Investigating the interaction of the PKB PH domain with inositol phosphate-based compounds

By Sarah Ann Jacqueline Rosen

A thesis submitted to Imperial College London in candidature for the degree of Doctor of Philosophy of Imperial College.

Department of Chemistry
Imperial College London
Exhibition Road
London
SW7 2AZ

July 2013

Declaration of Originality

I, Sarah Ann Jacqueline Rosen, declare that this thesis, and all the work presented in it, is my own and that it has been generated by me as the result of my own original research, unless otherwise stated.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

Protein kinase B (PKB) plays a key role in the phosphoinositide 3-kinase pathway, one of the most frequently activated proliferation pathways in cancer. A key stage in this pathway is PKB's translocation to the plasma membrane, which is driven by direct interaction of PKB's pleckstrin homology (PH) domain with the inositol phosphate head-groups of phosphoinositide lipids PtdIns(3,4,5)P_3 and PtdIns(3,4)P_2.

In this thesis a computational approach has been applied to study the interaction between PKB's PH domain and the inositol phosphate head-groups of phosphoinositide lipids. Herein the first full set of parameters for these inositol phosphates has been generated using a clearly defined algorithmic approach. The parameters have been applied in a total of 6 μs of molecular dynamics (MD) simulations to investigate the interaction between inositol phosphates and the PKB PH domain. The simulations successfully mirror, and additionally rationalise, PKB's experimental interactions and behaviour. As well as investigating the native system, a PKB mutant that has been reported in multiple human cancers has also been explored. This PKB mutant constitutively targets the plasma membrane due to its broadened phosphoinositide selectivity. The atomic-level view available from MD simulations has helped elucidate the molecular mechanism behind this.

Information obtained about the PKB PH domain's binding interface has been used to design inositol phosphate derivatives to inhibit PKB activation. MD simulations have allowed a number of inositol phosphate derivatives to be rapidly screened for their inhibitory behaviour. The predicted behaviour of a select number of these derivatives has been further assessed using biochemical techniques. Gratifyingly, the computational and biochemical results are shown to be in good agreement.

For my family

"Concern for man and his fate must always form the chief interest of all technical endeavours. Never forget this in the midst of your diagrams and equations."

Albert Einstein

Acknowledgements

I'd like to thank Ian Gould for his help, great enthusiasm, and creating a fun and productive work environment. Thanks also go to Piers Gaffney and David Mann for their assistance and advice in the early years of my project. I am grateful to Rudiger Woscholski and Lok Mak for their help in re-introducing me to the world of biochemistry. It has been great to have the opportunity to collaborate with Aine Slowey and Stuart Conway at Oxford University. Thank you for all your hard work, and trips down to London.

A special thank you is also due to all members of room 537. Richard, Lucy and Alex…you were there from the very beginning, and you honestly made my PhD experience. In times when I wanted to smash my computer you offered chocolate, a welcome break and most importantly a lot of laughs. Richard – congratulations, you managed to put up with my tears, my stress, my distractions and my (sometimes silly) questions for 4 years. You not only became a fantastic work-mate who I could bounce my ideas off of, but you also became a great (and very patient) friend who I will miss travelling the world with. Callum, Pietro and Angela – you have also been responsible for some of my favourite moments in 537, and I thank you so much for your friendship, laughter, advice…and chocolate. Additionally I'd like to thank the honorary members of room 537 (Shaggy and Jemima) for making my PhD full of so much fun.

Outside of Imperial College I owe thanks to many others for their continual support, patience and unconditional love. Mum, Dad, Simon, Grandma, Grandpa and Auntie Lilian - you have had to deal with me at my worst, yet you have never been anything other than supportive. You have always made me feel that I can do absolutely anything I put my mind to, and it is your unwavering faith in me and my abilities that has helped me through my PhD. I am fortunate to have so many amazing role models in my life.

Mark – I owe you extra thanks. You don't need to put up with me when an experiment goes wrong, but you always do. You have provided love, support and moments of sheer silliness to see me through my PhD. You have frequently allowed me to put my PhD first, and I can't thank you enough for everything you've put into our relationship. I can't wait to become your wife this year, and look forward to a life of happiness together. I also thank all the Fenton family for their continual support, and understanding my late arrival to Friday night dinners for many years!

Publications and External Presentations

Publications

A Theoretical Investigation of Inositol 1,3,4,5-tetrakisphosphate, S. A. J. Rosen, P. R. J. Gaffney and I. R. Gould, *Phys. Chem. Chem. Phys*., 2011, 13, 1070-1081.

Understanding the Relative Affinity and Specificity of the Pleckstrin Homology Domain of Protein Kinase B for Inositol Phosphates, S. A. J. Rosen, P. R. J. Gaffney, B. Spiess and I. R. Gould, *Phys. Chem. Chem. Phys*., 2012, 14, 929-936.

Oral Presentations

243rd ACS National Meeting, San Diego, USA, 25-29 March 2012: "Understanding the interactions of inositol phosphates with protein kinase B using long time-scale simulations"

MGMS Young Modellers' Forum, London, England, 25 November 2011: "Flipping inositol phosphates: A molecular dynamics approach to understanding the selectivity of protein kinase B"

Poster Presentations

243rd ACS National Meeting, San Diego, USA, 25-29 March 2012: "Flipping inositol phosphates: A molecular dynamics approach to understanding the selectivity of protein kinase B"

Ninth Triennial Congress of the World Association of Theoretical and Computational Chemists, Santiago de Compostela, Spain, 17-22 July 2011: "A tough Akt to follow: Understanding and predicting interactions of protein kinase B"

39th ACS National Meeting, San Francisco, USA, 21-25 March 2010: "A theoretical investigation of inositol 1,3,4,5-tetrakisphosphate"

Table of Contents

Abstract 3

Acknowledgements 6

Publications and External Presentations 7

List of Abbreviations 14

Chapter 1 – Introduction to Biological System 16

1. 1 Phosphoinositides and Inositol Phosphates 16

1. 1. 1 Structure 16

1. 1. 2 Charge State 18

1. 1. 3 Function 18

1. 2 Protein Kinase B 20

1. 2. 1 Structure 20

1. 2. 1. 1 Pleckstrin Homology Domain 21

1. 2. 2 PKBα PH domain-Phosphoinositide Interaction 22

1. 2. 2. 1 Conformational Change 24

1. 2. 3 Crystal Structures 25

1. 2. 4 Function 25

1. 2. 4. 1 Phosphoinositide 3-Kinase Pathway 25

1. 2. 5 PKB and Cancer 28

1. 3 Background to Project/Previous Research 29

1. 4 Overview of Thesis 31

Chapter 2 - Quantum Mechanics 33

2. 1 Introduction to Quantum Mechanics 33

2. 2 Schrödinger Equation 34

2. 2. 1 The Born-Oppenheimer Approximation 34

2. 3 The Hartree Fock SCF Method 35

2. 3. 1 The Self-Consistent Field Approach 36

2. 3. 2 The Variational Principle 36

2. 3. 3 Electron Correlation Energy 37

2. 4 Møller-Plesset Perturbation Theory 37

2. 5 DFT Theory 38

2. 5. 1 The Exchange Correlation Term 39

2. 6 Basis Set 40

Chapter 3 - Molecular Mechanics **42**
3. 1 Introduction to Molecular Mechanics **42**
3. 2 The Force Field **42**
3. 2. 1 Bonded Interactions 44
3. 2. 2 Non-Bonded Interactions 45
3. 2. 2. 1 Van der Waals Interactions 45
3. 2. 2. 2 Electrostatic Interactions 45
3. 2. 2. 3 Non-bonded Cut-offs 47
3. 2. 2. 4 Long-Range Forces 47
3. 2. 3 Atom Types and Parameters 48
3. 3 The Solvent Model **48**
3. 4 Periodic Boundary Conditions **49**
3. 5 Energy Minimisation **50**
3. 5. 1 First Derivative Methods 51
3. 5. 2 Second Derivative Methods 51
3. 6 Molecular Dynamics Simulations **52**
3. 6. 1 Time-step 53
3. 6. 2 MD Protocol 54
3. 6. 3 Length of Simulation 55
3. 7 MM-PBSA **56**
3. 7. 1 Molecular Mechanical Energy 58
3. 7. 2 Solvation Energy 58
3. 7. 2. 1 Polar Solvation Energy 58
3. 7. 2. 2 Non-polar Solvation Energy 59
3. 7. 3 Entropy 59

Chapter 4 - Parameterisation of Ins(1,3,4,5)P_4 ..... **60**
4. 1 Introduction **60**
4. 2 Previous Parameters for Inositol Phosphates **60**
4. 3 Parameterisation Protocol **64**
4. 4 Program Details **65**
4. 5 Initial Structure **66**
4. 6 DFT Gas Phase Investigation **66**
4. 6. 1 Protocol for Charge Derivation 66
4. 6. 1. 1 Atom Type HY 69
4. 6. 2 Validation of Parameters 70
4. 6. 2. 1 MM Energy Minimisation 70

4. 6. 3 Results and Discussion 70
4. 6. 3. 1 Energetic Properties 70
4. 6. 3. 2 Structural Properties 72
4. 7 Comparison of DFT and MP2 Levels of Theory 73
4. 7. 1 Protocol for Geometry Optimisation 73
4. 7. 2 Results and Discussion 73
4. 7. 2. 1 Energetic Properties 73
4. 7. 2. 2 Structural Properties 76
4. 8 Comparison of ff99SB and GAFF Force Fields 76
4. 8. 1 Protocol for Parameter Assignment 76
4. 8. 2 Results and Discussion 77
4. 8. 2. 1 Energetic Properties 77
4. 9 Calculations of Ins(1,3,4,5)P_4 in Implicit Solvent 80
4. 9. 1 Implicit Solvation Models 80
4. 9. 1. 1 The PCM Method 80
4. 9. 1. 2 The Generalised Born Method 81
4. 9. 2 Protocol for Charge Derivation 82
4. 9. 3 Validation of Parameters 82
4. 9. 3. 1 Quantum Mechanical *versus* Molecular Mechanical Properties 82
4. 9. 3. 2 Theory *versus* Experiment 83
4. 9. 4 Results and Discussion 83
4. 9. 4. 1 Energetic Properties 83
4. 9. 4. 2 Structural Properties 85
4. 9. 4. 3 Theory *versus* Experiment 85
4. 10 Conclusion 88
Chapter 5 – MD Simulations of the Native PKB PH Domain with Inositol Phosphates 91
5. 1 Introduction 91
5. 2 The PKB PH Domain-Inositol Phosphate Interaction 91
5. 2. 1 PKB's Selectivity 93
5. 2. 2 Binding Mode of Inositol Phosphates/Phosphoinositides 93
5. 3 The Computational Approach 94
5. 4 Program Details 94
5. 5 Parameterisation of Inositol Tris-phosphates 94
5. 5. 1 Validation of Parameters 98
5. 6 MD Protocol 98
5. 6. 1 Initial Configuration 98

5. 6. 2 Minimisation, Equilibration and Production Run 98
5. 6. 3 MM-PBSA Calculations 99
5. 6. 3. 1 Reasoning for the MM-PBSA Method 100
5. 6. 3. 2 Entropy 100
5. 6. 3. 3 Dielectric Constant 100
5. 6. 3. 4 Weighted Free Energy Calculations 101
5. 6. 4 Hydrogen Bond Interactions 101
5. 6. 5 RMS Deviations and Fluctuations 101
5. 6. 6 Cluster Analysis 102
5. 7 Results and Discussion 103
5. 7. 1 Free Energies of Binding 103
5. 7. 1. 1 PKB PH Domain with Inositol Phosphates 103
5. 7. 1. 2 PKB PH Domain with Flipped Inositol Phosphates 111
5. 7. 1. 3 PKB PH Domain with Phosphoinositide 4, 5-bisphosphate 113
5. 7. 2 Orientation and Molecular Behaviour 115
5. 8 Conclusion 119
Chapter 6 - MD Simulations of the Mutant PKB PH Domain with Inositol Phosphates 121
6.1 Introduction 121
6. 1. 1 Residue 17 121
6. 1. 2 Membrane Localisation of Native PKB 123
6. 1. 3 Membrane Localisation of E17K PKB 123
6. 1. 3. 1 Current Understanding 124
6. 2 Program Details 124
6. 3 MD Protocol 124
6. 3. 1 Initial Configuration 124
6. 3. 1. 1 E17K Crystal Structure 125
6. 3. 2 Minimisation, Equilibration and Production Run 126
6. 3. 3 Free Energy Calculations 126
6. 3. 3. 1 Weighted Free Energy Calculations 128
6. 3. 4 Hydrogen Bond Interactions 128
6. 4 Results and Discussion 128
6. 4. 1 E17K PKB PH Domain with Ins(1,4,5)P_3 128
6. 4. 1. 1 Free Energies of Binding 128
6. 4. 1. 2 Orientation and Molecular Behaviour 129
6. 4. 2 E17K PKB PH Domain with 3-phosphorylated Inositol Phosphates 130
6. 4. 2. 1 E17K PKB PH Domain with Ins(1,3,4,5)P_4 130

6. 4. 2. 2 E17K PKB PH Domain with Ins(1,3,4)P_3 132
6. 5 Conclusion 132
Chapter 7 - Inositol Phosphate Derivatives 134
7. 1 Introduction 134
7. 1. 1 Previous Examples of Inositol Phosphate Derivatives 135
7. 1. 2 A Collaboration with the Conway Group 139
7. 2 Screening 5-Position Ins(1,3,4,5)P_4 Derivatives 141
7. 2. 1 Program Details 143
7. 2. 2 Parameterisation of 5-position Ins(1,3,4,5)P_4 Derivatives 143
7. 2. 3 Minimisation, Equilibration and Production Run 144
7. 2. 4 Free Energy Calculations 145
7. 2. 5 Hydrogen Bond Interactions 145
7. 2. 6 Results and Discussion 146
7. 3 Detailed Computational Investigation of Selected Inositol Phosphate Derivatives 152
7. 3. 1 Parameterisation of Selected Ins(1,3,4,5)P_4 Derivatives 153
7. 3. 1. 1 Sulfamate-based derivatives 153
7. 3. 2 Minimisation, Equilibration and Production Run 154
7. 3. 3 MM-PBSA Calculations 155
7. 3. 4 Results and Discussion 157
7. 3. 4. 1 D-*myo*-Ins(1,3,4)P_3-5-*O*-acetamidoethylphosphate ester 158
7. 3. 4. 2 D-5-*O*-Sulfamoyl-*myo*-Ins(1,3,4)P_3 160
7. 3. 4. 3 D-*myo*-Ins(1,4,5)P_3-3-*O*-methylphosphate ester 161
7. 3. 4. 4 D-3-*O*-Sulfamoyl-myo-Ins(1,4,5)P_3 163
7. 4 *In Vitro* Investigation of Selected Inositol Phosphate Derivatives 165
7. 4. 1 Issues with the Inositol Phosphate Derivatives 165
7. 4. 2 Expression of PKB 166
7. 4. 3 Protein Lipid Overlay (PLO) Assay 167
7. 4. 3. 1 General Method 168
7. 4. 3. 2 Linearity Tests 169
7. 4. 3. 3 Competition-based Assays 171
7. 4. 3. 4 Adaptations to the Competitive PLO Assay 173
7. 4. 4 Isothermal Titration Calorimetry (ITC) 175
7. 4. 4. 1 Method 176
7. 4. 4. 2 Results and Discussion 177
7. 4. 5 Intrinsic Tryptophan Fluorescence Spectroscopy 179
7. 4. 6 Enzyme-Linked ImmunoSorbant Assay (ELISA) 180

7. 4. 6. 1 Linearity Tests....181
7. 4. 6. 2 Optimisation of the ELISA assay....183
7. 4. 6. 3 Competitive ELISA assays with the Ins(1,3,4,5)P_4 Derivatives....184
7. 4. 7 *In Vitro* Phosphorylation Assay....187
7. 4. 7. 1 Method....187
7. 4. 7. 2 Results....187
7. 5 Comparison of *In Vitro* and Computational Results....189
7. 6 Conclusion....190
Chapter 8 – Summary and Future Work....192
8. 1 Inositol Phosphate Parameters....193
8. 2 Parameterisation of Phosphoinositide Lipids....194
8. 3 Full PKB Simulations....195
8. 4 Selectivity of Ins(1,3,4,5)P_4 Derivatives....195
8. 5 Conclusion....196
Chapter 9 – Materials and Methods....197
9. 1 General....197
9. 2 Protein Expression....197
9. 3 SDS PAGE....199
9. 4 Bradford Assay....200
9. 5 Buffer Solutions....200
References....201

Electronic Supplementary Information provided on disc.

List of Abbreviations

ANTH domain	AP180 N-Terminal Homology domain
B3LYP	Becke 3-parameter; Lee, Yang, Parr
CDK inhibitor	Cyclin-Dependent Kinase inhibitor
DAG	Diacyl Glycerol
DFT	Density Functional Theory
ENTH Domain	Epsin N-terminal Homology Domain
FEP	Free Energy Perturbation
FYVE Domain	Fab1, YOTB/ZK632.12, Vac1, and EEA1 domain
GB	Generalised Born
GST	Glutathione S-Transferase
HF	Hartree Fock
Ins(1,3,4,5)P_4	Inositol 1,3,4,5-tetrakisphosphate
Ins(1,3,4)P_3	Inositol 1,3,4-trisphosphate
Ins(1,4,5)P_3	Inositol 1,4,5-trisphosphate
MD	Molecular Dynamics
MEP	Molecular Electrostatic Potential
MM	Molecular Mechanics
MM-PBSA	Molecular Mechanics-Poisson Boltzmann Surface Area
MP	Møller Plesset
PDB	Protein Data Bank
PDK1	Phosphoinositide-Dependent Kinase
PH domain	Pleckstrin Homology domain
PKB	Protein Kinase B
PME	Particle Mesh Ewald
PtdIns(3,4,5)P_3	Phosphatidylinositol 3,4,5-trisphosphate
PtdIns(3,4)P_2	Phosphatidylinositol 3,4-bisphosphate
PtdIns(4,5)P_2	Phosphatidylinositol 4,5-bisphosphate
PtdInsP_ns	Phosphatidylinositol phosphate
PTEN	Phosphatase and Tensin Homolog
QM	Quantum Mechanics
RESP	Restrained Electrostatic Potential

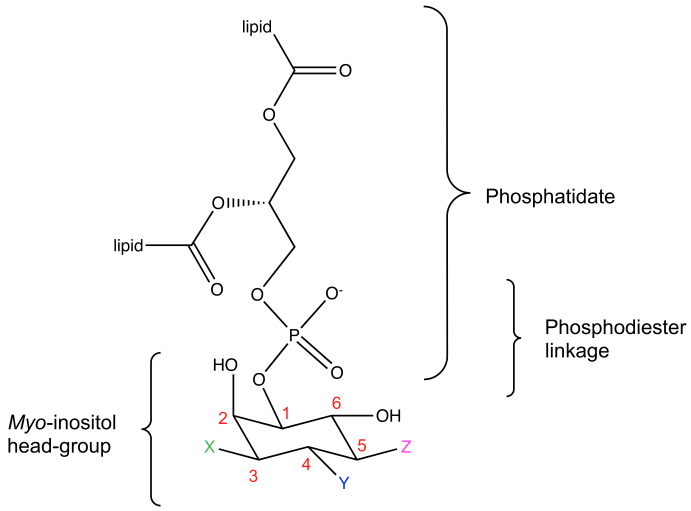
RMS	Root Mean Square
RMSD	Root Mean Square Deviation
SCF	Self Consistent Field
STO	Slater Type Orbital
TI	Thermodynamic Integration
vdW	Van der Waals

Chapter 1 – Introduction to Biological System

1. 1 Phosphoinositides and Inositol Phosphates

1. 1. 1 Structure

Phosphoinositides ($PtdInsP_n$s) are anionic glycerophospholipids present on the cytosolic face of eukaryotic cell membranes. They consist of a D-*myo*-inositol phosphate head-group covalently linked to a phosphatidate tail *via* a phosphodiester linkage at the 1-position of the inositol ring, Figure 1.



Substitution	Name
X=Y=Z=OH	PtdIns
X=$OPO(OH)_2$,Y=Z=OH	PtdIns(3)P
Y=$OPO(OH)_2$, X=Z=OH	PtdIns(4)P
Z=$OPO(OH)_2$, X=Y=OH	PtdIns(5)P
X=Y=$OPO(OH)_2$, Z=OH	PtdIns(3,4)P_2
X=Z=$OPO(OH)_2$, Y=OH	PtdIns(3,5)P_2
Y=Z=$OPO(OH)_2$, X=OH	PtdIns(4,5)P_2
X=Y=Z=$OPO(OH)_2$	PtdIns(3,4,5)P_3

Figure 1: The seven naturally occurring phosphoinositides. The D-*myo*-inositol head-group is covalently linked to a hydrophobic phosphatidate tail. The phosphatidate tail consists of glycerol bound to two long lipophilic fatty acid esters (referred to as diacyl glycerol) plus the phosphodiester. All seven naturally occurring phosphoinositides are illustrated, plus phosphatidylinositol (PtdIns).

