/PIP system. Density map of orientation (R_{ZZ}) vs. PIP binding angle (φ) in the 'tightly' and 'loosely' bound regions for each PH/PIP system, averaged between 'crystal' and 'simulation' repeats. The contour profile of the combined data set (from Figure 4.11) is shown in grey for reference and binding modes labelled using the same numbering as Figure 4.11. Note that there is not always an exact overlap of an interaction mode between an individual domain and the average result, likely due to differences in the loop structure of each domain affecting binding angle and orientation. Additional modes are labelled X-Z. Representative structures of each mode for each system are shown in Figure 4.13.

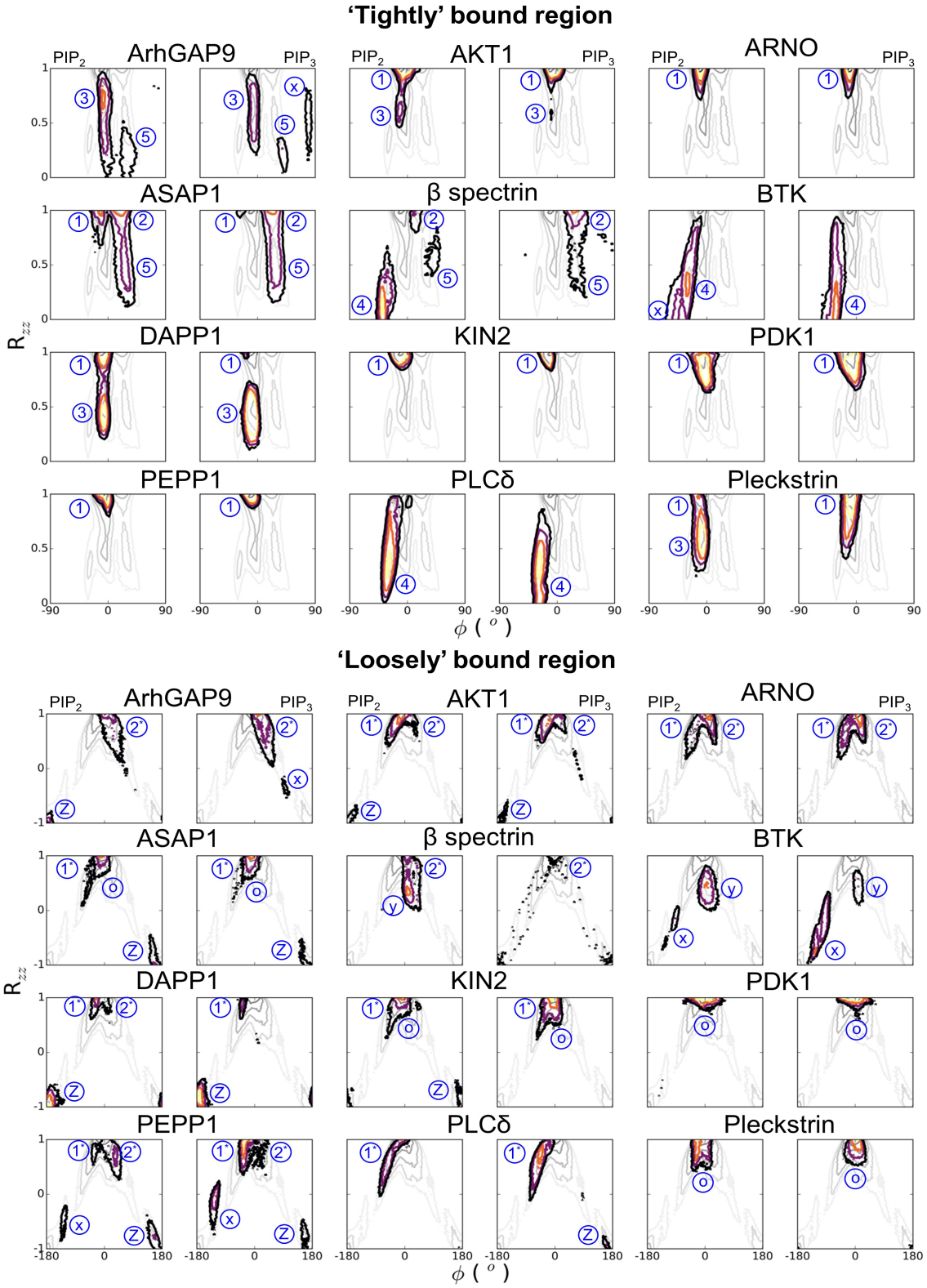


Figure 4.12: Binding mode landscapes for each PH/PIP system. *Caption on previous page.*

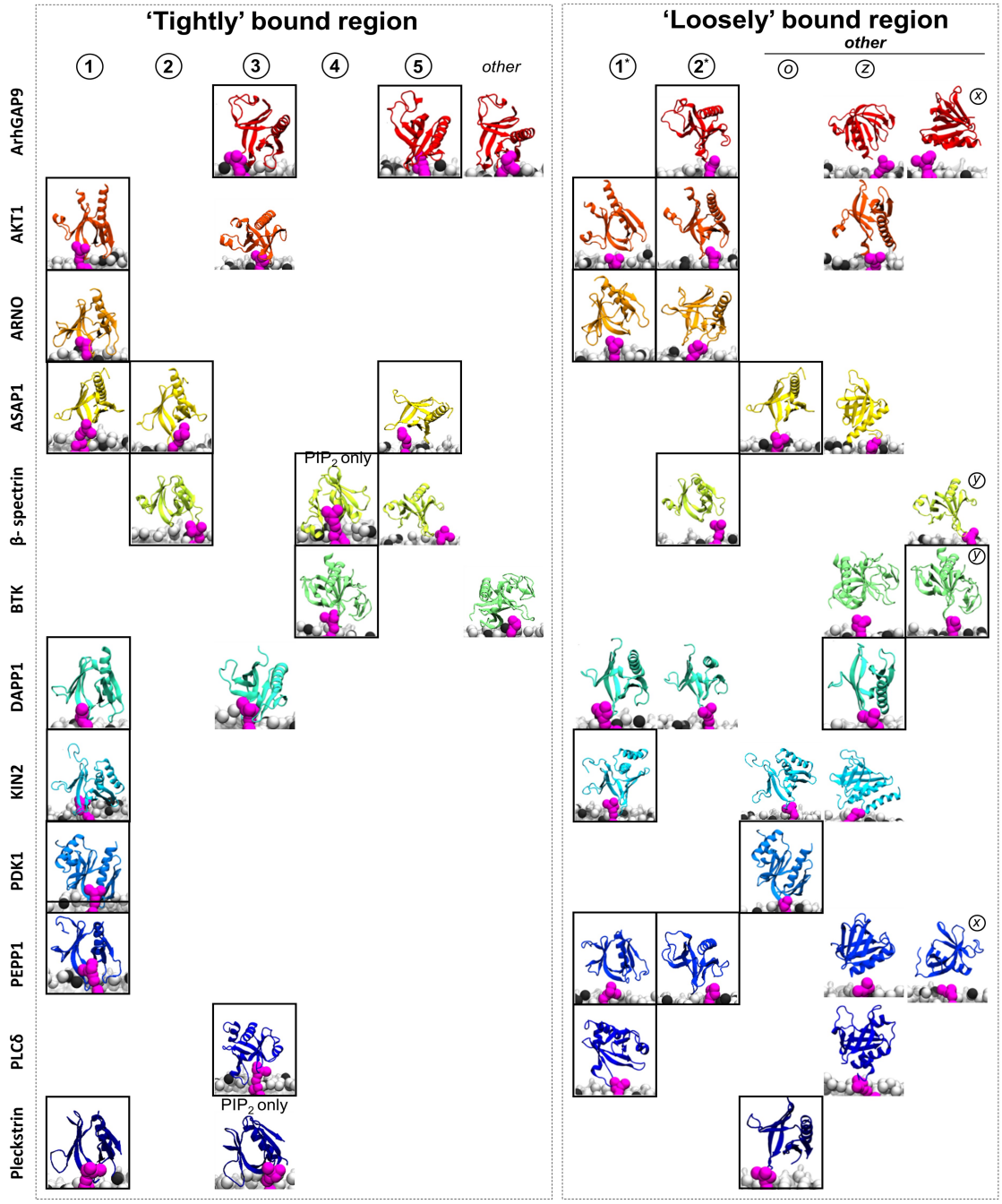


Figure 4.13: Representative structures of PIP-binding modes for each PH domain. Representative snapshots of the binding modes labelled in Figure 4.12. Boxed structures indicate the favoured binding mode(s).

4.3.4.2 - *What determines binding mode?*

The 'loose' binding modes can be distinguished by binding site alone; determinants of this have been discussed in the previous section. Similarly, modes 1, 3 and 4

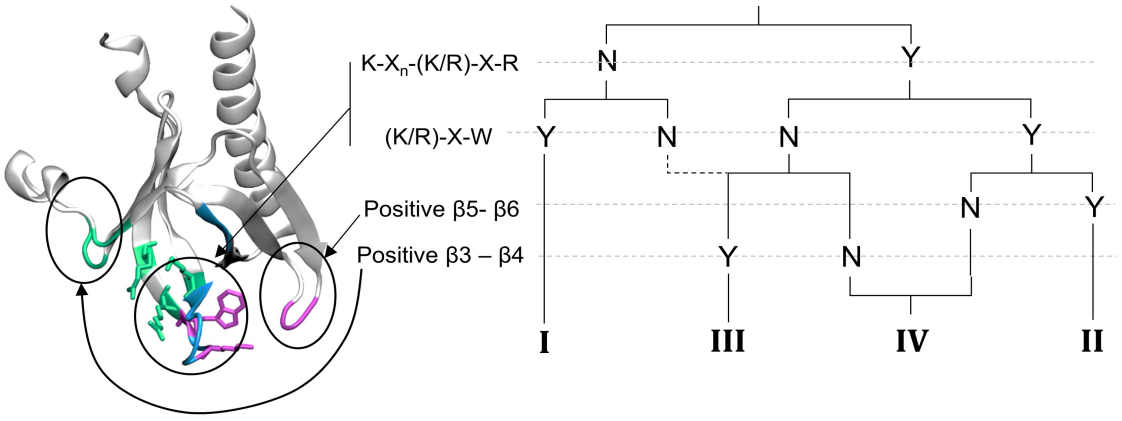


Figure 4.14: Determining binding site from sequence. Flowchart indicating the classification of PH domains into the four groups identified in Figure 4.9 based on the presence of certain sequence features, namely presence of motifs in the β1-β2 loop and charge distribution in flanking loops.

vs. modes 2 and 5 are distinguished by binding at the C vs. A site, also discussed above. The factors that lead to the different orientations within each of these groups is however difficult to pin down definitively due to the low number of domains exhibiting some of these interaction modes; nonetheless, some trends can be identified relating to the distribution of positively charged residues relevant loops. The β1-β2 loop in general contains the largest number of important PIP-interacting residues; in the 'canonical' orientation, a C-site bound PIP will be positioned along the inside of the β1-β2 curve, thus maximising these interactions, while an A-site PIP will be positioned between the β1-β2 loop and β3-β4 loop, the other main contributor of A-site interactions (Figure 4.10); this likely explains why these canonical orientations (modes 1 and 2, respectively) are most commonly seen.

Additional charges along the 'front' face of the PH domain (at the end of the β1-β2 loop for AhrGAP9, β-spectrin and pleckstrin-C, or the β5-β6 loop for DAPP1, AKT1, ASAP1; Figure 4.10) result in binding modes in which the PH domain rotates forward so PIP is in better contact with these residues, leading to 'mode 3' and 'mode 5' for C- and A-site interactions, respectively (Figure 4.11). Meanwhile, PH domains with a large number of positive/polar residues in the β3-β4 loop (e.g. PLCδ1, BTK and β-spectrin; Figure 4.10) bind in an orientation that allows PIP to align with this loop

when at the C-site - requiring a ∼90° rotation relative to the canonical orientation and thus leading to 'mode 4' (Figure 4.11). This mode is not observed for kindlin-2, despite also possessing strongly interacting residues in the β3-β4; this may be due to the longer, less structured nature of this loop in kindlin-2 (Figure 4.1), which might therefore be expected to be more flexible and able to contribute these interactions without requiring a change in orientation. PLCδ1 was also noted to form significant contacts with the β3-β4 loop in unrestrained MD binding simulations [104].

4.3.4.3 - *What determines selectivity/affinity?*

The identity of the PIP-binding site appears to be a good indicator of affinity, with the PH domains showing higher affinities being those with more C-site interactions (e.g. Group III; Figure 4.15 A), while a greater degree of A-site over other (non-canonical) interactions correlates with weaker binding (Figure 4.15 B). There did not appear to be a similar trend involving specific binding mode/membrane orientation. The generally greater strength of C-site interactions is likely due to the more 'closed' nature of this site, such that the PIP molecule sits inside the curve of the β1-β2 loop thus allowing a greater number of interacting residues. In contrast the A-site, on the opposite side of this loop, is more open. The features predicting which group a PH domain belongs to, discussed in Section 4.3.4.1, may thus be considered as predictors for overall affinity of a PH domain for PIPs.

Of the three domains that exhibit selective binding in this study (i.e. β-spectrin, DAPP1 and AKT1), selectivity of β-spectrin can most clearly be linked with the binding mode. For PIP_3, β-spectrin shows largely A-site interactions when tightly bound, while PIP_2 is able to interact more strongly with the C-site (Figure 4.9), seen in the PMF profiles as an additional deeper well (Figure 4.5). β-spectrin binds PIP_3 in mode '4', in which the PIP molecule is aligned along the β3-β4 loop (Figure 4.11). In β-spectrin, this loop extends away from the C-site at a larger angle than for other domains that bind in this orientation, leading to a relatively shallow binding site likely less accommodating for the larger headgroup of PIP_3. Consequently, PIP_2 interactions are favoured over PIP_3 for this domain.

DAPP1 was observed to be highly PIP_3 selective. DAPP1 interacts with PIP when tightly bound with two possible orientations ('1' and '3' in Figure 4.11); the latter

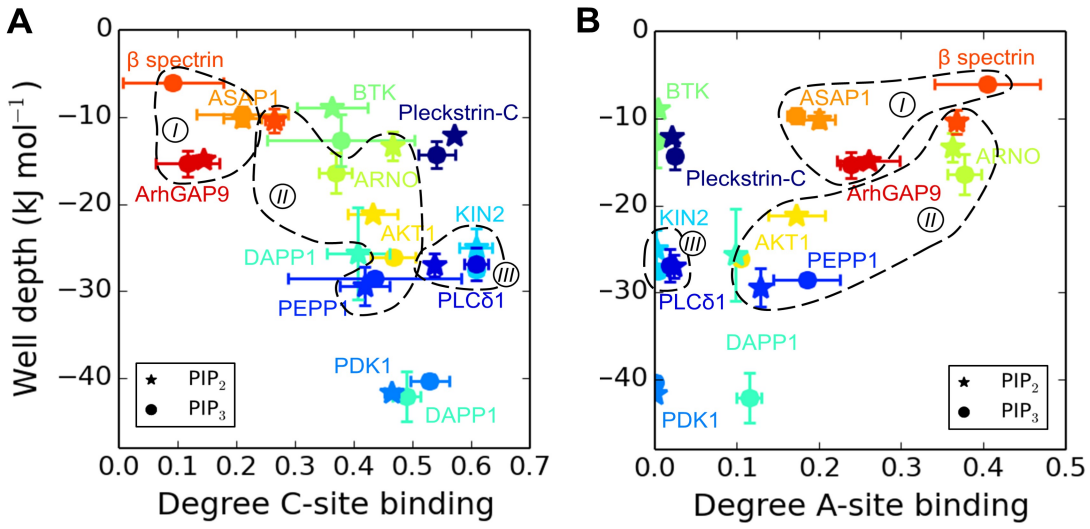


Figure 4.15: Relation between binding site and affinity. Average well depth for each PH-PIP system against **(A)** degree of C-site binding (as a fraction of all interactions) and **(B)** degree of A-site binding (as a fraction of all non-C site interactions). Domains belonging to Groups I-III are indicated.

allows additional electrostatic interactions and is observed to a greater extent with PIP_3, perhaps due to the higher charge, leading to the observed selectivity. AKT1 was also observed to be PIP_3 selective, though there is no clear difference in binding mode between PIP_2 and PIP_3. This selectivity may also be simply the result of the increased charge of the PIP_3 molecule. It is not clear why DAPP1 exhibits a much greater degree of selectivity than any other domain.

Overall, given the low degree of selectivity observed here, it remains difficult to isolate clear trends which determining PIP_2 vs. PIP_3 selectivity across the PH family, though interactions with PIP_3 do appear in general to be more energetically favourable (Figure 4.6 B, Figure 4.7 B), as would be expected given its greater charge. Determinants of selective binding may be revealed following investigation of a more extended set of PH domains with MD simulations.

4.3.5 - *Additional PH/PIP systems*

4.3.5.1 - *The E17K-AKT1 mutant*

To further examine the sensitivity of PMFs to the details of PH domain structure and the degree of agreement of the calculated PMFs with available experimental data,

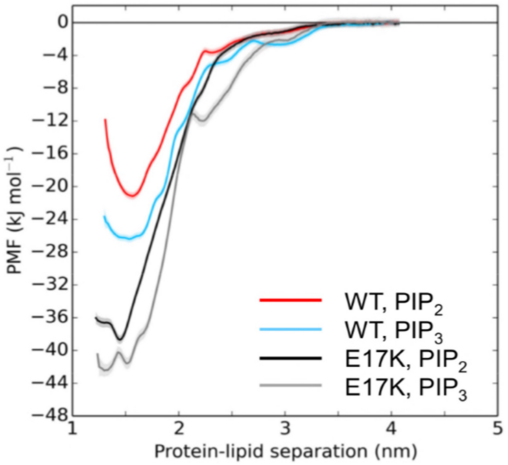


Figure 4.16: Effect of the AKT1-PH E17K mutation. PMF profiles for wild-type AKT1 (from the 'crystal' initial configuration) and AKT1-PH-E17K binding to PIP_2 (red and black, respectively) or PIP_3 (blue and grey, respectively). Bootstrap errors are shown.

the effect of a single amino acid mutation on interactions with PIPs was explored. The oncogenic E17K mutation in the AKT1 PH domain was selected due to its clinical relevance, and as it has been characterised both in vitro and in vivo [16]. This mutation is found in patients with breast and colorectal cancers, and results in a substantive increase in the affinity of the AKT1 PH domain for PIP_2 and PIP_3. PMF profiles were calculated for AKT1-E17K-PH binding PIP_2 and PIP_3, using the method outlined in Section 4.2. Only one set of US/CG simulations for each system was performed. Comparing with results from wildtype AKT1, these calculations suggest that the mutation does indeed result in a large decrease, around -20 kJ mol^{-1}, in the free energy of binding AKT1-PH domain to PIP_2 and PIP_3 - i.e. a significant increase in affinity, consistent with experimental reports (Figure 4.16).

4.3.5.2 - *An additional test for initial system configuration bias: C-site ArhGAP9*

Two US/CG simulations were performed for each system to assess the bias introduced by the choice of initial configuration; agreement between profiles suggested this bias was not significant. However, while the 'crystal' and 'simulation' initial structures used possessed different membrane orientations, all pairs shared the binding site (A or C).

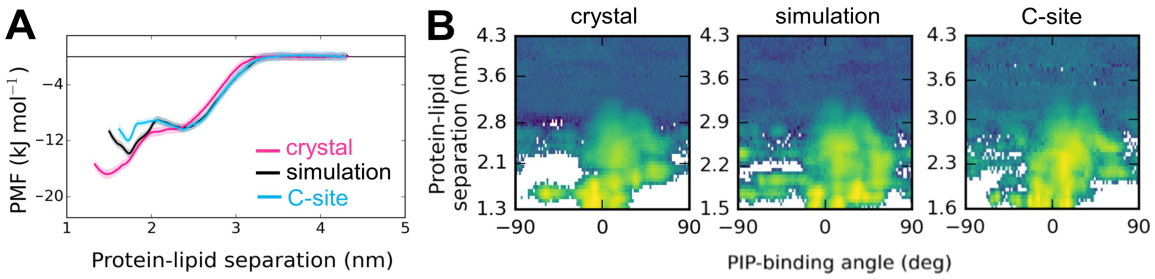


Figure 4.17: Effect of initial PIP binding site for ArhGAP9. (A) PMF profiles for ArhGAP9 binding to PIP_3 and **(B)** energy landscapes of separations vs. binding angle, based on US simulations from initial configurations with PIP bound in the A-site ('crystal' and 'simulation' as shown in Figure 4.5/Figure 4.8) or the C-site. Snapshots of initial structures are shown in Figure 4.4.

Structural analysis suggested that A vs C transitions at low protein-lipid separations are rarer that reorientations on the membrane, and as such the initial PIP interaction site may be more likely to bias US results that choice of orientation. To test this, an additional US/CG simulation was carried out using an initial structure in which the ArhGAP9 PH domain was bound to PIP at the C site. The resulting PMF profiles closely matched those generated from the original initial structures in which PIP started in the A site (Figure 4.17). This demonstrates that, at least in the case of ArhGAP9, transition between A and C sites are possible and occur on sufficient timescales such that the initial bound site does not affect results.

4.4 - Further comments of dual binding sites: role of the A site?

This study has demonstrated A-site interactions for several PH domains, most clearly for ArhGAP9, ASAP1, and β-spectrin but also for AKT1, ARNO and PEPP1, further expanding the list of PH domains exhibiting 'dual binding' as identified through experiment [123] and primarily computational studies [104, 127, 130]. While only single-PIP interactions were examined in this chapter, the majority of A or C binding modes observed (modes 1, 2, 3 and 5; Figure 4.11) present both sites towards the membrane, indicating simultaneous binding of two PIP molecules - one in each site - is possible. When investigated in Chapter 3, such dual binding greatly increased the calculated binding free energy of GRP1-PH to the membrane; a further increase

was seen on addition of more PIPs, reaching the range seen experimentally. A similar effect would be expected for the domains studied here and explains the discrepancy observed between the computational and experimentally reported binding energies.

The A site would thus seem to play an important role in allowing the high-affinity binding characteristic of many PH domains. It is noted that fewer C-site interactions correlate with weaker binding to a single PIP, but also stronger A-site interactions - the A-site may 'compensate' for weaker C-site interactions to insure high-affinity binding is still achieved. The presence of favourable A-site modes at greater protein-lipid separation may also assist in the recruitment the PH domain to the membrane, providing a more accessible transition state before the tight, productive binding mode. The presence of two distinct interaction sites may also ensure binding in a 'productive' orientation.

It is also possible that the two sites may preferentially interact with different lipids, including non-PIP species, allowing more targeted recruitment to membranes only where both species are present. Indeed, of those domains studied here, experimental studies have suggested that the A-sites of AKT1 [171] and PDK1 [162] may bind phosphatidylserine. Large-scale studies have suggested this 'coincidence detection' plays an important role in the targeting of a large number of PH and lipid binding domains in general [136, 172].

4.5 - Caveats of the US/CG method: an updated discussion

Some concerns of the US/CG methodology were introduced in Section 3.5; namely, the use of a coarse-grained force field, and the potential for incomplete sampling. Further to Chapter 3, the computational results here regarding PIP selectivity are generally in qualitative agreement with available experimental results, and quantitative disagreements are likely explained by the different lipid environments used. Testing of an additional single amino-acid mutation highlighted that the coarse-grained force field is still sensitive on a residue scale. An atomistic umbrella sampling approach, recently demonstrated for a PH/PIP interaction [127] may offer a higher resolution picture of free energy landscapes, but is likely to considerably complicate the second issue - that of achieving adequate sampling and convergence. PIP/protein interactions may

vary from initial PDB-derived configurations over timescales large in comparison to the standard length of MD simulations; a recent study of potassium channel/PIP interactions [173], for example, saw changes in hydrogen bonding pattern at greater than 100 ns time scales. Sampling and convergence issues are likely for atomistic PMFs, limiting the usefulness of this approach particularly when considering multiple systems as presented here.

The reasonable results obtained here also suggests that adequate sampling/convergence was achieved, particularly the agreement between multiple initial structures, differing by both membrane orientation or binding site presented in Section 4.3.3. It is still possible that orthogonal coordinates are varying on timescales larger than the simulation time, and lack of a bias due to initial binding site seen for ArhGAP9 may not be representative of other domains. However, exhaustive tests of sampling and convergence undermine the aim being able to efficiently analyse multiple protein/membrane systems, and the work here was deemed suitable for the current time.

Another issue raised here is that of choice of reaction coordinate and representative structural coordinates - while an obvious choice, the protein-lipid separation reaction coordinate used here does not itself distinguish between different interaction sites or modes, the feature of interest, and in both Chapter 3 and here multiple measures have had to be used to confirm e.g. A vs. C site binding, and it is possible there are hidden modes that are being overlooked. A more sophisticated choice of reaction coordinate, taking in to account several factors, may directly and more efficiently reveal the different interaction modes of choice.

4.6 - Conclusions

An umbrella sampling/coarse-grained molecular dynamics approach was used to calculate one dimensional free energy profiles for the binding of 12 different PH domains to model membranes containing either a PI(4,5)P_2 or PI(3,4,5)P_3 lipid molecule, allowing a molecular energetic understanding of the complexities of the interactions of the PH domains with PIP molecules in membranes. The domains were found to exhibit a range of selectivities (consistent with existing results) and affinities, with several distinct binding modes observed that could be related to the structural

and sequence differences between the different PH domain species. The presence of A-site interactions was demonstrated for multiple PH domains and structural determinants predicting the A or C site binding identified, allowing classification of PH domains according to the relative strengths of each interaction mode; this was found to correlate with binding affinity. This provides further evidence for the role of the A site, initially thought to occur in a minority of PH domains but now revealed to play a much wider role, in allowing high-affinity membrane association was proposed.

Comparison of structural and energetic data from a single, large-scale study such as this allows streamlined and reliable comparison between systems allowing the factors controlling interaction strength and structure to be determined. Although further work is required to confirm the links between PH structure and PIP-interaction energetics proposed here, this study provides a first systematic comparison of the molecular energetics of protein-membrane interactions across a family of related proteins, revealing unexpected complexities and providing a proof-of-principle for the application of this approach to other families of peripheral membrane proteins. in order to identify how interaction energetics is determined by structure.