The inositol head-group is responsible for the main properties and interactions of phosphoinositide lipids. The hydroxyl groups on the inositol ring can be phosphorylated at the 3-, 4- and 5-positions in different combinations to form a family of seven phosphoinositides, illustrated in Figure 1.[1] The levels of these phosphoinositides in the cell are regulated by a variety of kinases and phosphatases, which reversibly phosphorylate and dephosphorylate the hydroxyl groups on the inositol ring. Dysregulation of these enzymes is associated with a large number of human diseases, highlighting the importance of phosphoinositides in a variety of cellular functions, Figure 2.[2]

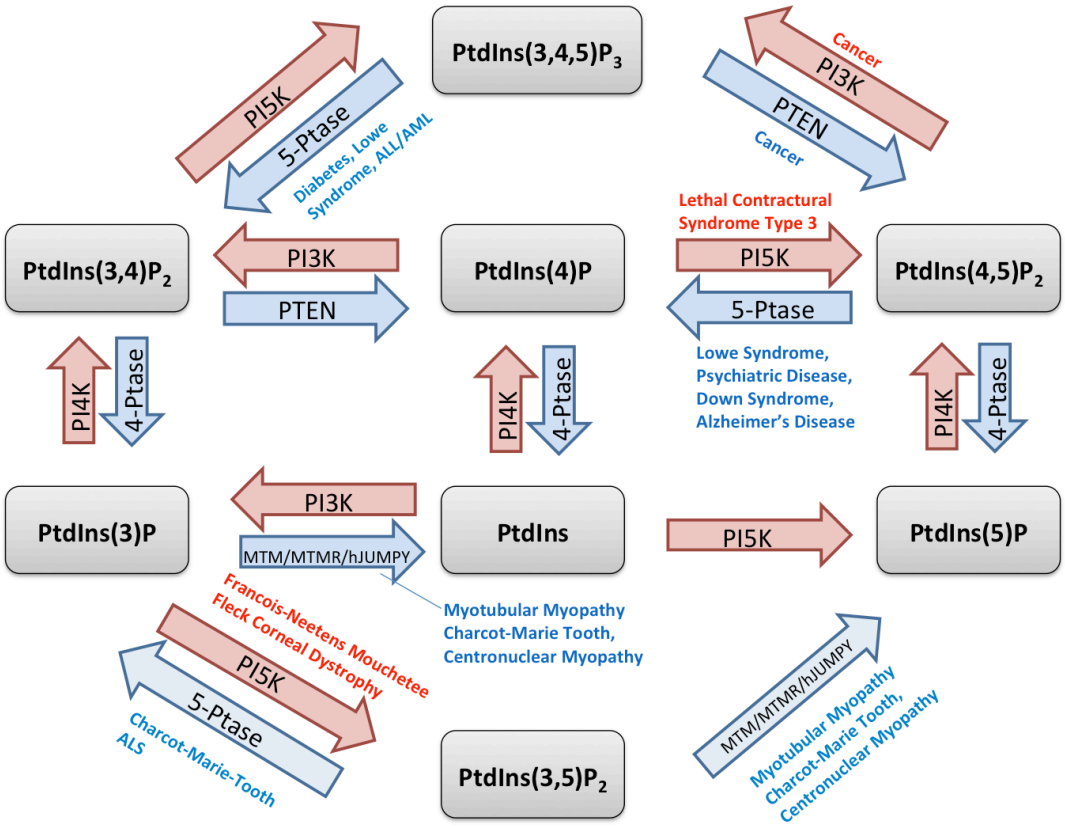


Figure 2: The inter-conversion of the naturally occurring phosphoinositides. Diagram adapted from McCrea and De Camilli.[2] Diseases associated with the kinases and phosphatases regulating the phosphoinositides are illustrated. Kinases and phosphatases are displayed with red and blue arrows respectively. PI3K, PI4K and PI5K represent phosphoinositide 3-kinases, 4-kinases and 5-kinases respectively. 4-Ptase and 5-Ptase represent phosphoinositide 4-phosphatases and 5-phosphatases respectively. MTM and MTMR refer to myotubularin and myotubularin related proteins. hJUMPY is a 3-phosphatase. AML and ALL represent acute myeloid leukemia and acute lymphoblastic leukemia respectively. ALS represents amyotrophic lateral sclerosis.

Whilst inositol phosphates form the hydrophilic head-groups of phosphoinositide lipids, they also exist as soluble ligands found in the cytosol of the cell. These are also interconverted by families of kinases and phosphatases.

1. 1. 2 Charge State

At physiological pH phosphomonoesters are negatively charged and adopt either a fully deprotonated or singly deprotonated state, as shown for methyl phosphate in Figure 3. However, which of these protonation states the phosphomonoester groups of phosphoinositides and inositol phosphates adopt is difficult to assign. This is due to the extremely complex ionisation behaviour of these molecules, discussed in detail in Chapters 4 and 5.[3-5]

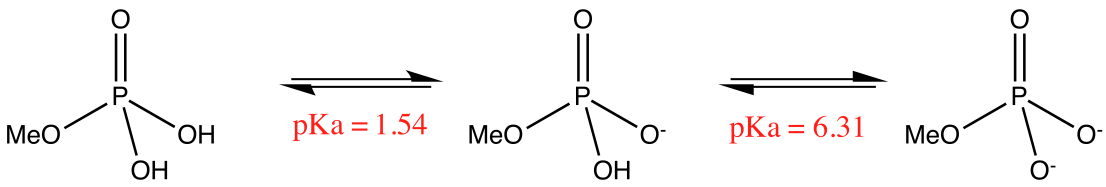


Figure 3: The pK_a values of the first and second deprotonations of methyl phosphate.[6] The pK_a value for the second deprotonation is close to physiological pH.

1. 1. 3 Function

Phosphoinositide lipids comprise approximately 1% of phospholipids in the plasma membrane.[7] Despite being only a minor component of the membrane, they are involved in a number of critical cell processes. They provide docking sites for a variety of proteins that play important roles in signal transduction, cytoskeletal dynamics and membrane trafficking.[8, 9]

Proteins are recruited to the plasma membrane *via* their phosphoinositide-binding domains of which there are several, including FYVE, ENTH, ANTH, PX and PH

domains.[7, 10] These domains have a characteristic cationic pocket lined with basic amino acids that interact with the anionic inositol phosphate head-groups of phosphoinositide lipids, Figure 4.

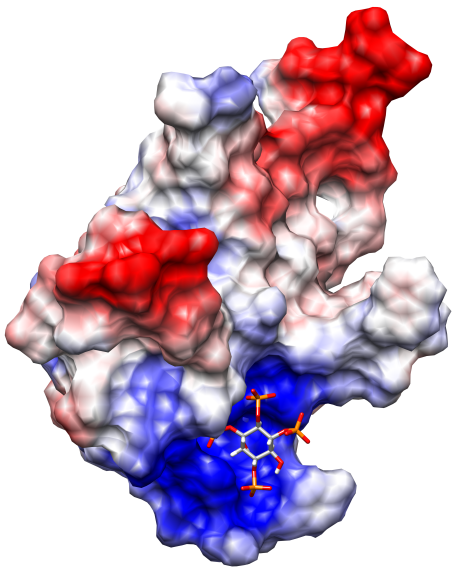


Figure 4: The cationic pocket of the PKB PH domain interacting with the inositol phosphate head-group of PtdIns(3,4,5)P_3. The electrostatic surface potential of the PKB PH domain is illustrated, with blue areas (+10 kcal mol^{-1} e^{-1}) representing positively charged residues and red areas (-10 kcal mol^{-1} e^{-1}) representing negatively charged residues. PDB accession code 1UNQ.[11]

Phosphatidylinositol 4,5-bisphosphate [PtdIns(4,5)P_2] is the most abundant polyphosphorylated phosphoinositide in mammalian cells. As well as acting as a docking site for intracellular proteins, PtdIns(4,5)P_2 can also be cleaved to form the second messengers Ins(1,4,5)P_3 and diacylglycerol (thereby providing a link between the families of phosphoinositides and inositol phosphates).[12, 13]

Under basal conditions the levels of the 3-phosphorylated phosphoinositides, PtdIns(3,4,5)P_3 and PtdIns(3,4)P_2, are much lower than that of PtdIns(4,5)P_2. These 3-phosphorylated phosphoinositides only constitute approximately 0.0001% of the cell membrane.[7, 14] In contrast, PtdIns(4,5)P_2 is thought to typically be 500 times more abundant.[7, 14]

1. 2 Protein Kinase B

1. 2. 1 Structure

Protein kinase B (PKB) is a serine/threonine kinase at the focus of many cell signalling pathways. There are 3 known mammalian isoforms of PKB (PKBα, PKBβ, PKBγ) which all show a high sequence homology. Each of these isoforms consists of 3 functional domains: an N-terminal pleckstrin homology (PH) domain, a kinase domain, and a C-terminal regulatory domain containing a hydrophobic motif.

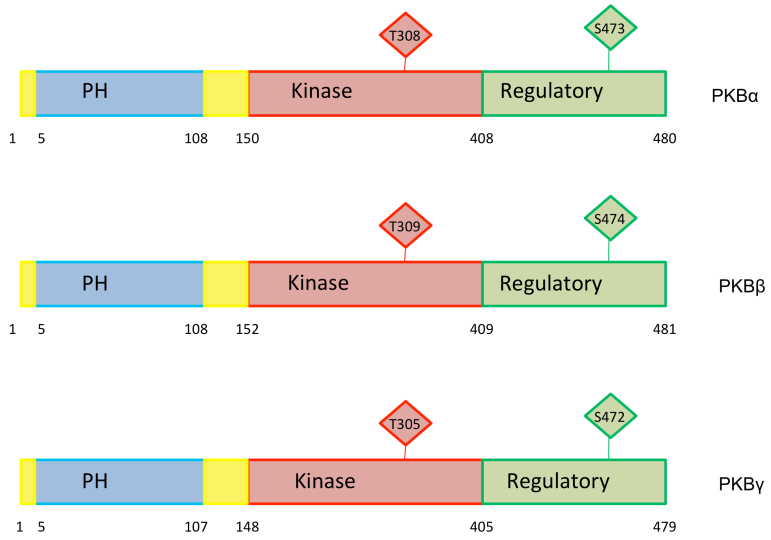


Figure 5: The three PKB isoforms. Each isoform consists of three domains: an N-terminal phosphoinositide binding PH domain, a kinase domain, and a C-terminal regulatory domain. The threonine and serine residues highlighted undergo phosphorylation to produce the active protein.

All three PKB isoforms share common roles, as well as exhibiting individual functionality. Whilst PKBα and PKBβ are expressed throughout the body, PKBγ is predominantly expressed in the brain, heart and kidney.[15, 16] The isoform considered throughout this work is PKBα, which is frequently found to be elevated in human cancers.[17]

1. 2. 1. 1 Pleckstrin Homology Domain

Pleckstrin homology domains are one of the most widely investigated lipid binding domains and are found in a large variety of proteins.[18] They consist of approximately 100-120 amino acids and, despite their low sequence homology, form a conserved secondary and tertiary structure.

Pleckstrin homology domains are constructed from two perpendicular β-sheets, referred to as a β-sandwich, capped by a C-terminal amphipathic helix at one corner, Figure 6. The perpendicular β-sheets are constructed from 3 and 4 antiparallel β-strands respectively. Three of the loops linking the β-strands are of particular interest due to their variable length and structure in different proteins, referred to as variable loops. These loops typically contain numerous cationic residues, forming a surface that can interact with poly-anionic ligands.[19] They are therefore thought to be partly responsible for the different specificities of various PH domains.[20] PH domains typically fall into four different categories with regards to their *in vitro* affinities for different phosphoinositides:[21]

- Group 1: PH domains specifically bind to PtdIns(3,4,5)P_3.
- Group 2: PH domains have a high affinity for both PtdIns(4,5)P_2 and PtdIns(3,4,5)P_3.
- Group 3: PH domains bind to PtdIns(3,4,5)P_3 and PtdIns(3,4)P_2 with greatest affinity.
- Group 4: PH domains that exhibit a low binding affinity for phosphoinositides, but are able to bind non-phosphoinositide ligands.

The PKB PH domain falls into group 3.

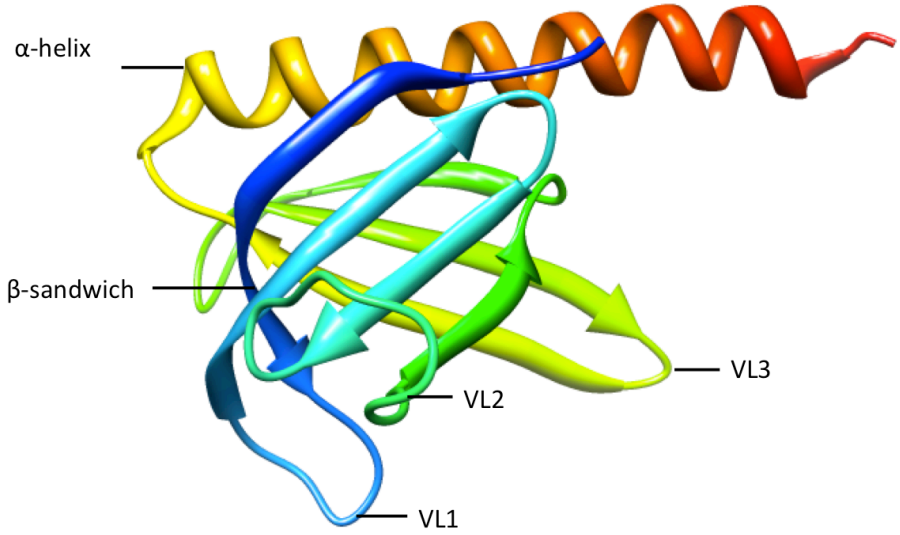


Figure 6: Crystal structure of the PKBα PH domain. The α–helix, β-sandwich, and variable loops 1, 2 and 3 (VL1, VL2 and VL3) are highlighted. PDB accession code 1UNQ.[11]

1. 2. 2 PKBα PH domain-Phosphoinositide Interaction

It is the distinct arrangement of hydroxyl and phosphomonoester substitution on the inositol ring that is responsible for the specificity of many phosphoinositide-protein interactions.[7] Experimental studies investigating PKB's affinity for phosphoinositides reflect its preference for 3-phosphorylated phosphoinositides [PtdIns(3,4)P_2 and PtdIns(3,4,5)P_3], over PtdIns(4,5)P_2.[22-25] Mutagenesis and crystallographic studies have helped explain this preference, illustrating that interactions of phosphoinositides with the PKB PH domain occur principally *via* phosphate groups 3 and 4 on their inositol phosphate head-groups.[11, 26]

The crystal structure of the PKB PH domain in a complex with Ins(1,3,4,5)P_4 [the head-group of PtdIns(3,4,5)P_3] illustrates that phosphate group 3 interacts with Lys14, Arg23, Arg25 and Asn53 and phosphate group 4 with Lys14, Asn53 and Arg86, Figure 7. Additionally phosphate group 1 contributes to binding by interacting with

Arg23. In contrast, phosphate group 5 is solvent exposed in the crystal structure and is hence considered to make no significant interactions.

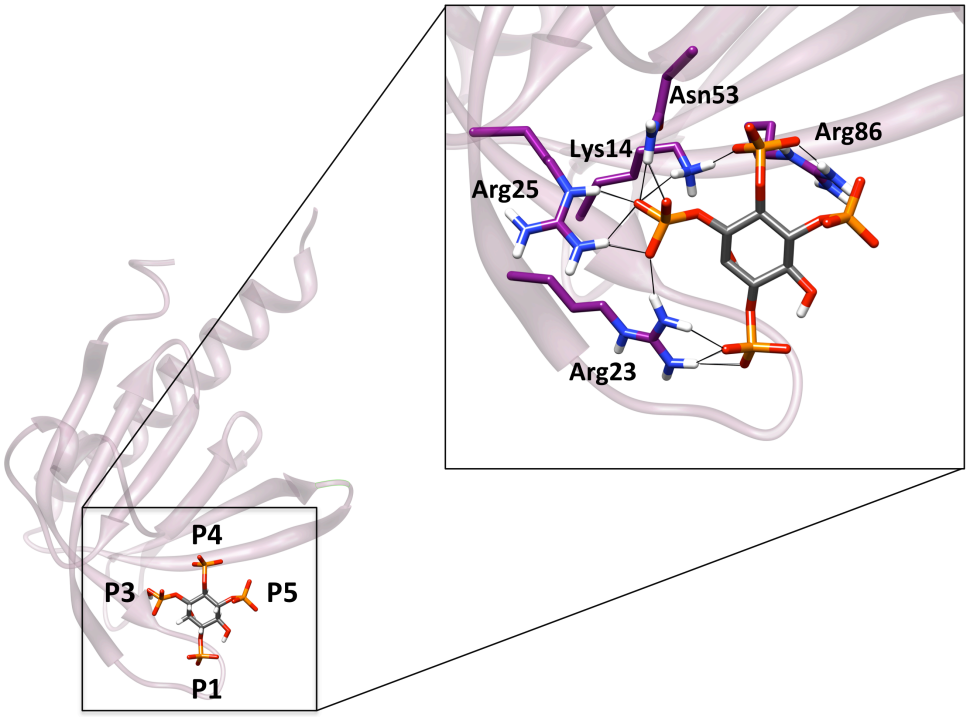


Figure 7: The crystal structure of Ins(1,3,4,5)P_4 in a complex with the PKB PH domain (PDB accession code 1UNQ).[11] The binding mode of Ins(1,3,4,5)P_4 in the PKB PH domain, with its 5-phosphate group pointing away from the binding site, is illustrated. A zoomed image shows the hydrogen bond interactions apparent in the binding site, indicated with black lines.

This binding mode of Ins(1,3,4,5)P_4 with the 5-phosphate group solvent exposed is also assumed for its parent lipid PtdIns(3,4,5)P_3. This is due to PtdIns(3,4,5)P_3's similar *in vitro* affinity for the PKB PH domain compared to PtdIns(3,4)P_2.[22, 26, 27] Furthermore, this binding mode for phosphoinositides is further consistent with the inability of PtdIns(4,5)P_2 to bind.

1. 2. 2. 1 Conformational Change

Upon binding the inositol phosphate head-groups of phosphoinositide lipids, the PKB PH domain illustrates a significant conformational change. This is displayed in Figure 8 which highlights the difference in structure of the PKB PH domain in its *apo*-state and bound to Ins(1,3,4,5)P_4.[11]

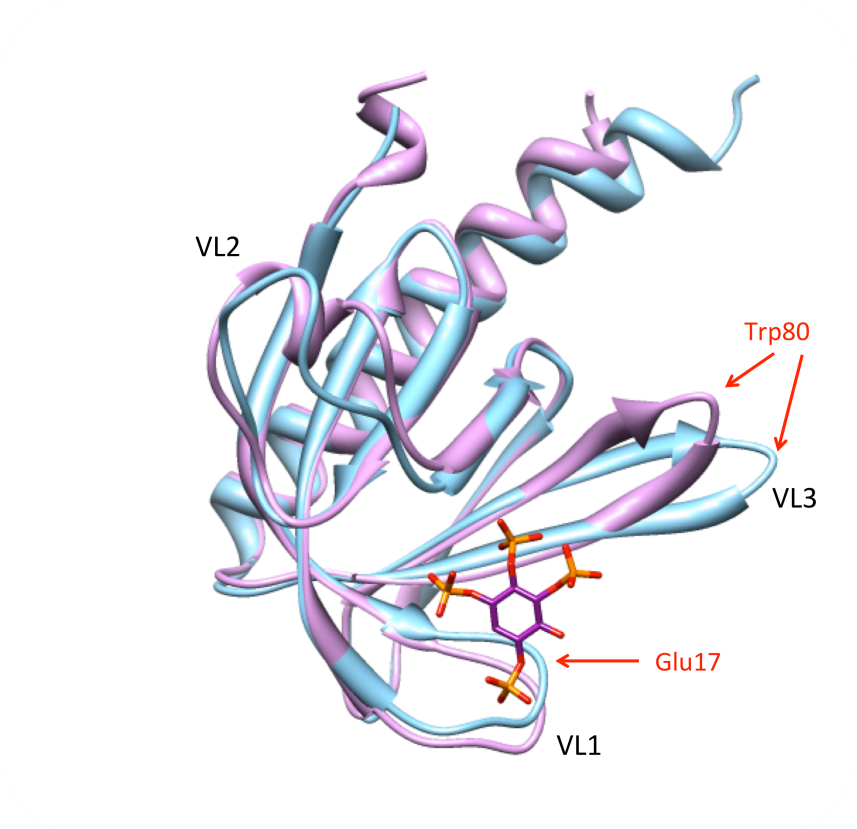


Figure 8: Comparison of the crystal structures of the PKB PH domain in its *apo*-state (blue) and bound to Ins(1,3,4,5)P_4 (magenta). PDB accession codes 1UNP and 1UNQ respectively.[11] Variable loops (VL) 1, 2 and 3 are indicated.