Chapter 5

Interactions of PTEN with membranes containing PIP_3 lipids

5.1 - Background

A PH domain, by itself, represents a relatively simple protein/PIP interaction; there are many examples of more complicated system, for example presenting multiple binding domains, of which a well-studied example is the PIP_3 phosphatase PTEN. Introduced in Section 1.1.3.2, PTEN is a negative regulator of the cell survival/growth-promoting Akt/PKB signalling pathway [29]; as such, loss-of-function of PTEN has been associated with many cancers (see e.g. [174], is the primary cause of a set of inheritable syndromes collectively known as PTEN hamartoma tumour syndrome (PHTS) [35], and has been linked with other disorders such as autism [36]. Even a slight downregulation of PTEN expression can have significant implications for the formation of tumours [37]. Determining the molecular details of the membrane association of PTEN in its role as a PIP_3 phosphatase is vital to understanding the role of healthy PTEN and the effect of mutations, particularly in the context of potential strategies for the treatment and prevention of these diseases.

A crystal structure of PTEN is available in the PDB (ID 1D5R) and has been used for MD simulations investigating PTEN/membrane interactions in several studies

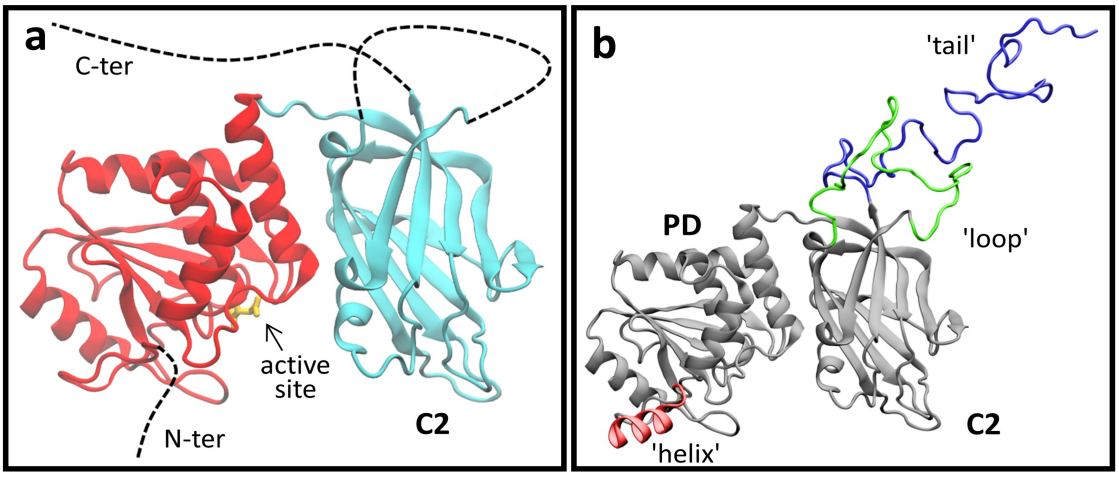


Figure 5.1: Structure of PTEN. (A) Crystal structure of PTEN (PDB: 1D5R) in cartoon representation, showing the phosphatase domain PD (residues 14-184) in red, C2 (residues 186-351, minus the missing 'loop') in cyan, and a tartrate molecule bound at the active site in yellow. The N-terminal, C-terminal and loop regions missing from the structure are indicated by dotted lines **(B)** Initial model of full-length PTEN, showing the helix (residues 1-14), loop (281-312) and tail (351-402) modelled regions in pink, green and blue, respectively.

[42, 44, 63]. This structure includes the catalytic phosphatase domain (PD) and the membrane-binding C2 domain, but is missing the 13 residue N-terminal region, a 31 residue loop within the C2 domain, and a 52 residue C-terminal tail [39] (Figure 5.1 A). While the PD/C2 core has been demonstrated as the 'minimal functional unit' required for membrane binding and phosphatase activity [40], these other regions should not be overlooked: the C-terminal tail appears to play an important role in regulation of PTEN [40] and the N-terminal region is thought to be involved in PI(4,5)P_2-mediated membrane binding and activation [49]. However, the finer details of these are not yet clear, and it is further possible that these regions may have other as-of-yet unexplored roles. The length-scales offered by MD simulations may allow detailed investigation of their behaviour; indeed, an MD study has been performed with these missing regions modelled in [44], but through only a single simulation, starting with PTEN close to the membrane. As such, molecular details of the behaviour and potential role of the N-terminal, loop and tail regions in membrane association of PTEN remain unclear.

The near-atomistic structural details offered by MD may further be key to decoding the nature of clinically relevant mutations, the effects of many of which are unclear. The most common amino-acid substitutions cluster at the catalytic site, as expected, but a significant number are also observed at the PD/C2 interface. It has been proposed that optimal orientation of the catalytic site on the membrane is assisted by the C2 domain and so hindered upon disruption of this interface [39]. A recent MD study of several PTEN mutants provided evidence for interface disruption in solution [175]. Further investigation in a membrane environment is required to confirm the effects of these, and other mutants.

The membrane binding of the PTEN crystal structure was previously investigated using MD in a 'multiscale' approach [42]. This approach makes use of a coarse-grain (CG) model to capture the longer-timescale membrane binding process and obtain initial models of PTEN-membrane binding modes, followed by conversion to a full atomistic model and refinement to obtain an accurate binding pose.

The work presented in this chapter aimed to generate a full-length model of PTEN and to use this model with a multiscale approach, in order to fully explore and obtain an accurate model of PTEN/membrane association, including the behaviour and role of the N-terminal, C2 loop and C-terminal tail. Such a thorough investigation of *full-length* PTEN has not yet been covered by previous simulation studies. This work also aimed to characterise the dynamics of wild-type PTEN in view of identifying the potential molecular effects of pathogenic mutations. This will shed further light on the exact role of PTEN in health and disease, significant especially in the context of management of PTEN-related diseases.

5.2 - Methods

A summary of the simulations performed in this chapter is provided as Table 5.1.

5.2.1 - *Structural details*

Modelling full-length PTEN - An atomistic structure of PTEN was obtained from the PDB (1D 1D5R) (PTEN-crystal). The sequences of the three regions missing from this structure (residues 1-13, 281-311 and 351-402, termed 'helix', 'loop' and

Table 5.1: Summary of simulations performed for Chapter 5. Details of the simulations performed in this chapter. Several repeats were performed under each simulation condition, as indicated under 'Simulation time': repeat coarse-grain binding simulations different in the initial orientation of the protein relative to the membrane, while repeat atomistic simulations different in the seed used for generating initial velocities. 'Interdomain restraints' indicates the presence of elastic network restraints between the PD and C2 domains of PTEN in coarse-grain simulations.
PTEN-c - crystal structure; *PTEN-h* - crystal structure + helix model region; *PTEN-lh* - crystal structure + helix + loop model regions, *PTEN-tlh* - crystal structure + helix + loop + tail model regions (i.e. full length PTEN)

Type	Resolution	PTEN model	PTEN mutation	Interdomain restraints?	Simulation time
binding	coarse-grain	PTEN-h	WT	Y	10 x 5 µs
binding	coarse-grain	PTEN-lh	WT	Y	10 x 5 µs
binding	coarse-grain	PTEN-tlh	WT	Y	10 x 10 µs
binding	coarse-grain	PTEN-lh	WT	N	5 x 5 µs
binding	coarse-grain	PTEN-lh	R173C	N	5 x 5 µs
bound	atomistic	PTEN-tlh	WT	-	3 x 0.5 µs
solution	atomistic	PTEN-c	WT	-	3 x 0.5 µs
solution	atomistic	PTEN-tlh	WT	-	2 x 0.5 µs, 1 x 1 µs
solution	atomistic	PTEN-c	R173C	-	1 x 0.5 µs

'tail', respectively) were submitted to the secondary structure prediction servers Predict Protein [176], JPRED [177], Psipred [178], RaptorX [179] and Phyre [180]. Based on these predictions, the missing regions were added to the crystal structure using Modeller [181], specifying residues 2-11 as α-helical. Loop optimisation was performed, generating twenty structures; the full-length structure (PTEN-tlh) chosen for simulation was that where the modelled regions were judged to extend sufficiently from the core domains to avoid biasing simulations, while minimising the simulation box to be required. An R173C mutant PTEN was generated using the Mutator plug-in of VMD [105].

CG PTEN models - Coarse-grained Martini forcefield representations [61, 78] of PTEN wild-type or R173C mutant, containing all regions missing from the crystal structure (PTEN-tlh), were generated from the corresponding full-length atomistic structures using an in-house script. Partial models containing the region covered by

the crystal structure and the helix region alone (PTEN-h) or helix and loop regions (PTEN-lh) were also generated by manually deleting the appropriate residues. Models of wild-type PTEN-lh and PTEN-R173C-lh without lattice restraints between the PD and C2 domains were generated by manually deleting inter-domain restraints (between a residue in the range 1-184 and a residue in the range 189-350) from the appropriate parameter file.

CG membrane - A Martini coarse-grain membrane was generated from a preformed bilayer of xy dimensions 15×15 nm^2 consisting of 342 POPC molecules in each leaflet. An in-house script was used to exchange lipids to achieve a composition of POPC:POPS:PIP_3 78:20:2. This gave a total of three PIP_3 molecules per leaflet.

Setup for atomistic solution simulations - PTEN structures were placed in simulation boxes of size $11 \times 11 \times 11$ nm^3 (PTEN-c) or $12 \times 14 \times 12$ nm^3 (PTEN-tlh). The system was solvated with SPC water and 0.15 M neutralising Na^+/Cl^- ions added. Energy minimisation via steepest decent was carried out, and the system equilibrated with the protein backbone atoms restrained for 1.5 ns.

Setup for CG binding simulations - The protein was centred 12 nm above (along the +z axis) the membrane in a random orientation. The protein-membrane system was solvated with box size $15 \times 15 \times 25$ nm^3 and neutralising Na^+/Cl^- ions added to a concentration of 0.15 M. Energy minimisation via steepest decent was carried out and the system equilibrated for 5 ns with the backbone particles of the protein restrained.

Setup for AT membrane bound simulations - A snapshot from CG simulations in which PTEN-tlh was bound to the membrane in the productive orientation was converted to atomistic representation using an in-house script [182]. The system was solvated with box size $15 \times 15 \times 14$ nm^3 and neutralising Na^+/Cl^- ions added to a concentration of 0.15 M. Energy minimisation vis steepest decent was carried out and the system equilibrated for 1.5 ns with the backbone atoms of the protein restrained.

5.2.2 - *Simulation details*

All simulations were run using the GROMACS 4.6.5 [82] software for the times listed in Table 5.1. Additional details are provided below. Repeat CG binding

simulations differed by initial orientation of PTEN relative to the membrane, while repeat atomistic simulations used different seeds for assigning initial velocities.

Atomistic simulations - Atomistic simulations were carried out with the GROMOS 43a1 force field [76] and a timestep of 2 fs, writing particle coordinates every 20 ps. Lennard-Jones interactions were shifted to zero between 0.9 - 1.2 nm. Long range electrostatic interactions were treated using the Particle-Mesh Ewald method [86]. The nearest neighbour list was updated every 10 steps. Temperature and pressure were maintained at 310 K and 1 bar, respectively, using a v-rescale thermostat (coupling constant 1 ps) and Parrinello-Rahman barostat [88] (coupling constant 1 ps, compressibility 5 x 10^{-6} bar^{-1}; semi-isotrpoic for membrane simulations). Bond lengths were constrained using the LINCS algorithm [84].

CG simulations - CG simulations were carried out with the Martini 2.1 force field [61, 78]. The timestep was 20 fs and particle coordinates were written out every 400 ps. Coulomb and Lennard-Jones interactions were shifted to zero between 0 - 1.2 nm and 0.9 - 1.2 nm, respectively. The nearest neighbour list was updated every 10 steps. Temperature and pressure were maintained at 323 K and 1 bar, respectively, using a Berendsen thermostat [87] (coupling constant 1 ps) and barostat (semi-isotropic; coupling constant 1 ps, compressibility 5 x 10^{-6} bar^{-1}). Bond lengths were constrained using the LINCS algorithm [84].

5.2.3 - *Analysis*

Electrostatic surfaces - Snapshots of full-length PTEN were obtained from PTEN-tlh solution simulations. Atomistic PTEN-h and PTEN-lh structures were generated by removing the appropriate regions, and the electrostatic surfaces of PTEN-tlh, PTEN-lh, PTEN-h and the crystal structure calculated using the APBS Tools [183] plugin of PyMol.

Model region conformations - PTEN structures from each simulation frame were aligned to a reference structure and coordinates of representative residues of each of the helix (residue 4), loop (residue 296) and tail (residue 402) regions, if present, were recorded. Coordinates of every residue in each region were additionally measured every 100 frames.

Energy landscapes - A reference structure was selected from simulation matching the 'productively bound' structure suggested by previous studies (e.g. [42, 44, 63]). Following backbone alignment (using only the non-modelled regions, i.e. residues 14-280 and 312-351) in the xy plane, the orientation of PTEN relative to the reference structure in each frame was measured as the R_{ZZ} component of the rotation matrix, obtained using the GROMACS *g_rotmat* tool. Membrane distance was taken as the z component of the protein and lipid centres of mass, determined using the GROMACS *g_dist* tool. Data were combined to form a density landscape on distance vs. orientation, which was converted to an approximate energy surface using Equation 2.15.

Protein-lipid contacts - The GROMACS *g_mindist* tool was used to obtain the minimum distance of each residue from the nearest PIP_3 (or POPC) phosphate bead in each simulation frame. The fraction of frames in which each residue was in contact with a given lipid species, determined as when the minimum distance was below 0.6 nm, was then calculated.

Assigning binding mode - The average number of residues in contact with PIP and the average membrane orientation R_{ZZ} were calculated across CG binding simulations using a 200ns sliding window. PTEN was determined to be 'membrane bound' if the average number of contacts was greater than two, and productively bound if the average R_{ZZ} was above 0.9.

Radial distribution functions - The radial distribution function (RDF) about membrane-bound PTEN of each lipid species in each leaflet was estimated using the GROMACS *g_rdf tool*.

PIP binding density - Taking only frames in which PTEN was productively bound, backbone alignment with the representative productively-bound structure (used in obtaining energy landscapes, above) was performed and the xy location of the 1' phosphate of each PIP in the bound leaflet recorded. A probability landscape for PIP location relative to PTEN was generated combining the results from all coarse-grain binding simulations.

Relative domain orientation - A reference structure was generated with the long axis of PTEN aligned to the y axis and the membrane-binding interface oriented towards the -z axis (see Figure 5.11. Backbone alignment with the C2 domain (residues 189 to 350) was performed and the rotation matrix of the PD domain (residues 14-184) relative to that in the reference structure was measured for each frame using the GROMACS *g_rotmat* tool. The rotation about each axis was then calculated from the appropriate components of the rotation matrix: $\theta = -\sin R_{ZX}$, $\psi = \text{atan2}(R_{XY}, R_{ZZ})$, $\varphi = \text{atan2}(R_{YZ}, R_{XX})$.

5.3 - Results and discussion

5.3.1 - *Modelling full-length PTEN*

To generate a full-length model of PTEN, the potential secondary structure of the three regions missing from the crystal structure (residues 1-14, 281-311 and 352-403) was investigated by submitting the PTEN sequence to several secondary structure prediction servers (Predict Protein [176], JPRED [177], Psipred [178], RaptorX [179] and Phyre [180]). All servers predicted an α-helical structure in the N-terminal region, subsequently referred to as the 'helix' region, while the other two regions - the C2 'loop' and C-terminal 'tail' - gave inconsistent or generally 'disordered' predictions (Figure 5.2). These results are generally consistent with the experimental predictions for a disordered loop and tail (e.g. [39]) and N-terminal helix (e.g. [49]), and were used to guide modelling of the missing regions on to the crystal structure of PTEN as outlined in Section 5.2.1.

As the exact interactions of the helix, loop and tail with the C2 and PD domains is unknown, these 'missing' regions were initially modelled extending away from the PTEN core (Figure 5.1 B). Three atomistic 'solution' simulations of 0.5 µs were performed as outline in Section 5.2.2 to investigate the dynamics of the full-length PTEN in an aqueous environment. In all simulations, the three modelled regions quickly (within ∼1 ns) 'collapsed' onto the core region; while the exact conformation taken after this collapse differed between simulations, it each case was generally stable for the remainder of the simulation (Figure 5.3 A). A notable exception, the 'helix' is observed to unfold in one of the three simulations (Figure 5.3 A *iii*), suggesting the

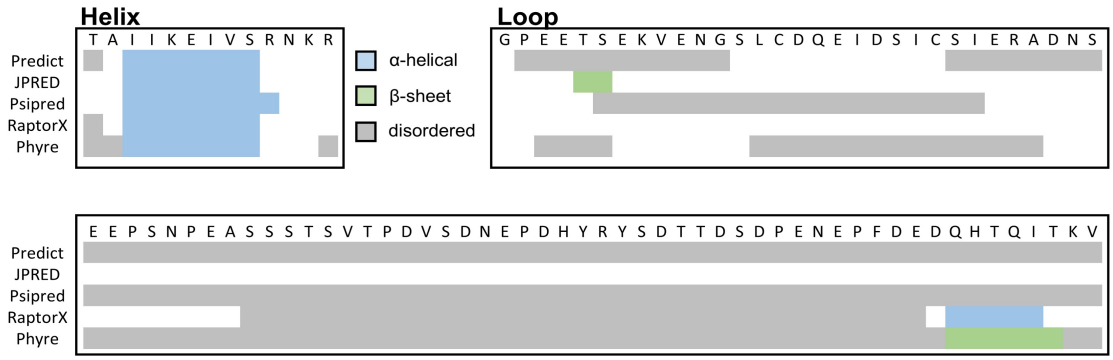


Figure 5.2: Secondary structure prediction of missing PTEN regions. Comparison of results from several secondary structure prediction servers for **(A)** the N-terminal region (residues 1-14), **(B)** the C2 loop (resides 281-311) and **(C)** the C-terminal tail (residues 352-403). Residues predicted to be α-helical, β-sheet or disordered are indicated by blue, green and grey, respectively.

helix may be unstable, at least in the absence of a bilayer with which it may interact. This region remains helical in the other two simulations even after one was extended to 1 μs (Figure 5.3 A *i-ii*).

Although the conformations of the helix, loop and tail regions differ between the final snapshots of the three in-solutions repeats, their effects on the electrostatic surface of PTEN remain generally the same (Figure 5.4): addition of the helix slightly increases the positive surface at the 'front' of the PD domain, while the loop and tail increase the negative surface on the 'front' and 'rear' of the C2 domain, respectively. Note that the profiles for PTEN-h and PTEN-lh shown in Figure 5.4 are calculated for structures following manual deletion of the tail and/or loop of the corresponding PTEN-tlh snapshot, and so may not represent the electrostatic surface of equilibrated models. Nonetheless, these results indicate that the loop and tail in particular make a significant contribution to the electrostatic surface of PTEN.

5.3.2 - *Binding modes and the binding process: CG binding simulations*

To investigate the PTEN-membrane complex formation and in particular the role of the helix, loop and tail regions in directing this process, CG-MD simulations were performed in which the protein was initially displaced from a bilayer containing PIP_3 lipids and allowed to freely diffuse through the surrounding aqueous environment

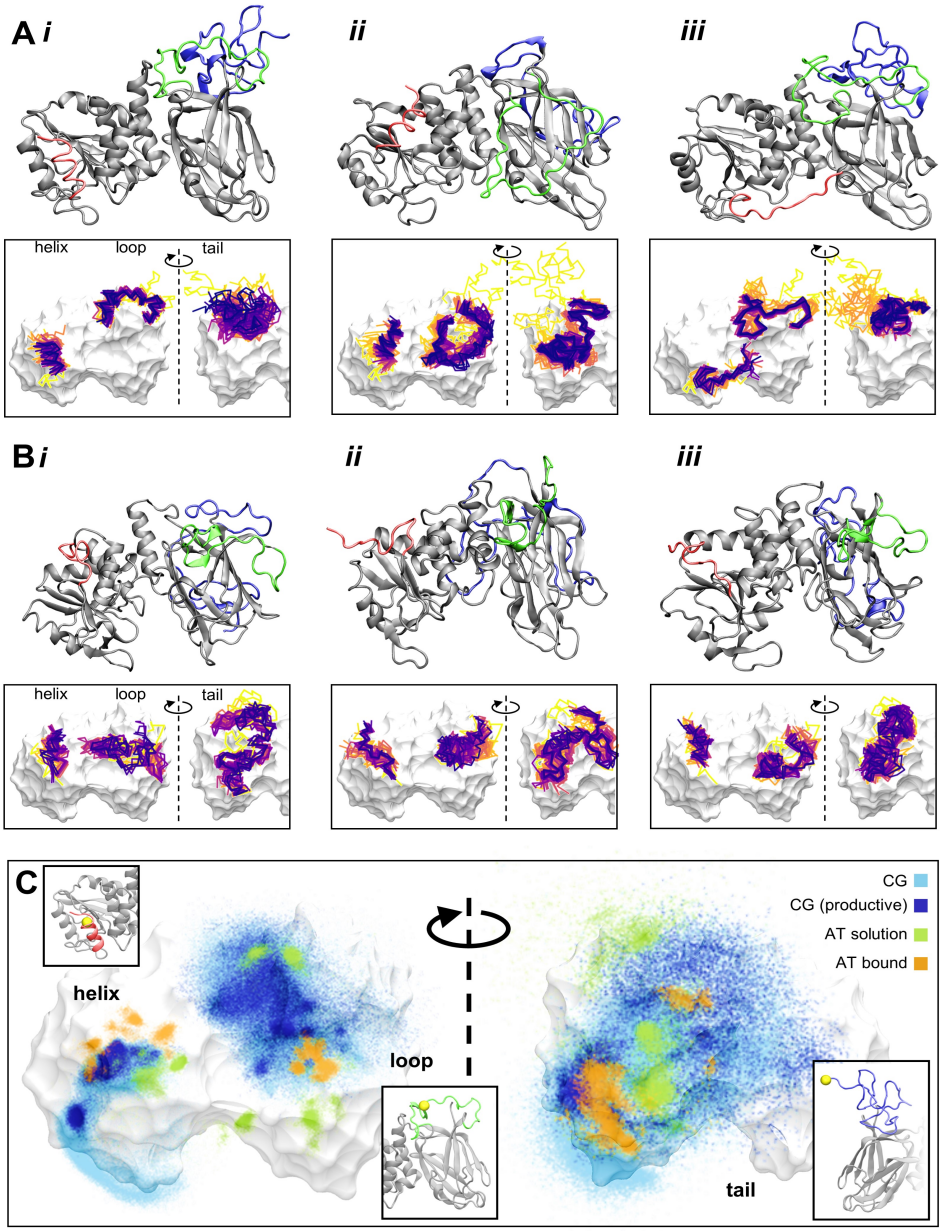


Figure 5.3: Dynamics of the helix, loop and tail regions. Locations of backbone atoms of the helix, loop and tail regions throughout atomistic **(A)** in-solution and **(B)** membrane-bound PTEN-tlh simulations, for each of three repeats. Conformations throughout simulations are coloured from yellow to blue (lower panel). Snapshots taken form the last frame of each simulations are shown in cartoon representation (upper panel); the helix, loop and tail regions in pink, green and purple respectively. **(C)** Locations of representative atoms (see insets) of the helix (residue 4), loop (residue 293) and tail (residue 402) regions thorough atomistic solution (green), atomistic bound (orange) and coarse-grained (cyan, with productively-bound structure in bark blue) simulations.

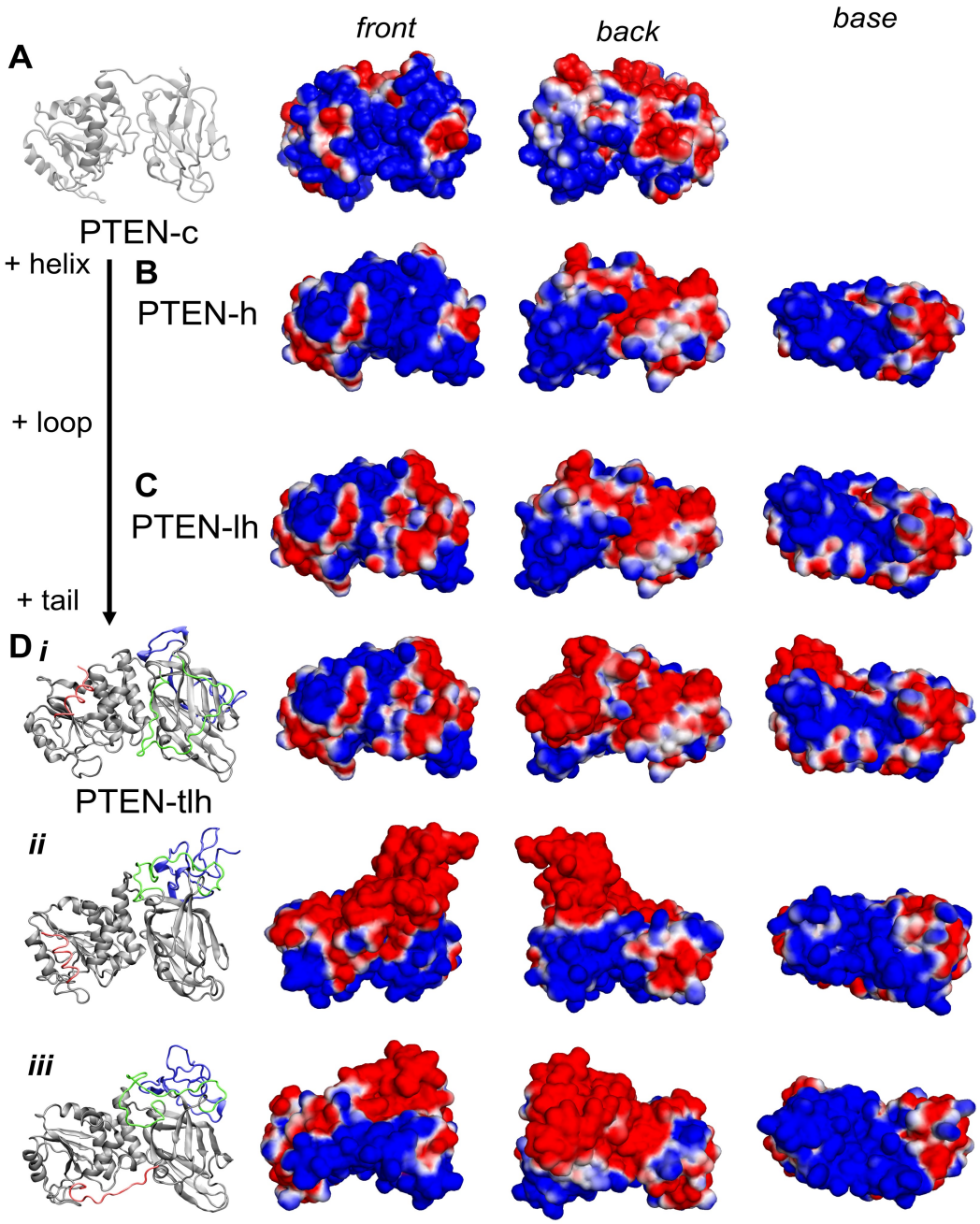


Figure 5.4: Electrostatic surface profile of PTEN. The electrostatic surface of **(A)** the crystal structure of PTEN (PDB ID 1D5R), **(B)** PTEN-h, **(C)** PTEN-lh and **(D)** PTEN-tlh. PTEN-tlh surfaces are calculated from the final snapshots of three repeat atomistic solution simulations. PTEN-h and PTEN-lh surfaces are calculated from appropriate selections of the snapshot shown in **D** *i*.

and/or interact with the bilayer (outlined in Section 5.2.2). Ten simulations were performed for each PTEN models containing the helix (PTEN-h), helix and loop (PTEN-lh) or all three modelled regions (PTEN-tlh).