Binding of Ins(1,3,4,5)P_4 to the PKB PH domain is shown to interrupt a network of hydrogen bonds and cause significant movement of the variable loops. Differences in the position of variable loops 1, 2, and 3 are evident, shifting approximately 2.5, 7.6 and 7.4 Å respectively.[11] Upon binding Ins(1,3,4,5)P_4, loop 1 moves away from the PKB binding site as residue Glu17 is repelled from the highly anionic ligand, highlighted in Figure 8. Loop 2 forms a short α-helix, the purpose of which has not yet been elucidated. Loop 3 makes an especially large shift upon Ins(1,3,4,5)P_4

binding. The significant movement of the hydrophobic Trp80 residue on the tip of loop 3 (highlighted in Figure 8) has led to numerous suggestions of its role in PKB activity. A recent co-crystal structure of the PKB PH and kinase domains shows that loop 3 is in fact buried in the kinase domain in its inactive state. Upon binding to phosphoinositides, these domains are shown to separate allowing phosphorylation and activation of PKB.[28]

1. 2. 3 Crystal Structures

There are a number of crystal structures of the PKB PH domain currently available. High resolution crystal structures of the PKB PH domain in its *apo*-state, in a complex with Ins(1,3,4,5)P_4, and in a complex with a sulfate molecule were all resolved in a single study in 2003 (PDB accession codes 1UNP, 1UNQ and 1UNR respectively).[11] Only a year earlier, the crystal structure of PKB's PH domain in a complex with Ins(1,3,4,5)P_4 was solved for the first time (PDB accession code 1H10).[26] Since studying the native system crystal structures of a mutant PKB PH domain, as well as structures with alternative ligands, have also been resolved.[29, 30]

1. 2. 4 Function

PKB acts as a regulator for a variety of downstream proteins that play critical roles in cell survival, proliferation, growth and metabolism.[31, 32] These include Bad, Caspase-9 and FOXO transcription factors, which regulate the cell's apoptotic machinery,[33-35] and GSK3 and the CDK (cyclin dependent kinase) inhibitors p21 and p27 associated with cell cycle regulation.[36-38]

1. 2. 4. 1 Phosphoinositide 3-Kinase Pathway

The phosphoinositide 3-kinase (PI3K) intracellular pathway is one of the most frequently activated proliferation pathways in cancer, and is illustrated in Figure 9.[32] This pathway commences with the agonist stimulation of receptor tyrosine kinases and G-protein coupled receptors. Stimulation of these receptors results in the recruitment and activation of class I phosphoinositide 3-kinases (PI3Ks). Active

PI3Ks phosphorylate PtdIns(4,5)P_2 at the 3-hydroxyl group to form PtdIns(3,4,5)P_3, which is subsequently dephosphorylated by 5-phosphatases to form PtdIns(3,4)P_2. The dramatic increase in the concentration of the 3-phosphorylated phosphoinositides from their low basal levels allows them to function as membrane anchors for PKB.[7, 14] PKB is recruited to the plasma membrane *via* direct interaction of its PH domain with the inositol phosphate head-groups of PtdIns(3,4,5)P_3 and PtdIns(3,4)P_2. This interaction induces a conformational change in PKB, which facilitates phosphorylation of its residues Thr308 and Ser473 by PDK1 (3-phosphoinositide-dependent kinase 1) and mTORC2 (mammalian target of rapamycin complex 2) respectively.[11, 39-41] As highlighted in Figure 5, residue Thr308 is located in the activation loop of the kinase domain and Ser473 in the hydrophobic motif of the regulatory domain.

PDK1 phosphorylates PKB by co-localising with it at the plasma membrane; PDK1's PH domain also binds to 3-phosphorylated phosphoinositides. Additionally, it has recently been proposed that PDK1 and PKB are in equilibrium between an associated and dissociated state in the cytoplasm, prior to cell stimulation.[42]

Once activated, PKB dissociates from the plasma membrane and phosphorylates serine and threonine residues on a range of downstream cytosolic targets. The PI3K pathway is negatively regulated by PTEN (phosphatase and tensin homolog), which antagonises PI3K activity by dephosphorylating PtdIns(3,4,5)P_3 back to PtdIns(4,5)P_2.

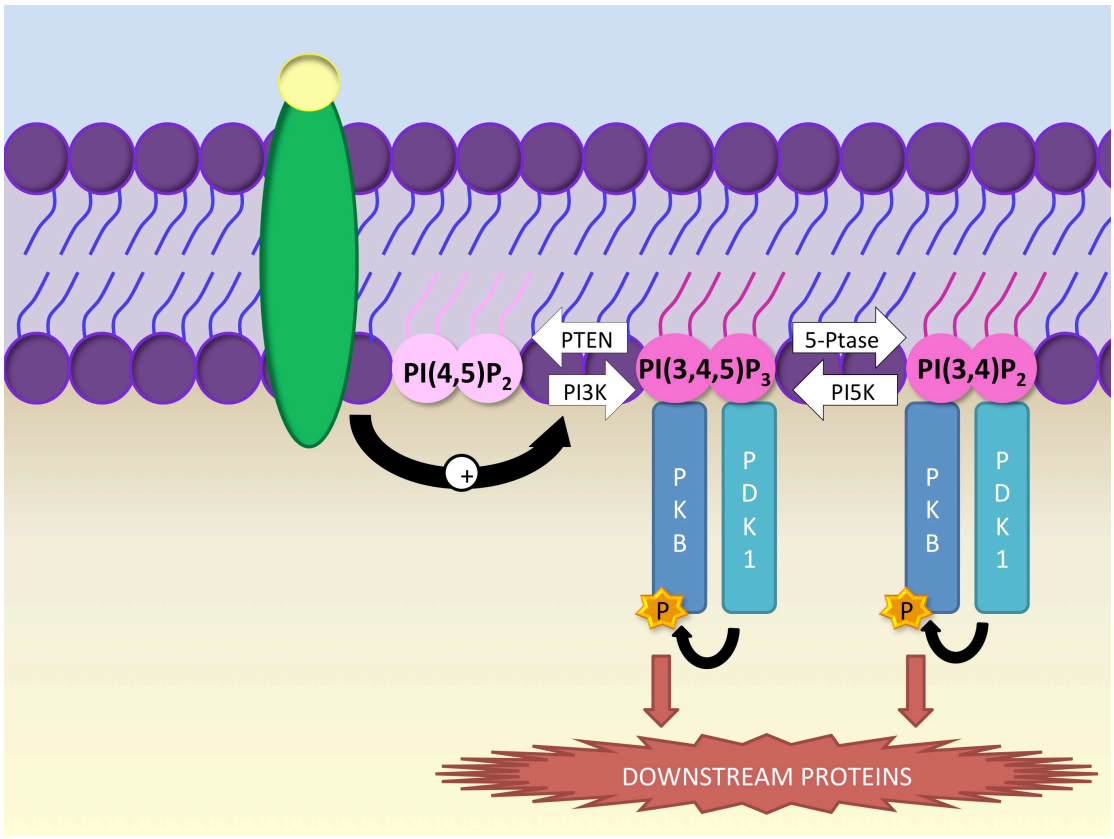


Figure 9: The PI3K pathway. The mechanism of activation for PKB is illustrated. The second phosphorylation step by mTORC2 has been omitted for ease of visualisation. The yellow circle represents an agonist for the green cell surface receptor.

1. 2. 5 PKB and Cancer

A number of proteins involved in the PI3K pathway are associated with a broad range of cancers, including class I PI3Ks, PTEN and PKB.[43] Increased PKB expression or activation has been noted in multiple cancer types, and in some cases has an incidence of over 50%.[44-47] Whilst this is primarily due to aberrant upstream proteins of PKB, an oncogenic PKB mutant has also been reported.[29, 48] This mutant has been reported in human breast, lung, ovarian, and colorectal cancers.[29, 49-51] The mutation lies at the binding interface of PKB's PH domain with phosphoinositides, and is thought to cause PKB to constitutively target the plasma membrane leading to over-activation.

As the PI3K pathway is one of the most frequently activated proliferation pathways in cancer, PKB presents itself as a therapeutic target at its heart. Current drugs that target PKB fall into three main categories: phosphoinositide mimetics, ATP-competitive and allosteric inhibitors.[52] The majority of inhibitors that compete with ATP are non-selective against the different PKB isoforms and closely related kinases, due to the similarities of their ATP binding sites.[53] PKB's PH domain therefore presents an alternative site to target that has proved popular in the past few years.[54] Inhibitors that target the PKB PH domain include phosphatidylinositol ether lipid analogues (PIAs), alkylphospholipids (APLs), sulfonamides and purine/pyrimidine analogues.[54-58] A number of these are discussed in further detail in Chapter 7.

Currently no targeted PKB therapeutics have reached the mass market. The most clinically advanced PKB inhibitor is perifosine, which currently has orphan drug status in the EU and the US to treat multiple myeloma. Perifosine is an alkylphospholipid that interferes with PKB's recruitment to the plasma membrane.[59]

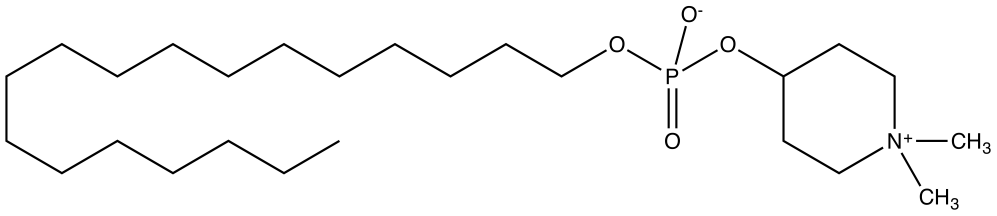


Figure 10: Schematic drawing of perifosine.

1. 3 Background to Project/Previous Research

The importance of understanding the interaction between PKB's PH domain and the inositol phosphate head-groups of phosphoinositide lipids has been highlighted. In this thesis a computational approach, using MD simulations, has been applied to study this system.

An advantage of using MD simulations to study a biological system is its efficiency compared to experimental techniques. For our system of interest computational techniques are of further convenience due to the micromolar affinity of PKB PH domain-inositol phosphate interactions.[22] This relatively low binding affinity, in comparison to drug molecules in the nanomolar range, means that large quantities of protein and inositol phosphates are required for *in vitro* techniques. Producing large quantities of these is both expensive and time-consuming, particularly so for inositol phosphate-based ligands which are often difficult and time consuming to synthesise.[60]

The lack of *in vitro* experiments investigating the interaction between PKB and inositol phosphates/phosphoinositides is highlighted by the repeated reference to Frech *et al.*'s work in which tryptophan fluorescence spectroscopy was used to determine the binding affinities of a number of inositol phosphates and phosphoinositide ligands for the PKB PH domain.[22] Despite the fact this work was carried out in 1997, and some useful data points are missing, there have been few additional investigations in the field.

Whilst a computational approach does not provide an alternative to experimental work, it can serve to explain and support experimental data. Additionally computational simulations can help direct experimental efforts and limit the number of *in vitro* experiments required. The popularity of *in silico* techniques in drug design is highlighted by the increasing number of discoveries made using computer-aided approaches.[52]

Docking studies have proved fairly popular for use in targeting the PKB PH domain, however there are few examples of studying this system using MD simulations.[52] This

can largely be attributed to the fact that until 2012 force fields did not contain parameters for inositol phosphates or phosphoinositides. In 2012 the Charmm carbohydrate force field was extended to enable the modelling of phosphates linked to carbohydrates, by transferring parameters from methylphosphates.[61] However, when investigating the interaction of PKB's PH domain with Ins(1,3,4,5)P_4 the authors of the parameters noted a rotation of the inositol ring from its crystal structure orientation, and loss of interactions *via* the 4-phosphate group.[61] As experimental studies suggest that interactions *via* the 4-phosphate group are of critical importance for binding, it is vital that the model replicates this behaviour. This work therefore highlights the importance of parameterising highly charged inositol phosphates thoroughly and algorithmically. Prior to the publication of the Charmm parameters, other groups also developed parameters for PtdIns(4,5)P_2 to study its interaction with gelsolin,[62] and in a lipid membrane.[63]

Whilst the behaviour of phosphoinositides in a lipid membrane have been briefly explored, there are very few examples of modelling their interaction with PH domains.[64] Sansom *et al.* are one of the main contributors in this field. Examples of their work include using MD simulations to study the interaction of PLCδ1's PH domain with PtdIns(4,5)P_2 and Ins(1,4,5)P_3,[65] and GRP1's PH domain with PtdIns(4,5)P_2.[66, 67] However, relatively little work has been carried out on PKB's PH domain. The only example, other than that already mentioned, of using MD simulations to study the interaction of the PKB PH domain with phosphoinositides/inositol phosphates is by Rong *et al.*[68] However this work was carried out before the crystal structure of the PKB PH domain-Ins(1,3,4,5)P_4 complex was solved and therefore is based on a homology model. Additionally as this study was carried out over a decade ago, simulations were only run for 200 ps.

1. 4 Overview of Thesis

This thesis consists of 8 further chapters (and a disc containing the Electronic Supplementary Information). The contents of these chapters are summarised below:

Chapter 2 – Quantum Mechanics provides background information about quantum mechanics (QM) theory. This chapter explains the different QM methods and basis sets applied in this thesis.

Chapter 3 – Molecular Mechanics describes the force field approach to studying biological systems. The Amber force field is explained in detail.

Chapter 4 - Parameterisation of Inositol 1,3,4,5-Tetrakisphosphate describes the parameterisation of Ins(1,3,4,5)P_4 for use in MD simulations. The protocol used in this work is in keeping with current Amber force fields. A number of variations within this protocol are considered.

Chapter 5 – MD Simulations of the Native PKB PH Domain with Inositol Phosphates describes the use of MD simulations to study the interaction of Ins(1,3,4)P_3, Ins(1,4,5)P_3 and Ins(1,3,4,5)P_4 [the head-groups of PtdIns(3,4)P_2, PtdIns(4,5)P_2 and PtdIns(3,4,5)P_3 respectively] with the PKB PH domain. The free energies of binding and the molecular behaviour of these complexes are validated by comparison to experimental studies. Dynamic information from the MD simulations helps build upon information from the crystal structure.

Chapter 6 - MD Simulations of the Mutant PKB PH Domain with Inositol Phosphates describes the investigation of a mutant form of the PKB PH domain found in numerous cancer types. Previous experimental work suggests that this mutant has an altered phosphoinositide selectivity resulting in it constituitively targeting the plasma membrane. MD simulations have been applied to help explain the altered behaviour of this mutant.

Chapter 7 - Inositol Phosphate Derivatives uses information from the previous chapters to help design inositol phosphate derivatives. These derivatives are designed to target the PKB PH domain and prevent its translocation to the plasma membrane. A computational approach is used to screen a number of potential inositol phosphate derivatives. Promising derivatives were synthesised by the Conway group at Oxford University, who also assisted in the design of these ligands. Once computational predictions had been made about the behaviour/binding affinity of the inositol phosphate ligands, biochemical techniques were then applied to validate these predictions.

Chapter 8 – Summary and Future Work summarises the findings within this thesis. Additionally, possible improvements and adaptations to the work carried out are considered. Numerous different directions this work could now take and further experiments (both computational and biochemical) that would be useful and interesting to carry out are suggested. Preliminary work for extending this study is also detailed.

Chapter 9 – Materials and Methods contains further details for the biochemical techniques applied in this work.

Chapter 2 - Quantum Mechanics

2. 1 Introduction to Quantum Mechanics

The development of quantum mechanics began when various phenomena, such as black body radiation and the photoelectric effect, could not be explained by classical Newtonian mechanics.[69, 70] These phenomena cast doubt onto the nature of an electron, and led to the idea that electrons were able to behave as both particles and waves, referred to as 'wave-particle duality'. Schrödinger developed an equation to deal with this, using the concept of a wavefunction, which now forms the foundation of quantum mechanics.[71]

Quantum mechanics (QM) is an aspect of molecular modelling that explicitly represents electrons in its calculations, offering detailed structural and energetic descriptions of molecules. QM calculations can therefore be applied for a multitude of purposes. These include investigating the formation/breaking of bonds and studying excited states of a system. Additionally QM calculations are frequently employed to optimise the geometry of a system to a local energy minimum. QM geometry optimisation calculations have been carried out in this work to determine the equilibrium structures of various inositol phosphate-based compounds.

2. 2 Schrödinger Equation

The Schrödinger equation is used to describe the quantum state of a system. For a time–independent polyelectronic system, the Schrödinger equation can be written as:

$$E\psi = -\frac{\hbar^2}{2}\sum_{n}^{nuclei}\frac{1}{M_n}\nabla_n^2\psi - \frac{\hbar^2}{2m_e}\sum_{i}^{electrons}\nabla_i^2\psi - \sum_{n}^{nuclei}\sum_{i}^{electrons}\frac{Z_n e^2}{4\pi\varepsilon_0 r_{in}}\psi$$

$$+ \sum_{i}^{electrons}\sum_{j>i}^{electrons}\frac{e^2}{4\pi\varepsilon_0 r_{ij}}\psi + \sum_{n}^{nuclei}\sum_{m>n}^{nuclei}\frac{Z_m Z_n e^2}{4\pi\varepsilon_0 r_{mn}}\psi$$

Equation 1: The time-independent Schrödinger equation. ψ is the wavefunction, which forms the fundamental basis of quantum mechanics. E is the total molecular energy of the system. The first two terms represent the kinetic energy of the nuclei and electrons respectively, where ∇^2 denotes differentiation with respect to the particle co-ordinates and $\hbar$ is Planck's constant divided by 2π. M_n is the mass of a nucleus with charge Z, and m_e is the mass of an electron of charge e. Subscripts i and j refer to electrons and m and n to nuclei. The last three terms of this equation describe the potential energy due to electrostatic interactions between nuclei and electrons, electrons and electrons, and nuclei and nuclei. Distances separating nuclei and electrons, pairs of electrons, and pairs of nuclei are given by r_{in}, r_{ij} and r_{mn} respectively.

2. 2. 1 The Born-Oppenheimer Approximation

Due to its complexity, the Schrödinger equation can only be solved exactly for the simplest of systems. For all poly-electronic atoms or molecules, mathematical approximations to the true solution of Schrödinger's equation are required.[72]

The Born-Oppenheimer approximation allows the motions of nuclei and electrons to be separated. The mass of a nucleus is much greater than the mass of an electron. Nuclei therefore move very slowly compared to electrons. The nuclei can therefore be considered fixed as the electrons carry out their motions, allowing the nuclear kinetic energy (in equation 1) to be ignored. The Schrödinger equation can thus be simplified

to:

$$E_{el}\psi_{el} = -\frac{\hbar^2}{2m_e}\sum_{i}^{electrons}\nabla_i^2\psi_{el} - \sum_{n}^{nuclei}\sum_{i}^{electrons}\frac{Z_n e^2}{4\pi\varepsilon_0 r_{in}}\psi_{el} + \sum_{i}^{electrons}\sum_{j>i}^{electrons}\frac{e^2}{4\pi\varepsilon_0 r_{ij}}\psi_{el}$$

Equation 2: The electronic Schrödinger equation. Where E_{el} is the electronic energy. The energy due to the coulombic interactions between nuclei and nuclei can then be added to this electronic energy to calculate the total energy.

2. 3 The Hartree Fock SCF Method

The Hartree Fock self-consistent field (HF SCF) method is an approximate method for solving the time-independent Schrödinger equation for a poly-electronic system.[73] For this method, the electron-electron repulsion term is neglected: The electron-electron repulsion is treated in an average way in which each electron is considered to move in the mean field of the other electrons of the system.[74] Thus for a single electron, the Schrödinger equation can be written as:

$$E_1\phi_1 = -\frac{\hbar^2}{2m_e}\nabla_1^2\phi_1 - \sum_{n}\frac{Z_n e^2}{4\pi\varepsilon_0 r_{1n}}\phi_1 + \sum_{j} v_{1j}(r_1)\phi_1$$

Equation 3: The one electron Schrödinger equation for electron 1. $v_{1j}(r_1)$ is the electron-electron repulsion energy averaged over all the other electrons.[74]

By treating the electron-electron repulsions in an average way, the poly-electronic wavefunction can be written as the product of one-electron wavefunctions. This is referred to as the Hartree product:

$$\psi_{el}(r) = \phi_1(r_1)\phi_2(r_2) \dots \phi_n(r_n)$$

Equation 4: The electronic wavefunction as a product of single-electron wavefunctions.