5.3.2.1 - *Binding landscapes*

To identify membrane-interaction modes, density landscapes of distance against orientation of PTEN relative to the membrane were constructed (Figure 5.5 A). Orientation was measured relative to a reference structure, selected to match the 'productive binding' structure suggested by previous studies (e.g. [42, 44, 63]) in which both domains contact the membrane and the active site is oriented towards the bilayer surface ('mode 1' in Figure 5.5 B). A binding mode with an R_{ZZ} of 1 therefore corresponds to the productive binding mode. As expected, this mode is the most favoured for each of the three PTEN models (Figure 5.5 A).

A range of secondary membrane-associated orientations are observed for PTEN-h, with mode '3', in which PTEN binds in an 'upside down' orientation (see Figure 5.5 B), being the most favoured of these (Figure 5.5 A *i*). PTEN-lh and PTEN-tlh, meanwhile, demonstrate only one secondary orientation (Figure 5.5 A *ii-iii*): mode '2', in which PTEN interacts with the membrane via the PD domain, with the 'helix' lying parallel to the membrane surface (see Figure 5.5 B). These observed binding modes are consistent with the electrostatic profiles presented in Section 5.3.1: PTEN-c and PTEN-h both possess a positive surface across the 'front' of the proteins both core domains and the 'back' of the C2 domain in addition to the canonical membrane binding interface on the base of the protein (Figure 5.4 A). These regions allow non-specific interactions with anionic lipids in the bilayer in the range of non-productive binding orientations seen. Upon addition of the loop and helix, these positive patches are greatly reduced, with only the 'front' face of the PD domain remaining (Figure 5.4 D). Mode '2' corresponds to binding via this face, and is the only secondary mode observed for PTEN-lh and PTEN-tlh.

It is noted that while the distribution of possible distances and orientations appears well covered, control of sampling is not as rigorous as for the landscapes constricted from US simulations presented in Chapters 3 and 4. The landscapes presented in

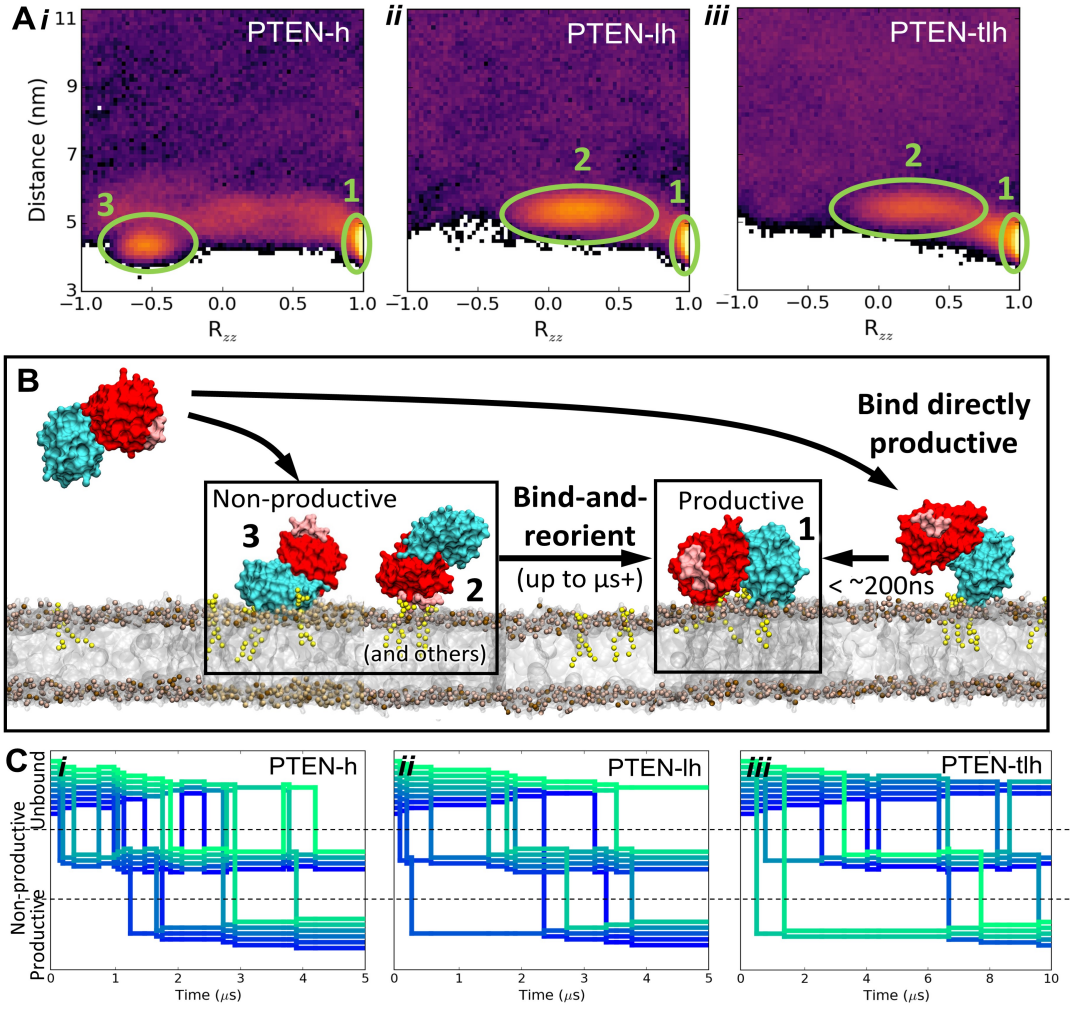


Figure 5.5: Modes of interaction of PTEN with the membrane. **(A)** Distance-orientation density landscapes from coarse-grained binding simulations of *(i)* PTEN-h, *(ii)* PTEN-lh and *(iii)* PTEN-tlh, indicating the presence of preferred binding modes. Orientation is measured as the R_{ZZ} component of the rotation matrix of PTEN relative to a reference structure corresponding to 'mode 1' in **B**. **(B)** Representative snapshots of favourable binding orientations. Arrows indicate the possible orders of binding events observed. **(C)** 'Binding paths' taken by each of the 10 repeats for *(i)* PTEN-h, *(ii)* PTEN-lh and *(iii)* PTEN-tlh. At each timepoint PTEN is identified as 'productively bound', 'non-productively bound' or 'unbound'.

Figure 5.5 may not accurately depict the underlying energy landscape; nonetheless, they give a good indication of the presence and favourability of different binding modes.

For each model, the productive-binding mode '1' was observed to occur either directly

(reached within 200 ns of first contact, usually via the C2 domain) or following a reorientation step from a non-productive orientation such as modes 2 or 3 (Figure 5.5 B, C). Once reached, the productive orientation remained stable for the rest of the simulation in all cases; on the other hand, PTEN was occasionally observed to unbind from the membrane when in other interaction modes (Figure 5.5 C). PTEN-tlh was overall less likely to bind and stay bound, with only 5 of 10 cases were bound after the initial 5µs. Even following extension of simulations to 10µs, PTEN remained bound in only 7 of 10 repeats.

However, when binding does occur in PTEN-tlh simulations, it is more likely to be in the directly productive orientation (3 of 7 repeats (or 3 of 9 initial binding events), compared to 3 of 10 repeats (or 3 of14 binding events) with PTEN-h (Figure 5.5 C) and 0 of 4 for PTEN-c in a previous study [42]. Overall, these observations suggest that the helix, loop and tail regions contribute toward directing the productive binding of PTEN. For the loop and tail regions, this may largely be due to the elimination of non-productive modes through altering the electrostatic profile of the protein, as discussed in Section 5.3.1. Further analysis may reveal more direct roles of these and the helix region in membrane binding.

5.3.2.2 - *PIP interactions*

Membrane association of PTEN was found to result in clustering of PIP_3 lipids. This can clearly be seen in the radial distribution function of PIP_3 about PTEN (Figure 5.6 A; note that PIP_3 in the unbound leaflet is unaffected) and is consistent with previous computational results (e.g. [42, 63]). PIP_3 residue contact patterns while productively bound were consistent across all three models (Figure 5.7 A *i-iii*). Mapping these onto the atomistic PTEN structure clearly shows the membrane-binding interface (Figure 5.7 B), in agreement with that identified in previous studies [39, 41, 42, 44, 63, 184].

The density map of PIP_3 positions relative to PTEN while productively bound (Figure 5.6 B) reveals several favoured binding locations. Note that there were only three PIP molecules present in each membrane in these simulations; given the degree of clustering suggested by this a previous studies, it is likely that if more PIPs were

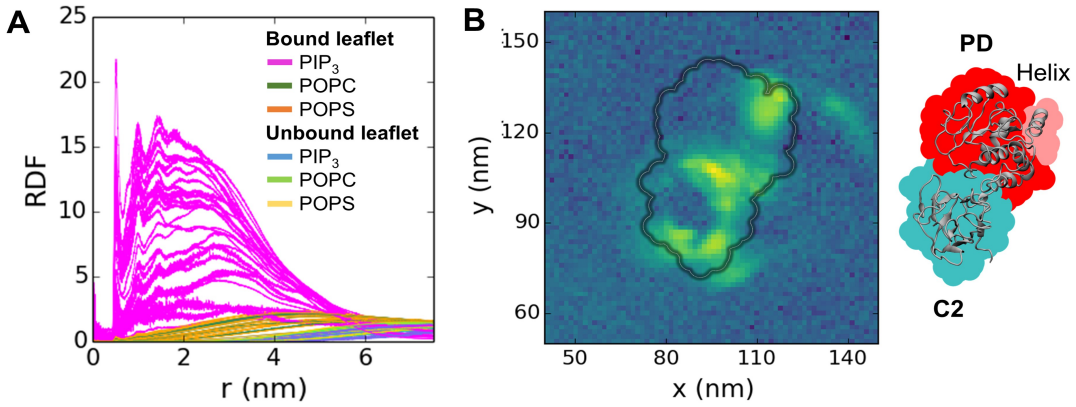


Figure 5.6: Clustering of PIP_3 around PTEN. (A) Radial distribution functions about productively-bound PTEN centre-of-mass for each lipid species in the bound and unbound leaflets. **(B)** Density map of locations of PIP_3 molecules in the bound leaflet relative to PTEN while it is bound in the productive orientation, averaged over all coarse-grained binding simulations.

included, additional binding locations would be observed. Here, PIP binding is observed to cluster in three locations: around the C2 domain (expected given its nature as a PIP-binding domain), in the PD domain (specifically, at the base of the helix region) and in the cleft between the PD and C2 domains.

This last location is in the proximity of the catalytic site. While no PIP_3 contacts are observed with the catalytic resides C124 and R130, a number of contacts are observed with neighbouring residues (∼125-129; the 'P-loop') (Figure 5.7 A) and PIP is observed to approach to around 1nm from of the position of a bound tartrate in the PTEN crystal structure, suggested to represent the binding site [39] (Figure 5.7 C). Given the height of the suggested active site relative to the membrane surface, it is possible that some form of protein structural change is required to bring the site closer the membrane before binding at the active site is seen.

Of interest, the PIP-binding site in the PD domain at the base of the N-terminal helix is consistent with experimental studies suggesting that $PI(4,5)P_2$ binds in this area as an allosteric activator [49]. This interaction site is reflected in a high number of contacts around residues 12-15 (Figure 5.7 A). Being 'completed' by the addition of the 'helix' model region, this likely contributes to the increased favourability of

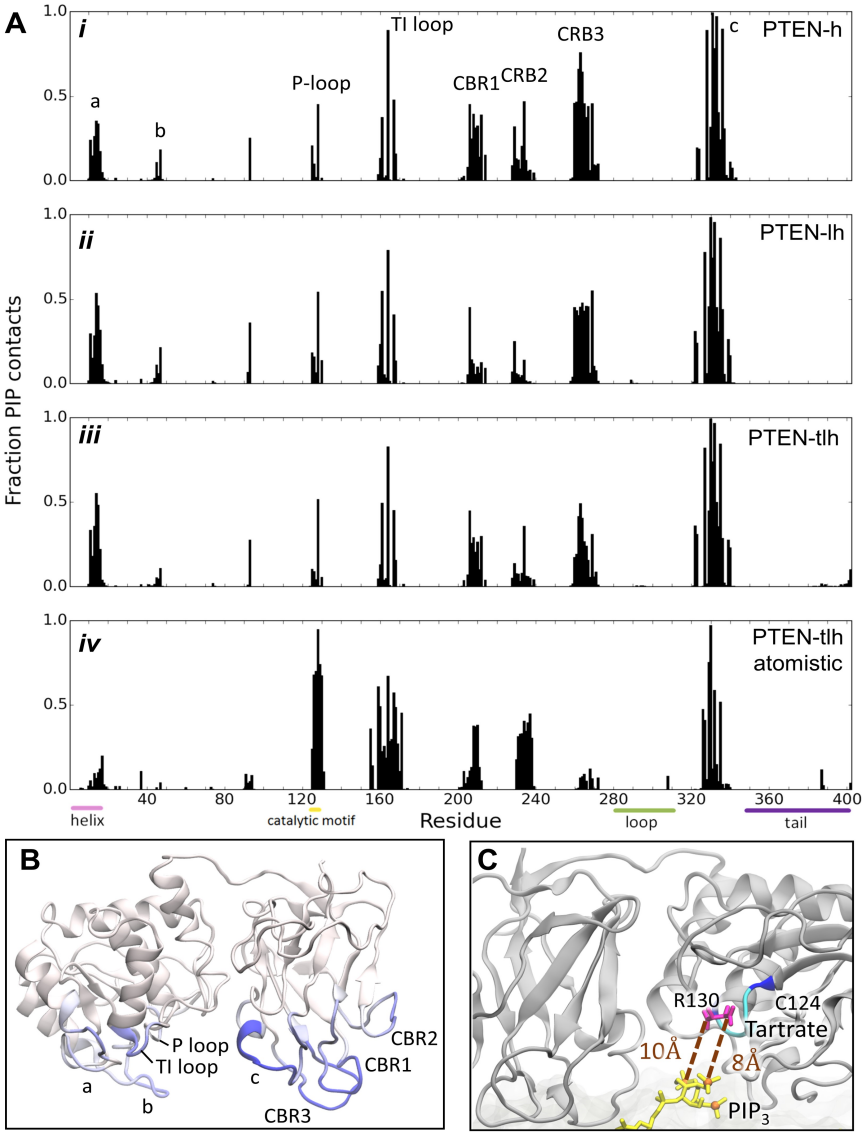


Figure 5.7: PIP_3-PTEN interactions. **(A)** Average fraction of frames in which each residue is in contact with PIP_3 while PTEN is productively bound, from *(i)* PTEN-h, *(ii)* PTEN-lh and *(iii)* PTEN-tlh coarse-grain simulations and *(iv)* PTEN-tlh atomistic bound simulations. **(B)** Fraction of PIP_3 contacts per residue, averaged over all coarse-grain simulations, mapped on to the crystal structure, coloured from white (no contacts) to blue. **(C)** Snapshot, converted to atomistic representation, showing the closest PIP_3 is is observed to approach the active site throughout coarse-grain simulations. The relative position of the tartrate molecule in the crystal structure is shown in pink; the catalytic residues C124 and R130 are shown in blue, and neighbouring residues 125-129 in cyan. PIP_3 is shown in yellow with the 3', 4' and 5' phosphorous atoms in orange. This structure was used as the initial structure of atomistic bound simulations.

direct productive binding observed for helix-containing PTEN models compared to the crystal structure, described in Section 5.3.2.1.

5.3.2.3 - *Conformation of the helix, loop and tail regions*

The three modelled regions show much greater variation throughout coarse-grained simulations compared to atomistic solution simulations. This is understandable given the much larger timescales covered, and the inability in coarse-grain simulations to form hydrogen bonds, which may limit the formation of new, stable interactions. Helix stability is difficult to assess from coarse-grained simulations due to the presence of an elastic network model to maintain secondary structure. Tracing the free end of the helix (at residue 4) suggests it generally lies flat against the PD domain, and while productively bound tends to be oriented roughly vertically (Figure 5.3 C *i*), pointing away from the membrane; this may allow better access to the apparent PIP binding site at its base.

Tracing representative atoms from each of the loop and tail regions (residues 296 and 402 respectively, i.e. those furthest in sequence from the core region) indicates these regions vary greatly in conformation, and in particular may extend towards the membrane binding interface (Figure 5.3 C *ii-iii*). Indeed, an (albeit small) number of PIP_3 contacts are observed within the tail region (Figure 5.7 C *iii*). Further refinement is required to assess if these interactions are important.

5.3.3 - *The productive binding mode: atomistic refinement*

5.3.3.1 - *Changes to the binding mode: penetration and PIP contacts*

In order to obtain a more accurate model of the productive binding mode, a snapshot was taken from CG simulations in which a PIP_3 molecule was bond in the C2/PD cleft (near the active site, as shown in Figure 5.7 C), converted to full-atom resolution and simulated (Figure 5.8). Three atomistic membrane-bound simulations of 0.5 µs each were performed as outline in Section 5.2.2. Notably, insertion of the C2 domain in to the membrane is observed to increase in each simulation (Figure 5.8 C). Average density profiles along z (the membrane normal) clearly show the C2 domain extending into the phosphate headgroup layer of the membrane (Figure 5.9 A). With

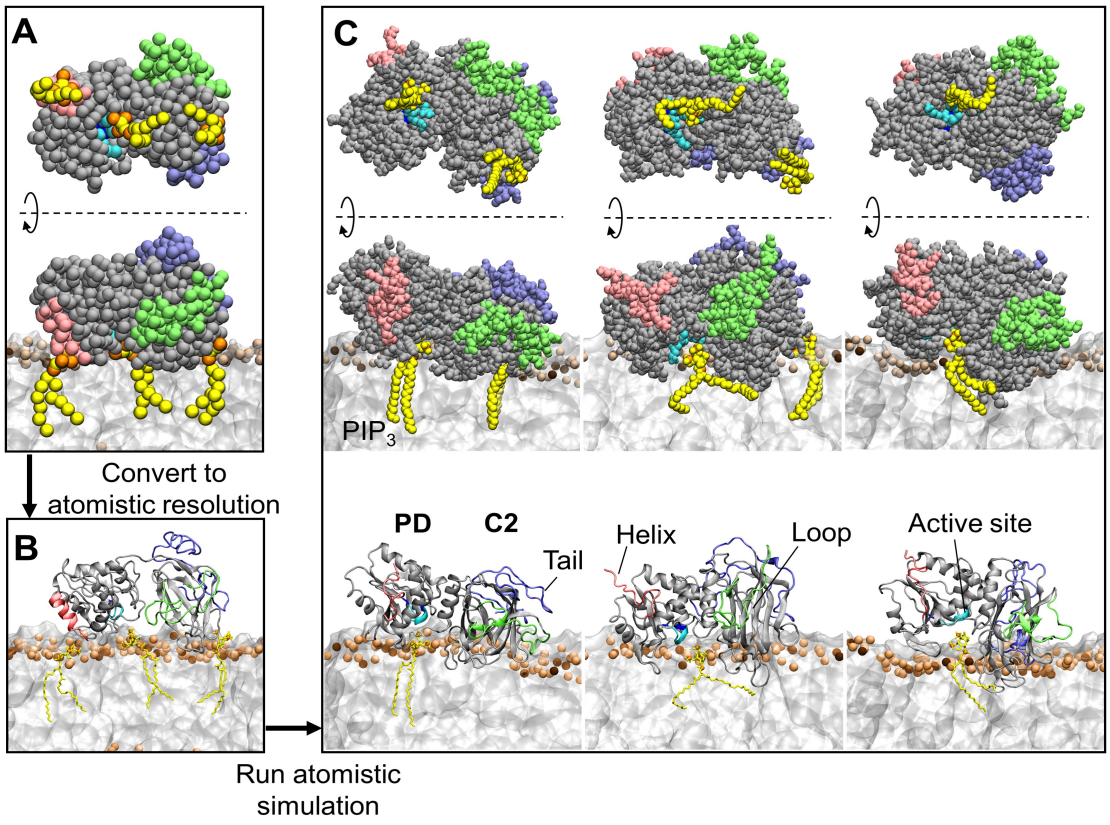


Figure 5.8: Atomistic refinement of the productive binding pose. (A) A coarse-grained model of the productive binding mode is obtained from coarse-grained simulations, **(B)** converted to atomistic representation and **(C)** simulated for 0.5 µs. The final snapshot from each of three repeat atomistic simulations is shown.

The PTEN core region is shown in grey, and the helix, loop and tail regions in pink, green and purple, respectively. The catalytic (C123 and R130) and neighbouring (124-129) residues are shown in dark and light blue, respectively; PIP_3 molecules are shown in yellow, with phosphorous /phosphate beads in orange; POPC/POPS phosphates are shown in light and dark brown, respectively.

this increased insertion, the active site is brought closer to the membrane surface and so is more accessible to PIP_3, which now approaches within ∼0.5 nm of the active site. The PIP_3 initially located in the PD/C2 cleft remains bound here in all three cases, with a high degree of contacts with P-loop residues observed (Figure 5.7 A *iv*), though again no contacts with the catalytic residues are seen.

Aside from the slight decrease in protein-membrane separation as the C2 domain inserts further into the membrane, the binding mode is unchanged, remaining in the

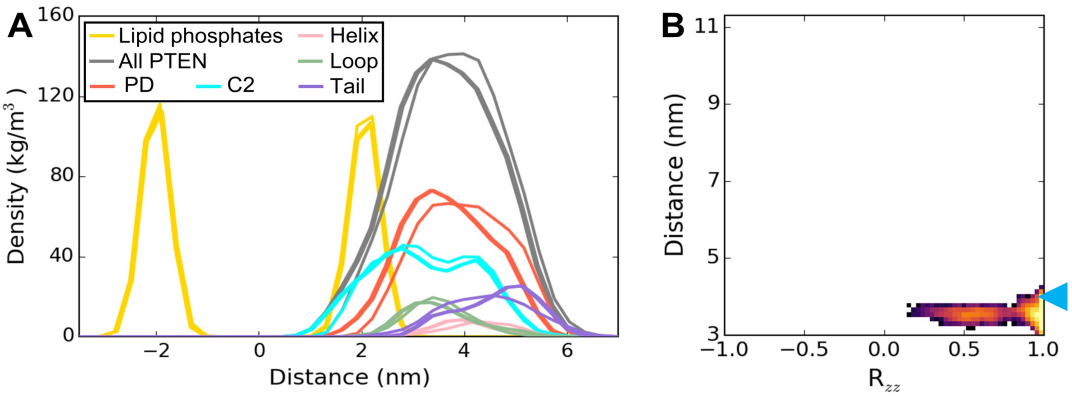


Figure 5.9: Membrane penetration and orientation of the productive binding mode. **(A)** Density profiles along the membrane normal of PTEN components relative to the lipid phosphate headgroups, averaged over each of three atomistic membrane bound simulations. **(B)** Landscape of orientation (R_{ZZ}) vs. distance for atomistic membrane bound simulations, constructed as for coarse-grained landscapes in Figure 5.5. The blue arrow indicates the configuration of the initial structure converted from a coarse-grain snapshot.

productive orientation (Figure 5.9 B), though the PIP_3 lipid initially bound at the base of the helix is not retained in any of the three simulations. The PIP_3 lipid initially bound to the C2 domain is also observed to unbind in one of the repeats (Figure 5.8 C). This is reflected in a reduction of PIP_3 contacts in these regions (Figure 5.7 A *iv*). Contacts within the CBR3 loop also decrease, likely as this loop is buried by insertion to the membrane. The unbinding of these PIP_3 lipids may be due to the shift in PTEN orientation as C2 membrane penetration is increased. Given the degree of PIP_3 clustering observed both here and in previous studies [42, 63], it is expected that, given extension of these simulations, these PIPs would eventually be observed to rebind the stabilised PTEN.

5.3.3.2 - *The helix, loop and tail regions*

The helix was observed to be unstable and 'unravel' in all atomistic membrane-bound three simulations, though it does remain associated with the surface of the PD domain (Figure 5.3 B).Along with the one solution simulation in which the helix was unstable, this suggests that despite the consensus of secondary structure prediction, this N-terminal region is unstructured. It is possible that other factors not considered here

may contribute to the stability of this helix - for example, it has been suggested that the N-terminal region forms an amphipathic helix that sits parallel to the membrane surface, and may only be stable upon contact with the membrane[49]. It is also possible that the initial atomistic structure was biased, due to the choice of coarse-grain snapshot or conversion process, in such a way that favoured helix instability.

PIP_3 contacts during coarse-grain binding simulations are consistent with the presence of a $PI(4,5)P_2$ allosteric binding site in the N-terminal region has been suggested experimentally [49]. While the simulations performed here utilised PIP_3 rather than $PI(4,5)P_2$, these lipids are expected to behave similarly under the coarse-grained resolution. Further simulations are required to confirm and further examine this site under atomistic resolution.

After an initial equilibration, the conformations of the loop and tail regions in atomistic bound simulations remain relatively unchanged on the timescale explored (Figure 5.3 B), with both remaining in proximity with the membrane (Figure 5.8 C). A small number of PIP_3 contacts are observed within the tail region, as well as some in the loop region (Figure 5.7 A *iv*). The formation of contacts between these regions and the membrane (PIP_3 or other lipids) may further contribute to the binding of PTEN.

Phosphorylation of residues toward the ends of the tail region (S380, T382, T383 and S385) has been shown to lower membrane binding, with conflicting reports on whether this is due to the phosphorylated tail obstructing the membrane-binding interface predominantly at the PD [40, 41, 184], the C2 domain [46, 50–52], or a PTEN dimer formation interface [47, 48]. The distribution of the tail throughout coarse-grain and atomistic simulations supports that binding to the membrane-binding interface, at least at the C2 domain, is possible (Figure 5.3). Further simulations, with the tail phosphorylated, are required to confirm this and to better determine the exact role of the tail in preventing membrane binding.

In contrast to the interaction of the tail with the membrane seen here, a previous computational study of full-length PTEN found that while the tail did bind near the membrane-binding interface in solution, it extended away from the membrane while bound [44]. This extension was required to fully explain density profiles of membrane-bound PTEN obtained from neutron scattering experiments [45]. The dif-

ference in tail configurations observed between these studies may reflect environmental differences, or, given the large distribution of tail conformations observed during coarse-grained simulations (Figure 5.3 C), may reflect different slices of an extended conformation landscape that is not is not fully explored in the atomistic simulation timescales. It is possible that movement of the tail away from the membrane may also be observed here if the atomistic simulations were are sufficiently extended or enhanced sampling approaches used.

5.3.4 - *PTEN mutations and the role of the PD/C2 interface*

To identify mutations of interest, I performed a literature survey for reports of clinical missense mutation (Appendix B). The locations of these mutations are summarised and mapped on to the initial PTEN full length model in Figure 5.10. For some residues, the likely effect of their mutation leading to reduced or eliminated PTEN function can be readily determined: for example, R130 (the most commonly mutated residue) is a catalytic residue, and the K13, R14 and R15 residues are observed to form PIP_3 contacts above and so are likely important for productive membrane binding. In line with previous observations (e.g. [39]), many mutated residues are observed around the PD/C2 interface, including S170, D252 and notably R173, the second most frequently mutated residue overall (Figure 5.10; Appendix B). As mentioned in Section 5.1, maintenance of the PD/C2 interface has been proposed to be important for directing productive binding, but the exact effects of mutations here have not been well examined.

5.3.4.1 - *Relative domain orientation*

To assess changes at the PD/C2 interface, a set of 'relative domain orientation' angles were defined, aligning the C2 domain and measuring the rotation about each axis of the PD domain relative to that in the crystal structure (see inset, Figure 5.11). Relative orientations of the PD domain throughout both PTEN-tlh in-solution simulations and three additional in-solution simulations performed with the crystal structure (PTEN-c) demonstrate good agreement, close to that observed in the crystal structure (Figure 5.11 A *i-ii*).

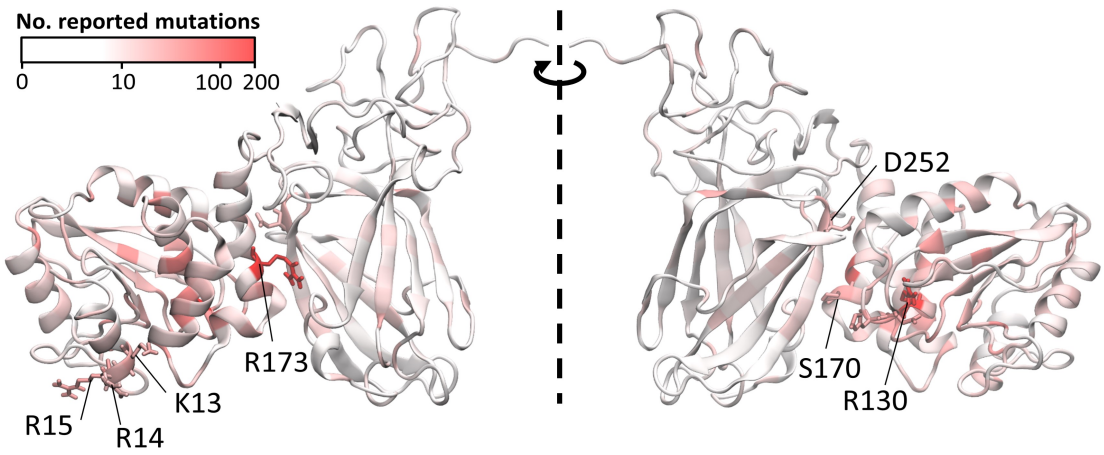


Figure 5.10: Clinical missense mutations of PTEN. Cartoon representation of full-length PTEN, coloured by mutation frequency per residue from white (no mutations) to red. Residues of interest are shown explicitly.

Atomistic membrane-bound simulations, on the other hand, showed more pronounced rotation about the z' and x' axes (Figure 5.11 A *iii*). These results suggest that there may be some conformational change occurring as part of the binding process. These rotations appear in general to bring the PD and C2 domains closer together at the membrane surface near the location of the active site (Figure 5.11 B), though retention of the PIP_3 lipid bound at the PD/C2 cleft suggests there is still sufficient room to bind in this here. This rotation may thus serve to create a more favourable environment to draw PIP_3 lipids to the active site by increasing the number of residues available for contact at a given time.

Change in inter-domain orientation was previously observed with membrane binding of the PTEN-c structure [42], though by 10° in the -ψ direction, opposite the reorientations observed here. This may again reflect a larger underlying orientation landscape that is not fully explored in the timescales of the atomistic simulations. Indeed, coarse-grained simulations, discussed below, show a much greater range of movement.

The coarse-grained simulations performed here make used of an elastic network model in order to maintain protein secondary structure. The elastic restraints are based on proximity of beads in the initial structure; given the proximity of the PD and C2

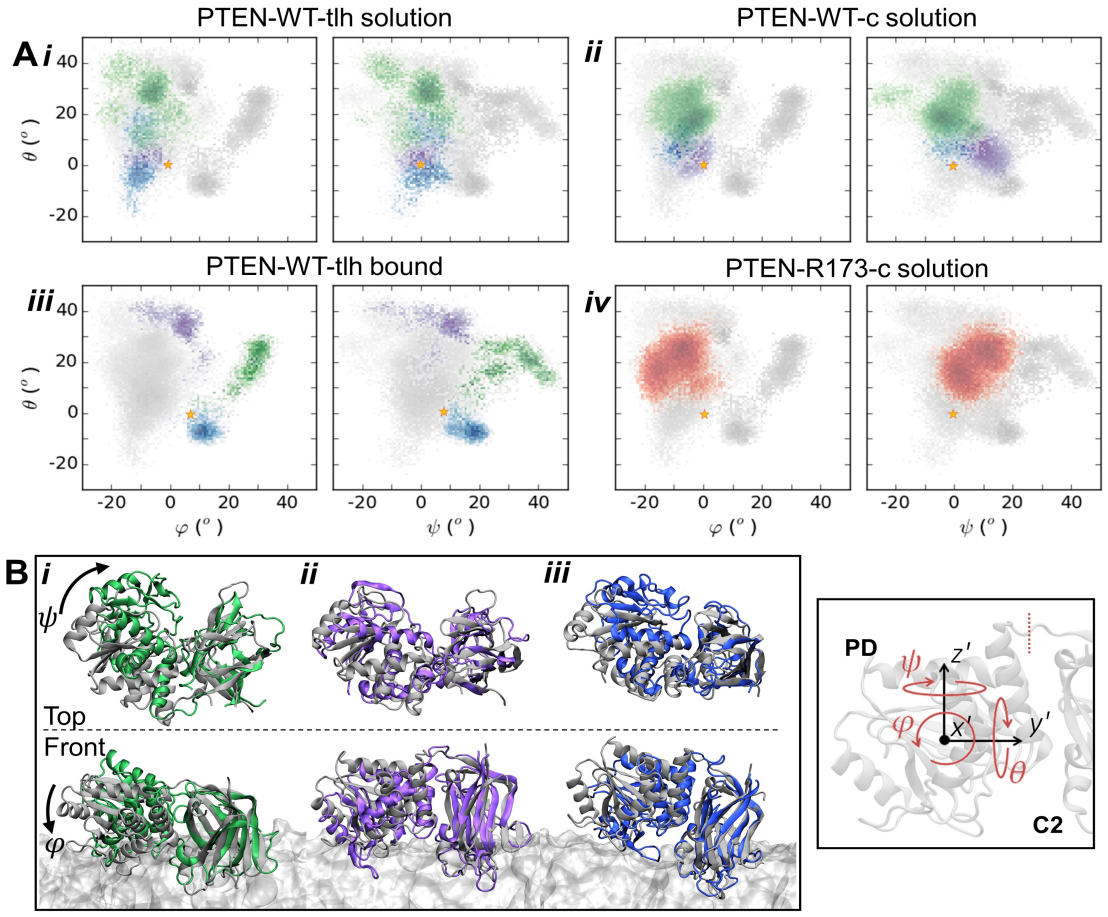


Figure 5.11: Relative orientation of the PD and C2 domains. **(A)** Distribution of relative orientation of the PD and C2 domains throughout wild-type (i) PTEN-tlh solution, (ii) PTEN-c solution and (iii) PTNE-tlh membrane bound (three repeats each, in green, blue and purple); and mutant (iv) PTEN-R173C-c solution simulations (red). Orientation is measured at rotation about each axis of the PD domain, after alignment by the C2 domain (see inset). An orange star indicates the starting configuration for each simulation set. For comparison, the combined landscape from all atomistic simulations is shown in the background in grey for each set of simulations. **(B)** Final snapshots from wild-type PTEN-tlh membrane bound simulations (green, purple or blue) with the crystal structure (grey) aligned by the C2 domain, indicating the shift in relative domain orientation. Helix, loop and tail regions are omitted for clarity.

domains, restraints are introduced across the interface, complicating investigation of relative PD/C2 orientation. To address this, an additional set of 5 coarse grained binding simulations were performed using a PTEN-lh model in which restraints between the PD and C2 domains were removed. The resulting landscape is similar

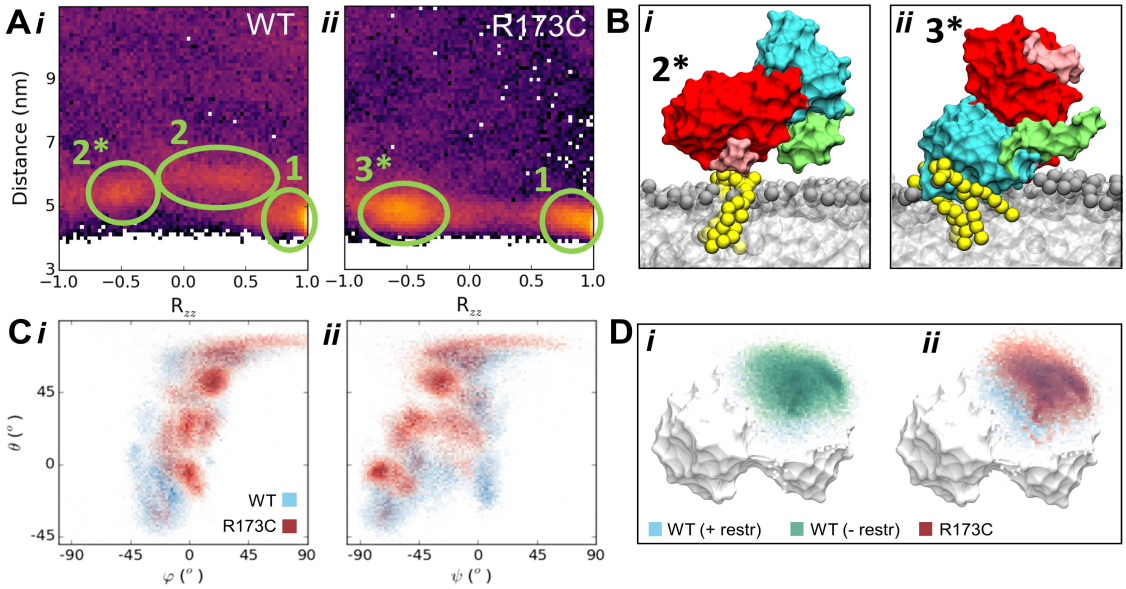


Figure 5.12: Coarse-grain investigations of the R173C mutant. From coarse-grained binding simulations of PTEN-lh in which restraints between the PD and C2 domains have been removed: **(A)** Distance/orientation density landscape for *(i)* wild-type and *(ii)* R173C PTEN, constructed as for those in Figure 5.5. **(B)** Representative snapshots of additional binding modes indicated in **(A)**: *(i)* mode 2* and *(ii)* mode 3*. **(C)** Distribution of relative domain orientation for WT (blue) and R173C (red) PTEN, measured as in Figure 5.11. **(D)** Distribution of residues in the loop (residue 280-312) throughout simulations for *(i)* wild-type (green) and *(ii)* R173C mutant (red) PTEN. The corresponding distribution from initial wild-type PTEN-lh, with inter-domain restraints present, is shown in blue for reference.

to that of the initial PTEN-lh mode (Figure 5.12 A *i*). Notably, while the productive mode '1' remains the preferred orientation, the relative level of mode '2' is decreased, while a new interaction mode '2*', a variation of mode '2' also interacting with PIP via the helix region, but with a different overall orientation (Figure 5.12 B *i*), is observed. It is not clear what causes these changes in the energy landscape; more binding simulations may be required to better determine the extent of changes. Nonetheless, the removal of PD/C2 restraints does not appear to significantly perturb the binding landscape.

Investigating the relative domain orientation throughout coarse-grain simulations (Figure 5.12 shows variation over a much larger range of angles than seen in atomistic simulations (Figure 5.11 A). Given the limitations of the coarse-grain model, partic-

ularly at representing changes in protein structure, it is unclear to what extent this better reflects the full underlying energy landscape, but nonetheless will provide a useful comparison for the effects of mutations of relative domain orientation.

5.3.4.2 - *The R173C mutation*

As mentioned, the PD/C2 interface residue R173 is the second most frequently mutated residue (Figure 5.10; Appendix B). Cystein is the most common substitution identified. R173C was therefore chosen for investigation via simulation. A 500 ns atomistic in-solution simulation was performed using the R173C mutant of PTEN-c. Domain orientation throughout PTEN-R173-c simulations shows greater fluctuation than wild-type atomistic solution simulations (Figure 5.11 A *iii*). Although these orientations do not fall outside the overall range observed for wild-type PTEN, this indicate that the R173C mutation destabilises the domain interface, as expected and consistent with previous computational solution simulations [175].

To investigate the effect of the R173C mutation on membrane binding, 5 coarse-grain binding simulations were performed using an R173C mutant of the PTEN-lh model in which PD/C2 restraints had been removed. The distance-orientation landscape indicates that productive binding is still observed (Figure 5.12 A *ii)*. Interestingly, a secondary binding mode '3*' (Figure 5.12 B *ii*), similar to mode '3' for PTEN-h (involving interaction with the membrane via the C2 domain, while the PD domain extends away from the membrane; Figure 5.5 B), is seen. This mode is observed to a comparable extent to the productive binding mode, and is not observed for previous coarse-grain PTEN-lh simulations (Figure 5.12 A *i*).

The range of domain orientations observed is again greater than that seen in atomistic simulations, and notably shifted towards smaller ψ and larger φ angles compared to the wild-type (Figure 5.12 C). As mentioned above, it is unclear how well the angle range explored reflects physiological dynamics given the coarse-grain model, but comparing the wild-type and mutant results nonetheless suggests some effect on inter-domain dynamics by the R173C mutation. The difference in angle, particularly the increased φ, appear in general to bring the PD and C2 domain closer at the 'front' edge of the interface, perhaps due to replacement of the bulky R173 sidechain with the smaller cysteine.

In Section 5.3.4.1, an increase in ψ was observed with wild-type membrane binding. The *decrease* in ψ compared to wild-type PTEN of the R173C mutant as observed here may thus lead to the reduction in productive binding observed in Figure 5.12 A. While the lack of mode '2' may be largely due to the smaller ensemble of simulation (only 5 simulations were performed, compared to 10 for initial coarse-grained binding simulations presented in Section 5.3.2.1 leading to limited exploration of the underlying binding landscape, it is not immediately clear how what mechanism leads to the increase favourability of mode '3*', though the above results suggest the relative domain orientation may play a role.

Looking at the distribution of residues in the 'loop' region throughout simulations, PTEN-WT-lh without inter-domain restraints shows a similar distribution to that seen in initial PTEN-lh simulations (green and blue in Figure 5.12 D, respectively). PTEN-R173C-lh, however, shows a preference for conformations located relatively higher on the 'front' face of PTEN (red, Figure 5.12 D *ii*). This moves the loop further from the core PTEN region and so may reduce the extent to which the non-productive-binding positive regions are obstructed, discussed in Section 5.3.2.1, leading to an increase in these non-productive interaction modes. The loop is observed to extend across towards the interface and PD domain (Figure 5.3; Figure 5.12 D). The change in this region as relative domain orientation is shifted may thus lead to the different loop conformations observed as, for example, interactions are lost and binding towards the interface becomes less favourable. Further investigation is required to explore and confirm these details.

5.4 - Further discussion and conclusions

In this chapter, modelling of full-length PTEN and its use in multiscale simulations allowed investigation of the membrane-binding of this protein, obtaining details of binding modes and landscapes, the binding process and dynamics. PIP_3 contacts, clustering and the productive binding mode were consistent with previous studies. Results reveal that three regions missing from the crystal structure play a role in directing productive binding through introduction of binding sites and changes in the electrostatic surface, emphasising the importance of considering all regions of a protein to accurately simulate its behaviour, even if small. Further evidence for the

potential role of the C-terminal tail in regulating binding and of the N-terminal region in PIP-binding was also seen, though this N-terminal region may not be helical as previously thought. Initial results also demonstrate the importance of the PD/C2 interface, with changes in domain orientation observed on membrane binding and alteration of inter-domain orientation and membrane binding landscapes with the clinically relevant R173C mutation, located in the PD/C2 interface.

Overall, these results demonstrate the advantages of multiscale simulations, utilising the increased timescales of coarse-grained simulations to better explore conformational landscapes, and the increased resolution of atomistic simulations to capture details such as membrane insertion. Further multiscale simulations may first be applied to PTEN to confirm results suggested here, in particular the changes of domain orientation on membrane binding and effect of the R173C mutation, and may next consider the range of other clinically-relevant PTEN mutations. The results presented here and those of future studies will contribute towards understanding the behaviour of healthy and mutated PTEN and may ultimately aide the development of drugs and in turn the management of PTEN-related disorders.

Chapter 6

Conclusions

6.1 - Summary

The work presented in this thesis involved use of MD simulation to study the interactions between canonical peripheral proteins and membranes containing PIP lipids. These simulations allowed high-resolution, dynamic information of the protein/membrane systems to be obtained and even quantified. In general, results were consistent with and provided further confirmation of previous experimental results, and in addition offered new insights into the molecular and energetic details of these interactions.

In Chapters 3 and 4, a pipeline combining the MD techniques of US and CG modelling was set up, tested, and used to investigate PIP interactions of PH domains. This work demonstrates how MD tools may be combined to form pipelines for the large-scale study and comparison of biomolecular systems, allowing the features that characterise certain behaviours to be better identified. By allowing thorough investigation of the binding energy landscape and providing directly linked structures of near-atomistic resolution, these binding features could be readily identified on a residue scale. The presence of interactions at a secondary PIP-binding site and simultaneous binding of multiple PIPs were suggested from simulations results to be more important to the membrane binding of PH domains than had been suggested previously through experimental studies.

In Chapter 5, a multiscale MD approach was used to investigate membrane binding of the tumour suppressor PTEN. This work demonstrated the benefits of a multiscale approach, combining the longer-timescale advantage of coarse-grained simulations (able to capture membrane binding events and favourability of different interaction modes) and the resolution advantage of atomistic simulations (able to capture more subtle details such as degree of membrane insertion). Results showed an increase in directed productive binding when a full rather than a truncated model of PTEN was used, emphasising the importance of carefully considering the protein model used when studying its behaviour. The high resolution of MD simulations also allowed direct measurement of protein dynamics and the effects of a single amino-acid substitution, further suggesting the importance of the PTEN interdomain interface in productive membrane binding and its potential disruption by the clinically-relevant R173C mutation, helping to bridge the gap between sequence modifications and changes in PTEN activity.

6.2 - Future directions

It is clear from these results that computational simulations are applicable towards understanding protein/membrane interactions in both health and disease, with many possible applications that may be pursued in future work. Firstly, several of the results presented in this thesis would benefit from the added confirmation and clarity that may be provided by additional or repeated simulations, for example the study via US/CG of more PH domains to better identify sequence elements that can be used to predict favourable membrane PIP-interaction sites and ultimately PIP affinity and selectivity; or further simulations with R173C and membrane bound PTEN to confirm the changing landscapes of interdomain orientation and relation to productive binding.

Having developed and successfully used the US/CG pipeline to compare PIP binding across PH domains, this method may next be applied to understanding the many other important membrane-binding proteins and domain families [7], such as the PIP-binding C2 and FYVE domains. PTEN also represents a target of interested, allowing the method to be tested on a more complicated system involving multiple domains and potentially providing quantification of the effects of various mutations.

The clinical relevance of PTEN makes this protein in general an ideal target for future simulations. Better understanding of how this protein and its mutated variants behave may contribute to the management of PTEN-related diseases through assessment of risk and development of treatments. Only one mutation was investigated here; many other of the large number of clinically-relevant mutations are poorly understood and may benefit from an MD investigation.

An alternate translation variant of PTEN (PTEN-long) [38] possesses an extended N-terminus region and is observed to be secreted and able to enter and function in other cells. This presents an exciting therapeutic opportunity as a drug-delivery system for PTEN. However, the way in which this variant may differ from regular PTEN need to first be understood, with already some suggestion that membrane binding is altered in PTEN-long [41]. PTEN has also been observed to form dimers and this has been suggested to contribute to the high sensitivity to mutation, with mutants acting in a dominant-negative manner to sequester healthy PTEN [47, 48]. Formation and membrane binding of dimers has not yet been well investigated. The details of the allosteric activation of PTEN by $PI(4,5)P_2$ [185] are also currently poorly understood, but may be revealed through use of appropriate simulations.

Future efforts may also focus on improving the methodologies used. As discussed, the issues of sampling and reduced accuracy due to resolution remain particular limitations of US/CG simulations. Improved error assessment may better identify areas of poor sampling [186] to be addressed, and modifications to the US method such as replica exchange have been demonstrated to improve sampling [187]. It is also worth considering the choice of reaction coordinate while linear distance as used here is simple to set up and easy to intuit, further structural analysis was required to properly differentiate the binding modes encountered. Use of more detailed reaction coordinate combining several factors may allow different states to be more readily distinguished and reduce hidden minima in the one-dimensional energy profile (see e.g. [187]).

As computational power and simulation efficiency are improved, atomistic simulations will become more feasible for studying the binding process through free simulations or US, but are likely limited to single systems for the near future. A quantum

mechanical treatment through QM/MM simulations [188] may offer an even higher level of accuracy; full quantum mechanical treatment is currently limited on the order of a few residues surrounding a binding site, but may nonetheless be useful to understand, for example, the catalytic process of PTEN.

From another direction, there are also many opportunities for the improvement of experimental approaches to studying protein/membrane interactions. Many of the methods used are not quantified, lack a native membrane environment or are not suitable for a high-throughput approach, and may show poor agreement. More recently developed methods, such as PLIF (protein-lipid interaction by fluorescence) [189] and atomic force microscopy (AFM) [190] may offer an easier approach to quantify protein/lipid interactions in physiological conditions.

This thesis represents only a small selection of the ways in which computational MD simulations may be used to improve our understanding of the nature of interactions between proteins and membranes in the cell. Together with experimental investigations, MD will continue to play an integral role in characterising these vital associations in both the healthy cell and in relation to the large number of disorders associated with their disruption.

Appendix A

Overview of selected code written for this thesis

This appendix presents a brief overview of a selection of the code written during this thesis to automate the setup, running and analysis of simulations.

A.1 - The `auxiliary` module of MDAnalysis

MDAnalysis is an open-source Python library for the analysis of molecular dynamics simulations (see main body refs. [118, 119]). The purpose of the auxiliary module was to allow 'external' timeseries data to be read alongside a corresponding trajectory within an MDAnalysis `Universe`.

A file containing timeseries data, for example the reaction coordinate value data obtained from umbrella sampling simulations, may be added to a trajectory using the `add_auxiliary` method: (`u.trajectory.add_auxiliary(`*auxname*, *auxfile*`)`). Multiple auxiliaries can be added to each `Universe`. The module will then align the timeseries with the trajectory; as each new simulation frame is loaded, the corresponding value of the auxiliary data is also loaded and is available as e.g. `u.trajectory.ts.aux.auxname`. Optionally, the nearest auxiliary timepoint or an average of nearest timepoints may be stored with the frame instead of only a directly matching auxiliary value. Utilities are also available to allow e.g. looping through only those simulation frames with a corresponding auxiliary data (using `u.iter_as_aux(`*auxname*`)`, or loop directly through the auxiliary data (using `trajectory.iter_auxiliary(`*auxname*`)`).

This code is part of MDAnalysis from version 0.16.0, which is available online from *http://www.mdanalysis.org*. Further details on usage and options available can be found in the MDAnalysis documentation,
http://www.mdanalysis.org/docs/documentation_pages/auxiliary_modules.html.

A.2 - Other scripts and modules

The following Python scripts and modules were used for the analysis carried out in this thesis; they are not part of any formal distribution and have not been thoroughly tested and documented, but are nonetheless available online from *https://github.com/fiona-naughton/US-analysis*.

wham.py *- a wrapper for the Grossfield WHAM*

Provides a `GrossfieldWHAM` class as an interface from Python to the bash-fronted Grossfield WHAM [90]. An instance is generated providing the reaction coordinate timeseries data (`source_data`), set of window position (`x0s`), restraint contant (`ks`) and temperature: `wham = GrossfieldWHAM(`*source_data, ks, x0s, temp*`)`. Trajectory pull force data may be used in place of reaction coordinate data.

Appropriate input files can then be generated using `wham.write_input_files()`, and WHAM run using `wham.run_wham([`*wham options*`])`. The arguments normally required for running the Grossfield WHAM are here optional as default values have been provided. This includes the start and end reaction coordinate values, which will be automatically set from the timeseries data if not specified. The ability to select a start and end time is also added, and windows without data in the reaction coordinate/time range will be automatically excluded to allow for a successful Grossfield WHAM run. Following successful run of WHAM, the output profile and window F-values can be read in using `wham.read_output()`, optionally converting to a chosen energy unit.

bundle.py *- group US simulations and metadata for MDAnalysis*

This module was written to allow the grouping of trajectories, for example a set of umbrella sampling windows, within MDAnalysi, along with various useful metadata. It provides a `make_bundle()` function that takes a set of provided trajec-

tories and corresponding auxiliary data (e.g. reaction coordinate data) and metadata (e.g. window restraints) and generates a 'bundle' using MDSynthesis (main body ref. [120]): `bundle = make_bundle(`*trajectoryfiles, topology, auxs, data, temp*`)`; auxiliaries are provided as as `auxs={auxname: [auxfile1, ...], ...}` and metadata as `data={dataname: [datavalue1, ...], ...}`. A bundle can also be generated from a list of existing `Universe`s. This bundle contains the MDAnalysis `Universe` for each trajectory with any specified auxiliaries added as `bundle[`*i*`].universe`, where *i* is the index or name of the trajectory, and the metadata as `bundle.categories[`*dataname*`]`.