Equation 4 does not satisfy the Pauli exclusion principle and must therefore be adjusted to include electron spin. The Pauli exclusion principle states that if two electrons occupy an atomic orbital their spins must be anti-symmetric.[75] Electron spin is incorporated into the product of the one-electron wavefunctions by use of a Slater determinant, which expresses the many-electron wavefunction as a single determinant of spin orbitals.[76]

2. 3. 1 The Self-Consistent Field Approach

In order to calculate the average repulsion energy (v_{1j}) experienced by electron 1 in equation 3, all of the remaining electrons' orbitals must be known. This therefore poses a circular problem. An initial guess for these 'remaining' orbitals is made, referred to as zeroth-order orbitals. The one-electron equation is then solved for ϕ_1, providing a first order approximation. This value for ϕ_1 is included in the first order approximation for ϕ_2 and so on for all the electrons in the system. The first order approximations are then used to calculate second order approximations. The iterations are continued until the computed energy is self- consistent.[74]

2. 3. 2 The Variational Principle

HF theory utilises the variational principle, which states that the energy of any normalised anti-symmetric wavefunction will always be greater than the energy of the exact wavefunction.[76] The variational principle therefore provides a way to calculate the upper limit to the true energy of a system.

2. 3. 3 Electron Correlation Energy

The difference in energy between the true energy of the system (E_0) and the HF calculated energy (E_{HF}) is known as the correlation energy (E_{corr}).

$$E_{corr} = E_0 - E_{HF}$$

Equation 5: The correlation energy expressed as the difference between the true energy of the system and the HF calculated energy.

The correlation energy is due to the HF method neglecting electron correlation. It should be noted that a certain amount of electron correlation is accounted for in HF theory in the exchange term, which describes correlation for electrons with the same spin. However, correlation due to electrons with opposite spin is not included.[76, 77]

2. 4 Møller-Plesset Perturbation Theory

Møller-Plesset perturbation theory adds electron correlation effects by adding higher excitations to HF theory as a non-iterative correction.[76] This method is based on the Rayleigh-Schrödinger perturbation theory, in which the true Hamiltonian operator ($\hat{\mathrm{H}}$) is divided into two parts: a 'zeroth-order' Hamiltonian ($\hat{\mathrm{H}}_0$) and a small perturbation ($\widehat{\mathrm{V}}$):

$$\hat{\mathrm{H}} = \hat{\mathrm{H}}_0 + \lambda\widehat{\mathrm{V}}$$

Equation 6: Calculation of the true Hamiltonian operator $\hat{\mathrm{H}}$ from the 'zeroth-order' Hamiltonian ($\hat{\mathrm{H}}_0$) and a small perturbation ($\widehat{\mathrm{V}}$). λ is a parameter that can vary between 0 to 1. When λ equals 0, $\hat{\mathrm{H}}$ is equal to $\hat{\mathrm{H}}_0$. When λ is 1, $\hat{\mathrm{H}}$ equals the true Hamiltonian operator.

The eigenfunctions (ψ_i) and corresponding energies (E_i) of the true Hamiltonian operator $\hat{\mathrm{H}}$ are expressed in powers of λ. Corrections to the HF energy are thus

expressed as an infinite series, where $E_i^{(1)}$ is the first-order correction and $E_i^{(2)}$ the second order correction to the energy, and so on.[72]

$$\psi_i = \psi_i^{(0)} + \lambda\psi_i^{(1)} + \lambda^2\psi_i^{(2)} + \cdots = \sum_{n=0} \lambda^n \psi_i^{(n)}$$

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \cdots = \sum_{n=0} \lambda^n E_i^{(n)}$$

Equation 7: The eigenfunctions (ψ_i) and corresponding eigenvalues (E_i) of $\hat{H}$ expressed as powers of λ.

Many levels of MP theory exist, each offering increased orders of perturbation/correction. The most commonly used of these is MP2, which includes first and second order perturbations/corrections to the 'zeroth-order' Hamiltonian.

The advantage of Møller-Plesset perturbation theory is that it adds electron correlation effects in the quantum mechanical calculations. However, MP methods are not variational, and can therefore result in energies lower than the true energy of the system.[72]

2. 5 DFT Theory

Since the late 1980's and 1990's there has been an increased interest in density functional theory (DFT) methods.[72] This growth in popularity is largely due to its relatively low computational cost in calculating the electron correlation energy for a system, compared to post-HF methods.

DFT methods are based on a theorem by Kohn and Hohenberg which states that the ground-state energy, and properties of a system, can be defined by a unique functional of the electron density.[78] Kohn and Sham suggested that this functional should be approximated by partitioning it into four main terms:[79]

$$E[\rho(r)] = \sum_i \int \psi_i(r)\left(-\frac{\nabla^2}{2}\right)\psi_i(r)dr + \frac{1}{2}\iint \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|}dr_1 dr_2 + E_{XC}[\rho(r)] - \sum_n \int \frac{Z_n}{|r - R_n|}\rho(r)dr$$

Equation 8: The energy of a system within the Kohn-Sham scheme, split into four terms. The first term describes the kinetic energy, the second the electron-electron Coulombic energy, the third the contributions from exchange and correlation, and the final term the potential energy of the electron-nuclear interactions. $\rho(r)$ is the electron density. r_1 and r_2 are the electronic positions and R_n the nuclear position.

2. 5. 1 The Exchange Correlation Term

The exchange correlation term, $E_{XC}[\rho(r)]$, in equation 8 accounts for the energy due to electron-electron interactions not covered by the electron-electron Coulombic term. As there is no explicit functional form for the exchange-correlation term, a variety of approximations have been defined. Typically $E_{XC}[\rho(r)]$ is split into separate exchange and correlation terms.

In this thesis a hybrid functional, B3LYP (Becke, three-parameter, Lee-Yang-Parr), has been applied.[80] The exchange correlation term for B3LYP is a weighted sum/hybrid of contributions from the HF exchange term and the Becke 88 exchange functional, the VWN III correlation functional and the correlation functional of Lee, Yang and Parr.[80, 81]

DFT methods are significantly less computationally expensive than post-HF methods. The use of a functional of the electron density reduces the many-body problem of N electrons with 3N spatial co-ordinates to 3 spatial co-ordinates.[82] Despite its reduced cost, the exact form of the functional used to describe the exchange correlation term is unknown. Therefore approximate functionals for describing these terms are applied. An issue with the use of approximate functionals is that the exact electron exchange

and correlation effects covered are unclear. In contrast HF, and post-HF methods, treat the exchange contribution by essentially exact means.[72]

2. 6 Basis Set

A basis set is a mathematical description of a system's orbitals used in QM calculations. Linear combinations of basis functions (also referred to as atomic functions) are used to form these basis sets. The choice of basis set for studying a system is of critical importance. A larger, more complex basis set will give a better approximation of the orbitals, as fewer restrictions are applied to the location of the electrons.[76] However the accuracy of a basis set has to be balanced against its computational cost.

Basis functions were first developed by John C. Slater in 1930.[83] Unfortunately, some of the integrals involved in calculations for Slater-type orbitals (STOs) are complicated and computationally expensive. Boys noted that these STOs can be approximated by linear combination of Gaussian-type orbitals (GTOs), resulting in a large reduction in computational cost.[84] Today numerous basis sets are composed of these GTOs, and are discussed below.

Minimal basis sets are composed of the minimum number of basis functions required to describe each atom. For example, carbon can be described using five basis functions: 1s, 2s, 2px, 2py, 2pz. Minimal basis sets use fixed-size orbitals, and hence are often not sufficient for many calculations.

In comparison to minimal basis sets, split valence basis sets have an increased number of basis functions for each atom. The valence shell orbitals are described using two (or more) sizes of basis function, whilst the core orbital/s is represented by a single basis function. For example, a carbon atom is described by nine basis functions (1s, 2s, 2s', 2px, 2py, 2pz, 2px', 2py', 2pz') for a double split valence basis set such as 3-21g.[85]

Further flexibility of the atomic orbitals can be added by accounting for excited states, using polarised basis sets. Basis functions of higher angular momentum than required to describe the ground state are included. These additional orbitals are denoted by an asterisk, where 6-31g* adds d functions to heavy atoms and 6-31g** adds d functions to heavy atoms plus p functions to hydrogen atoms.[86]

Chapter 3 - Molecular Mechanics

3. 1 Introduction to Molecular Mechanics

Despite recent advances in quantum mechanical methods many systems are too large, and hence computationally expensive, to be studied using this approach. Molecular mechanics (MM) methods, also referred to as force field methods, present a practical computational approach to study such systems at the atomic level.

MM methods ignore electronic motions, and calculate the energy of a system as a function of the nuclear positions. The separation of the electronic and nuclear motions is facilitated by the Born Oppenheimer approximation, previously described in section 2.2.1.[87] The neglect of electronic motions in MM methods significantly reduces the computational cost of the calculations, allowing larger systems to be studied.

3. 2 The Force Field

A force field is defined by its potential energy function and an associated set of parameters. For most force fields the potential energy function takes the form of:

$$E_{potential} = E_{bond} + E_{angle} + E_{dihedral} + E_{non-bond}$$

Equation 9: A general potential energy function. The potential energy of a system is typically described as the sum of the bond, angle, dihedral and non-bonded energy contributions. These terms are illustrated in Figure 11.

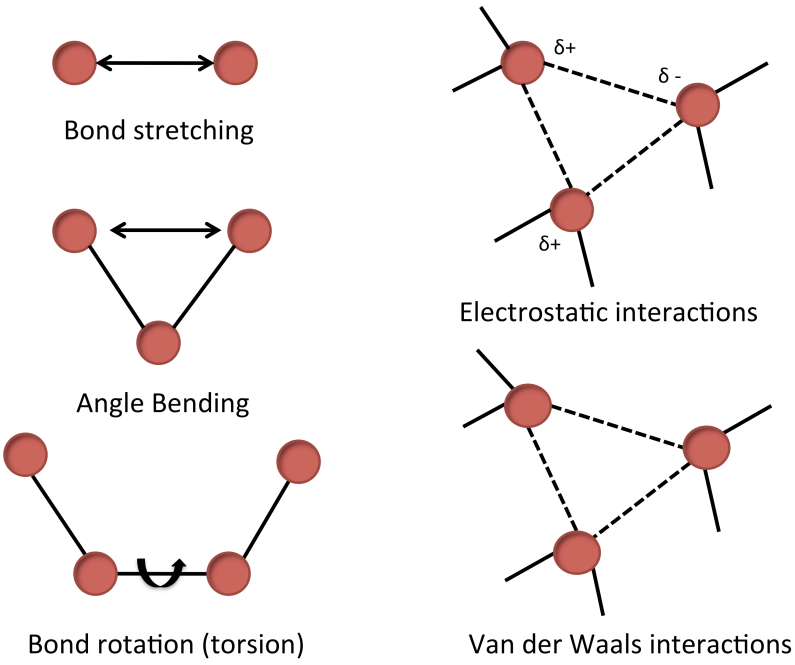


Figure 11: Representation of the different terms in the potential energy function. The non-bonded energy contributions are split into electrostatic and van der Waals interactions. Diagram adapted from Leach *et al.*[72]

Numerous different force fields exist, each developed to deal with different types of biological systems.[88-90] In this work the Amber ff99SB force field and General Amber Force Field (GAFF) extension have been applied.[91, 92] The ff99SB force field includes amino acid and nucleic acid parameter sets, whilst the GAFF extension allows organic molecules to be studied. These Amber force fields use the following functional form:

$$V(r^n) = \sum^{bonds} K_r(r - r_{eq})^2 + \sum^{angles} K_\theta(\theta - \theta_{eq})^2 + \sum^{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j}^{vdWs} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right] + \sum_{i<j}^{electrostatics} \frac{q_i q_j}{\varepsilon_r R_{ij}}$$

Equation 10: The potential energy function for the Amber force field. The potential energy V is expressed as function of the positions r of n atoms. r_{eq}, θ_{eq}, K_r, K_θ, V_n, n, γ, A_{ij} and B_{ij} are parameters defined by the force field. Scale factors of 1/2 and 1/1.2 are applied for 1-4 vdW and 1-4 electrostatic interactions respectively.

3. 2. 1 Bonded Interactions

The first three terms in equation 10 represent interactions between atoms covalently attached *via* 1 to 3 bonds.

The first term describes the energy involved in stretching a bond between two atoms, approximated by a harmonic potential obeying Hooke's law. The energy contribution from bond stretching is proportional to the square of the displacement of the bond length (r) from its equilibrium bond length (r_{eq}). Similarly, the second term also utilises Hooke's law to describe angle bending, with the energy contribution varying due to distortion of the bond angle (θ) from its equilibrium value (θ_{eq}). K_r and K_θ are the force constants for the bond and angle terms respectively.

These two terms are regarded as having 'hard' degrees of freedom, as a large amount of energy is required to cause a significant deviation from their equilibrium geometries. The 'softer' torsional and non-bonding terms are responsible for the majority of the variation in structure and energies observed in biological systems.

The torsional/dihedral energy is described by the third term in equation 10 and describes the barriers to rotation about a bond. A Fourier series expansion is used to describe the torsional potential, where V_n is the relative barrier height to rotation, n the multiplicity (number of minima in a 360° rotation), ϕ the dihedral/torsion angle and γ is the phase factor (which determines the location of the minima). Most dihedral angles in Amber are described adequately by a single two- or three-fold term ($n = 2, 3$), however a few dihedral angles require more terms to accurately represent the torsional profile. Improper torsions describe out-of-plane bending and are accounted for in the same way as regular dihedral angles, except that the four atoms are not attached in sequence.

3. 2. 2 Non-Bonded Interactions

Non-bonded interactions are considered to act 'through-space'. In Amber force fields, non-bonded interactions are deemed to be of paramount importance.[93] However, these interactions are also the most time consuming to evaluate, as the number of possible non-bonded interactions is equal to roughly the square of the number of atoms in the system. The non-bonded interactions can be split into van der Waals and electrostatic interactions, discussed in the following sections.

3. 2. 2. 1 Van der Waals Interactions

In the Amber force field, van der Waals (vdW) interactions are described by a Lennard Jones 12-6 function, illustrated as the penultimate term in equation 10.

$$V(r) = \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right]$$

Equation 11: A Lennard Jones 12-6 function. A_{ij} and B_{ij} equal εr_m^{12} and $2\varepsilon r_m^6$ respectively, where ε is the well depth and r_m is the interatomic distance at which the energy is a minimum.

VdW interactions are considered to arise from a combination of attractive and repulsive forces. At longer distances, attractive forces dominate. This attractive force varies as R_{ij}^{-6}, where R is the interatomic distance between atoms i and j. The attractive forces are due to the formation of instantaneous dipoles that can induce a second dipole in a nearby atom, thus resulting in an attractive inductive effect. At short distances, repulsive forces dominate, which vary as R_{ij}^{-12} in this equation.[72]

3. 2. 2. 2 Electrostatic Interactions

Different elements have different associated electronegativities, resulting in the unequal charge distribution in a molecule. For current Amber force fields, the charge distribution is represented by partial atomic charges. The energy of the electrostatic

interactions between a pair of partial atomic charges is calculated using Coulomb's law, shown in the final term of equation 10.

$$V(r) = \frac{q_i q_j}{\varepsilon_r R_{ij}}$$

Equation 12: Electrostatic interactions calculated using Coulomb's law. q_i and q_j are the partial atomic charges of atoms i and j, R_{ij} is the distance between them, and ε_r is the relative permittivity of the medium.

3. 2. 2. 2. 1 The RESP Procedure

The restrained electrostatic potential (RESP) procedure is used to derive the charge parameters for the Amber ff99SB force field.[94-96] In the RESP procedure the molecular electrostatic potential (MEP) of a geometry-optimised molecule is computed. Charge values are then subsequently fitted to atom centred points to best reproduce this MEP. The RESP procedure is "a powerful, general and algorithmic method to determine atomic charges for force field purposes".[94]

The process begins with evaluating the electrostatic potential at a large number of points defined by shells at 1.4, 1.6, 1.8 and 2.0 times the van der Waals radii of the atoms.[97] The value of the electrostatic potential at each point is calculated using the Hartree Fock method with a 6-31g* basis set. This level of theory is chosen as it exaggerates the dipole moment of most residues by 10-20%, which is consistent with the TIP3P model of water.[96]

A least squares fitting procedure is used to fit charges to each atomic centre in a two-stage approach. In the first stage, charges for the entire molecule are fitted with weak hyperbolic restraints (0.0005 au) on non-hydrogen atoms. If the molecule contains methyl or methylene groups, these groups are re-fitted in a second stage with stronger restraints (0.001 au) on non-hydrogen atoms, whilst the remaining atoms of the molecule are frozen to their values from the first stage. This two-stage approach allows charges on buried atoms (for example, alkyl carbons) to be reduced without

significantly affecting the fit to the electrostatic potential.[94]

3. 2. 2. 3 Non-bonded Cut-offs

Non-bonded interactions are the most computationally expensive terms of force field calculations. The number of possible non-bonded interactions in a system is equal to roughly the square of the number of atoms in the system, $N(N-1)/2$. To dramatically increase the speed of calculations, only atoms within a given cut-off distance of each other are generally considered. This approximation is justified as the size of vdWs interactions decrease rapidly with distance, falling to zero at ~8 Å. Therefore the Lennard Jones potential can be truncated without introducing significant errors into the calculation. The use of periodic boundary conditions (section 3.4) limits the maximum cut-off distance to half the size of the smallest cell dimension, as each atom should only 'see' one image of the remaining atoms in the system. This is referred to as minimum image convention.

Unfortunately electrostatic interactions cannot be truncated in the same manner, as they are longer ranged than vdW interactions. The long-range electrostatic interactions are considered in the following section.

3. 2. 2. 4 Long-Range Forces

Long-range electrostatic interactions cannot be handled in the same manner as vdW interactions, as their range can often exceed beyond half the box length. The particle mesh Ewald (PME) method is applied to deal with these interactions.[98] Within the defined cut-off distance electrostatic interactions are calculated in real/direct space. However long-range interactions beyond the defined cut-off distance are calculated in reciprocal space, which reduces the cost of these calculations to $N \ln N$.

3. 2. 3 Atom Types and Parameters

A good parameter set is critical to the success of a force field. In order to be practical a parameter set should be relatively small and transferrable, yet still be able to accurately represent a system. In force field calculations, atom types are used to describe/characterise atoms in similar environments. Parameters are then provided for interactions between each atom type, rather than between each individual atom, simplifying the number of parameters needed to describe a system. Each force field has a finite number of atom types that are assigned to each atom dependent on its element, hybridisation, aromaticity and atoms it is bonded to.

In this work, atom types from the ff99SB and GAFF force fields have been applied. Atom types from the ff99SB force field are typically used to describe protein and nucleic acid systems. The GAFF force field was developed as a more general purpose force field, providing a broad set of parameters that can easily be transferred, covering most of organic chemical space.[92] Both ff99SB and GAFF atom types have been considered for application to inositol phosphate-based ligands in this work.

3. 3 The Solvent Model

MD simulations are usually performed in the presence of solvent. For biological systems, such as proteins, the solvent typically used is water. Water models fall into two main categories: implicit and explicit water models.

Implicit models represent solvent as a continuous medium instead of individual explicit water molecules. Whilst implicit models can exert some properties of water such as dampening the electrostatic interactions of the solute, some effects such as hydrogen bonding are not described adequately. Further discussion of implicit solvent models is provided in Chapter 4.

For the MD simulations in this thesis explicit water models have been used throughout. Although explicit water models are more computationally expensive than implicit models, there are numerous approaches to reduce the cost. Simple rigid

solvent molecules reduce the degrees of freedom available to water molecules, and thus also reduce the number of interactions to be calculated. The molecules are kept rigid such that no bond or angle deformation can occur. A commonly applied example of a rigid solvent model is the TIP3P model used in this work.[99] The TIP3P model is referred to as a 3-site model as it uses three interaction sites for electrostatic interactions: Charges are places on the central oxygen atom and balancing charges on the two hydrogen atoms. The vdW interaction between two water molecules is calculated using a single site located on the oxygen atom. Additionally periodic boundary conditions, discussed in the following section, also help to lower the cost associated with explicit water models.

3. 4 Periodic Boundary Conditions

Periodic boundary conditions allow simulations to model bulk fluid behaviour using a relatively small number of solvent molecules. A periodic array of boxes (each containing the particles being simulated) is constructed in 3 dimensions around the central box. This periodic array is illustrated in two-dimensions in Figure 12. If a particle diffuses out of the central box and into a neighbouring box, it re-enters the box from the opposite site as its image. The number of particles within a box therefore remains constant.

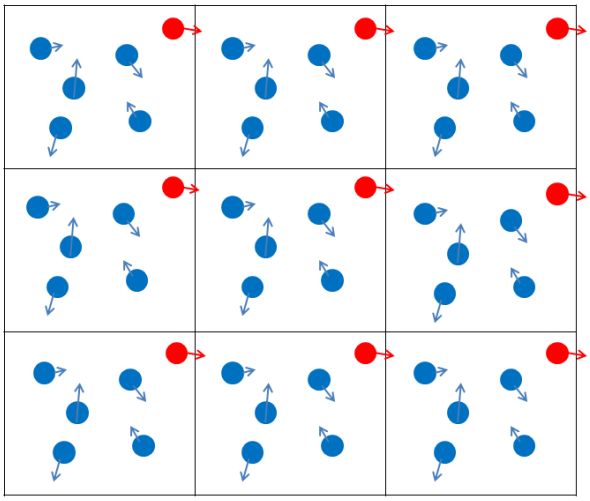


Figure 12: A schematic representation of periodic boundary conditions, shown here in two dimensions. Each box is identical and repeated in all in directions so that an atom leaving a box reappears in the neighbouring box, highlighted by the red atom.