This module can be used to group umbrella sampling simulations and the relevant metadata, which may then be used, for example, with the above to run WHAM analysis. The resulting F-values can also be stored alongside the trajectories in the bundle. These can then be used together with the following to generate profiles or averages with other system coordinates calculated with MDAnalysis.

unbias.py - *energy landscapes along and averages of other coordinates*

Provides an `UnbiasSystem` class. An instance is loaded with the appropriate metadata and reaction coordinate timeseries file (`rcoord`) for each window: `unbias=UnbiasSystem(`*Fvals, rcoords, ks, x0s*`)`. An unbiased probability histogram over multiple coordinates of the system, may then be calculated following Equation 2.28, given a list `data` of the timeseries of each of those coordinates for each window, using `unbias.unbias(data)`. Lower-dimension histograms along selected coordinates can then be obtained using `unbias.get_hist(`*axes*`)`, where `axes` is a list of the indicies of the coordinates of interest. These probability histograms may then be used to generate energy landscapes.

The average value of a coordinate as the reaction coordinates changes can also be calculated following Equation 2.29 using `unbias.average(data)`, where `data` is the list of the timeseries of that coordinate for each window.

Appendix B

Clinically relevant PTEN missense mutations

B.1 - A summary of clinically reported missense mutations of PTEN

Clinical studies reporting the identification of one or more missense mutations of PTEN were located and compiled to form a list of clinically relevant PTEN mutations, shown below. From this, the relative frequency of mutation of each residue was calculated and can be seen projected onto the PTEN structure in the main thesis body in Figure 5.10. The individuals from whom these mutations were identified presented with any of variety of phenotypes, including the presence of cancers or other disorders, as indicated in the following table. The relevant clinical studies identified from this literature search are listed following the table in Section A.2.

Abbreviations and acronyms: AA - anaplastic astrocytoma; ASD - autism spectrum disorder; BC - breast cancer; BL - Burkitt's lymphoma; BRR - Bannayan-Riley-Ruvalcaba syndrome; CCSK - clear cell sarcoma of the kidney; CD - Cowden disease; CRC - colorectal cancer; EC - endometrial cancer; EH - endometrial hyperplasia; GC - gastric cancer; Glb. - glioblastoma; Gls. - gliosarcoma; HNC - head and neck cancer; LC - lung cancer; LDD - Lhermitte-Duclos disease; Macr. - macrocephaly; MM - malignant melanoma; Odg. - oligodendroglia; OC - ovarian cancer; PC - prostate cancer; PTHS - PTEN hamartoma tumour syndrome; RCC - renal cell carcinoma; SCLS - small cell lung cancer; TC - thyroid cancer; UBC - urinary bladder cancer; VC - vulvar cancer

Resid.	Mut-ation	Num. cases	Phenotype (refs.)
T2	I	1	CCSK [1]
A3	S	1	EC [2]
	D	2	CRC [3][4]
I5	L	1	Glb. [5]
K6	R	1	VC [6]
	I	1	Glb. [7]
	N	2	EC [8] EC [9]
	E	1	GC [10]
I8	T	1	MM [11]
S10	R	1	BC [12]
	N	2	LBCL [13] LC [14]
R11	K	2	UBC [15] EC [16]
	I	1	EC [17]
K13	E	4	Glb. [18] [19] (x2) EC [20]
	Q	1	EC [21]
R14	G	3	BL [22] VC [6] CRC [23]
	M	1	CRC [24]
	S	2	HNC [25]
R15	K	2	EC [26] UBC [15]
	I	3	EC [27] (x2) [28]
	S	1	CRC [29]
	T	1	HNC [25]
Y16	C	2	LBCL [30] Glb. [31]
Q17	P	1	Gls. [32]
D19	N	1	MM [33]
G20	E	1	EC [34]
F21	S	1	EC [8]
D22	E	1	ASD [35]
L23	V	1	EC [8]
	F	1	Glb. [36]
D24	N	2	EC [37] CRC [29]
	H	1	RCC [38]
	Y	5	EC [26] [39] (x2) OC [40] BRR [41]
	V	1	GC [10]
	G	5	Glb. [19] [31] EC [8][16] [42]
	E	2	Glb. [31] GC [10]
L25	F	2	EC [43] Glb. [7]

Resid.	Mut-ation	Num. cases	Phenotype (refs.)
T26	P	2	EC [20] [44]
Y27	D	2	GC [10] BC [45]
	S	1	Glb. [46]
	C	4	Glb. [31] EC [39] [42] [47]
I28	M	2	CRC [3] EC [48]
Y29	S	1	GC [10]
	A	1	MM [11]
P30	L	1	PHTS [49]
	Q	1	EC [16]
N31	H	1	CRC [3]
	K	1	EC [16]
I33	S	1	EC [16]
	T	1	EC [17]
A34	D	2	BRR [50] EC [8]
	V	1	EC [16]
M35	L	1	BC [51]
	V	2	Glb. [52] EC [47]
	R	1	Glb. [53]
	I	1	LC [44]
G36	R	5	CD [54] LDD [55] EH [56] Glb. [31] [57]
	E	3	EC [58] (x2) Glb. [59]
	V	1	Glb. [31]
F37	L	1	EC [17]
	S	2	VC [6] UBC [15]
P38	S	2	MM [3][11]
	R	1	EC [39]
	H	1	ASD [60]
	L	1	MM [61]
	F	2	MM [27]
A39	T	1	Glb. [62]
	P	1	CD [45]
	V	1	GC [10]
L42	P	1	Gbl. [46]
	R	2	Glb. [19] [63]
E43	K	1	MM [11]
G44	D	3	Odg. [18] Macr. [64] VC [29]
V45	A	1	HNC [9]

continued on following page

continued from previous page

Resid.	Mut-ation	Num. cases	Phenotype (refs.)
Y46	H	2	EC [65] Glb. [19]
	C	1	CRC [66]
R47	G	1	Glb. [59] EC [42]
	M	2	EC [16] GC [10]
	K	1	Glb. [19]
N48	D	1	Gls. [32]
	I	1	UBC [67]
	K	1	CD [68]
N49	I	1	Glb. [69]
D52	N	1	MM [33]
	G	1	VC [6]
V53	G	1	CRC [66]
	D	1	CD [70]
V54	L	1	OC [2]
R55	G	1	PC [71]
	S	1	BC [72]
	W	1	UBC [15]
F56	V	2	EC [42] OC [40]
	C	1	EC [73]
L57	S	1	Odg. [74]
	W	2	Glb. [75] [52]
D58	Y	1	Glb. [76]
S59	P	1	BC [77]
	L	1	LC [3]
K60	T	1	CRC [23]
H61	R	97	CRC [78] (x2) AA [79] Glb. [53] [80] [81] SCLS [40]
	L	1	BL [82]
	Q	1	Glb. [83]
	D	3	BC [84] Macr. [85] BRR [86]
K62	R	1	CRC [87]
	I	1	GC [88]
	N	1	EC [8]
N63	Y	1	CRC [89]
Y65	N	1	EC [21]
	D	1	Glb. [31]
	C	1	CRC [87]
K66	Q	1	EC [21]
	E	2	EC [16] CRC [29]
	N	5	Gls. [90] EC [8][16] [42] HNC [9]
I67	Q	1	GC [88]
	R	2	CD [50] EC [91]
	T	1	EC [16]
Y68	N	3	EC [39] CRC [24] ASD [60]
	H	7	CD [92] RCC [93] EC [20] [58] Glb. [31] [76] BRR [50]
	C	4	Glb. [59] EC [16] [39] CRC [89]
L70	P	1	EC [16]
	V	1	ASD [94]
	I	1	EC [47]
	F	1	EC [37]
C71	Y	3	EC [16] [21] Glb. [95]
	W	1	MM [61]
E73	D	1	CRC [4]
	V	1	EC [16]
T78	A	1	Glb. [63]
A79	T	2	Odg. [18] MM [96]
K80	E	2	Glb. [81] LDD [55]
	N	1	Glb. [97]
F81	C	3	CRC [89] EC [26] [98]
	L	1	EC [37]
N82	T	2	GC [88] BC [99]
R84	G	1	VC [6]
V85	A	2	GC [100] (x2)
	G	1	GC [29]
A86	V	1	EC [16]
Q87	P	1	EC [16]
Y88	N	1	OC [101]
	H	3	MM [102] LDD [55] (x2)
	S	1	Gls. [32]
	C	5	BC [103] Glb. [7] (x2) EC [39] [47]
F90	S	1	EC [37]

continued on following page

continued from previous page

Resid.	Mut-ation	Num. cases	Phenotype (refs.)
E91	Q	2	PC [104] GC [105]
	A	1	Glb. [31]
D92	N	2	Glb. [63] BC [106]
	H	2	BC [107] SCLC [40]
	Y	1	EC [8]
	A	2	Glb. [31] PHTS [49]
	G	3	EC [16] [58] [60]
	V	2	EC [8][16]
	E	6	Glb. [59] [83] (x2) EC [16] [47] [48]
H93	D	1	EC [37]
	Y	5	CRC [4] CD [34] EC [20] [60]
	L	1	BC [12]
	R	7	Glb. [52] EC [16] [17] [39] [58] ASD [108] NHC [109]
	Q	1	Glb. [36]
N94	Y	1	EC [21]
P95	S	3	Glb. [31] EC [8] MM [102]
	L	3	EH [110] EC [16] [37]
	G	1	CRC [29]
P96	S	2	LC [100] Glb. [52]
	Q	2	Glb. [31] CD [111]
	T	1	EC [16]
	L	4	Glb. [7][19] [29] [59]
	A	1	EC [16]
T98	G	1	Glb. [112]
E99	K	1	GC [10]
I101	N	1	PC [71]
	T	5	Glb. [31] (x2) [36] Odg. [74] CRC [29]
	M	1	RCC [93]
C105	R	1	EC [73]
	G	1	Glb [52] MM [11]
	Y	3	BRR [50] AA [31] [113]
	S	1	Glb. [114]
	F	2	Glb. [31] [115]
	W	2	BC [106] Glb. [80]
D107	N	1	BC [106]
D107	Y	3	EC [116] [91] Glb. [117]
	A	1	Gls. [32]
	V	1	CD [118]
L108	P	4	MM [11][61] CRC [119] PHTS [49]
	R	1	BC [107]
Q110	P	1	EC [48]
W111	R	1	Glb. [52]
L112	V	5	Glb. [31] EC [16] [42] [73] [84]
	P	3	CD [92] [120] Glb. [19]
	Q	1	MM [121]
	R	1	Glb. [19]
S113	R	1	EH [122]
D115	N	1	EC [123]
D116	G	1	TC [124]
H118	R	1	TC [124]
	L	2	MM [88] [61]
	P	1	ASD [125]
V119	F	1	Glb. [7]
	D	1	EC [116]
A120	T	2	GC [89]
A121	P	3	BC [87] Glb. [18] [117]
I122	N	1	EC [8]
	S	2	Glb. [65] EC [58]
H123	D	2	CD [111] EC [16]
	Y	2	EC [37] [42]
C124	S	6	EC [42] [60] AA [114] [126] TC [127]
	R	6	CD [37] [50] (x2) EC [16] [58]
	Y	1	GC [88]
K125	E	3	EC [8]Glb.[19] GC [87]
	T	1	CRC [4]
	N	1	Glb. [128]
A126	T	4	Glb. [36] EC [8][126] [129]
	P	3	AA [97] EC[8]PHTS[130]
	S	3	EC [16] [42] CRC [89]
	D	2	EC [42] GC [10]

continued on following page

continued from previous page

Resid.	Mut-ation	Num. cases	Phenotype (refs.)
A126	V	2	EC [39] [58]
	L	1	OC [131]
G127	R	5	EC [16] (x2) [39] Glb. [19] GC [10]
	E	1	MM [62]
	V	1	EC [16]
K128	G	1	PHTS [132]
	Q	2	SCLC [40] EC [16]
	N	5	MM [7] CRC [119] EC [28] [42] BC [12]
G129	R	5	Glb. [19] [80] CRC [78] CD [49] TC [124]
	E	5	CD [133] (x2) EC [42] CRC [119] SCLC [40]
	V	4	EC [16] (x2) [84] [134]
R130	G	95	EC [8](x5) [16] (x25) [17] (x2) [20] (x5) [26] [37] [37] [39] (x12) [42] (x14) [43] (x2) [47] (x2) [48] [73] (x6) [84] (x2) [135] (x4) [136] [134] (x2) Glb. [7][95] EH [122] OC [131] (x2) [137] [138] (x2) LC [139]
	Q	78	EC [8][16] (x17) [19] [73] (x4) [17] [27] [28] [37] [39] (x6) [42] (x6) [47] [58] (x3) [91] (x2) [107] (x2) [126] (x3) [140] [141] (x2) AA [97] CD [128] [142] BC [12] [72] [106] [128] CRC [4][23] [66] [89] (x3) [37] [143] OC [144] GC [10] SCLC [40] PC [145] MM [61] [121] (x2) TC [124] (x2) EH [110]
	P	9	EC [16] (x3) [42] [47] [106] Glb. [7] BC [103] [12]
	L	9	EC [39] [42] [91] [124] [116] CD [50] EH [56] Glb. [19] MM [11]
T131	P	1	OC [101]
	A	1	Glb. [59]
	I	2	OC [146] PHTS [49]
G132	S	3	Glb. [7][29] EC [39]
	C	1	MM [147]
	D	8	Glb. [19] [59] [90] EC [12] [16] [28] [101] PHTS [130]
	V	2	Glb. [81] OC [138]
	R	1	EC [43]
M134	T	2	EC [89] CRC [119]
M134	R	2	BRR [99] EC [58]
	I	2	EC [16] Macr. [148]
I135	V	3	BRR [50] EC [84] PC [71]
	K	1	Glb. [36]
C136	R	9	Glb. [18] [52] CD [149] PHTS [150] BRR [118] MM [11] EC [16] [39] [58]
	Y	4	BC [107] CD [151] EC [42] [129]
	F	3	Glb. [81] EC [16]
T138	D	2	OC [138] CRC [89]
	C	1	EC [16]
L139	V	1	EC [39]
	F	1	Glb. [152]
L140	F	1	EC [8]
H141	Y	1	Glb. [153]
	R	1	LDD [154]
R142	W	5	EC [16] [17] [26] [43] CRC [78]
	P	1	PHTS [49]
	Q	6	TC [124] [155] LC [156] CRC [23] OC [34]
G143	D	1	EC [58]
K144	Q	5	CRC [23] [24] [89] (x3)
	T	2	EC [16] [47]
F145	I	2	CRC [66] EC [16]
	C	1	EC [17]
L146	V	1	EC [27]
K147	R	1	TC [124]
A148	T	1	Glb. [31]
Q149	R	1	EC [43]
E150	Q	1	CRC [87]
	G	1	PC [71]
	D	1	EC [34]
A151	D	2	TC [124] PHTS [49]
L152	V	1	Glb. [53]
	R	1	EC [16]
F154	V	1	CRC [89]
	L	4	EC [39] (x2) [43] [48]
Y155	N	1	PHTS [49]
	H	3	Glb. [18] SCLC [157] EC [21]

continued on following page

continued from previous page

Resid.	Mutation	Num. cases	Phenotype (refs.)
G156	C	9	BC [103] TC [124] CRC [119] PHTS [49] Glb. [19] [95] EC [16] [73] [58]
	R	1	Glb. [158]
E157	G	1	ASD [64]
R159	K	2	TC [124] Glb. [59]
	S	3	CRC [159] (x2) EC [39]
T160	S	1	BC [72]
	I	2	Glb. [59] MM [102]
R161	K	2	BC [72] (x2)
K164	R	1	EC [43]
G165	R	7	EC [16] (x3) Glb. [83] [21] [76] [160]
	E	6	CD [161] SCLC [162] Odg. [163] Glb. [42] [31] RCC CRC [24]
	V	2	CD [50] RCC [164]
T167	A	2	Glb. [113] MM [165]
	S	1	Glb. [59]
	I	1	Glb. [59]
P169	H	1	Glb. [50]
S170	G	1	Glb. [5]
	N	3	Glb. [166] [81] EC [167]
	I	3	EC [17] Glb. [36] [76]
	R	1	BRR [22]
Q171	E	3	Glb. [42] [31] EC [17]
	K	1	SCLC [40]
	R	1	Glb. [168]
	P	1	Glb. [36]
	H	1	Glb. [18]
R173	C	39	EC [8][65] [16] [20] (x3) [27] (x2) [28] [39] (x2) [43] (x2) [47] [141] [58] (x5) [82] [91] [126] Glb. [9][18] [29] (x2) [42] [53] [59] [114] (x2) [169] [163] MM [3]CRC [89] (x2) EH [56]
	H	19	Glb. [18] [42] (x2) [94] (x2) [52] [79] [170] [160] [31] EC [8][16] [58] [67] (x2) OC [144] AA [114] CRC [89] [78]
	P	1	LLD [171]
Y174	N	2	EC [42] (x2)
Y176	C	1	ASD [125]
S179	I	1	EC [21]
L181	P	2	CRC [66] (x2)
K183	N	1	CRC [89]
N184	1	1	Odg. [170]
D187	Y	1	CRC [89]
Y188	S	1	GC [88]
P190	L	1	OC [138]
V191	A	1	EH [56]
L193	P	1	Glb. [59]
M198	K	2	EC [42] PHTS [130]
E201	I	1	Macr. [64]
P204	T	1	GC [88]
	S	1	Glb. [81]
	Q	1	EC [17]
	L	1	CRC [89]
C211	Y	1	AA [170]
P213	H	1	EC [43]
	R	1	PC [145]
F215	S	3	BC [103] EC [16] [17]
V216	M	1	EC [34]
V217	I	1	MM [33]
	D	1	BL [22]
	A	1	CRC [87]
Y225	C	1	HNC [96]
G230	R	1	TC [172]
	E	1	EH [173]
R234	Q	2	TC [124] Glb [174]
F238	L	1	EC [48]
F241	L	1	Glb. [59]
	S	2	CRC [175] ASD [108]
P244	S	1	MM [176]
P246	L	4	Glb. [31] [36] BRR [50] LDD [55]
L247	F	1	SCLC [40]
P248	H	1	SCLC [40]
G251	C	2	Glb. [31]
	V	2	HNC [9]MM [11]
D252	Y	3	Glb. [36] EC [16] CRC [89]
	G	2	Glb. [177] PHTS [49]
	V	2	EC [48] TC [111]

continued on following page

continued from previous page

Resid.	Mutation	Num. cases	Phenotype (refs.)
I253	N	3	AA [81] BC [178] [179]
	S	1	Glb. [19]
V255	A	1	ASD [60]
	G	1	EC [16]
E256	K	1	BC [72]
	Q	1	CRC [89]
K257	L	1	Glb [180]
M264	R	1	MM [11]
M270	T	1	EC [16]
E271	S	1	Glb. [59]
H272	Y	1	PC [71]
	P	1	BL [22]
	R	2	EC [43] MM [11]
F272	S	1	BC [181]
W274	R	1	CRC [29]
	G	1	AA [31]
V275	A	1	EC [42]
	L	1	EC [16]
N276	S	1	ASD [125]
T277	I	1	Glb. [59]
	R	1	PHTS [49]
P283	S	1	EC [17]
E284	K	1	EC [98]
K289	E	1	CD [182]
E291	N	1	EC [183]
R308	C	1	MM [102]
E314	K	1	OC [2]
L318	F	1	TC [124] EC [48]
N323	K	1	CRC [87]
	T	1	EC [16]
L325	R	1	EC [16]
D326	G	1	AA [74]
	N	1	ASD [35]
D331	H	1	HNC [9]
	G	2	EC [34] HNC [184]
R335	L	1	CD [185]
	G	1	EC [16]
Y336	C	1	CRC [78]
S338	T	1	EC [20]
P339	S	2	TC [124]
	L	2	MM [96] LC [183]
F341	V	5	Glb. [95] EC [16] [20] BC [179] (x2)
	C	2	EC [39] CRC [66]
K342	N	3	LBCL [13] PHTS [49] EC [16]
V343	I	1	CD [143]
K344	R	1	PC [71]
L345	Q	1	Glb. [186]
F347	L	1	CD [13]
T348	I	2	EH [56] PC [71]
P345	R	1	BC [179]
S364	C	1	Glb. [80]
V369	G	1	PC [71]
E373	K	1	UBC [108]
D375	N	1	Glb. [5]
T382	S	1	PC [71]
T383	I	1	CCSK [1]
H397	Y	1	GC [139]
T398	S	1	Glb. [187]
Q399	R	1	EC [58]
T401	I	1	Glb. [5]

B.2 - References

B1. Little, S. E., Bax, D. A., Rodriguez-Pinilla, M., Natrajan, R., Messahel, B., Pritchard-Jones, K., Vujanic, G. M., Reis-Filho, J. S. & Jones, C. Multifaceted dysregulation of the epidermal growth factor receptor pathway in clear cell sarcoma of the kidney. *Clinical cancer research : an official journal of the American Association for Cancer Research* **13,** 4360–4 (2007).

B2. Sato, N, Nishida, M & Noguchi, M. [An approach to early genetic alterations in precancerous cells]. *Human cell* **13,** 103–8 (2000).

B3. Jaiswal, B. S. *et al.* Somatic mutations in p85alpha promote tumorigenesis through class IA PI3K activation. *Cancer cell* **16,** 463–74 (2009).

B4. Seshagiri, S. *et al.* Recurrent R-spondin fusions in colon cancer. *Nature* **488,** 660–4 (2012).

B5. Fan, X., Aalto, Y., Sanko, S. G., Knuutila, S., Klatzmann, D. & Castresana, J. S. Genetic profile, PTEN mutation and therapeutic role of PTEN in glioblastomas. *International journal of oncology* **21,** 1141–50 (2002).

B6. Holway, A. H. *et al.* Somatic mutation of PTEN in vulvar cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **6,** 3228–35 (2000).

B7. Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science (New York, N.Y.)* **321,** 1807–12 (2008).

B8. Rudd, M. L., Price, J. C., Fogoros, S., Godwin, A. K., Sgroi, D. C., Merino, M. J. & Bell, D. W. A unique spectrum of somatic PIK3CA (p110alpha) mutations within primary endometrial carcinomas. *Clinical cancer research : an official journal of the American Association for Cancer Research* **17,** 1331–40 (2011).

B9. Dulak, A. M. *et al.* Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nature genetics* **45,** 478–86 (2013).

B10. Bass, A. J. *et al.* Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513,** 202–209 (2014).

B11. Akbani, R. *et al.* Genomic Classification of Cutaneous Melanoma. English. *Cell* **161,** 1681–1696 (2015).

B12. Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).

B13. Grønbaek, K, Zeuthen, J, Guldberg, P, Ralfkiaer, E & Hou-Jensen, K. Alterations of the MMAC1/PTEN gene in lymphoid malignancies. *Blood* **91,** 4388–90 (1998).

B14. Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455,** 1069–75 (2008).

B15. Wang, D. S., Rieger-Christ, K, Latini, J. M., Moinzadeh, A, Stoffel, J, Pezza, J. A., Saini, K, Libertino, J. A. & Summerhayes, I. C. Molecular analysis of PTEN and MXI1 in primary bladder carcinoma. *International journal of cancer. Journal international du cancer* **88,** 620–5 (2000).

B16. Kandoth, C. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497,** 67–73 (2013).

B17. Mutter, G. L. Histopathology of genetically defined endometrial precancers. *International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists* **19,** 301–9 (2000).

B18. Duerr, E. M. *et al.* PTEN mutations in gliomas and glioneuronal tumors. *Oncogene* **16,** 2259–64 (1998).

B19. Bleeker, F. E. *et al.* Mutational profiling of kinases in glioblastoma. *BMC cancer* **14,** 718 (2014).

B20. Minaguchi, T *et al.* PTEN mutation located only outside exons 5, 6, and 7 is an independent predictor of favorable survival in endometrial carcinomas. *Clinical cancer research : an official journal of the American Association for Cancer Research* **7,** 2636–42 (2001).

B21. Yokoyama, Y, Wan, X, Shinohara, A, Takahashi, S, Takahashi, Y, Niwa, K & Tamaya, T. Expression of PTEN and PTEN pseudogene in endometrial carcinoma. *International journal of molecular medicine* **6,** 47–50 (2000).

B22. Love, C. *et al.* The genetic landscape of mutations in Burkitt lymphoma. *Nature genetics* **44,** 1321–5 (2012).

B23. Hühns, M., Salem, T., Schneider, B., Krohn, M., Linnebacher, M. & Prall, F. PTEN mutation, loss of heterozygosity, promoter methylation and expression in colorectal carcinoma: two hits on the gene? *Oncology reports* **31,** 2236–44 (2014).

B24. Cancer Genome Atlas. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–7 (2012).

B25. Lawrence, M. S. *et al.* Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517,** 576–582 (2015).

B26. Ikeda, Y. *et al.* Genome-wide single nucleotide polymorphism arrays as a diagnostic tool in patients with synchronous endometrial and ovarian cancer. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society* **22,** 725–31 (2012).

B27. Bussaglia, E, del Rio, E, Matias-Guiu, X & Prat, J. PTEN mutations in endometrial carcinomas: a molecular and clinicopathologic analysis of 38 cases. *Human pathology* **31,** 312–7 (2000).

B28. Catasus, L., Gallardo, A., Cuatrecasas, M. & Prat, J. PIK3CA mutations in the kinase domain (exon 20) of uterine endometrial adenocarcinomas are associated with adverse prognostic parameters. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **21,** 131–9 (2008).

B29. Berg, M. *et al.* DNA sequence profiles of the colorectal cancer critical gene set KRAS-BRAF-PIK3CA-PTEN-TP53 related to age at disease onset. *PloS one* **5,** e13978 (2010).

B30. Sakai, A, Thieblemont, C, Wellmann, A, Jaffe, E. S. & Raffeld, M. PTEN gene alterations in lymphoid neoplasms. *Blood* **92,** 3410–5 (1998).

B31. Schmidt, E. E., Ichimura, K, Goike, H. M., Moshref, A, Liu, L & Collins, V. P. Mutational profile of the PTEN gene in primary human astrocytic tumors and cultivated xenografts. *Journal of neuropathology and experimental neurology* **58,** 1170–83 (1999).