3. 5 Energy Minimisation

Molecular simulations are often concerned with structures at the minima of the potential energy surface (PES). Energy minimisation, also referred to as geometry optimisation, is a technique used in both quantum mechanical and force field approaches. Although inadequate to provide details on the dynamics of a system, energy minimisation provides an efficient method to refine molecular structures to a local energy minimum. In this work, energy minimisation methods have been employed to generate suitable starting structures for MD simulations. Additionally, minimisation calculations have also been applied for the comparison of QM and MM descriptions of inositol-based ligands. Below some of the methods for finding minimum energy structures are discussed.[72, 100]

Minima are stationary points on the PES corresponding to stable states of a system. Minima are defined as points on the PES where the first derivative of the potential energy function is zero with respect to all the co-ordinates and the second derivatives are all positive. This can be expressed for the function $V(x_1, x_2, x_3 \ldots, x_n)$:

$$\frac{\partial V}{\partial x_i} = 0; \qquad \frac{\partial^2 V}{\partial x_i^2} > 0$$

Equation 13: The first and second derivatives of the potential energy function (V) at a minimum. The variables x_i are the co-ordinates of the atoms.

Locating the minimum for a system with 3N co-ordinates is extremely difficult. Therefore minima are determined by using numerical methods that gradually change the co-ordinates of a system to produce a series of configurations with increasingly lower energies.

3. 5. 1 First Derivative Methods

The first derivative (i.e. the gradient) of the potential energy with respect to a given co-ordinate indicates the direction in which the energy minimum lies. The magnitude of the gradient indicates the steepness of the local slope. The energy of a system can be gradually minimised/lowered by adjusting the position of each atom in response to the force acting on it at each step (the force is equal to the negative of the gradient). Steepest descent and conjugate gradient methods are first-order minimisation techniques that are frequently used, and have been applied in this work.[72]

The steepest descent method always moves in a direction negative to the gradient (parallel to the net force), towards a minimum. This minimum then provides the starting point for the next step. A line search is carried out in the direction perpendicular to the previous line search until a minimum along this line is found, and so forth. This method is typically used for the beginning stages of a minimisation, as it is suited for systems that are far from their equilibrium. However, for systems close to equilibrium this method does not tend to converge well. Conjugate gradient methods attempt to improve upon this method. In steepest descent the directions of successive steps are orthogonal, which can result in oscillatory behaviour in narrow valleys. Instead conjugate gradient methods use a combination of the current gradient as well as the previous search direction to decide the direction of the next line search. Removing the restriction of movement in only an orthogonal direction improves the performance of the minimisation when close to a minimum.

3. 5. 2 Second Derivative Methods

Second derivative methods use both the first derivatives (i.e. the gradients) and the second derivatives of the potential energy to locate a minimum. The second derivatives of the energy describe the curvature of the function. [72]

The Newton-Raphson method is an example of a second derivative method. This method is efficient when close to a minimum, as the surface is essentially quadratic.

However for large systems it is impractical as it involves the calculation and inversion of the Hessian matrix, which is very computationally demanding.

3. 6 Molecular Dynamics Simulations

Molecular dynamics simulations are employed to study how the dynamics of a system evolves over time. MD simulations generate successive configurations of a system over time by integrating Newton's second law of motion:

$$\frac{F}{m} = \frac{d^2r}{dt^2}$$

Equation 14: Newton's second law of motion. This equation describes the motion of a particle, where r is the position of a particle of mass m, operated on by a force F.

The force is determined from the first derivative of the potential energy expression described by the force field equation.

$$F = -\frac{dV}{dr}$$

Equation 15: Force is proportional to the negative derivative of potential energy V with respect to position r.

The differential equations that arise from Newton's second law are solved using numerical integration methods, such as finite difference methods. In finite difference methods, the integration is carried out in many small steps, each separated by a time-step δt.

The Leap-frog algorithm is one of the commonly used finite difference methods, which is applied in Amber.[101] In this algorithm, the positions and velocities are

calculated at different points in the trajectory. Positions are calculated at times t, $t + \delta t$ *etc*., whilst velocities are calculated half way through these time intervals, $t + \frac{1}{2}\delta t$. This is illustrated in equation 16.

$$\boldsymbol{v}\left(t + \frac{1}{2}\partial t\right) = \boldsymbol{v}\left(t - \frac{1}{2}\partial t\right) + \partial t\boldsymbol{a}(t)$$

$$\boldsymbol{r}(t + \partial t) = \boldsymbol{r}(t) + \partial t\boldsymbol{v}\left(t + \frac{1}{2}\partial t\right)$$

Equations 16: Equations illustrating the Leap-frog algorithm. Velocities are calculated from the current acceleration and the velocity from half a time-step behind. Positions are calculated from the previous position and the half time-step velocity.

3. 6. 1 Time-step

The choice of time-step is important for a simulation. If a time-step is too short, the amount of phase space explored will be insufficient. However if the time-step is too large, atoms can become unreasonably close between two steps as the force is assumed to be constant throughout a time-step. This therefore causes instabilities in the simulation to arise. To avoid these inaccuracies, a time-step is limited to one order of magnitude smaller than the highest frequency motion in the system. The highest frequency motions are due to bond stretches, often involving hydrogen atoms, with a period of approximately 10 fs. The time-step for a simulation is thus theoretically limited to 1 fs. However in practice, the SHAKE algorithm allows constraints on high frequency motions (i.e. on bonds involving H atoms) therefore permitting a larger time-step of 2 fs to be used.[102]

3. 6. 2 MD Protocol

Molecular dynamics simulations are carried out in multiple stages. The starting structure for a simulation is often taken from crystallographic or NMR data. The starting structure is then adapted by adding hydrogen atoms, missing atoms, counter-ions and solvent. Following this, the solvated system is then energy minimised to remove any close contacts or high energy conformations.

Prior to MD, velocities are assigned to atoms by applying a Maxwell-Boltzmann distribution of velocities based on the system's temperature. The force acting on each atom is then calculated and used to generate new positions and velocities for each atom using the integration method discussed previously.

To begin the MD simulation, the system is gradually heated from 0 K to the desired temperature (usually 300 K) at constant volume. Following this, a period of equilibration at constant pressure and temperature is typically applied. The production MD simulation is then carried out. In this work the production MD simulations were carried out at constant temperature and pressure, as is typically used for simulations of biomolecules as it replicates experimental conditions. It is from this production run that the properties and behaviour of the equilibrated system can be determined.

3. 6. 3 Length of Simulation

The period of equilibration and production MD can vary depending on the system being studied. It is important that the system is properly equilibrated so that the production run represents the dynamics of the system at equilibrium.

The length of the production MD simulation is also an important consideration. This largely depends on the computational resources available and the motions of the biological system being investigated. Different motions occur across a number of different timescales, illustrated in Figure 13. Whilst microsecond simulations are becoming increasingly popular due to recent computational advances, the ability of force fields to capture realistic and stable simulations over this long timescale is still a big question in the computational community.[103-105]

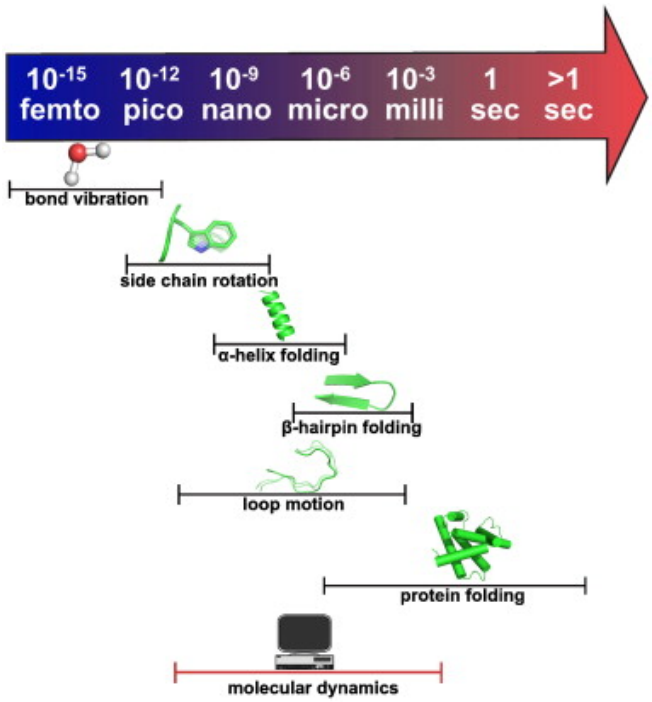


Figure 13: Comparison of motions in proteins and their timescale, taken from Werner *et al.*[106] Reprinted from Advanced Drug Delivery Reviews, 64, T. Werner, M. B. Morris, S. Dastmalchi, W. B. Church, Structural modelling and dynamics of proteins for insights into drug interactions, 323-343, Copyright 2012, with permission from Elsevier.

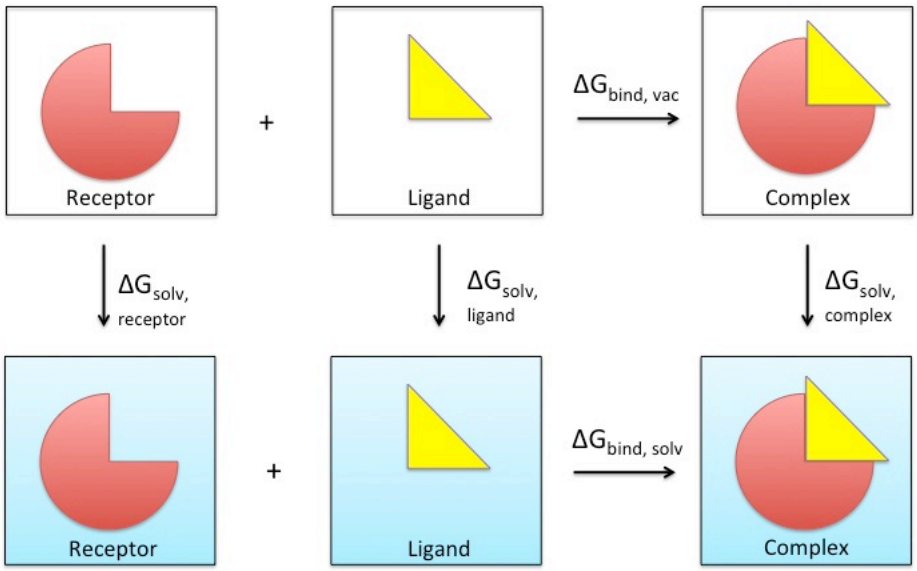
3. 7 MM-PBSA

The MM-PBSA (Molecular Mechanics-Poisson Boltzmann Surface Area) approach was developed by Kollman and Case for calculating the free energy of binding for biomolecular systems.[107, 108] MM-PBSA is classified as an 'end point' post-processing free energy approach. It merely requires conformational ensembles of the bound complex and of the unbound receptor and ligand separately – the end-points of the binding interaction, equation 17. This sets it apart from 'pathway' free energy methods, which simulate the free energy change of a chemical mutation through a series of alchemical intermediates. This additional simulation leads pathway free energy methods, such as Free Energy Perturbation (FEP) or Thermodynamic Integration (TI), to have a far greater computational cost than end-point methods, albeit with suggested improvements in accuracy.[109, 110] However, the cost of pathway techniques makes them intractable for the large-scale simulations performed in this study.

$$\Delta G_{bind} = G_{complex} - (G_{receptor} + G_{ligand})$$

Equation 17: The free energy of binding (ΔG_{bind}) calculated as the difference in the free energies of the bound (complex) and un-bound (receptor and ligand) species.

The MM-PBSA method utilises a thermodynamic cycle to calculate the free energy of binding, illustrated in Figure 14. A thermodynamic cycle is required because fluctuations in the solvation energy due to solvent-solvent interactions can be an order of magnitude greater than the energy contributions from the protein and ligand, if the binding free energy is calculated directly from a solvated trajectory. This can therefore result in large errors in the calculated free energy of binding. To circumvent this problem the free energy of binding is calculated indirectly as illustrated in Figure 14.



$$\Delta G_{bind,solv} = \Delta G_{bind,vac} - \Delta G_{solv,receptor} - \Delta G_{solv,ligand} + \Delta G_{solv,complex}$$

Figure 14: A thermodynamic cycle for calculating the free energy of binding indirectly. The free energy of binding is the sum of the gas phase interaction energy ($\Delta G_{bind,vac}$) and the difference in solvation free energies between the complex and the un-bound receptor and ligand.

The gas phase interaction energy ($\Delta G_{bind,vac}$) in the thermodynamic cycle can be decomposed into molecular mechanical and entropic contributions. The free energy of binding can therefore be written as:

$$\Delta G_{bind,solv} = \Delta E_{MM} + \Delta\Delta G_{solv} - T\Delta S$$

Equation 18: Calculation of the free energy of binding. The free energy is split into contributions from the gas phase MM energy (ΔE_{MM}), the solvation free energy ($\Delta\Delta G_{solv}$) and the entropy (ΔS), where T is the temperature in Kelvin.

3. 7. 1 Molecular Mechanical Energy

The first term in equation 18 represents the gas phase molecular mechanical (MM) energy. This is calculated using the same force field as applied in the MD simulations. It can be further split into the following terms:

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdW}$$

Equation 19: Calculation of the gas phase molecular mechanical energy. The MM energy includes contributions from the internal (ΔE_{int}), electrostatic (ΔE_{ele}) and van der Waals (ΔE_{vdW}) energies.

3. 7. 2 Solvation Energy

The second term of equation 18 represents the solvation free energy. The replacement of the explicit solvation of the simulation with an implicit model improves the speed of the free energy calculation and reduces errors from noise created by the random motion of explicit water molecules.[111] The solvation free energy is split into polar and non-polar contributions. These are calculated using the Poisson Boltzmann/surface area (PBSA) approach.

3. 7. 2. 1 Polar Solvation Energy

The polar solvation energy is the free energy change of transferring a molecule from gas phase ($\varepsilon = 1$) to water ($\varepsilon = 80$), and is estimated by implicit solvent models. In this study the Poisson Boltzmann (PB) implicit solvent model has been applied. The PB model is generally considered to be the gold standard for implicit solvation calculations.[112, 113] A variant of the MM-PBSA approach is the MM-GBSA method, in which the Generalised Born implicit solvent model is used in place of the Poisson Boltzmann model.[114] The Generalised Born (GB) model offers a less theoretically rigorous, but also less computationally expensive, representation of implicit solvation.

The performance of MM-PBSA and MM-GBSA methods in predicting free energies

of binding was recently assessed by Hou *et al.*[115] MM-PBSA was shown to perform better at calculating absolute free energies than MM-GBSA. However, MM-PBSA's ability to out-perform MM-GBSA at calculating relative free energies was system dependent.

3. 7. 2. 2 Non-polar Solvation Energy

The non-polar contribution to the solvation free energy is primarily due to the free energy required for cavity formation of the solute. It is assumed to be proportional to the solvent-accessible surface area (SASA):

$$G_{SASA} = \gamma . SASA$$

Equation 20: The non-polar contribution to the solvation free energy. *SASA* is the solvent-accessible surface area of the molecules and γ is the solvent surface tension.

3. 7. 3 Entropy

The final term in equation 18 illustrating the entropic contribution to binding is usually calculated using either quasiharmonic or normal mode analysis. Both of these techniques are extremely computationally demanding, due to the large number of minimised snapshots required for quasiharmonic analysis and the extensive minimisation of snapshots required for normal mode analysis.[111, 115] As such, entropic contributions are often neglected due to their computational expense.

Chapter 4 - Parameterisation of Ins(1,3,4,5)P_4

4. 1 Introduction

There are no existing parameters in the Amber force field for inositol phosphates, or their parent phosphoinositide lipids. Therefore, in order to perform MD simulations involving these molecules, suitable parameters have to be generated. In this chapter the parameterisation of inositol 1,3,4,5-tetrakisphosphate has been explored. A significant portion of the work within this chapter has been submitted as a paper.[116]

4. 2 Previous Parameters for Inositol Phosphates

Until recently force fields did not contain parameters for inositol phosphates. However in 2012 the Charmm carbohydrate force field was extended to enable the modelling of phosphates linked to carbohydrates.[61] The bonded and non-bonded parameters for these phosphate groups were generally taken from prior versions of the force field.[61] Partial charges on the phosphate groups were transferred from previous parameter development on methylphosphates. Unfortunately when applied in test case simulations to study the interactions of inositol phosphates, the parameters sometimes failed to replicate the experimental behaviour of these ligands.[61]

There are few other examples of modelling inositol phosphates or phosphoinositides in the literature. Those that do exist simplify the problem of the net charge and ionisation state of these molecules. The phosphate groups of phosphoinositides/inositol phosphates are often assumed to all be fully deprotonated.[62, 63, 68] However *in vitro* data suggests that full deprotonation of all of these groups, within such a small molecule, is extremely unlikely and unfavourable.[3, 4] In other efforts a more realistic approach has been taken, where a charge state for the phosphoinositides/inositol phosphates has been chosen from experimental data.[65, 117] Here some of the phosphate groups are fully deprotonated, whilst others are mono-protonated. The issue with this approach is that the charge state for these molecules is uncertain, illustrated by the net charge of -3, -4 and -5 all predicted for PtdIns(4,5)P_2 from different *in vitro* experiments.[118-120] The variability in these results

is due to the fact that the net charge of inositol phosphates and phosphoinositides can vary strongly around physiological pH due to their complex ionisation behavior.[3-5] Additionally, local pH and interactions with proteins can further add to this variability.[118]

Many of the roles of inositol phosphates are intimately linked to their ionisation state.[3] Therefore in this work, parameters are generated for all possible ionisation states of Ins(1,3,4,5)P_4. The Ins(1,3,4,5)P_4 isomers fall into five different charge subsets/states ranging from -4 to -8. The *in vitro* probabilities of these charge states at varying pH are displayed in Figure 15. Within these five different charge states, there are sixteen possible ionisation states of Ins(1,3,4,5)P_4, illustrated in Figure 16. The naming convention created for these isomers is as follows: All fully deprotonated phosphate groups are followed by the letter 'O' and mono-protonated phosphate groups by the letter 'H'. For example, for isomer 5O134H phosphate group 5 is fully deprotonated and phosphate groups 1, 3 and 4 are mono-protonated. The *in vitro* probability of the different ionisation states within each charge state is also provided. This probability data is taken from the *in vitro* work of Borkovec and Spiess.[3]

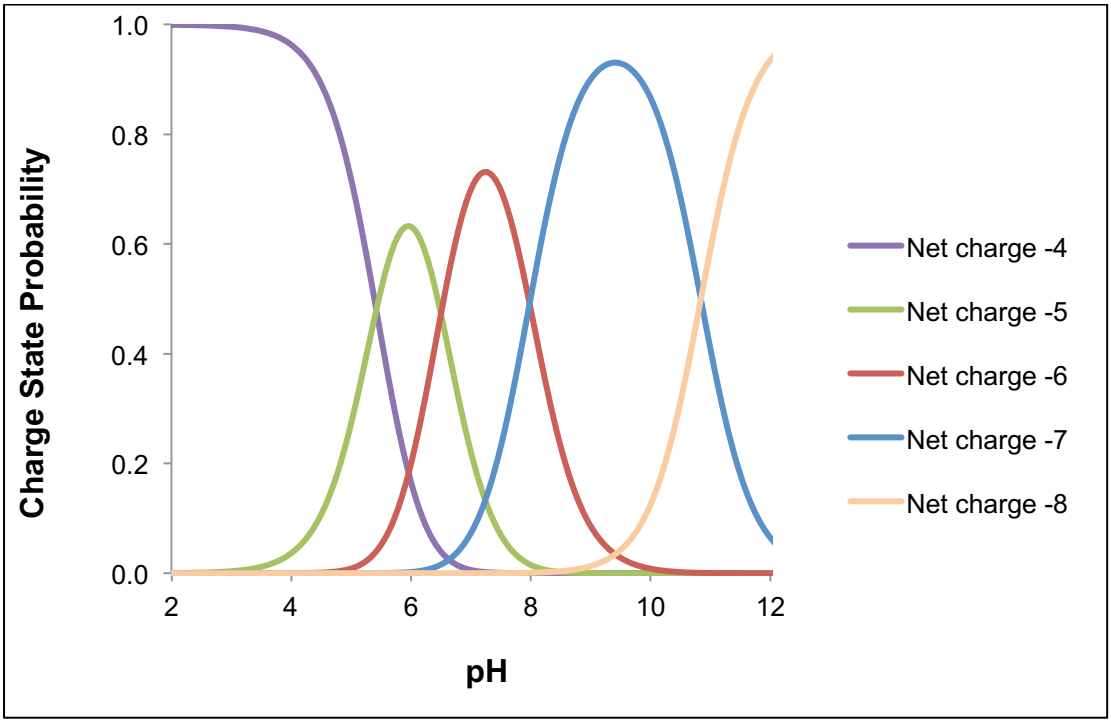


Figure 15: Variation of the charge state probabilities of Ins(1,3,4,5)P_4 as a function of pH, in 0.2 M KCl. Results from Borkovec and Spiess.[3]

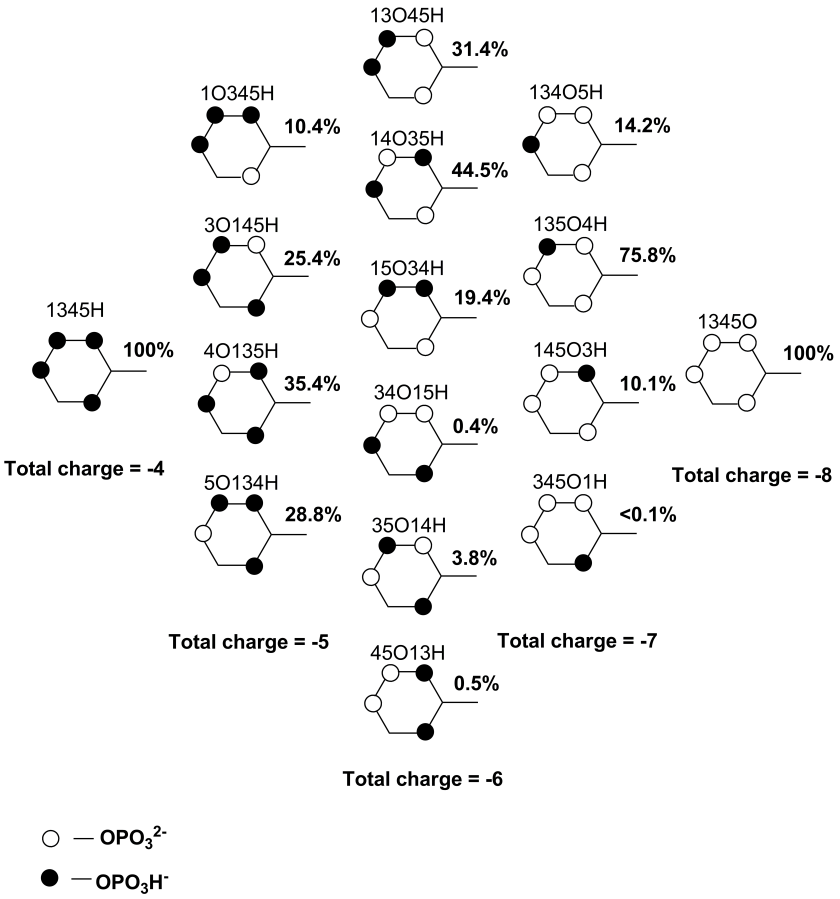


Figure 16: The sixteen Ins(1,3,4,5)P_4 isomers. Diagram adapted from Borkovec and Spiess.[3] The horizontal line indicates the axial hydroxyl group at position 2. Circles represent phosphate groups 1, 3, 4 and 5. Filled circles represent mono-protonated phosphate groups, and empty circles represent fully deprotonated phosphate groups. The name given to, net charge of, and *in vitro* probability of each isomer are provided. For this naming convention, all fully deprotonated phosphate groups are followed by the letter 'O' and mono-protonated phosphate groups followed by the letter 'H'.