B32. Reis, R. M., Könü-Lebleblicioglu, D, Lopes, J. M., Kleihues, P & Ohgaki, H. Genetic profile of gliosarcomas. *The American journal of pathology* **156,** 425–32 (2000).

B33. Celebi, J. T., Shendrik, I, Silvers, D. N. & Peacocke, M. Identification of PTEN mutations in metastatic melanoma specimens. *Journal of medical genetics* **37,** 653–7 (2000).

B34. Lin, W. M., Forgacs, E, Warshal, D. P., Yeh, I. T., Martin, J. S., Ashfaq, R & Muller, C. Y. Loss of heterozygosity and mutational analysis of the PTEN/MMAC1 gene in synchronous endometrial and ovarian carcinomas. *Clinical cancer research : an official journal of the American Association for Cancer Research* **4,** 2577–83 (1998).

B35. Buxbaum, J. D. *et al.* Mutation screening of the PTEN gene in patients with autism spectrum disorders and macrocephaly. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics* **144,** 484–491 (2007).

B36. Liu, W, James, C. D., Frederick, L, Alderete, B. E. & Jenkins, R. B. PTEN/MMAC1 mutations and EGFR amplification in glioblastomas. *Cancer research* **57,** 5254–7 (1997).

B37. Tashiro, H, Blazes, M. S., Wu, R, Cho, K. R., Bose, S, Wang, S. I., Li, J, Parsons, R & Ellenson, L. H. Mutations in PTEN are frequent in endometrial carcinoma but rare in other common gynecological malignancies. *Cancer research* **57,** 3935–40 (1997).

B38. Shuch, B. *et al.* Germline PTEN mutation Cowden syndrome: an underappreciated form of hereditary kidney cancer. *The Journal of urology* **190,** 1990–8 (2013).

B39. Peterson, L. M., Kipp, B. R., Halling, K. C., Kerr, S. E., Smith, D. I., Distad, T. J., Clayton, A. C. & Medeiros, F. Molecular characterization of endometrial cancer: a correlative study assessing microsatellite instability, MLH1 hypermethylation, DNA mismatch repair protein expression, and PTEN, PIK3CA, KRAS, and BRAF mutation analysis. *International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists* **31,** 195–205 (2012).

B40. Kurose, K, Zhou, X. P., Araki, T, Cannistra, S. A., Maher, E. R. & Eng, C. Frequent loss of PTEN expression is linked to elevated phosphorylated Akt levels, but not associated with p27 and cyclin D1 expression, in primary epithelial ovarian carcinomas. *The American journal of pathology* **158,** 2097–106 (2001).

B41. Celebi, J. T., Tsou, H. C., Chen, F. F., Zhang, H, Ping, X. L., Lebwohl, M. G., Kezis, J & Peacocke, M. Phenotypic findings of Cowden syndrome and Bannayan-Zonana syndrome in a family associated with a single germline mutation in PTEN. *Journal of medical genetics* **36,** 360–4 (1999).

B42. Byron, S. A., Gartside Michael G.Wellens, C. L., Mallon Mary A. aand Powell, M. A., Goodfellow, P. J. & Pollock, P. M. *Cancer Research* **63,** 10005.

B43. Murayama-Hosokawa, S *et al.* Genome-wide single-nucleotide polymorphism arrays in endometrial carcinomas associate extensive chromosomal instability with poor prognosis and unveil frequent chromosomal imbalances involved in the PI3-kinase pathway. *Oncogene* **29,** 1897–908 (2010).

B44. Kanaya, T., Kyo, S., Sakaguchi, J., Maida, Y., Nakamura, M., Takakura, M., Hashimoto, M., Mizumoto, Y. & Inoue, M. Association of mismatch repair deficiency with PTEN frameshift mutations in endometrial cancers and the precursors in a Japanese population. *American journal of clinical pathology* **124,** 89–96 (2005).

B45. Halachmi, N *et al.* Somatic mutations of the PTEN tumor suppressor gene in sporadic follicular thyroid tumors. *Genes, chromosomes & cancer* **23,** 239–43 (1998).

B46. Davies, M. P. *et al.* Mutation in the PTEN/MMAC1 gene in archival low grade and high grade gliomas. *British journal of cancer* **79,** 1542–8 (1999).

B47. Akiyama-Abe, A *et al.* Loss of PTEN expression is an independent predictor of favourable survival in endometrial carcinomas. *British journal of cancer* **109,** 1703–10 (2013).

B48. Nagy, E, Gajjar, K. B., Patel, I. I., Taylor, S, Martin-Hirsch, P. L., Stringfellow, H. F., Martin, F. L. & Phillips, D. H. MGMT promoter hypermethylation and K-RAS, PTEN and TP53 mutations in tamoxifen-exposed and non-exposed endometrial cancer cases. *British journal of cancer* **110,** 2874–80 (2014).

B49. Bubien, V. *et al.* High cumulative risks of cancer in patients with PTEN hamartoma tumour syndrome. *Journal of medical genetics* **50,** 255–63 (2013).

B50. Marsh, D. J. *et al.* Germline PTEN mutations in Cowden syndrome-like families. *Journal of medical genetics* **35,** 881–5 (1998).

B51. Zhang, L., Vlad, A., Milcarek, C. & Finn, O. J. Human mucin MUC1 RNA undergoes different types of alternative splicing resulting in multiple isoforms. *Cancer immunology, immunotherapy : CII* **62,** 423–35 (2013).

B52. Schmidt, M. C. *et al.* Impact of genotype and morphology on the prognosis of glioblastoma. *Journal of neuropathology and experimental neurology* **61,** 321–8 (2002).

B53. Fults, D, Pedone, C. A., Thompson, G. E., Uchiyama, C. M., Gumpper, K. L., Iliev, D, Vinson, V. L., Tavtigian, S. V. & Perry, W. L. Microsatellite deletion mapping on chromosome 10q and mutation analysis of MMAC1, FAS, and MXI1 in human glioblastoma multiforme. *International journal of oncology* **12,** 905–10 (1998).

B54. Guldberg, P, thor Straten, P, Birck, A, Ahrenkiel, V, Kirkin, A. F. & Zeuthen, J. Disruption of the MMAC1/PTEN gene by deletion or mutation is a frequent event in malignant melanoma. *Cancer research* **57,** 3660–3 (1997).

B55. Zhou, M., Gu, L., Findley, H. W., Jiang, R. & Woods, W. G. PTEN reverses MDM2-mediated chemotherapy resistance by interacting with p53 in acute lymphoblastic leukemia cells. *Cancer research* **63,** 6357–62 (2003).

B56. Maxwell, G. L., Risinger, J. I., Tong, B, Shaw, H, Barrett, J. C., Berchuck, A & Futreal, P. A. Mutation of the PTEN tumor suppressor gene is not a feature of ovarian cancers. *Gynecologic oncology* **70,** 13–6 (1998).

B57. Nickel, G. C. *et al.* Characterizing mutational heterogeneity in a glioblastoma patient with double recurrence. *PloS one* **7,** e35262 (2012).

B58. Bilbao, C. *et al.* The relationship between microsatellite instability and PTEN gene mutations in endometrial cancer. *International journal of cancer. Journal international du cancer* **119,** 563–70 (2006).

B59. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455,** 1061–8 (2008).

B60. Klein, S., Sharifi-Hannauer, P. & Martinez-Agosto, J. A. Macrocephaly as a clinical indicator of genetic subtypes in autism. *Autism research : official journal of the International Society for Autism Research* **6,** 51–6 (2013).

B61. Gast, A., Scherer, D., Chen, B., Bloethner, S., Melchert, S., Sucker, A., Hemminki, K., Schadendorf, D. & Kumar, R. Somatic alterations in the melanoma genome: a high-resolution array-based comparative genomic hybridization study. *Genes, chromosomes & cancer* **49,** 733–45 (2010).

B62. Kraus, J. A., Felsberg, J, Tonn, J. C., Reifenberger, G & Pietsch, T. Molecular genetic analysis of the TP53, PTEN, CDKN2A, EGFR, CDK4 and MDM2 tumour-associated genes in supratentorial primitive neuroectodermal tumours and glioblastomas of childhood. *Neuropathology and applied neurobiology* **28,** 325–33 (2002).

B63. Izumoto, S, Ohnishi, T, Kanemura, H, Arita, N, Maruno, M, Moriuchi, T, Suzuki, S & Yoshimine, T. PTEN mutations in malignant gliomas and their relation with meningeal gliomatosis. *Journal of neuro-oncology* **53,** 21–6 (2001).

B64. Varga, E. A., Pastore, M., Prior, T., Herman, G. E. & McBride, K. L. The prevalence of PTEN mutations in a clinical pediatric cohort with autism spectrum disorders, developmental delay, and macrocephaly. *Genetics in medicine : official journal of the American College of Medical Genetics* **11,** 111–7 (2009).

B65. Weigelt, B., Warne, P. H., Lambros, M. B., Reis-Filho, J. S. & Downward, J. PI3K pathway dependencies in endometrioid endometrial cancer cell lines. *Clinical cancer research : an official journal of the American Association for Cancer Research* **19,** 3533–44 (2013).

B66. Danielsen, S. A., Lind, G. E., Bjørnslett, M., Meling, G. I., Rognum, T. O., Heim, S. & Lothe, R. A. Novel mutations of the suppressor gene PTEN in colorectal carcinomas stratified by microsatellite instability- and TP53 mutation- status. *Human mutation* **29,** E252–62 (2008).

B67. Integrated genomic analyses of ovarian carcinoma. *Nature* **474,** 609–15 (2011).

B68. Vega, A., Torres, J., Torres, M., Cameselle-Teijeiro, J., Macia, M., Carracedo, A. & Pulido, R. A novel loss-of-function mutation (N48K) in the PTEN gene in a Spanish patient with Cowden disease. *The Journal of investigative dermatology* **121,** 1356–9 (2003).

B69. Stechishin, O. D., Luchman, H. A., Ruan, Y., Blough, M. D., Nguyen, S. A., Kelly, J. J., Cairncross, J. G. & Weiss, S. On-target JAK2/STAT3 inhibition slows disease progression in orthotopic xenografts of human glioblastoma brain tumor stem cells. *Neuro-oncology* **15,** 198–207 (2013).

B70. De Lima, E. U., Soares, I. C., Danilovic, D. L. S. & Marui, S. New mutation in the PTEN gene in a Brazilian patient with Cowden's syndrome. *Arquivos brasileiros de endocrinologia e metabologia* **56,** 592–6 (2012).

B71. Dong, J. T., Li, C. L., Sipe, T. W. & Frierson, H. F. Mutations of PTEN/MMAC1 in primary prostate cancers from Chinese patients. *Clinical cancer research : an official journal of the American Association for Cancer Research* **7,** 304–8 (2001).

B72. Yang, P. [To promote the repairment of CNS by targeting PTen]. *Sheng li ke xue jin zhan [Progress in physiology]* **41,** 313–6 (2010).

B73. Konopka, B., Janiec-Jankowska, A., Czapczak, D., Paszko, Z., Bidziński, M., Olszewski, W. & Goluda, C. Molecular genetic defects in endometrial carcinomas: microsatellite instability, PTEN and beta-catenin (CTNNB1) genes mutations. *Journal of cancer research and clinical oncology* **133,** 361–71 (2007).

B74. Sasaki, H, Zlatescu, M. C., Betensky, R. A., Ino, Y, Cairncross, J. G. & Louis, D. N. PTEN is a target of chromosome 10q loss in anaplastic oligodendrogliomas and PTEN alterations are associated with poor prognosis. *The American journal of pathology* **159,** 359–67 (2001).

B75. Steck, P. A. *et al.* Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nature genetics* **15,** 356–62 (1997).

B76. Chiariello, E, Roz, L, Albarosa, R, Magnani, I & Finocchiaro, G. PTEN/MMAC1 mutations in primary glioblastomas and short-term cultures of malignant gliomas. *Oncogene* **16,** 541–5 (1998).

B77. Ueda, K, Nishijima, M, Inui, H, Watatani, M, Yayoi, E, Okamura, J, Yasutomi, M, Nakamura, Y & Miyoshi, Y. Infrequent mutations in the PTEN/MMAC1 gene among primary breast cancers. *Japanese journal of cancer research : Gann* **89,** 17–21 (1998).

B78. Peeters, M. *et al.* Massively parallel tumor multigene sequencing to evaluate response to panitumumab in a randomized phase III study of metastatic colorectal cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **19,** 1902–12 (2013).

B79. Maier, D, Zhang, Z, Taylor, E, Hamou, M. F., Gratzl, O, Van Meir, E. G., Scott, R. J. & Merlo, A. Somatic deletion mapping on chromosome 10 and sequence analysis of PTEN/MMAC1 point to the 10q25-26 region as the primary target in low-grade and high-grade gliomas. *Oncogene* **16,** 3331–5 (1998).

B80. Tunca, B., Bekar, A., Cecener, G., Egeli, U., Vatan, O., Tolunay, S., Kocaeli, H. & Aksoy, K. Impact of novel PTEN mutations in Turkish patients with glioblastoma multiforme. *Journal of neuro-oncology* **82,** 263–9 (2007).

B81. Zhou, X. P. *et al.* Germline mutations of p53 but not p16/CDKN2 or PTEN/MMAC1 tumor suppressor genes predispose to gliomas. The ANOCEF Group. Association des NeuroOncologues d'Expression Française. *Annals of neurology* **46,** 913–6 (1999).

B82. Butler, M. P., Wang, S. I., Chaganti, R. S., Parsons, R & Dalla-Favera, R. Analysis of PTEN mutations and deletions in B-cell non-Hodgkin's lymphomas. *Genes, chromosomes & cancer* **24,** 322–7 (1999).

B83. Yost, S. E., Pastorino, S., Rozenzhak, S., Smith, E. N., Chao, Y. S., Jiang, P., Kesari, S., Frazer, K. A. & Harismendy, O. High-resolution mutational profiling suggests the genetic validity of glioblastoma patient-derived pre-clinical models. *PloS one* **8,** e56185 (2013).

B84. Konopka, G & Bonni, A. Signaling pathways regulating gliomagenesis. *Current molecular medicine* **3,** 73–84 (2003).

B85. Reardon, W, Zhou, X. P. & Eng, C. A novel germline mutation of the PTEN gene in a patient with macrocephaly, ventricular dilatation, and features of VATER association. *Journal of medical genetics* **38,** 820–3 (2001).

B86. Sharma, M. R., Petty, E. M. & Lesperance, M. M. Airway obstruction caused by PTEN hamartoma (Bannayan-Riley-Ruvalcaba) syndrome. *Archives of otolaryngology–head & neck surgery* **133,** 1157–60 (2007).

B87. Nassif, N. T., Lobo, G. P., Wu, X., Henderson, C. J. A., Morrison, C. D., Eng, C., Jalaludin, B. & Segelov, E. PTEN mutations are common in sporadic microsatellite stable colorectal cancer. *Oncogene* **23,** 617–28 (2004).

B88. Wang, X., Bhattacharyya, D., Dennewitz, M. B., Kalinichenko, V. V., Zhou, Y., Lepe, R. & Costa, R. H. Rapid hepatocyte nuclear translocation of the Forkhead Box M1B (FoxM1B) transcription factor caused a transient increase in size of regenerating transgenic hepatocytes. *Gene expression* **11,** 149–62 (2003).

B89. Day, F. L. *et al.* PIK3CA and PTEN gene and exon mutation-specific clinicopathologic and molecular associations in colorectal cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **19,** 3285–96 (2013).

B90. Actor, B., Cobbers, J. M.J. L., Büschges, R., Wolter, M., Knobbe, C. B., Lichter, P., Reifenberger, G. & Weber, R. G. Comprehensive analysis of genomic alterations in gliosarcoma and its two tissue components. *Genes, chromosomes & cancer* **34,** 416–27 (2002).

B91. Cohn, D. E., Basil, J. B., Venegoni, A. R., Mutch, D. G., Rader, J. S., Herzog, T. J., Gersell, D. J. & Goodfellow, P. J. Absence of PTEN repeat tract mutation in endometrial cancers with microsatellite instability. *Gynecologic oncology* **79,** 101–6 (2000).

B92. Tsou, H. C. *et al.* The genetic basis of Cowden's syndrome: three novel mutations in PTEN/MMAC1/TEP1. *Human genetics* **102,** 467–73 (1998).

B93. Kondo, K *et al.* PTEN/MMAC1/TEP1 mutations in human primary renal-cell carcinomas and renal carcinoma cell lines. *International journal of cancer. Journal international du cancer* **91,** 219–24 (2001).

B94. Hobert, J. A., Embacher, R., Mester, J. L., Frazier, T. W. & Eng, C. Biochemical screening and PTEN mutation analysis in individuals with autism spectrum disorders and macrocephaly. *European journal of human genetics : EJHG* **22,** 273–6 (2014).

B95. Kato, H *et al.* Functional evaluation of p53 and PTEN gene mutations in gliomas. *Clinical cancer research : an official journal of the American Association for Cancer Research* **6,** 3937–43 (2000).

B96. Poetsch, M, Dittberner, T & Woenckhaus, C. PTEN/MMAC1 in malignant melanoma and its importance for tumor progression. *Cancer genetics and cytogenetics* **125,** 21–6 (2001).

B97. Rasheed, B. K., Stenzel, T. T., McLendon, R. E., Parsons, R, Friedman, A. H., Friedman, H. S., Bigner, D. D. & Bigner, S. H. PTEN gene mutations are seen in high-grade but not in low-grade gliomas. *Cancer research* **57,** 4187–90 (1997).

B98. Sun, H, Enomoto, T, Fujita, M, Wada, H, Yoshino, K, Ozaki, K, Nakamura, T & Murata, Y. Mutational analysis of the PTEN gene in endometrial carcinoma and hyperplasia. *American journal of clinical pathology* **115,** 32–8 (2001).

B99. Figer, A, Kaplan, A, Frydman, M, Lev, D, Paswell, J, Papa, M. Z., Goldman, B & Friedman, E. Germline mutations in the PTEN gene in Israeli patients with Bannayan-Riley-Ruvalcaba syndrome and women with familial breast cancer. *Clinical genetics* **62,** 298–302 (2002).

B100. Kim, Y. *et al.* Integrative and comparative genomic analysis of lung squamous cell carcinomas in East Asian patients. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **32,** 121–8 (2014).

B101. Wu, H., Weng, D. & Xing, H. Reversal of multidrug resistance and inhibition of phosphorylation of AKT in human ovarian cancer cell line by wild-type PTEN gene. *Journal of Huazhong University of Science and Technology. Medical sciences = Hua zhong ke ji da xue xue bao. Yi xue Ying De wen ban = Huazhong keji daxue xuebao. Yixue Yingdewen ban* **27,** 713–6 (2007).

B102. Abdel-Rahman, M. H., Yang, Y., Zhou, X.-P., Craig, E. L., Davidorf, F. H. & Eng, C. High frequency of submicroscopic hemizygous deletion is a major mechanism of loss of expression of PTEN in uveal melanoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **24,** 288–95 (2006).

B103. Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486,** 395–9 (2012).

B104. Suzuki, A *et al.* High cancer susceptibility and embryonic lethality associated with mutation of the PTEN tumor suppressor gene in mice. *Current biology : CB* **8,** 1169–78 (1998).

B105. Zhang, Q. & Claret, F. X. Phosphatases: the new brakes for cancer development? *Enzyme research* **2012,** 659649 (2012).

B106. Nassiri, I., Faghihi, M. & Tavassoli, M. Pharmacogenomic profiling of the PI3K/PTEN pathway in sporadic breast cancer. *Iranian biomedical journal* **13,** 79–86 (2009).

B107. Hollestelle, A., Elstrodt, F., Nagel, J. H. A., Kallemeijn, W. W. & Schutte, M. Phosphatidylinositol-3-OH kinase or RAS pathway mutations in human breast cancer cell lines. *Molecular cancer research : MCR* **5,** 195–201 (2007).

B108. Butler, M. G. *et al.* Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *Journal of medical genetics* **42,** 318–21 (2005).

B109. Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486,** 400–4 (2012).

B110. Levine, R. L., Cargile, C. B., Blazes, M. S., van Rees, B, Kurman, R. J. & Ellenson, L. H. PTEN mutations and microsatellite instability in complex atypical hyperplasia, a precursor lesion to uterine endometrioid carcinoma. *Cancer research* **58,** 3254–8 (1998).

B111. Bussaglia, E. *et al.* PTEN mutations in eight Spanish families and one Brazilian family with Cowden syndrome. *The Journal of investigative dermatology* **118,** 639–44 (2002).

B112. Rasheed, A. *et al.* Molecular markers of prognosis in astrocytic tumors. *Cancer* **94,** 2688–97 (2002).

B113. Raffel, C, Frederick, L, O'Fallon, J. R., Atherton-Skaff, P, Perry, A, Jenkins, R. B. & James, C. D. Analysis of oncogene and tumor suppressor gene alterations in pediatric malignant astrocytomas reveals reduced survival for patients with PTEN mutations. *Clinical cancer research : an official journal of the American Association for Cancer Research* **5,** 4085–90 (1999).

B114. Zainuddin, N, Jaafar, H, Isa, M. N. & Abdullah, J. M. Presence of allelic loss and PTEN mutations in malignant gliomas from Malay patients. *The Medical journal of Malaysia* **59,** 468–79 (2004).

B115. Furnari, F. B., Lin, H, Huang, H. S. & Cavenee, W. K. Growth suppression of glioma cells by PTEN requires a functional phosphatase catalytic domain. *Proceedings of the National Academy of Sciences of the United States of America* **94,** 12479–84 (1997).

B116. Risinger, J. I., Hayes, K, Maxwell, G. L., Carney, M. E., Dodge, R. K., Barrett, J. C. & Berchuck, A. PTEN mutation in endometrial cancers is associated with favorable clinical and pathologic characteristics. *Clinical cancer research : an official journal of the American Association for Cancer Research* **4,** 3005–10 (1998).

B117. Wang, S. I., Puc, J, Li, J, Bruce, J. N., Cairns, P, Sidransky, D & Parsons, R. Somatic mutations of PTEN in glioblastoma multiforme. *Cancer research* **57,** 4183–6 (1997).

B118. Paparo, L., Rossi, G. B., Delrio, P., Rega, D., Duraturo, F., Liccardo, R., Debellis, M., Izzo, P. & De Rosa, M. Differential expression of PTEN gene correlates with phenotypic heterogeneity in three cases of patients showing clinical manifestations of PTEN hamartoma tumour syndrome. *Hereditary cancer in clinical practice* **11,** 8 (2013).

B119. Ngeow, J. *et al.* Prevalence of germline PTEN, BMPR1A, SMAD4, STK11, and ENG mutations in patients with moderate-load colorectal polyps. *Gastroenterology* **144,** 1402–9, 1409.e1–5 (2013).

B120. Sutphen, R, Diamond, T. M., Minton, S. E., Peacocke, M, Tsou, H. C. & Root, A. W. Severe Lhermitte-Duclos disease with unique germline mutation of PTEN. *American journal of medical genetics* **82,** 290–3 (1999).

B121. Xing, F *et al.* Concurrent loss of the PTEN and RB1 tumor suppressors attenuates RAF dependence in melanomas harboring (V600E)BRAF. *Oncogene* **31,** 446–57 (2012).

B122. Orbo, A, Kaino, T, Arnes, M, Kopp, M & Eklo, K. Genetic derangements in the tumor suppressor gene PTEN in endometrial precancers as prognostic markers for cancer development: a population-based study from northern Norway with long-term follow-up. *Gynecologic oncology* **95,** 82–8 (2004).

B123. Trédan, O. *et al.* Predicting everolimus treatment efficacy in patients with advanced endometrial carcinoma: a GINECO group study. *Targeted oncology* **8,** 243–51 (2013).

B124. Hou, P. *et al.* Genetic alterations and their relationship in the phosphatidylinositol 3-kinase/Akt pathway in thyroid cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **13,** 1161–70 (2007).

B125. Orrico, A, Galli, L, Buoni, S, Orsi, A, Vonella, G & Sorrentino, V. Novel PTEN mutations in neurodevelopmental disorders and macrocephaly. *Clinical genetics* **75,** 195–8 (2009).

B126. Abdullah, J. M., Zainuddin, N., Sulong, S., Jaafar, H. & Isa, M. N. Molecular genetic analysis of phosphatase and tensin homolog and p16 tumor suppressor genes in patients with malignant glioma. *Neurosurgical focus* **14,** e6 (2003).

B127. Wang, S., Liu, F., Li, X. & Dong, P. [Research of the expression of PTEN in squamous cell carcinoma of larynx by the tissue chip technology]. *Lin chuang er bi yan hou tou jing wai ke za zhi = Journal of clinical otorhinolaryngology, head, and neck surgery* **21,** 1081–4 (2007).

B128. Kurose, K, Araki, T, Matsunaka, T, Takada, Y & Emi, M. Variant manifestation of Cowden disease in Japan: hamartomatous polyposis of the digestive tract with mutation of the PTEN gene. *American journal of human genetics* **64,** 308–10 (1999).

B129. Di Lorito, A., Rosini, S., Falò, E., Gustapane, S., Gomes, M., Costa, J. L. & Schmitt, F. C. Molecular alterations in endometrial archived liquid-based cytology. *Diagnostic cytopathology* **41,** 492–6 (2013).

B130. Heindl, M. *et al.* Autoimmunity, intestinal lymphoid hyperplasia, and defects in mucosal B-cell homeostasis in patients with PTEN hamartoma tumor syndrome. *Gastroenterology* **142,** 1093–1096.e6 (2012).

B131. Obata, K, Morland, S. J., Watson, R. H., Hitchcock, A, Chenevix-Trench, G, Thomas, E. J. & Campbell, I. G. Frequent PTEN/MMAC mutations in endometrioid but not serous or mucinous epithelial ovarian tumors. *Cancer research* **58,** 2095–7 (1998).

B132. Ellison, J. & Driscoll, D. Novel human pathological mutations. Gene symbol: PTEN. Disease: PTEN Hamartoma Tumour Syndrome. *Human genetics* **127,** 124 (2010).

B133. Liaw, D *et al.* Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nature genetics* **16,** 64–7 (1997).

B134. Konopka, B., Paszko, Z., Janiec-Jankowska, A. & Goluda, M. Assessment of the quality and frequency of mutations occurrence in PTEN gene in endometrial carcinomas and hyperplasias. *Cancer letters* **178,** 43–51 (2002).

B135. Turbiner, J., Moreno-Bueno, G., Dahiya, S., Sánchez-Estevez, C., Hardisson, D., Prat, J., Oliva, E. & Palacios, J. Clinicopathological and molecular analysis of endometrial carcinoma associated with tamoxifen. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **21,** 925–36 (2008).