4. 3 Parameterisation Protocol

Herein we have extended the Amber force field to study Ins(1,3,4,5)P_4. The parameterisation protocol that was used in this work is in keeping with the Amber ff99SB force field. A simplified view of this parameterisation protocol is illustrated in Figure 17.

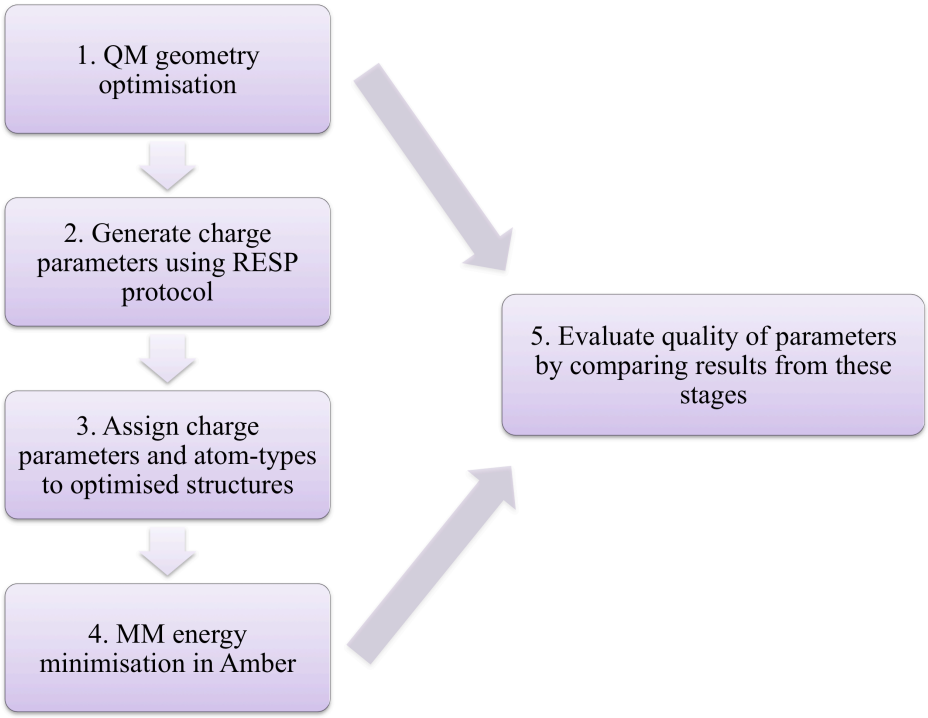


Figure 17: A schematic summary of the parameterisation protocol that was applied in this study.

As illustrated in Figure 17, QM calculations were used to optimise the geometry of each Ins(1,3,4,5)P_4 isomer. The molecular electrostatic potential for each optimised isomer was then calculated and subsequently fitted to atom centred points using the RESP protocol, thereby generating the charge parameters. Additional parameters for the bond, angle, torsion and vdW terms were assigned using pre-defined atom types.

The quality of the parameters was evaluated by comparing structural and energetic properties of the Ins(1,3,4,5)P_4 isomers from the QM geometry optimisation stage, from which the parameters were generated, to an MM energy minimisation stage, in which the parameters were applied. This allowed assessment of whether the point charge model and atom types were able to replicate Ins(1,3,4,5)P_4's QM behaviour, which acts as a benchmark.

Variations within this protocol were considered in order to decide the most appropriate method to obtain the parameters for this highly charged ligand. Different QM levels of theory (DFT and MP2) were investigated for the geometry optimisation of the Ins(1,3,4,5)P_4 isomers. Additionally atom types from two Amber force fields (ff99SB and GAFF) were considered. The calculations in this protocol were carried out in gas phase, as is customary for the Amber force field.

In addition to the calculations in gas phase, calculations of the Ins(1,3,4,5)P_4 isomers were also carried out in implicit solvent. These calculations allowed comparison and validation of the theoretical calculations with experimental data.

4. 4 Program Details

Gaussian 03[121] and Gaussian 09[122] were applied for the geometry optimisation calculations and to calculate the molecular electrostatic potential of the Ins(1,3,4,5)P_4 isomers. Graphical representations were produced with Gaussview.[123]

The Amber10[120] and Amber11[124] suite of programs, implementing the ff99SB force field (and GAFF extension), were used for the MM calculations. Parameter and topology files were created using the LEaP module. The RESP program was used to fit the QM calculated MEP to atom-centred points. Minimisation calculations were performed with Sander. Graphical representations were produced with VMD and Chimera.[125, 126]

4. 5 Initial Structure

The X-ray crystal structure of Ins(1,3,4,5)P_4 in a complex with the PKB PH domain (PDB accession code 1UNQ) provided the initial configuration for Ins(1,3,4,5)P_4.[11] In this crystal structure Ins(1,3,4,5)P_4 adopts a chair conformation, with all phosphate groups occupying equatorial positions. The hydroxyl group at the 6-position is also equatorial, whilst the 2-hydroxyl group is axial. This conformation is referred to as *myo*-Ins(1,3,4,5)P_4. The chair conformation of the cyclohexane ring of Ins(1,3,4,5)P_4 has been confirmed by NMR data.[127]

In the crystal structure, phosphate group 3 has a planar arrangement. It was therefore necessary to adapt phosphate group 3 towards a tetrahedral arrangement similar to that of the 1-, 4- and 5-phosphate groups. In order to produce the sixteen different Ins(1,3,4,5)P_4 isomers, hydrogen atoms were bonded to the four phosphate groups in the appropriate combinations.

4. 6 DFT Gas Phase Investigation

4. 6. 1 Protocol for Charge Derivation

The sixteen Ins(1,3,4,5)P_4 isomers were geometry optimised until they reached a local minimum on the potential energy surface. The isomers were first optimised using the DFT-B3LYP/3-21g* level of theory with Gaussian03, and then further optimised using DFT-B3LYP/6-31g**. These electronic structure calculations were carried out in gas phase. The DFT-B3LYP level of theory was initially chosen for this stage as it accounts for electron correlation at a significantly lower computational cost than post-HF electron correlation methods. Furthermore previous research suggests this level of theory is appropriate and comparable to post-HF methods for studying highly charged phosphorylated species.[128, 129]

The molecular electrostatic potential for each of the optimised isomers was calculated using the HF/6-31g* level of theory. The RESP algorithm, in Amber 10, was employed to fit the QM calculated MEP to atom-centred points. As there are no

degenerate hydrogen atoms in Ins(1,3,4,5)P_4, a one stage RESP protocol was applied.

The charge parameters obtained, and atom types assigned, for the Ins(1,3,4,5)P_4 isomers are provided in section S.3 of the Electronic Supplementary Information. Parameters for the -4 Ins(1,3,4,5)P_4 are illustrated in Figure 18 and Table 1.

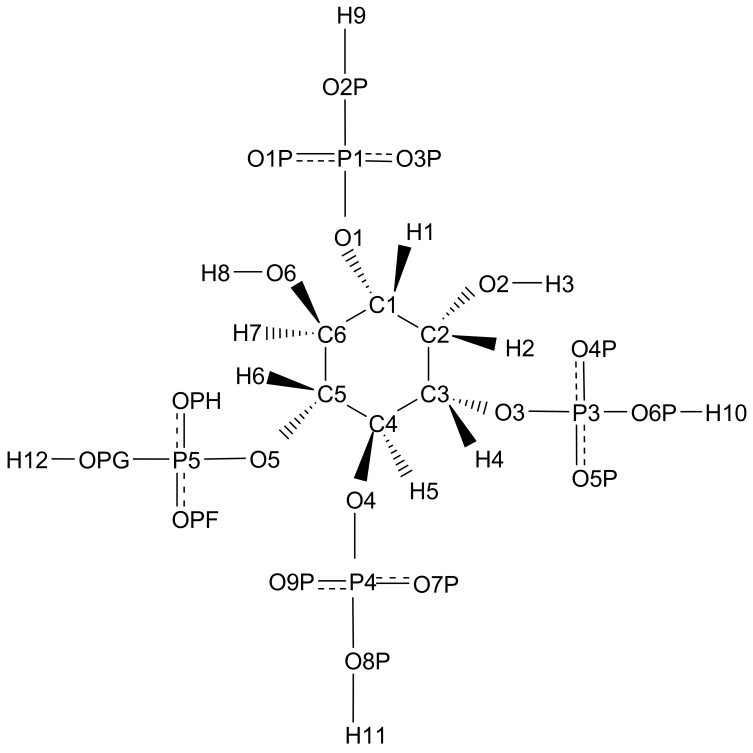


Figure 18: Atomic labelling scheme for the -4 isomer, 1345H of Ins(1,3,4,5)P_4.

Table 1: Assigned atom types and charge parameters for the -4 1345H isomer of Ins(1,3,4,5)P_4.

Atom	ff99SB atom type	Charge parameters	Atom	ff99SB atom type	Charge parameters
C1	CT	0.024812	P4	P	1.241543
O1	OS	-0.467569	O7P	O2	-0.832113
C2	CT	0.024704	O8P	OH	-0.727447
O2	OH	-0.582613	O9P	O2	-0.804811
C3	CT	0.042999	P5	P	1.208422
O3	OS	-0.503174	OPF	O2	-0.829258
C4	CT	0.020635	OPG	OH	-0.702063
O4	OS	-0.473885	OPH	O2	-0.775938
C5	CT	0.049358	H1	HC	0.122736
O5	OS	-0.448672	H2	HC	0.127450
C6	CT	0.077708	H3	HO	0.419241
O6	OH	-0.585574	H4	HC	0.155901
P1	P	1.263615	H5	HC	0.168002
O1P	O2	-0.824174	H6	HC	0.132010
O2P	OH	-0.708093	H7	HC	0.120269
O3P	O2	-0.842697	H8	HO	0.387273
P3	P	1.287731	H9	HY	0.381048
O4P	O2	-0.830816	H10	HY	0.484520
O5P	O2	-0.853312	H11	HY	0.444755
O6P	OH	-0.769119	H12	HY	0.376596

Atom types were assigned from the ff99SB force field by analogy to nucleotides, and an MD study of PtdIns(4,5)P_2.[62, 91, 93] Additional angle parameters were required for atom type HC and were assigned values from the literature, Table 2.[130]

Table 2: Angle parameters for atom type HC. K_θ and θ_{eq} represent the angular force constant (in kcal mol^{-1} radian^{-2}) and the equilibrium angle value (in °) respectively.

Angle Parameter	K_θ	θ_{eq}
HC-CT-OH	35.0	109.5
HC-CT-OS	35.0	109.5

4. 6. 1. 1 Atom Type HY

A new atom type was developed to overcome issues with the hydrogen atoms of protonated phosphate groups. During energy minimisation calculations in Amber these hydrogen atoms, when assigned atom type HO,[91] were shown to overlap with an oxygen atom of the same phosphate group. This effect is due to a strong electrostatic attraction between the large negative charge on the oxygen atom and positive charge on the hydrogen atom when in close proximity. As atom type HO has no vdW radius, the angle term is the only contribution to oppose this attraction, with the HO-OH-P and OH-P-O2 terms being only 45 kcal mol^{-1} radian^{-2}.[91, 131] This problem is well documented on the Amber mailing list.[132]

A new atom type HY was therefore developed with the same bond, angle and torsion parameters as atom type HO, but with additional vdW parameters. The vdW parameters were assigned by analogy to other hydrogen atom types in the ff99SB force field, with a vdW radius of 1.000 Å and a well depth of 0.0157 kcal mol^{-1}, as for a hydrogen bonded to a sulphur atom, HS.[91]

4. 6. 2 Validation of Parameters

The quality of the parameters was evaluated by applying them to the Ins(1,3,4,5)P_4 isomers in an MM energy minimisation stage.

4. 6. 2. 1 MM Energy Minimisation

The Ins(1,3,4,5)P_4 isomers, with parameters assigned, were minimised by 1500 cycles of steepest descent followed by 2500 cycles of conjugate gradient minimisation using Sander in Amber10. The isomers were then further minimised using the Newton-Raphson method in Amber until a root-mean-square gradient of $1x10^{-9}$ kcal mol^{-1} $Å^{-1}$ was reached. These minimisation calculations were carried out in gas phase. For all cycles, the Ins(1,3,4,5)P_4 isomers were minimised using an 999 Å cut-off.

4. 6. 3 Results and Discussion

In the following sections the quality of the parameters generated for the Ins(1,3,4,5)P_4 isomers is evaluated by comparing the structural and energetic properties of the isomers from the QM geometry optimisation stage and the MM energy minimisation stage.

4. 6. 3. 1 Energetic Properties

The Ins(1,3,4,5)P_4 isomers were ranked, within each charge subset, according to their energy from the QM geometry optimisation stage and MM energy minimisation stage, Table 3.

Table 3: A comparison of the energetic ordering of the Ins(1,3,4,5)P_4 isomers from the QM geometry optimisation stage and MM energy minimisation stage. The isomers are ranked according to their energy within the charge subsets. The lowest energy isomers of each charge subset are given the ranking 1, and the remaining isomers ranked accordingly. Energy differences to the lowest energy isomer of each subset are stated. Absolute energies for the lowest energy isomers of each subset are provided in section S.1 of the Electronic Supplementary Information.

Charge	Isomer	QM geometry optimised energetic ordering	QM energy difference in kcal mol^{-1}	MM energy minimised energetic ordering	MM energy difference in kcal mol^{-1}
-4	1345H	1	0.00	1	0.00
-5	5O134H	1	0.00	1	0.00
	1O345H	2	1.03	2	0.87
	4O135H	3	10.59	3	4.46
	3O145H	4	10.74	4	6.78
-6	13O45H	1	0.00	1	0.00
	15O34H	2	2.44	2	0.36
	45O13H	3	23.50	3	2.95
	34O15H	4	27.69	4	7.44
	14O35H	5	28.43	5	7.46
	35O14H	6	35.31	6	9.54
-7	135O4H	1	0.00	2	0.43
	145O3H	2	9.95	3	3.18
	134O5H	3	12.30	1	0.00
	345O1H	4	21.51	4	5.93
-8	1345O	1	0.00	1	0.00

A comparison of the energetic ordering of the Ins(1,3,4,5)P_4 isomers from the QM and MM stages shows them to be in very good agreement. For the -5 charge subset, the energetic ordering is the same for both stages. Furthermore the calculated energy differences, although not identical, show similar patterns in magnitude: The energy difference between the lowest energy isomer, 5O134H, and 1O345H is extremely small (approximately 1 kcal mol^{-1}). The 3O145H and 4O135H isomers are then higher in energy, and of a similar order to each other.

The -6 subset also has identical energetic orderings for the two stages. Again similar patterns in the relative energies for this subset are apparent: The 13O45H isomer is the lowest in energy for this subset, with the 15O34H ranked close 2nd. The 45O13H isomer is then ranked 3rd, followed by the 34O15H and 14O35H isomers, which are approximately 5 kcal mol^{-1} higher in energy.

Unfortunately there is a slight discrepancy between the results for the -7 subset of isomers. The 134O5H isomer, which is lowest in energy in the MM calculations, is ranked 3rd lowest in the QM calculations. Analysis of the structures does not allude to a reason for this difference in ordering. However for a small system with this extremely high charge, a 3 kcal mol^{-1} energy difference between the three lowest energy isomers in the MM calculations is small and thus some interchanging in the ordering could be expected here.

4. 6. 3. 2 Structural Properties

The structures of the geometry optimised Ins(1,3,4,5)P_4 isomers were compared against those following energy minimisation in Amber. On visual inspection the Ins(1,3,4,5)P_4 isomers displayed no significant structural rearrangements between these two stages. Further analysis supported this, showing that the bond, angle and dihedral parameters of the isomers were generally within 1–2, 5 and 10% respectively for both stages.

In this gas phase investigation, the lower energy isomers uniformly adopted chair-like conformations. However, it should be noted that the three highest energy isomers of the -6 and -7 subsets optimised/minimised to conformations with some boat-like character. It is the large electrostatic repulsion between the highly charged phosphate groups that triggers these distortions, allowing the neighbouring phosphate groups to increase the distance between one another, Figure 19.

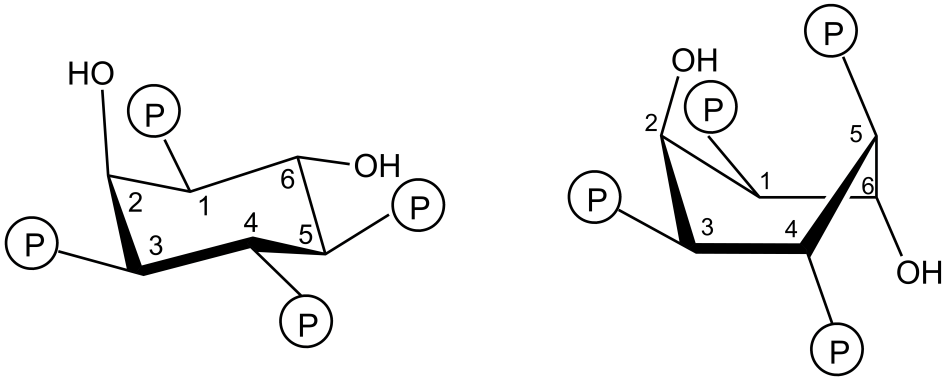


Figure 19: Inositol 1,3,4,5-tetrakisphosphate in the chair and boat conformation.

4. 7 Comparison of DFT and MP2 Levels of Theory

The MP2 level of theory was also considered for the geometry optimisation of the Ins(1,3,4,5)P_4 isomers. In this section, results using the MP2 level of theory are evaluated against those for the DFT-B3LYP method in section 4.6.

4. 7. 1 Protocol for Geometry Optimisation

The Ins(1,3,4,5)P_4 isomers were geometry optimised using the MP2/3-21g* level of theory, followed by MP2/6-31g**, with Gaussian03. These electronic structure calculations were carried out in gas phase. The energetic ordering and structural conformations of the Ins(1,3,4,5)P_4 isomers from the DFT and MP2 optimisations were then compared.