B136. Hirai, Y, Tanaka, N, Furuta, R, Kawaguchi, T, Sakamoto, M, Shirahama, S & Noda, T. Somatic mutations of the PTEN/NMAC1 gene associated with frequent chromosomal loss detected using comparative genomic hybridization in endometrial cancer. *Gynecologic oncology* **83,** 81–8 (2001).

B137. Oliva, E, Sarrió, D, Brachtel, E. F., Sánchez-Estévez, C, Soslow, R. A., Moreno-Bueno, G & Palacios, J. High frequency of beta-catenin mutations in borderline endometrioid tumours of the ovary. *The Journal of pathology* **208,** 708–13 (2006).

B138. Kinross, K. M. *et al.* An activating Pik3ca mutation coupled with Pten loss is sufficient to initiate ovarian tumorigenesis in mice. *The Journal of clinical investigation* **122,** 553–7 (2012).

B139. Werner, D., Atmaca, A., Pauligk, C., Pustowka, A., Jäger, E. & Al-Batran, S.-E. Phase I study of everolimus and mitomycin C for patients with metastatic esophagogastric adenocarcinoma. *Cancer medicine* **2,** 325–33 (2013).

B140. Koul, D., Shen, R., Garyali, A., Ke, L. D., Liu, T.-J. & Yung, W. K. A. MMAC/PTEN tumor suppressor gene regulates vascular endothelial growth factor-mediated angiogenesis in prostate cancer. *International journal of oncology* **21,** 469–75 (2002).

B141. Simpkins, S. B., Peiffer-Schneider, S, Mutch, D. G., Gersell, D & Goodfellow, P. J. PTEN mutations in endometrial cancers with 10q LOH: additional evidence for the involvement of multiple tumor suppressors. *Gynecologic oncology* **71,** 391–5 (1998).

B142. Villeneuve, H., Tremblay, S., Galiatsatos, P., Hamel, N., Guertin, L., Morency, R. & Tischkowitz, M. Acinic cell carcinoma of the retromolar trigone region: expanding the tumor phenotype in Cowden syndrome? *Familial cancer* **10,** 691–4 (2011).

B143. Lynch, E. D. *et al.* Inherited mutations in PTEN that are associated with breast cancer, cowden disease, and juvenile polyposis. *American journal of human genetics* **61,** 1254–60 (1997).

B144. Saito, M *et al.* Allelic imbalance and mutations of the PTEN gene in ovarian cancer. *International journal of cancer. Journal international du cancer* **85,** 160–5 (2000).

B145. Grasso, C. S. *et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487,** 239–43 (2012).

B146. Schöndorf, T., Ebert, M. P., Hoffmann, J., Becker, M., Moser, N., Pur, S., Göhring, U.-J. & Weisshaar, M.-P. Hypermethylation of the PTEN gene in ovarian cancer cell lines. *Cancer letters* **207,** 215–20 (2004).

B147. Zmajkovicova, K., Jesenberger, V., Catalanotti, F., Baumgartner, C., Reyes, G. & Baccarini, M. MEK1 is required for PTEN membrane recruitment, AKT regulation, and the maintenance of peripheral tolerance. *Molecular cell* **50,** 43–55 (2013).

B148. Busa, T., Chabrol, B., Perret, O., Longy, M. & Philip, N. Novel PTEN germline mutation in a family with mild phenotype: difficulties in genetic counseling. *Gene* **512,** 194–7 (2013).

B149. Kubo, Y, Urano, Y, Hida, Y, Ikeuchi, T, Nomoto, M, Kunitomo, K & Arase, S. A novel PTEN mutation in a Japanese patient with Cowden disease. *The British journal of dermatology* **142,** 1100–5 (2000).

B150. Jenny, B., Radovanovic, I., Haenggeli, C.-A., Delavelle, J., Rüfenacht, D., Kaelin, A., Blouin, J.-L., Bottani, A. & Rilliet, B. Association of multiple vertebral hemangiomas and severe paraparesis in a patient with a PTEN hamartoma tumor syndrome. Case report. *Journal of neurosurgery* **107,** 307–13 (2007).

B151. Scala, S, Bruni, P, Lo Muzio, L, Mignogna, M, Viglietto, G & Fusco, A. Novel mutation of the PTEN gene in an Italian Cowden's disease kindred. *International journal of oncology* **13,** 665–8 (1998).

B152. He, J *et al.* Glioblastomas with an oligodendroglial component: a pathological and molecular study. *Journal of neuropathology and experimental neurology* **60,** 863–71 (2001).

B153. Kraus, J. A. *et al.* Molecular genetic alterations in glioblastomas with oligodendroglial component. *Acta neuropathologica* **101,** 311–20 (2001).

B154. Perren, A, Komminoth, P, Saremaslani, P, Matter, C, Feurer, S, Lees, J. A., Heitz, P. U. & Eng, C. Mutation and expression analyses reveal differential subcellular compartmentalization of PTEN in endocrine pancreatic tumors compared to normal islet cells. *The American journal of pathology* **157,** 1097–103 (2000).

B155. Sozopoulos, E., Litsiou, H., Voutsinas, G., Mitsiades, N., Anagnostakis, N., Tseva, T., Patsouris, E. & Tseleni-Balafouta, S. Mutational and immunohistochemical study of the PI3K/Akt pathway in papillary thyroid carcinoma in Greece. *Endocrine pathology* **21,** 90–100 (2010).

B156. Stjernström, A., Karlsson, C., Fernandez, O. J., Söderkvist, P., Karlsson, M. G. & Thunell, L. K. Alterations of INPP4B, PIK3CA and pAkt of the PI3K pathway are associated with squamous cell carcinoma of the lung. *Cancer medicine* **3,** 337–48 (2014).

B157. Jin, G. *et al.* PTEN mutations and relationship to EGFR, ERBB2, KRAS, and TP53 mutations in non-small cell lung cancers. *Lung cancer (Amsterdam, Netherlands)* **69,** 279–83 (2010).

B158. Tortosa, A, Ino, Y, Odell, N, Swilley, S, Sasaki, H, Louis, D. N. & Henson, J. W. Molecular genetics of radiographically defined de novo glioblastoma multiforme. *Neuropathology and applied neurobiology* **26,** 544–52 (2000).

B159. Karoui, M *et al.* Loss of heterozygosity on 10q and mutational status of PTEN and BMPR1A in colorectal primary tumours and metastases. *British journal of cancer* **90,** 1230–4 (2004).

B160. Tohma, Y, Gratas, C, Biernat, W, Peraud, A, Fukuda, M, Yonekawa, Y, Kleihues, P & Ohgaki, H. PTEN (MMAC1) mutations are frequent in primary glioblastomas (de novo) but not in secondary glioblastomas. *Journal of neuropathology and experimental neurology* **57,** 684–9 (1998).

B161. Nelen, M. R. *et al.* Novel PTEN mutations in patients with Cowden disease: absence of clear genotype-phenotype correlations. *European journal of human genetics : EJHG* **7,** 267–73 (1999).

B162. Peifer, M. *et al.* Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nature genetics* **44,** 1104–10 (2012).

B163. Jeuken, J. W., Nelen, M. R., Vermeer, H, van Staveren, W. C., Kremer, H, van Overbeeke, J. J. & Boerman, R. H. PTEN mutation analysis in two genetic subtypes of high-grade oligodendroglial tumors. PTEN is only occasionally mutated in one of the two genetic subtypes. *Cancer genetics and cytogenetics* **119,** 42–7 (2000).

B164. Kucejova, B. *et al.* Interplay between pVHL and mTORC1 pathways in clear-cell renal cell carcinoma. *Molecular cancer research : MCR* **9,** 1255–65 (2011).

B165. Tsao, H, Zhang, X, Fowlkes, K & Haluska, F. G. Relative reciprocity of NRAS and PTEN/MMAC1 alterations in cutaneous melanoma cell lines. *Cancer research* **60,** 1800–4 (2000).

B166. Peraud, A, Watanabe, K, Schwechheimer, K, Yonekawa, Y, Kleihues, P & Ohgaki, H. Genetic profile of the giant cell glioblastoma. *Laboratory investigation; a journal of technical methods and pathology* **79,** 123–9 (1999).

B167. Kurose, K, Bando, K, Fukino, K, Sugisaki, Y, Araki, T & Emi, M. Somatic mutations of the PTEN/MMAC1 gene in fifteen Japanese endometrial cancers: evidence for inactivation of both alleles. *Japanese journal of cancer research : Gann* **89,** 842–8 (1998).

B168. Tate, G., Suzuki, T., Endo, Y. & Mitsuya, T. A novel mutation of the PTEN gene in a Japanese patient with Cowden syndrome and bilateral breast cancer. *Cancer genetics and cytogenetics* **184,** 67–71 (2008).

B169. Wakimoto, H. *et al.* Targetable signaling pathway mutations are associated with malignant phenotype in IDH-mutant gliomas. *Clinical cancer research : an official journal of the American Association for Cancer Research* **20,** 2898–909 (2014).

B170. James, C. D. *et al.* Tumor suppressor gene alterations in malignant gliomas: histopathological associations and prognostic evaluation. *International journal of oncology* **15,** 547–53 (1999).

B171. Kirches, E, Steiner, J, Schneider, T, Vorwerk, C. K., Scherlach, C, Holtkamp, N, Keilhoff, G, Eng, C & Mawrin, C. Lhermitte-Duclos disease caused by a novel germline PTEN mutation R173P in a patient presenting with psychosis. *Neuropathology and applied neurobiology* **36,** 86–9 (2010).

B172. Pita, J. M., Figueiredo, I. F., Moura, M. M., Leite, V. & Cavaco, B. M. Cell cycle deregulation and TP53 and RAS mutations are major events in poorly differentiated and undifferentiated thyroid carcinomas. *The Journal of clinical endocrinology and metabolism* **99,** E497–507 (2014).

B173. Brachtel, E. F., Sánchez-Estevez, C., Moreno-Bueno, G., Prat, J., Palacios, J. & Oliva, E. Distinct molecular alterations in complex endometrial hyperplasia (CEH) with and without immature squamous metaplasia (squamous morules). *The American journal of surgical pathology* **29,** 1322–9 (2005).

B174. Staal, F. J. T. *et al.* A novel germline mutation of PTEN associated with brain tumours of multiple lineages. *British journal of cancer* **86,** 1586–91 (2002).

B175. Shin, K.-H., Kim, J.-M., Rho, K. S., Park, K.-H., Oh, J.-E. & Min, B.-M. Inactivation of the PTEN gene by mutation, exonic deletion, and loss of transcript in human oral squamous cell carcinomas. *International journal of oncology* **21,** 997–1001 (2002).

B176. Reifenberger, J, Wolter, M, Boström, J, Büschges, R, Schulte, K. W., Megahed, M, Ruzicka, T & Reifenberger, G. Allelic losses on chromosome arm 10q and mutation of the PTEN (MMAC1) tumour suppressor gene in primary and metastatic malignant melanomas. *Virchows Archiv : an international journal of pathology* **436,** 487–93 (2000).

B177. Martinez, R., Rohde, V. & Schackert, G. Different molecular patterns in glioblastoma multiforme subtypes upon recurrence. *Journal of neuro-oncology* **96,** 321–9 (2010).

B178. Hu, Y., Sun, H., Owens, R. T., Wu, J., Chen, Y. Q., Berquin, I. M., Perry, D., O'Flaherty, J. T. & Edwards, I. J. Decorin suppresses prostate tumor growth through inhibition of epidermal growth factor and androgen receptor pathways. *Neoplasia (New York, N.Y.)* **11,** 1042–53 (2009).

B179. Kurose, K., Gilley, K., Matsumoto, S., Watson, P. H., Zhou, X.-P. & Eng, C. Frequent somatic mutations in PTEN and TP53 are mutually exclusive in the stroma of breast carcinomas. *Nature genetics* **32,** 355–7 (2002).

B180. Reis, R. M., Herva, R, Brandner, S, Koivukangas, J, Mironov, N, Bär, W, Kleihues, P & Ohgaki, H. Second primary glioblastoma. *Journal of neuropathology and experimental neurology* **60,** 208–15 (2001).

B181. Boyault, S., Drouet, Y., Navarro, C., Bachelot, T., Lasset, C., Treilleux, I., Tabone, E., Puisieux, A. & Wang, Q. Mutational characterization of individual breast tumors: TP53 and PI3K pathway genes are frequently and distinctively mutated in different subtypes. *Breast cancer research and treatment* **132,** 29–39 (2012).

B182. Chi, S. G. *et al.* Mutational abrogation of the PTEN/MMAC1 gene in gastrointestinal polyps in patients with Cowden disease. *Gastroenterology* **115,** 1084–9 (1998).

B183. Trigka, E. A. *et al.* A detailed immunohistochemical analysis of the PI3K/AKT/mTOR pathway in lung cancer: correlation with PIK3CA, AKT1, K-RAS or PTEN mutational status and clinicopathological features. *Oncology reports* **30,** 623–36 (2013).

B184. Poetsch, M., Lorenz, G. & Kleist, B. Detection of new PTEN/MMAC1 mutations in head and neck squamous cell carcinomas with loss of chromosome 10. *Cancer genetics and cytogenetics* **132,** 20–4 (2002).

B185. Sawada, T, Hamano, N, Satoh, H, Okada, T, Takeda, Y & Mabuchi, H. Mutation analysis of the PTEN / MMAC1 gene in Japanese patients with Cowden disease. *Japanese journal of cancer research : Gann* **91,** 700–5 (2000).

B186. Wang, Z. J., Taylor, F, Churchman, M, Norbury, G & Tomlinson, I. Genetic pathways of colorectal carcinogenesis rarely involve the PTEN and LKB1 genes outside the inherited hamartoma syndromes. *The American journal of pathology* **153,** 363–6 (1998).

B187. Zhao, Y., Huang, Q., Wang, A., Dong, J. & Lan, Q. Mutation of PTEN in glioma stem/progenitor cells: a case report. *Cancer genetics and cytogenetics* **195,** 183–5 (2009).

References

1. Van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nature reviews. Molecular cell biology* **9,** 112–24 (2008).
2. Shevchenko, A. & Simons, K. Lipidomics: coming to grips with lipid diversity. *Nature Reviews Molecular Cell Biology* **11,** 593–598 (2010).
3. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141,** 1117–34 (2010).
4. Cho, W. & Stahelin, R. V. Membrane-Protein Interactions in Cell Signaling and Membrane Trafficking. *Annual Review of Biophysics and Biomolecular Structure* **34,** 119–151 (2005).
5. Stahelin, R. V. Lipid binding domains: more than simple lipid effectors. *Journal of Lipid Research* **50,** S299–S304 (2009).
6. Carlton, J. G. & Cullen, P. J. Coincidence detection in phosphoinositide signaling. *Trends in cell biology* **15,** 540–7 (2005).
7. Stahelin, R. V., Scott, J. L. & Frick, C. T. Cellular and molecular interactions of phosphoinositides and peripheral proteins. *Chemistry and physics of lipids* **182,** 3–18 (2014).
8. Kutateladze, T. G. Translation of the phosphoinositide code by PI effectors. *Nature chemical biology* **6,** 507–13 (2010).
9. Lemmon, M. A. Membrane recognition by phospholipid-binding domains. *Nature reviews. Molecular cell biology* **9,** 99–111 (2008).
10. Moravcevic, K., Oxley, C. L. & Lemmon, M. A. Conditional peripheral membrane proteins: facing up to limited specificity. *Structure (London, England : 1993)* **20,** 15–27 (2012).
11. Kutateladze, T. G. Phosphatidylinositol 3-phosphate recognition and membrane docking by the FYVE domain. *Biochimica et biophysica acta* **1761,** 868–77 (2006).

12. Cho, W. & Stahelin, R. V. *Membrane binding and subcellular targeting of C2 domains* 2006.

13. DiNitto, J. P. & Lambright, D. G. Membrane and juxtamembrane targeting by PH and PTB domains. *Biochimica et biophysica acta* **1761,** 850–67 (2006).

14. Mayer, B. J., Ren, R., Clark, K. L. & Baltimore, D. *A putative modular domain present in diverse signaling proteins* 1993.

15. Letunic, I. & Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research* (2017).

16. Carpten, J. D. *et al.* A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature* **448,** 439–444 (2007).

17. Lindvall, J. M. *et al.* Bruton's tyrosine kinase: cell biology, sequence conservation, mutation spectrum, siRNA modifications, and expression profiling. *Immunological Reviews* **203,** 200–215 (2005).

18. Lemmon, M. Pleckstrin homology domains: not just for phosphoinositides: Figure 1. *Biochemical Society Transactions* **32,** 707–711 (2004).

19. Isakoff, S. J., Cardozo, T., Andreev, J., Li, Z., Ferguson, K. M., Abagyan, R., Lemmon, M. A., Aronheim, A. & Skolnik, E. Y. Identification and analysis of PH domain-containing targets of phosphatidylinositol 3-kinase using a novel in vivo assay in yeast. *EMBO Journal* **17,** 5374–5387 (1998).

20. Lietzke, S. E., Bose, S, Cronin, T, Klarlund, J, Chawla, A, Czech, M. P. & Lambright, D. G. Structural basis of 3-phosphoinositide recognition by pleckstrin homology domains. *Molecular cell* **6,** 385–94 (2000).

21. Furukawa, Y., Kawasoe, T., Daigo, Y., Nishiwaki, T., Ishiguro, H., Takahashi, M., Kitayama, J. & Nakamura, Y. Isolation of a novel human gene, ARHGAP9, encoding a rho-GTPase activating protein. *Biochemical and biophysical research communications* **284,** 643–9 (2001).

22. Zhang, T. T., Li, H., Cheung, S. M., Costantini, J. L., Hou, S., Al-Alwan, M. & Marshall, A. J. *Phosphoinositide 3-kinase-regulated adapters in lymphocyte activation* 2009.

23. Roy, S. A. B., Langlois, M.-J., Carrier, J. C., Boudreau, F., Rivard, N. & Perreault, N. Dual regulatory role for phosphatase and tensin homolog in specification of intestinal endocrine cell subtypes. *World journal of gastroenterology : WJG* **18,** 1579–89 (2012).

24. Lemmon, M. A. & Ferguson, K. M. Molecular determinants in pleckstrin homology domains that allow specific recognition of phosphoinositides. *Biochemical Society transactions* **29,** 377–84 (2001).

25. He, M., Abdi, K. M. & Bennett, V. Ankyrin-G palmitoylation and ??II-spectrin binding to phosphoinositide lipids drive lateral membrane assembly. *Journal of Cell Biology* **206,** 273–288 (2014).

26. Balla, T. Inositol-lipid binding motifs: signal integrators through protein-lipid and protein-protein interactions. *Journal of cell science* **118,** 2093–104 (2005).

27. Balla, T. Phosphoinositides: Tiny Lipids With Giant Impact on Cell Regulation. *Physiological Reviews* **93,** 1019–1137 (2013).

28. Jackson, S. G., Zhang, Y., Haslam, R. J. & Junop, M. S. Structural analysis of the carboxy terminal PH domain of pleckstrin bound to D-myo-inositol 1,2,3,5,6-pentakisphosphate. *BMC structural biology* **7,** 80 (2007).

29. Cantley, L. C. The Phosphoinositide 3-Kinase Pathway. *Science* **296** (2002).

30. Wymann, M. P. & Schneiter, R. Lipid signalling in disease. *Nature Reviews Molecular Cell Biology* **9,** 162–176 (2008).

31. Milella, M. *et al.* PTEN: Multiple Functions in Human Malignant Tumors. *Frontiers in Oncology* **5,** 24 (2015).

32. Worby, C. a. & Dixon, J. E. PTEN. *Annual review of biochemistry* **83,** 641–69 (2014).

33. Steck, P. A. *et al.* Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nature genetics* **15,** 356–62 (1997).

34. Li, J *et al.* PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science (New York, N.Y.)* **275,** 1943–7 (1997).

35. Eng, C. PTEN: one gene, many syndromes. *Human mutation* **22,** 183–98 (2003).

36. Lv, J.-W., Cheng, T.-L., Qiu, Z.-L. & Zhou, W.-H. Role of the PTEN signaling pathway in autism spectrum disorder. *Neuroscience bulletin* **29,** 773–8 (2013).

37. Carracedo, A., Alimonti, A. & Pandolfi, P. P. PTEN level in tumor suppression: how much is too little? *Cancer research* **71,** 629–633 (2011).

38. Hopkins, B. D. *et al.* A Secreted PTEN Phosphatase That Enters Cells to Alter Signaling and Survival. *Science* **341,** 399–402 (2013).

39. Lee, J. O., Yang, H, Georgescu, M. M., Di Cristofano, A, Maehama, T, Shi, Y, Dixon, J. E., Pandolfi, P & Pavletich, N. P. Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association. *Cell* **99,** 323–34 (1999).

40. Das, S., Dixon, J. E. & Cho, W. Membrane-binding and activation mechanism of PTEN. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 7491–6 (2003).

41. Masson, G. R., Perisic, O., Burke, J. E. & Williams, R. L. The intrinsically disordered tails of PTEN and PTEN-L have distinct roles in regulating substrate specificity and membrane activity. *Biochemical Journal* **473,** 135–144 (2016).

42. Kalli, A. C. & Sansom, M. S. P. Interactions of peripheral proteins with model membranes as viewed by molecular dynamics simulations. *Biochemical Society Transactions* **42,** 1418–1424 (2014).

43. Kalli, A. C., Morgan, G. & Sansom, M. S. P. Interactions of the auxilin-1 PTEN-like domain with model membranes result in nanoclustering of phosphatidyl inositol phosphates. *Biophysical journal* **105,** 137–45 (2013).

44. Shenoy, S. S., Nanda, H. & Lösche, M. Membrane association of the PTEN tumor suppressor: electrostatic interaction with phosphatidylserine-containing bilayers and regulatory role of the C-terminal tail. *Journal of structural biology* **180,** 394–408 (2012).

45. Shenoy, S., Shekhar, P., Heinrich, F., Daou, M.-C., Gericke, A., Ross, A. H. & 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).

46. Nguyen, H. N., Afkari, Y, Senoo, H, Sesaki, H, Devreotes, P. N. & Iijima, M. Mechanism of human PTEN localization revealed by heterologous expression in Dictyostelium. *Oncogene* **33,** 5688–96 (2014).

47. Papa, A. *et al.* Cancer-associated PTEN mutants act in a dominant-negative manner to suppress PTEN protein function. *Cell* **157,** 595–610 (2014).

48. Heinrich, F., Chakravarthy, S., Nanda, H., Papa, A., Pandolfi, P. P., Ross, A. H., Harishchandra, R. K., Gericke, A. & Lösche, M. The PTEN tumor suppressor forms homodimers in solution. *Structure* **23,** 1952–1957 (2015).

49. Wei, Y., Stec, B., Redfield, A. G., Weerapana, E. & Roberts, M. F. Phospholipid binding sites of PTEN: Exploring the mechanism of PIP2 activation. *The Journal of biological chemistry* **290,** 1592–606 (2014).

50. Bolduc, D., Rahdar, M., Tu-Sekine, B., Sivakumaren, S. C., Raben, D., Amzel, L. M., Devreotes, P., Gabelli, S. B. & Cole, P. Phosphorylation-mediated PTEN conformational closure and deactivation revealed with protein semisynthesis. *eLife* **2,** e00691 (2013).

51. Odriozola, L., Singh, G., Hoang, T. & Chan, A. M. Regulation of PTEN activity by its carboxyl-terminal autoinhibitory domain. *The Journal of biological chemistry* **282,** 23306–15 (2007).

52. Rahdar, M., Inoue, T., Meyer, T., Zhang, J., Vazquez, F. & Devreotes, P. N. A phosphorylation-dependent intramolecular interaction regulates the membrane association and activity of the tumor suppressor PTEN. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 480–5 (2009).

53. Scott, J. L., Musselman, C. A., Adu-Gyamfi, E., Kutateladze, T. G. & Stahelin, R. V. Emerging methodologies to investigate lipid-protein interactions. *Integrative biology : quantitative biosciences from nano to macro* **4,** 247–58 (2012).

54. Harishchandra, R. K., Neumann, B. M., Gericke, A. & Ross, A. H. Biophysical methods for the characterization of PTEN/lipid bilayer interactions. *Methods (San Diego, Calif.)* **77-78,** 125–35 (2015).

55. Schmidt-Krey, I. & Rubinstein, J. L. Electron cryomicroscopy of membrane proteins: Specimen preparation for two-dimensional crystals and single particles. *Micron* **42,** 107–116 (2011).

56. Cho, W., Bittova, L. & Stahelin, R. V. Membrane Binding Assays for Peripheral Proteins. *Analytical Biochemistry* **296,** 153–161 (2001).

57. Dowler, S., Kular, G. & Alessi, D. R. Protein Lipid Overlay Assay. *Science Signaling* **2002,** pl6–pl6 (2002).

58. Falconer, R. J. & Collins, B. M. Survey of the year 2009: applications of isothermal titration calorimetry. *Journal of Molecular Recognition* **24,** 1–16 (2011).

59. Mozsolits, H. & Aguilar, M.-I. Surface plasmon resonance spectroscopy: An emerging tool for the study of peptide-membrane interactions. *Biopolymers* **66,** 3–18 (2002).

60. Ingólfsson, H. I., Lopez, C. A., Uusitalo, J. J., de Jong, D. H., Gopal, S. M., Periole, X. & Marrink, S. J. The power of coarse graining in biomolecular simulations. *Wiley interdisciplinary reviews. Computational molecular science* **4,** 225–248 (2014).

61. Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P. & Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of Chemical Theory and Computation* **4,** 819–834 (2008).

62. Kalli, A. C., Devaney, I. & Sansom, M. S. P. Interactions of phosphatase and tensin homologue (PTEN) proteins with phosphatidylinositol phosphates: insights from molecular dynamics simulations of PTEN and voltage sensitive phosphatase. *Biochemistry* **53,** 1724–32 (2014).

63. Lumb, C. N. & Sansom, M. S. P. Defining the membrane-associated state of the PTEN tumor suppressor protein. *Biophysical journal* **104,** 613–21 (2013).

64. Han, D. S., Golebiewska, U., Stolzenberg, S., Scarlata, S. F. & Weinstein, H. A dynamic model of membrane-bound phospholipase Cβ2 activation by G$\beta\gamma$ subunits. *Molecular pharmacology* **80,** 434–45 (2011).

65. Lai, C.-L., Srivastava, A., Pilling, C., Chase, A. R., Falke, J. J. & Voth, G. A. Molecular mechanism of membrane binding of the GRP1 PH domain. *Journal of molecular biology* **425,** 3073–90 (2013).

66. Lumb, C. N., He, J., Xue, Y., Stansfeld, P. J., Stahelin, R. V., Kutateladze, T. G. & Sansom, M. S. P. Biophysical and computational studies of membrane penetration by the GRP1 pleckstrin homology domain. *Structure (London, England : 1993)* **19,** 1338–46 (2011).

67. Jaud, S., Tobias, D. J., Falke, J. J. & White, S. H. Self-induced docking site of a deeply embedded peripheral membrane protein. *Biophysical journal* **92,** 517–24 (2007).

68. Lai, C.-L., Landgraf, K. E., Voth, G. A. & Falke, J. J. Membrane docking geometry and target lipid stoichiometry of membrane-bound PKCα C2 domain: a combined molecular dynamics and experimental study. *Journal of molecular biology* **402,** 301–10 (2010).

69. Manna, D., Bhardwaj, N., Vora, M. S., Stahelin, R. V., Lu, H. & Cho, W. Differential roles of phosphatidylserine, PtdIns(4,5)P2, and PtdIns(3,4,5)P3 in plasma membrane targeting of C2 domains. Molecular dynamics simulation, membrane binding, and cell translocation studies of the PKCalpha C2 domain. *The Journal of biological chemistry* **283,** 26047–58 (2008).

70. Chodera, J. D., Mobley, D. L., Shirts, M. R., Dixon, R. W., Branson, K. & Pande, V. S. *Alchemical free energy methods for drug discovery: Progress and challenges* 2011. arXiv: `NIHMS150003`.

71. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1,** 826–843 (2011).

72. Comer, J., Gumbart, J. C., Hénin, J., Lelièvre, T., Pohorille, A. & Chipot, C. The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask. *The Journal of Physical Chemistry B* **119,** 1129–1151 (2015).

73. Roux, B. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications* **91,** 275–282 (1995).

74. Leach, A. R. *Molecular modelling : principles and applications* 744 (Prentice Hall, 2001).

75. Brooks, B. R. *et al.* CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **30,** 1545–1614 (2009).

76. Scott, W. R. P. *et al.* The GROMOS Biomolecular Simulation Program Package. *The Journal of Physical Chemistry A* **103,** 3596–3607 (1999).

77. Berger, O., Edholm, O. & Jähnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophysical Journal* **72,** 2002–2013 (1997).

78. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *The journal of physical chemistry. B* **111,** 7812–24 (2007).

79. Ouyang, J. F. & Bettens, R. P. A. Modelling Water: A Lifetime Enigma. *CHIMIA* **69,** 104–111 (2015).

80. López, C. A., Sovova, Z., van Eerden, F. J., de Vries, A. H. & Marrink, S. J. Martini Force Field Parameters for Glycolipids. *Journal of Chemical Theory and Computation* **9,** 1694–1708 (2013).

81. Yesylevskyy, S. O., Schäfer, L. V., Sengupta, D. & Marrink, S. J. Polarizable water model for the coarse-grained MARTINI force field. *PLoS computational biology* **6,** e1000810 (2010).

82. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **4,** 435–447 (2008).

83. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26,** 1781–1802 (2005).

84. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G.E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18,** 1463–1472 (1997).

85. Van der Spoel, D., Lindahl, E., Hess, B. & Team. *GROMACS User Manual version 4.6.5* 2013.

86. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98,** 10089 (1993).

87. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. en. *The Journal of Chemical Physics* **81,** 3684 (1984).

88. Parrinello, M. Polymorphic transitions in single crystals: A new molecular dynamics method. en. *Journal of Applied Physics* **52,** 7182 (1981).

89. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **13,** 1011–1021 (1992).

90. Grossfield, A. WHAM: the weighted histogram analysis method, 2.0.9. 1–14 (2016).

91. Hub, J. S., de Groot, B. L. & van der Spoel, D. g_whamA Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *Journal of Chemical Theory and Computation* **6,** 3713–3720 (2010).

92. Doudou, S., Burton, N. A. & Henchman, R. H. Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. *Journal of Chemical Theory and Computation* **5,** 909–918 (2009).

93. Hedger, G., Shorthouse, D., Koldsø, H. & Sansom, M. S. P. Free Energy Landscape of Lipid Interactions with Regulatory Binding Sites on the Transmembrane Domain of the EGF Receptor. *The Journal of Physical Chemistry B* **120,** 8154–8163 (2016).

94. Sengupta, D. & Marrink, S. J. Lipid-mediated interactions tune the association of glycophorin A helix and its disruptive mutants in membranes. en. *Physical chemistry chemical physics : PCCP* **12,** 12987–96 (2010).

95. Aci-Sèche, S., Sawma, P., Hubert, P., Sturgis, J. N., Bagnard, D., Jacob, L., Genest, M. & Garnier, N. Transmembrane recognition of the semaphorin coreceptors neuropilin 1 and plexin A1: coarse-grained simulations. *PloS one* **9,** e97779 (2014).

96. Chng, C.-P. & Tan, S.-M. Leukocyte integrin $\alpha L\beta 2$ transmembrane association dynamics revealed by coarse-grained molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics* **79,** 2203–2213 (2011).

97. Hedger, G., Rouse, S. L., Domański, J., Chavent, M., Koldsø, H. & Sansom, M. S. P. Lipid-Loving ANTs: Molecular Simulations of Cardiolipin Interactions and the Organization of the Adenine Nucleotide Translocase in Model Mitochondrial Membranes. *Biochemistry* **55,** 6238–6249 (2016).

98. Knight, J. D. & Falke, J. J. Single-molecule fluorescence studies of a PH domain: new insights into the membrane docking reaction. *Biophysical journal* **96,** 566–82 (2009).

99. Clodi, M, Vollenweider, P, Klarlund, J, Nakashima, N, Martin, S, Czech, M. P. & Olefsky, J. M. Effects of general receptor for phosphoinositides 1 on insulin and insulin-like growth factor I-induced cytoskeletal rearrangement, glucose transporter-4 translocation, and deoxyribonucleic acid synthesis. *Endocrinology* **139,** 4984–90 (1998).

100. Ferguson, K. M., Kavran, J. M., Sankaran, V. G., Fournier, E., Isakoff, S. J., Skolnik, E. Y. & Lemmon, M. A. Structural basis for discrimination of 3-phosphoinositides by pleckstrin homology domains. *Molecular cell* **6,** 373–84 (2000).

101. Chen, H.-C., Ziemba, B. P., Landgraf, K. E., Corbin, J. A. & 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).

102. He, J., Haney, R. M., Vora, M., Verkhusha, V. V., Stahelin, R. V. & Kutateladze, T. G. Molecular mechanism of membrane targeting by the GRP1 PH domain. *Journal of lipid research* **49,** 1807–15 (2008).

103. Lumb, C. N. & 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).

104. Yamamoto, E., Kalli, A. C., Yasuoka, K. & Sansom, M. S. P. Interactions of Pleckstrin Homology Domains with Membranes: Adding Back the Bilayer via High-Throughput Molecular Dynamics. *Structure* **24,** 1421–1431 (2016).

105. Humphrey, W, Dalke, A & Schulten, K. VMD: visual molecular dynamics. *Journal of molecular graphics* **14,** 33–8, 27–8 (1996).

106. Kavran, J. M., Klein, D. E., Lee, A., Falasca, M., Isakoff, S. J., Skolnik, E. Y. & Lemmon, M. A. Specificity and Promiscuity in Phosphoinositide Binding by Pleckstrin Homology Domains. *Journal of Biological Chemistry* **273,** 30497–30508 (1998).

107. Dowler, S., Currie, R. A., Campbell, D. G., Deak, M., Kular, G., Downes, C. P. & Alessi, D. R. Identification of pleckstrin-homology-domain-containing proteins with novel phosphoinositide-binding specificities. *Biochemical Journal* **351,** 19 (2000).

108. Park, W. S., Heo, W. D., Whalen, J. H., O'Rourke, N. A., Bryan, H. M., Meyer, T. & Teruel, M. N. Comprehensive identification of PIP3-regulated PH domains from C. elegans to H. sapiens by model prediction and live imaging. *Molecular cell* **30,** 381–92 (2008).

109. Klarlund, J. K. Signaling by Phosphoinositide-3,4,5-Trisphosphate Through Proteins Containing Pleckstrin and Sec7 Homology Domains. *Science* **275,** 1927–1930 (1997).

110. Klarlund, J. K., Tsiaras, W, Holik, J. J., Chawla, A & Czech, M. P. Distinct polyphosphoinositide binding selectivities for pleckstrin homology domains of GRP1-like proteins based on diglycine versus triglycine motifs. *The Journal of biological chemistry* **275,** 32816–21 (2000).

111. Cronin, T. C., DiNitto, J. P., Czech, M. P. & Lambright, D. G. Structural determinants of phosphoinositide selectivity in splice variants of Grp1 family PH domains. *The EMBO journal* **23,** 3711–20 (2004).

112. DiNitto, J. P., Lee, M.-T., Malaby, A. W. & Lambright, D. G. Specificity and membrane partitioning of Grsp1 signaling complexes with Grp1 family Arf exchange factors. *Biochemistry* **49,** 6083–92 (2010).

113. Manna, D., Albanese, A., Park, W. S. & Cho, W. Mechanistic basis of differential cellular responses of phosphatidylinositol 3,4-bisphosphate- and phosphatidylinositol 3,4,5-trisphosphate-binding pleckstrin homology domains. *The Journal of biological chemistry* **282,** 32093–105 (2007).

114. Corbin, J. A., Dirkx, R. A. & Falke, J. J. GRP1 pleckstrin homology domain: activation parameters and novel search mechanism for rare target lipid. *Biochemistry* **43,** 16161–73 (2004).

115. Pilling, C., Landgraf, K. E. & 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,** 9845–56 (2011).

116. Eastwood, M. P. *et al.* Evaluating protein structure-prediction schemes using energy landscape theory. *IBM Journal of Research and Development* **45,** 475–497 (2001).

117. Chen, P.-C. & Kuyucak, S. Accurate determination of the binding free energy for KcsA-charybdotoxin complex from the potential of mean force calculations with restraints. *Biophysical journal* **100,** 2466–74 (2011).

118. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **32,** 2319–2327 (2011).

119. Gowers, R. J. *et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations* in (2016), 98–105.

120. Dotson, D., Gowers, R., Seyler, S., Linke, M. & Oliver, B. *MDSynthesis* 2016.

121. Hyvönen, M, Macias, M. J., Nilges, M, Oschkinat, H, Saraste, M & Wilmanns, M. Structure of the binding site for inositol phosphates in a PH domain. *The EMBO journal* **14,** 4676–85 (1995).

122. Ceccarelli, D. F. J., Blasutig, I. M., Goudreault, M., Li, Z., Ruston, J., Pawson, T. & Sicheri, F. Non-canonical interaction of phosphoinositides with pleckstrin homology domains of Tiam1 and ArhGAP9. *The Journal of biological chemistry* **282,** 13864–74 (2007).

123. Jian, X. *et al.* Molecular Basis for Cooperative Binding of Anionic Phospholipids to the PH Domain of the Arf GAP ASAP1. *Structure* **23,** 1977–1988 (2015).

124. Van Paridon, P. A., de Kruijff, B, Ouwerkerk, R & Wirtz, K. W. Polyphosphoinositides undergo charge neutralization in the physiological pH range: a 31P-NMR study. *Biochimica et biophysica acta* **877,** 216–9 (1986).

125. Kooijman, E. E., King, K. E., Gangoda, M. & Gericke, A. Ionization Properties of Phosphatidylinositol Polyphosphates in Mixed Model Membranes. *Biochemistry* **48,** 9360–9371 (2009).

126. Singh, G. & Tieleman, D. P. Using the Wimley-White Hydrophobicity Scale as a Direct Quantitative Test of Force Fields: The MARTINI Coarse-Grained Model. *Journal of chemical theory and computation* **7,** 2316–24 (2011).

127. Chan, K. C., Lu, L., Sun, F. & Fan, J. Molecular Details of the PH Domain of ACAP1BAR-PHProtein Binding to PIP-Containing Membrane. *Journal of Physical Chemistry B* **121,** 3586–3596 (2017).

128. Thomas, C. C., Deak, M., Alessi, D. R. & 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. *Current biology : CB* **12,** 1256–62 (2002).

129. Prashek, J., Truong, T. & Yao, X. Crystal structure of the pleckstrin homology domain from the ceramide transfer protein: Implications for conformational change upon ligand binding. *PLoS ONE* **8** (ed Saad, J.) e79590 (2013).

130. Buyan, A., Kalli, A. C. & Sansom, M. S. Multiscale Simulations Suggest a Mechanism for the Association of the Dok7 PH Domain with PIP-Containing Membranes. *PLoS Computational Biology* **12** (ed Briggs, J. M.) e1005028 (2016).

131. Salim, K, Bottomley, M. & Querfurth, E. Distinct specificity in the recognition of phosphoinositides by the pleckstrin homology domains of dynamin and Bruton's tyrosine kinase. *The EMBO journal* **15,** 6241–6250 (1996).

132. Zheng, J., Cahill, S. M., Lemmon, M. A., Fushman, D., Schlessinger, J. & Cowburn, D. Identification of the Binding Site for Acidic Phospholipids on the PH Domain of Dynamin: Implications for Stimulation of GTPase Activity. *Journal of Molecular Biology* **255,** 14–21 (1996).

133. Maroun, C. R., Moscatello, D. K., Naujokas, M. A., Wong, A. J., Holgado-madruga, M. & Park, M. CELL BIOLOGY AND METABOLISM : A Conserved Inositol Phospholipid Binding Site within the Pleckstrin Homology Domain of the Gab1 Docking Protein Is Required for Epithelial Morphogenesis A Conserved Inositol Phospholipid Binding Site within the Pleckstrin Ho. **274,** 31719–31726 (1999).

134. Gorai, S., Paul, D., Haloi, N., Borah, R., Santra, M. K. & Manna, D. Mechanistic insights into the phosphatidylinositol binding properties of the pleckstrin homology domain of lamellipodin. *Molecular bioSystems* **12,** 747–57 (2016).

135. Harlan, J. E., Hajduk, P. J., Yoon, H. S. & Fesik, S. W. Pleckstrin homology domains bind to phosphatidylinositol-4,5-bisphosphate. *Nature* **371,** 168–70 (1994).

136. Gallego, O. *et al.* A systematic screen for proteinlipid interactions in Saccharomyces cerevisiae. *Molecular Systems Biology* **6,** 430 (2010).

137. Liu, J., Fukuda, K., Xu, Z., Ma, Y. Q., Hirbawi, J., Mao, X., Wu, C., Plow, E. F. & Qin, J. Structural basis of phosphoinositide binding to kindlin-2 protein pleckstrin homology domain in regulating integrin activation. *Journal of Biological Chemistry* **286,** 43334–43342 (2011).

138. Frech, M., Andjelkovic, M., Ingley, E., Reddy, K. K., Falck, J. R. & Hemmings, B. A. High Affinity Binding of Inositol Phosphates and Phosphoinositides to the Pleckstrin Homology Domain of RAC/Protein Kinase B and Their Influence on Kinase Activity. *Journal of Biological Chemistry* **272,** 8474–8481 (1997).

139. Landgraf, K. E., Pilling, C. & 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,** 12260–9 (2008).

140. Uekama, N., Sugita, T., Okada, M., Yagisawa, H. & Tuzi, S. Phosphatidylserine induces functional and structural alterations of the membrane-associated pleckstrin homology domain of phospholipaseC-δ1. *FEBS Journal* **274,** 177–187 (2007).

141. Yates, L. A. *et al.* Structural and functional characterization of the kindlin-1 pleckstrin homology domain. *The Journal of biological chemistry* **287,** 43246–61 (2012).

142. Flesch, F. M., Yu, J. W., Lemmon, M. A. & Burger, K. N. J. Membrane activity of the phospholipase C-delta1 pleckstrin homology (PH) domain. *The Biochemical journal* **389,** 435–41 (2005).

143. Garcia, P., Gupta, R., Shah, S., Morris, A. J., Rudge, S. A., Scarlata, S., Petrova, V., McLaughlin, S. & Rebecchi, M. J. The pleckstrin homology domain of phospholipase C-.delta.1 binds with high affinity to phosphatidylinositol 4,5-bisphosphate in bilayer membranes. *Biochemistry* **34,** 16228–16234 (1995).

144. Currie, R. A., Walker, K. S., Gray, A., Deak, M., Casamayor, A., Downes, C. P., Cohen, P., Alessi, D. R. & Lucocq, J. Role of phosphatidylinositol 3,4,5-trisphosphate in regulating the activity and localization of 3-phosphoinositide-dependent protein kinase-1. *Biochemical Journal* **337,** 575–583 (1999).

145. Stephens, L *et al.* Protein kinase B kinases that mediate phosphatidylinositol 3,4,5-trisphosphate-dependent activation of protein kinase B. *Science* **279,** 710–714 (1998).

146. Franke, T. F., Kaplan, D. R., Cantley, L. C. & Toker, A. Direct regulation of the Akt proto-oncogene product by phosphatidylinositol-3,4-bisphosphate. *Science (New York, N.Y.)* **275,** 665–8 (1997).

147. Rao, V. R., Corradetti, M. N., Chen, J, Peng, J, Yuan, J, Prestwich, G. D. & Brugge, J. S. Expression cloning of protein targets for 3-phosphorylated phosphoinositides. *The Journal of biological chemistry* **274,** 37893–900 (1999).

148. Dowler, S., Currie, R. A., Downes, C. P. & Alessi, D. R. DAPP1: a dual adaptor for phosphotyrosine and 3-phosphoinositides. *Biochemical Journal* **342,** 7–12 (1999).

149. Rowland, M. M., Gong, D., Bostic, H. E., Lucas, N., Cho, W. & Best, M. D. Microarray analysis of Akt PH domain binding employing synthetic biotinylated analogs of all seven phosphoinositide headgroup isomers. *Chemistry and physics of lipids* **165,** 207–15 (2012).

150. Macia, E., Paris, S. & Chabre, M. Binding of the PH and polybasic C-terminal domains of ARNO to phosphoinositides and to acidic lipids. *Biochemistry* **39,** 5893–5901 (2000).

151. Venkateswarlu, K., Oatey, P. B., Tavaré, J. M. & Cullen, P. J. Insulin-dependent translocation of ARNO to the plasma membrane of adipocytes requires phosphatidylinositol 3-kinase. *Current Biology* **8,** 463–466 (1998).

152. Luo, R., Miller Jenkins, L. M., Randazzo, P. A. & Gruschus, J. Dynamic interaction between Arf GAP and PH domains of ASAP1 in the regulation of GAP activity. *Cellular Signalling* **20,** 1968–1977 (2008).

153. Kruljac-Letunic, A., Moelleken, J., Kallin, A., Wieland, F. & Blaukat, A. The tyrosine kinase Pyk2 regulates Arf1 activity by phosphorylation and inhibition of the Arf-GTPase-activating protein ASAP1. *The Journal of biological chemistry* **278,** 29560–70 (2003).

154. Baraldi, E., Carugo, K. D., Hyvönen, M., Surdo, P. L., Riley, A. M., Potter, B. V., O'Brien, R., Ladbury, J. E. & Saraste, M. Structure of the PH domain from Bruton's tyrosine kinase in complex with inositol 1,3,4,5-tetrakisphosphate. *Structure* **7,** 449–460 (1999).

155. Rameh, L. E. *et al.* A Comparative Analysis of the Phosphoinositide Binding Specificity of Pleckstrin Homology Domains. *Journal of Biological Chemistry* **272,** 22059–22066 (1997).

156. Fukuda, M., Kojima, T., Kabayama, H. & Mikoshiba, K. Mutation of the Pleckstrin Homology Domain of Bruton's Tyrosine Kinase in Immunodeficiency Impaired Inositol 1,3,4,5-Tetrakisphosphate Binding Capacity. *Journal of Biological Chemistry* **271,** 30303–30306 (1996).

157. Das, A., Base, C., Manna, D., Cho, W. & Dubreuil, R. R. Unexpected complexity in the mechanisms that target assembly of the spectrin cytoskeleton. *The Journal of biological chemistry* **283,** 12643–53 (2008).

158. Gong, D., Smith, M. D., Manna, D., Bostic, H. E., Cho, W. & Best, M. D. Microplate-based characterization of protein-phosphoinositide binding interactions using a synthetic biotinylated headgroup analogue. *Bioconjugate chemistry* **20,** 310–6 (2009).

159. Thomas, C. C., Dowler, S, Deak, M, Alessi, D. R. & van Aalten, D. M. 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. *The Biochemical journal* **358,** 287–94 (2001).

160. Legate, K. R., Takahashi, S., Bonakdar, N., Fabry, B., Boettiger, D., Zent, R. & Fässler, R. Integrin adhesion and force coupling are independently regulated by localized PtdIns(4,5)2 synthesis. *The EMBO journal* **30,** 4539–53 (2011).

161. Qu, H. *et al.* Kindlin-2 regulates podocyte adhesion and fibronectin matrix deposition through interactions with phosphoinositides and integrins. *Journal of cell science* **124,** 879–91 (2011).

162. Lucas, N. & Cho, W. Phosphatidylserine binding is essential for plasma membrane recruitment and signaling function of 3-phosphoinositide-dependent kinase-1. *The Journal of biological chemistry* **286,** 41265–72 (2011).

163. Komander, D., Fairservice, A., Deak, M., Kular, G. S., Prescott, A. R., Peter Downes, C, Safrany, S. T., Alessi, D. R. & van Aalten, D. M. F. Structural insights into the regulation of PDK1 by phosphoinositides and inositol phosphates. *The EMBO journal* **23,** 3918–28 (2004).

164. Hirose, K., Kadowaki, S., Tanabe, M., Takeshima, H. & Iino, M. Spatiotemporal Dynamics of Inositol 1,4,5-Trisphosphate That Underlies Complex Ca2+ Mobilization Patterns. *Science* **284,** 1527–1530 (1999).

165. Ferguson, K. M., Lemmon, M. A., Schlessinger, J. & Sigler, P. B. Structure of the high affinity complex of inositol trisphosphate with a phospholipase C pleckstrin homology domain. *Cell* **83,** 1037–1046 (1995).

166. Lemmon, M. A., Ferguson, K. M., O'Brien, R., Sigler, P. B. & Schlessinger, J. Specific and high-affinity binding of inositol phosphates to an isolated pleckstrin homology domain. *Proceedings of the National Academy of Sciences* **92,** 10472–10476 (1995).

167. Tall, E. *et al.* Phosphoinositide binding specificity among phospholipase C isozymes as determined by photo-cross-linking to novel substrate and product analogs. *Biochemistry* **36,** 7239–7248 (1997).

168. Várnai, P., Lin, X., Lee, S. B., Tuymetova, G., Bondeva, T., Spät, A., Rhee, S. G., Hajnóczky, G. & Balla, T. Inositol lipid binding and membrane localization of isolated pleckstrin homology (PH) domains. Studies on the PH domains of phospholipase C delta 1 and p130. *The Journal of biological chemistry* **277,** 27412–22 (2002).

169. Lenoir, M., Kufareva, I., Abagyan, R. & Overduin, M. Membrane and Protein Interactions of the Pleckstrin Homology Domain Superfamily. *Membranes* **5,** 646–663 (2015).

170. .

171. Huang, B. X., Akbar, M., Kevala, K. & Kim, H.-Y. Phosphatidylserine is a critical modulator for Akt activation. *The Journal of cell biology* **192,** 979–92 (2011).

172. Vonkova, I. *et al.* Lipid Cooperativity as a General Membrane-Recruitment Principle for PH Domains. *Cell reports* **12,** 1519–30 (2015).

173. Lacin, E., Aryal, P., Glaaser, I. W., Bodhinathan, K., Tsai, E., Marsh, N., Tucker, S. J., Sansom, M. S. P. & Slesinger, P. A. Dynamic role of the tether helix in PIP2-dependent gating of a G protein-gated potassium channel. *The Journal of general physiology* **149,** 799 (2017).

174. Simpson, L & Parsons, R. PTEN: life as a tumor suppressor. *Experimental cell research* **264,** 29–41 (2001).

175. Smith, I. N. & Briggs, J. M. Structural mutation analysis of PTEN and its genotype-phenotype correlations in endometriosis and cancer. *Proteins* **84,** 1625–1643 (2016).

176. Rost, B., Yachdav, G. & Liu, J. The PredictProtein server. *Nucleic acids research* **32,** W321–6 (2004).

177. Cole, C., Barber, J. D. & Barton, G. J. The Jpred 3 secondary structure prediction server. *Nucleic acids research* **36,** W197–201 (2008).

178. McGuffin, L. J., Bryson, K & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics (Oxford, England)* **16,** 404–5 (2000).

179. Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H. & Xu, J. Template-based protein structure modeling using the RaptorX web server. *Nature protocols* **7,** 1511–22 (2012).

180. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10,** 845–858 (2015).

181. Fiser, A. & Sali, A. Modeller: generation and refinement of homology-based protein structure models. *Methods in enzymology* **374,** 461–91 (2003).

182. Stansfeld, P. J. & Sansom, M. S. From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations. *Journal of Chemical Theory and Computation* **7,** 1157–1166 (2011).

183. Baker, N. A., Sept, D, Joseph, S, Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **98,** 10037–41 (2001).

184. Nguyen, H.-N. *et al.* A new class of cancer-associated PTEN mutations defined by membrane translocation defects. *Oncogene* (2014).

185. Wei, Y., Stec, B., Redfield, A. G., Weerapana, E. & Roberts, M. F. Phospholipid-binding sites of phosphatase and tensin homolog (PTEN): exploring the mechanism of phosphatidylinositol 4,5-bisphosphate activation. *The Journal of biological chemistry* **290,** 1592–606 (2015).

186. Zhu, F. & Hummer, G. Convergence and error estimation in free energy calculations using the weighted histogram analysis method. *Journal of computational chemistry* **33,** 453–65 (2012).

187. Domański, J., Hedger, G., Best, R. B., Stansfeld, P. J. & Sansom, M. S. P. Convergence and Sampling in Determining Free Energy Landscapes for Membrane Protein Association. *The Journal of Physical Chemistry B* **121,** 3364–3375 (2017).

188. Van der Kamp, M. W. & Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **52,** 2708–2728 (2013).

189. Ceccato, L., Chicanne, G., Nahoum, V., Pons, V., Payrastre, B., Gaits-Iacovoni, F. & Viaud, J. PLIF: A rapid, accurate method to detect and quantitatively assess protein-lipid interactions. *Science signaling* **9,** rs2 (2016).

190. Whited, A. M. & Park, P. S.-H. Atomic force microscopy: A multifaceted tool to study membrane proteins and their interactions with ligands. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1838,** 56–68 (2014).