4. 7. 2 Results and Discussion

4. 7. 2. 1 Energetic Properties

Interestingly, the energetic orderings of the Ins(1,3,4,5)P_4 isomers for the DFT and MP2 methods are generally shown to be in fairly close agreement, Table 4.

Table 4: A comparison of the energetic ordering of the Ins(1,3,4,5)P_4 isomers optimised using the DFT and MP2 levels of theory, in gas phase. The isomers are ranked according to their energy within the charge subsets. The lowest energy isomers of each charge subset are given the ranking 1, and the remaining isomers ranked accordingly. The energy differences to the lowest energy isomer of each subset are stated. Absolute energies for the lowest energy isomers of each subset are given in section S.1 of the Electronic Supplementary Information.

Charge	Isomer	QM geometry optimised energetic ordering for the DFT method	QM energy difference for the DFT method in kcal mol^{-1}	QM geometry optimised energetic ordering for the MP2 method	QM energy difference for the MP2 method in kcal mol^{-1}
-4	1345H	1	0.00	1	0.00
-5	5O134H	1	0.00	2	1.69
	1O345H	2	1.03	1	0.00
	4O135H	3	10.59	4	26.38
	3O145H	4	10.74	3	7.09
-6	13O45H	1	0.00	1	0.00
	15O34H	2	2.44	2	2.28
	45O13H	3	23.50	4	28.59
	34O15H	4	27.69	5	30.51
	14O35H	5	28.43	3	9.96
	35O14H	6	35.31	6	32.52
-7	135O4H	1	0.00	1	0.00
	145O3H	2	9.95	2	0.71
	134O5H	3	12.30	4	7.11
	345O1H	4	21.51	3	3.56
-8	1345O	1	0.00	1	0.00

For the -5 charge subset, the two isomers lowest in energy are switched in order for the optimisation methods. This is also true for the two highest energy isomers. However, the energy difference between these reversed pairs is negligible according to the DFT data. Interestingly, a more significant energy difference between the two highest energy isomers is apparent for the MP2 method. This difference in energy can be rationalised by studying the optimised structures of the 4O135H isomer. When optimised using the MP2 level of theory, the 4O135H isomer adopts a more chair-like

conformation in comparison to the DFT optimised isomer. This slight difference in the conformation of the inositol ring, results in a significant change to the position of the phosphate groups attached, and thereby its energy.

For the -6 subset, the only significant difference for the optimisation methods is the behaviour of the 14O35H isomer. For both methods the 14O35H isomer adopts a boat-like conformation, forming intra-molecular hydrogen bonds to offset the torsional strains associated with the strong electrostatic repulsion of the highly charged phosphate groups. However the DFT optimised isomer forms a chain of H bonds between 3-P-OH···2-O-H···5-O-P (Figure 20a) that is reversed for the MP2 optimised structure, forming interactions between 5-P-OH···2-O-H···3-O-P, (Figure 20b). This difference in structure and hydrogen bonding is the reason for the different energetic ranking of this isomer.

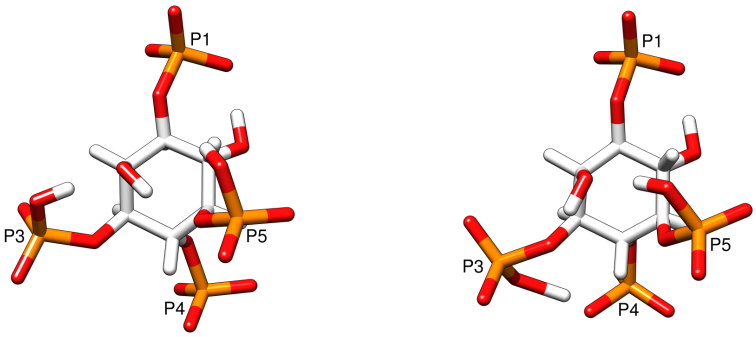


Figure 20: The 14O35H isomer of Ins(1,3,4,5)P_4 optimised using the DFT (a) and MP2 (b) level of theory. Phosphate groups 1, 3, 4 and 5 are labelled P1, P3, P4 and P5 respectively.

4. 7. 2. 2 Structural Properties

On visual inspection the optimised isomers displayed no large conformational differences for the DFT-B3LYP and MP2 methods, other than those already mentioned. Quantitative analysis showed that the bond, angle and dihedral parameters for the optimised structures are in very close agreement (generally within 1, 2-3 and 5% respectively) for both methods.

Additionally charge parameters obtained from both sets of optimised structures were in very close agreement, with charges for the phosphate groups generally varying by only 0-5% between both methods.

4. 8 Comparison of ff99SB and GAFF Force Fields

In a separate adaptation to the protocol, atom types from the GAFF force field were used as a comparison to the ff99SB atom types to determine whether they would be suitable for modelling the Ins(1,3,4,5)P_4 isomers. In previous investigations, both GAFF and ff99SB atom types have been applied to study PtdIns(4,5)P_2.[62, 133]

Atom types from the ff99SB force field are typically used to describe protein and nucleic acid systems, rather than small organic compounds. The GAFF force field was developed as a more general force field, to provide a broader set of parameters that can easily be transferred, covering most of organic chemical space.[92] The GAFF force field is compatible with ff99SB, facilitating the investigation of interactions between proteins and organic compounds.

4. 8. 1 Protocol for Parameter Assignment

GAFF atom types were assigned to the DFT optimised Ins(1,3,4,5)P_4 isomers from section 4.6. Atom types were assigned using the Antechamber program of Amber10. The initial atom types assigned by Antechamber resulted in a similar issue to that experienced with the ff99SB force field, with the hydrogen (atom type ho) and oxygen atoms of a protonated phosphate group overlapping during the energy

minimisation stage. Atom type hy was thus developed, with the same bond, angle and torsion parameters as atom type ho, but with a vdW radius of 1.000 Å and a well depth of 0.0157 kcal mol^{-1}. Atom type h1,[92] assigned by Antechamber for hydrogen atoms bonded to the inositol ring, was amended to hc, analogous to atom type HC for the ff99SB force field. The same RESP derived charge parameters were applied as in section 4.6.1.

The Ins(1,3,4,5)P_4 isomers, with parameters assigned, were minimised by 1500 cycles of steepest descent followed by 2500 cycles of conjugate gradient minimisation using Sander in Amber10. The isomers were then further minimised using the Newton-Raphson method in Amber until a root-mean-square gradient of $1x10^{-9}$ kcal mol^{-1} was reached. These minimisation calculations were carried out in gas phase. For all cycles, the Ins(1,3,4,5)P_4 isomers were minimised using an 999 Å cut-off.

4. 8. 2 Results and Discussion

The suitability of GAFF atom types for modelling Ins(1,3,4,5)P_4 is assessed in this section. Properties of the Ins(1,3,4,5)P_4 isomers are compared for the QM geometry optimisation stage (using the DFT-B3LYP method), and the MM energy minimisation stage with GAFF atom types.

4. 8. 2. 1 Energetic Properties

The energetic ordering of the DFT optimised isomers is in moderate agreement with that of the MM energy minimised isomers when assigned GAFF atom types, Table 5. The -6 subset shows promising results, with the three highest energy isomers being identical and the three lowest energy isomers being consistent although reordered within.

However, results for the -5 and -7 charge subsets are less convincing. The three lowest energy isomers from the -5 subset are re-ordered. Whilst the energy difference between these isomers is less than 0.5 kcal mol^{-1} according to the MM results, it is significantly larger according to the QM data (Table 5). Additionally, the -7 isomers

are also reordered with those ranked 2nd and 3rd switched.

Although the Amber model of the Ins(1,3,4,5)P_4 isomers is in moderate agreement with the QM data when GAFF atom types are assigned, results were significantly better with the ff99SB atom types (Table 3). The assignment of GAFF atom types not only alters the energetic ordering between the QM and MM stages, but additionally upsets the energetic differences between isomers within a subset. The significant difference between the results for GAFF and ff99SB atom types is unsurprising as supposedly equivalent parameters differ significantly in these force fields.[91, 92]

Table 5: A comparison of the energetic ordering of the Ins(1,3,4,5)P_4 isomers between the QM geometry optimisation stage and the MM energy minimisation stage using GAFF atom types. The isomers are ranked according to their energy within the charge subsets. The lowest energy isomers of each charge subset are given the ranking 1, and the remaining isomers ranked accordingly. The energy differences to the lowest energy isomer of each subset are stated. Absolute energies for the lowest energy isomers of each subset are given in section S.1 of the Electronic Supplementary Information.

Charge	Isomer	QM geometry optimised energetic ordering	QM energy difference in kcal mol^{-1}	MM energy minimised energetic ordering using GAFF atom types	MM energy difference using GAFF atom types in kcal mol^{-1}
-4	1345H	1	0.00	1	0.00
-5	5O134H	1	0.00	3	0.21
	1O345H	2	1.03	2	0.16
	4O135H	3	10.59	1	0.00
	3O145H	4	10.74	4	5.07
-6	13O45H	1	0.00	2	3.58
	15O34H	2	2.44	3	4.74
	45O13H	3	23.50	1	0.00
	34O15H	4	27.69	4	11.11
	14O35H	5	28.43	5	13.52
	35O14H	6	35.31	6	14.55
-7	135O4H	1	0.00	1	0.00
	145O3H	2	9.95	3	8.50
	134O5H	3	12.30	2	4.68
	345O1H	4	21.51	4	10.43
-8	1345O	1	0.00	1	0.00

4. 9 Calculations of Ins(1,3,4,5)P_4 in Implicit Solvent

In addition to the gas phase calculations in sections 4.6 to 4.8, calculations for the Ins(1,3,4,5)P_4 isomers were also performed in implicit solvent. These calculations allow comparison and validation of the theoretical calculations with *in vitro* data.

A consequence of including solvent in the model is that electrostatic interactions between the solute [Ins(1,3,4,5)P_4] and the solvent are taken into account. As a result intra-molecular electrostatic interactions for the solute are dampened, which could prove important when investigating the structural/energetic properties of such a small highly charged system. The solvent used in this study is water.

4. 9. 1 Implicit Solvation Models

Implicit solvation models represent solvent as a continuous medium, whereby dielectric parameters in the surrounding environment of a solute replicate those of a solvent. There are numerous implicit solvent models available, which can differ in their description of the solute, the solvent and the interface between them.[134] In this work the polarisable continuum model[135, 136] in Gaussian09 was applied for the QM geometry optimisation calculations and the generalised Born model in Amber for the MM energy minimisation calculations.[137-139]

4. 9. 1. 1 The PCM Method

The Polarisable Continuum Model (PCM) is frequently applied in QM calculations. In the PCM model the solute cavity is obtained from the overlapping vdW radii of its atoms. The polarisation of the dielectric medium is represented by an apparent surface charge that resides on this solute cavity.[140]

The vdW spheres that contribute to the cavity surface are divided into a number of small surface elements, with area ΔS. An initial charge for each surface element is calculated using the electric field gradient due to the solute:[141]

$$q_i = -\left(\frac{\varepsilon - 1}{4\pi\varepsilon}\right) E_i \Delta S$$

Equation 21: Calculation of the initial charge for each surface element. E_i is the electric field gradient at point i and ε is the dielectric constant of the medium.

The mutual polarisation of the surface charges is then calculated, resulting in an additional contribution to the electric field gradient.[141] The charges are modified iteratively until self-consistent. The potential due to the final set of surface charges $\phi_\sigma(r)$ is then added to the solute Hamiltonian as a perturbation, $H = H_0 + \phi_\sigma(r)$. A self-consistent field calculation is then initiated, which results in new values for the surface charges from the current wavefunction. This calculation is carried out iteratively until the surface charges (and solute wavefunction) are self-consistent.

4. 9. 1. 2 The Generalised Born Method

The generalised Born (GB) implicit solvent model was applied in the MM calculations. The GB model approximates the electrostatic contribution to the solvation free energy using the equation:

$$\Delta G_{el} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right) \sum_{i,j} \frac{q_i q_j}{f(r_{ij}, a_{ij})}$$

Equation 22: The Generalised Born equation for calculating the electrostatic solvation free energy. ε is the solvent dielectric, q_i and q_j are the point charges on atoms i and j respectively. $f(r_{ij}, a_{ij})$ is a function that depends upon the interparticle distances r_{ij} and the effective Born radii a_{ij}.

The effective Born radius of an atom reflects its degree of burial inside a molecule. It is officially defined as "the radius that would return the electrostatic energy of the system according to the Born equation if all other atoms in the molecule were uncharged".[72] The quality of the GB model is considered strongly dependent on the quality of the estimated effective Born radii.[112, 113]

4. 9. 2 Protocol for Charge Derivation

The sixteen Ins(1,3,4,5)P_4 isomers were geometry optimised in a solvent reaction field using the PCM model with the integral equation formalism (IEFPCM), in Gaussian09.[122, 135] Water was assigned as the solvent. As previous calculations in sections 4.6 and 4.7 suggested DFT-B3LYP was suitable for studying this system, this method was applied. The isomers were first optimised using the DFT-B3LYP/3-21g* level of theory, followed by a second optimisation with DFT-B3LYP/6-31g**. The MEP for each of the optimised isomers was calculated using the HF/6-31g* level of theory in gas phase, with Gaussian 09. A one stage RESP protocol was then employed to derive the atom-centred point charges. The RESP generated charge parameters for the solvent optimised Ins(1,3,4,5)P_4 isomers are illustrated in section S.3 of the Electronic Supplementary Information. The same atom types from the ff99SB force field were assigned as in the gas phase calculations in section 4.6.

4. 9. 3 Validation of Parameters

4. 9. 3. 1 Quantum Mechanical *versus* Molecular Mechanical Properties

The parameters (in section S.3 of the Electronic Supplementary Information) were applied to the solvent optimised Ins(1,3,4,5)P_4 structures for an energy minimisation stage in Amber. The Ins(1,3,4,5)P_4 isomers were minimised by 1500 cycles of steepest descent followed by 2000 cycles of conjugate gradient minimisation, in implicit solvent. The Hawkins, Cramer, Truhlar pairwise GB model[137, 138] with parameters described by Tsui and Case[139] (igb=1) was used as the solvent model. For all cycles, a 999 Å cut-off was applied. The quality of the parameters was initially assessed by comparing the energetic properties of the Ins(1,3,4,5)P_4 isomers between the QM geometry optimisation stage and the MM energy minimisation stage, both in implicit solvent.

4. 9. 3. 2 Theory *versus* Experiment

The results from the calculations in implicit solvent were also validated by comparison to experimental data. Borkovec *et al.* have previously determined the *in vitro* probabilities of the different Ins(1,3,4,5)P_4 isomers within each charge subset, using ^{31}P-NMR titration data, Figure 16.[3] The energetic ranking of the Ins(1,3,4,5)P_4 isomers from our calculations is assessed against these experimental probabilities.

4. 9. 4 Results and Discussion

4. 9. 4. 1 Energetic Properties

The calculations in implicit solvent were analysed by comparing the energetic ordering of the QM optimised and MM energy minimised Ins(1,3,4,5)$_4$ isomers, (columns 1 to 6 of Table 6).

The energetic ordering for the QM and MM stages are generally shown to be in good agreement. Exception to this is the -5 subset of isomers, for which the ordering is rearranged between the two stages. However, the small energy difference between these isomers (according to the QM calculations) provides some justification for this. The -6 subset of isomers shows an appreciably stronger agreement in the energetic ordering. The only discrepancies are reversals in the ordering of the isomers ranked 2nd and 3rd, and 5th and 6th in energy. These reversals are considered minor, as the differences in energy for the isomers concerned are relatively small (less than 1 and 4 kcal mol^{-1} respectively, according to the QM calculations). Additionally the energetic orderings for the -7 subset of isomers are in full agreement for both stages.

Although there are some discrepancies in the energetic ordering of the Ins(1,3,4,5)P_4 isomers from the QM and MM stages in implicit solvent, a good agreement is still apparent. These discrepancies can largely be attributed to the differences in modelling implicit solvent for both stages.

Table 6: A comparison between the energetic ordering of the Ins(1,3,4,5)P_4 isomers geometry optimised and energy minimised in implicit solvent with their *in vitro* probabilities. The isomers are ranked according to their energy/probability within the charge subsets. The lowest energy and most probable isomers of each charge subset were given the ranking 1, and the remaining isomers ranked accordingly. For the geometry optimisation and energy minimisation calculations, energy differences to the lowest energy isomer of each subset are stated. Absolute energies for the lowest energy isomers of each subset are given in section S.1 of the Electronic Supplementary Information.

Charge	Isomer	QM geometry optimised energetic ordering	QM energy difference in kcal mol^{-1}	MM energy minimised energetic ordering	MM energy difference in kcal mol^{-1}	*In vitro* probability ordering	*In vitro* probability in %[3]
-4	1345H	1	0.00	1	0.00	1	100
-5	4O135H	1	0.00	3	23.90	1	35.4
	5O134H	2	1.16	1	0.00	2	28.8
	3O145H	3	3.30	4	30.57	3	25.4
	1O345H	4	4.09	2	8.96	4	10.4
-6	14O35H	1	0.00	1	0.00	1	44.5
	15O34H	2	2.59	3	13.02	3	19.4
	13O45H	3	3.47	2	10.85	2	31.4
	35O14H	4	4.92	4	22.36	4	3.8
	34O15H	5	15.68	6	74.69	6	0.4
	45O13H	6	19.33	5	32.48	5	0.5
-7	135O4H	1	0.00	1	0.00	1	75.8
	134O5H	2	3.31	2	27.58	2	14.2
	145O3H	3	6.76	3	32.84	3	10.1
	345O1H	4	18.73	4	57.35	4	<0.1
-8	1345O	1	0.00	1	0.00	1	100

4. 9. 4. 2 Structural Properties

The sixteen solvent optimised isomers uniformly adopted chair-like conformations. In contrast, when optimised in gas phase using the same level of theory the higher energy isomers of the -6, -7 and -8 subsets displayed boat-like character. Furthermore, when optimised in the presence of solvent the Ins(1,3,4,5)P_4 isomers tended to form fewer intra-molecular hydrogen bonds than those in gas phase. These structural differences between the gas phase and solvent calculations inherently altered the energetic ordering of the Ins(1,3,4,5)P_4 isomers (see Tables 3 and 6). The pronounced difference in results from the gas phase and solvent calculations shows that the geometry and energy of the Ins(1,3,4,5)P_4 isomers are particularly dependent on their environment. These differences allow/force us to consider whether optimisation in solvent should be considered for parameterisation of highly charged ligands, as opposed to the typical gas phase protocol.

4. 9. 4. 3 Theory *versus* Experiment

The results from the calculations in implicit solvent were also validated by comparison to experimental data. Borkovec *et al.* have previously determined the *in vitro* probabilities of the different Ins(1,3,4,5)P_4 isomers within each charge subset, using ^{31}P-NMR titration data.[3]

The Ins(1,3,4,5)P_4 isomers were ordered according to their experimental probability. This ordering was then compared against the isomers' energetic ordering from the computational calculations in solvent (Table 6). The lowest energy and most probable isomers of each charge subset were given the ranking 1, and the remaining isomers ranked accordingly.

There is a strong agreement between the theoretical calculations and experimental data, with the lower energy Ins(1,3,4,5)P_4 isomers corresponding to those with higher probability. For the -5 and -7 charge subsets, the energetic ordering from the QM calculations fully agrees with the experimental probability ordering. There is also a strong agreement between the QM calculations and the experimental data for the -6 subset. A minor discrepancy for this subset is the reversal of the 2nd and 3rd lowest

energy/most probable isomers, 13O45H and 15O34H. However these isomers have less than a kcal mol^{-1} difference in energy according to the QM calculations. Furthermore, this inconsistency may be explained by an assumption made by Borkovec *et al.* in which the interaction between the 1- and 5- phosphate groups was fixed to a value obtained for inositol 1,2,3,5-tetrakisphosphate.[3] The additional reversal in the ordering of the least probable/highest energy isomers of this subset (34O15H and 45O13H) is considered insignificant as these have only a 0.1% difference in probability and a small energetic difference according to the QM calculations. The energetic ordering of the -6 charge subset according to the MM energy minimisation calculations fully agrees with the *in vitro* probability ordering.

The energetic/probability ordering of the Ins(1,3,4,5)P_4 isomers can largely be explained by the unfavourable electrostatic interactions of the highly charged, fully deprotonated phosphate groups (also referred to as phosphodianions). Additionally hydrogen bonding involving the mono-protonated phosphate groups also plays a critical role in stabilising Ins(1,3,4,5)P_4.

The isomers of the -5 charge subset are all shown to be in appreciable proportions in the *in vitro* data, and have similar energies according to the QM calculations. The fact that the 1O345H isomer is the least probable/has the highest energy of this subset can be explained by the isolation of the 1-position phosphodianion from the 3,4,5-phosphate motif. This isolation reduces the phosphodianion's ability to stabilise its high charge through hydrogen bonding with neighbouring phosphate groups. Conversely, the 4O135H isomer is the most probable of this subset, having two neighbouring phosphate groups to interact with.

Isomers of the -6 charge subset, with two fully deprotonated phosphate groups, vary significantly in probability and energy. These larger variations are thought to be mainly due to the unfavourable electrostatic interactions of the two phosphodianions. The 14O35H isomer is the most probable, comprising a 44.5% proportion of the -6 charge subset.[3] This isomer positions the two phosphodianions furthest apart across the inositol ring, minimising their interaction. The next most energetically stable/probable isomers are those with the phosphodianions separated by one position around the inositol ring (13O45H, 15O34H and 35O14H). The 35O14H isomer is

notably the least favourable of the three, due to both phosphodianions communicating *via* the central 4-phosphate group. The 34O15H and 45O13H isomers, with adjacent phosphodianions, are unsurprisingly the least stable/probable of the -6 subset.

Finally for the -7 subset, the 135O4H isomer is by far the most favourable/probable as the phosphodianions are all separated by one position around the inositol ring. The 134O5H and 145O3H isomers are then next, as these each have one pair of neighbouring phosphodianions. Lastly the 345O1H isomer has an extremely low probability (of less than 0.1%) and is very energetically unfavourable, due to the three neighbouring phosphodianions at the 3-, 4- and 5-positions.

The electrostatic basis for the ordering of the -6 and -7 isomers is further supported by examining the electrostatic contribution to the energy of these isomers (from the MM energy minimisation calculations), Table 7. A clear trend of increasing positive (unfavourable) electrostatic contributions to the total energy of these isomers with decreasing experimental probability is apparent.

Table 7: Analysis of the electrostatic contributions to the total energies of the -6 and -7 Ins(1,3,4,5)P_4 isomers. Electrostatic energies are provided from the MM energy minimised isomers in implicit solvent. Experimental probabilities of these isomers are provided to illustrate the trend.

Charge	Isomer	Electrostatic contribution to total computed energy[a] in kcal mol^{-1}	*In vitro* probability[b] in %
-6	14O35H	250.87	44.5
	13O45H	263.01	32.4
	15O34H	285.45	19.4
	35O14H	308.21	3.8
	45O13H	324.21	0.5
	34O15H	366.30	0.4
-7	135O4H	596.88	75.8
	134O5H	629.03	14.2
	145O3H	638.20	10.1
	345O1H	710.42	<0.1

[a] Electrostatic energies following energy minimisation in implicit solvent using Amber.

[b] Probabilities determined from ^{31}P-NMR titration data.[3]

4. 10 Conclusion

This chapter explores the parameterisation of the sixteen Ins(1,3,4,5)P_4 isomers. Parameters for these isomers were generated using a gas phase protocol analogous to that for the existing molecules in the ff99SB force field.[91]

Charge parameters generated from the DFT-B3LYP optimised isomers, combined with ff99SB atom types, proved to be a suitable option for modelling this highly charged system. The similar structural and energetic properties of the Ins(1,3,4,5)P_4 isomers from the QM optimisation stage (from which the parameters were generated) and the MM energy minimisation stage (in which the parameters were applied) suggests good quality parameters have been generated/assigned.

As an adaptation to this protocol, the computationally expensive MP2 level of theory was investigated for the geometry optimisation of the Ins(1,3,4,5)P_4 isomers. Despite its additional cost, the MP2 level of theory was shown to alter the structural and energetic results very little from the DFT-B3LYP results. Furthermore, the charge parameters obtained from the DFT and MP2 optimised isomers were in close agreement, particularly for the phosphate groups. The DFT level of theory was therefore confirmed as an efficient and suitable method for optimising Ins(1,3,4,5)P_4.

As a separate adaptation to the protocol, atom types from the GAFF force field were also assessed. Application of these atom types, in place of ff99SB atom types, in the energy minimisation stage resulted in a significantly poorer agreement with the QM DFT-B3LYP data. This is unsurprising as GAFF parameters are designed to be compatible with simple organic structures. In comparison, Ins(1,3,4,5)P_4 has a very complex and distinct structure, with highly charged vicinal phosphate groups. Therefore general atom types from the GAFF force field are not suitable to model this system.

Calculations for the Ins(1,3,4,5)P_4 isomers were also carried out in implicit solvent. These calculations provided a comparison to experimental data in solution. A remarkably good agreement between the more energetically favourable isomers from

the calculations and the more probable isomers according to experimental data is apparent. This validation against experimental data imparts a high degree of confidence in the computational calculations.

The calculations in implicit solvent were of additional interest due to the difference in the optimised structures for the solvent and gas phase calculations. As mentioned, the Ins(1,3,4,5)P_4 isomers all adopted chair like conformations in the implicit solvent calculations. In contrast, the higher charged, less energetically favourable isomers tended to contain some boat-like character in the gas phase calculations. NMR spectra show Ins(1,3,4,5)P_4 adopts a chair conformation in solution, further supported by the chair conformation of this inositol phosphate in its crystal structure with the PH domain of PKBα.[11, 127] This therefore led to the consideration of whether generating charge parameters for the less 'sensible' boat-like structures from the gas phase calculations was appropriate.

The difficulty in deciding an appropriate parameterisation protocol is that there is no absolute rule as to how this should be carried out. Traditionally, the ff99SB force field uses gas phase calculations to derive/generate the charge parameters. This is because the HF/6-31g* level of theory used to calculate the electrostatic potential tends to overestimate the dipole moment of most residues by 10-20%, incorporating the amount of polarisation expected in an aqueous solution.[94, 97] The concern of solvent optimisation is thus that solvent effects are over-accounted for.

Interestingly the parameters from the solvent optimised and gas phase optimised isomers do not differ as significantly as one may expect. Nevertheless, it is unknown how a small change in the charge parameters may alter Ins(1,3,4,5)P_4's behaviour and interactions. Therefore as a preliminary test we ran separate MD simulations of the Ins(1,3,4,5)P_4 isomers, with parameters from the gas phase and implicit solvent calculations, in a complex with the PKB PH domain. Results showed that the free energies of binding and hydrogen bonding patterns were remarkably similar for both sets of parameters, Table S2 in the Electronic Supplementary Information.

In this chapter numerous approaches to generate parameters for the Ins(1,3,4,5)P_4 isomers have been considered. Results show that the charge parameters generated

from the DFT-B3LYP gas phase optimisation, and ff99SB atom types, are suitable for modelling this system. This gas phase protocol is in-keeping with that for existing molecules in the ff99SB force field.[91]

As far as we are aware, the parameterisation of Ins(1,3,4,5)P_4 described within this chapter represents the first full set of parameters obtained for this molecule, with a clearly defined protocol.[116] The parameters developed provide a basis to study the interactions of Ins(1,3,4,5)P_4 using MD simulations. Additionally, the work within this chapter also provides a protocol to parameterise and study other inositol phosphates and related highly charged systems.

Chapter 5 – MD Simulations of the Native PKB PH Domain with Inositol Phosphates

5. 1 Introduction

Extension of the Amber force field to include parameters for inositol phosphates allows their numerous interactions to be studied using MD simulations. In this chapter parameters for the inositol tris-phosphates, Ins(1,3,4)P_3 and Ins(1,4,5)P_3, are generated using the protocol developed in Chapter 4. These inositol phosphate parameters are applied in MD simulations to investigate the interaction of Ins(1,3,4,5)P_4, Ins(1,3,4)P_3 and Ins(1,4,5)P_3 with the PKB PH domain. A significant portion of the work within this chapter has been submitted as a paper.[142]

5. 2 The PKB PH Domain-Inositol Phosphate Interaction

The PKB PH domain has a cationic pocket that is lined with basic amino acids. These amino acids are responsible for recognising and interacting with the highly anionic inositol phosphate head-groups of phosphoinositide lipids, Figure 21a. Structural and mutagenesis data show that phosphoinositide lipids interact with PKB principally *via* phosphate groups 3 and 4.[11, 26] Additionally phosphate group 1 is also thought to contribute to binding. In contrast, phosphate group 5 is presumed to make no significant interactions as it is solvent exposed in the crystal structure of Ins(1,3,4,5)P_4 in a complex with the PKB PH domain.[11, 26] This is illustrated in Figure 21b.

The main interactions of Ins(1,3,4,5)P_4 with the PKB PH domain are illustrated in Figure 21c. The crystal structure shows phosphate group 3 interacts with Lys14, Arg23, Arg25 and Asn53, phosphate group 4 with Lys14, Asn53 and Arg86 and phosphate group 1 with Arg23.

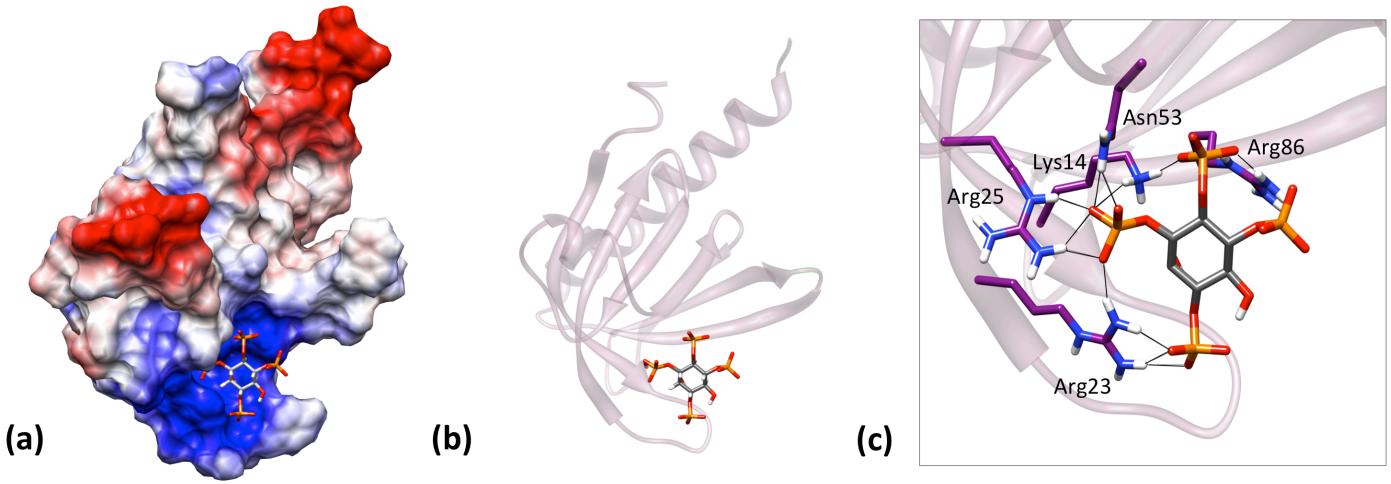


Figure 21: (a) The electrostatic potential map of the PKB PH domain interacting with Ins(1,3,4,5)P_4. Blue areas (+ 10 kcal mol^{-1} e^{-1}) represent positively charged residues and red areas (-10 kcal mol^{-1} e^{-1}) represent negatively charged residues. **(b) Crystal structure of the PKB PH domain in a complex with Ins(1,3,4,5)P_4** (PDB accession code 1UNQ),[21] **(c) A zoomed image of the crystal structure highlighting the interaction between residues in PKB's cationic pocket and Ins(1,3,4,5)P_4.** The side chains of amino acid residues lining the binding site are shown in dark magenta. Hydrogen bonds between Ins(1,3,4,5)P_4 and amino acid residues in the binding pocket are shown as black lines.

5. 2. 1 PKB's Selectivity

The distinct arrangement of phosphomonoester substitution on the inositol ring of phosphoinositides/inositol phosphates is responsible for the specificity of their interactions with receptor proteins.[7, 27] Numerous experimental studies investigating PKB's affinity for phosphoinositide lipids, have illustrated its preference for the 3-phosphorylated phosphoinositides [PtdIns(3,4,5)P_3 and PtdIns(3,4)P_2] over PtdIns(4,5)P_2.[22, 24, 25]

Interestingly, PKB does not appear to follow the same pattern of selectivity for inositol phosphates as for their parent phosphoinositide lipids.[22, 143] In a study by Frech *et al.,* Ins(1,4,5)P_3 and Ins(1,3,4,5)P_4 are shown to bind to the PKB PH domain with a similar affinity, despite their parent lipids [PtdIns(4,5)P_2 and PtdIns(3,4,5)P_3 respectively] displaying markedly different binding affinities to each other in this, and numerous other, investigations.[22-24] The relatively high affinity of Ins(1,4,5)P_3 for the PKB PH domain therefore requires explanation, as its parent lipid PtdIns(4,5)P_2 is not known to bind to PKB despite its relative abundance in resting cells.[144]

5. 2. 2 Binding Mode of Inositol Phosphates/Phosphoinositides

The binding mode of Ins(1,3,4,5)P_4 with its 3-phosphate group pointing directly into the PKB binding site, and 5-phosphate group solvent exposed, is also assumed for its parent lipid PtdIns(3,4,5)P_3 due to its similar binding affinity to PtdIns(3,4)P_2.[22, 26] This proposed binding mode for phosphoinositide lipids, in which the 5-phosphate group projects into solvent, is further consistent with the inability of PtdIns(4,5)P_2 to bind.

However, the relatively high *in vitro* binding affinity of Ins(1,4,5)P_3 for the PKB PH domain, suggests that this inositol phosphate ligand may adopt a different orientation in the binding site to its parent lipid PtdIns(4,5)P_2. This matter is difficult to investigate with experiments alone. Therefore in this chapter MD simulations are applied to explore the interaction of Ins(1,4,5)P_3, and its parent lipid PtdIns(4,5)P_2, with the PKB PH domain.

5. 3 The Computational Approach

This chapter describes a computational approach to investigate the interaction of the PKB PH domain with a number of inositol phosphates [Ins(1,3,4,5)P_4, Ins(1,3,4)P_3 and Ins(1,4,5)P_3]. MD simulations allow the molecular basis, dynamic behaviour and associated free energies of such interactions to be studied.

As highlighted in Chapter 4, a major difficulty in studying inositol phosphates using a computational approach is that their net charge/ionisation state in the cellular medium is currently uncertain.[5, 118] Therefore, for each inositol phosphate all possible isomers/ionisation states at physiological pH have been considered.[3, 145, 146]

5. 4 Program Details

Gaussian 09[122] was used for the geometry optimisation calculations and to generate the molecular electrostatic potential for the Ins(1,3,4)P_3 and Ins(1,4,5)P_3 isomers. The RESP program (in Amber) was used to derive the partial atomic charges.

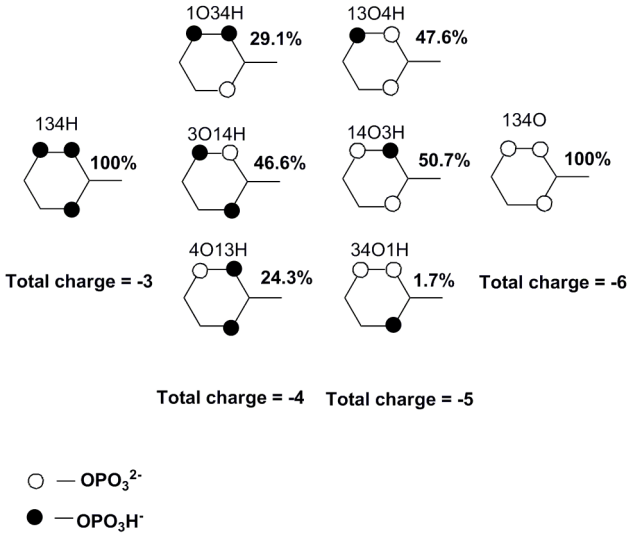
The Amber10[120] and Amber11[124] suite of programs implementing the ff99SB force field were used for energy minimisation, equilibration and production MD calculations. Parameter and topology files for the inositol phosphate-PKB PH domain complexes were creating using the LEaP module. Minimisation, equilibration and production MD calculations were performed with PMEMD. Analysis of hydrogen bonding, root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF) and cluster analysis were carried out using Ptraj. Free energy calculations, using the MM-PBSA method, were carried out using the Perl scripts in Amber. Graphical representations were produced with VMD and Chimera.[125, 126]

5. 5 Parameterisation of Inositol Tris-phosphates

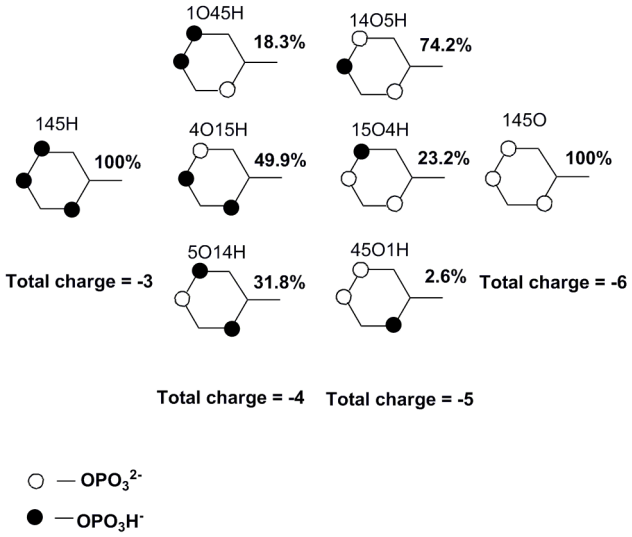
Parameters for Ins(1,3,4,5)P_4 were generated in Chapter 4. Herein parameters for Ins(1,3,4)P_3 and Ins(1,4,5)P_3 are generated using the same protocol.

The initial structures for Ins(1,3,4)P_3 and Ins(1,4,5)P_3 were adapted from that of Ins(1,3,4,5)P_4 previously studied. Phosphate groups 5 and 3 were substituted with hydroxyl groups for Ins(1,3,4)P_3 and Ins(1,4,5)P_3 respectively. Eight isomers/ionisation states with different levels and arrangements of protonation were investigated for both Ins(1,3,4)P_3 and Ins(1,4,5)P_3, illustrated in Figure 22. These isomers fall into four different charge subsets ranging from -3 to -6. The *in vitro* probabilities of these charge states at varying pH are displayed in Figure 23.[145, 146]

As for Ins(1,3,4,5)P_4, each isomer was geometry optimised until it reached a local minimum on the potential energy surface. The isomers were first optimised using the DFT-B3LYP/3-21g* level of theory and then further optimised using DFT-B3LYP/6-31g**. These calculations were carried out in gas phase. For each optimised structure the MEP was calculated using the HF/6-31g* level of theory. The RESP algorithm was employed to fit the QM calculated MEP to atom-centred points. The RESP generated point charges were assigned as the charge parameters. Atom types were assigned from the ff99SB force field, using analogous atom types to Ins(1,3,4,5)P_4. Parameters for Ins(1,3,4)P_3 and Ins(1,4,5)P_3 are supplied in sections S.4 and S.5 of the Electronic Supplementary Information.

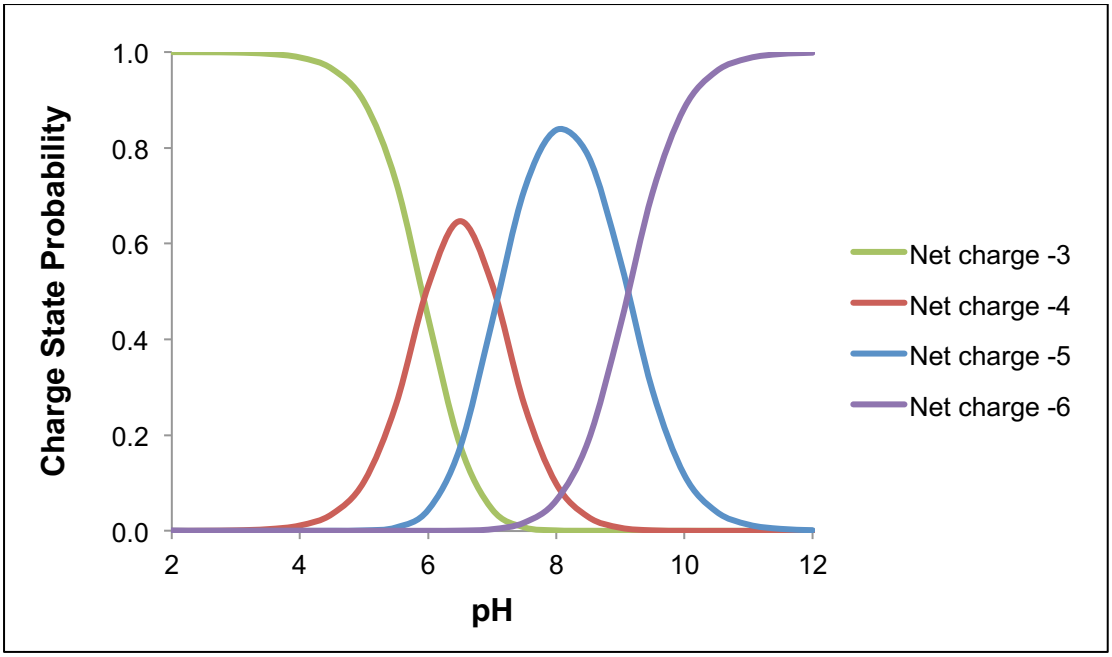


(a)

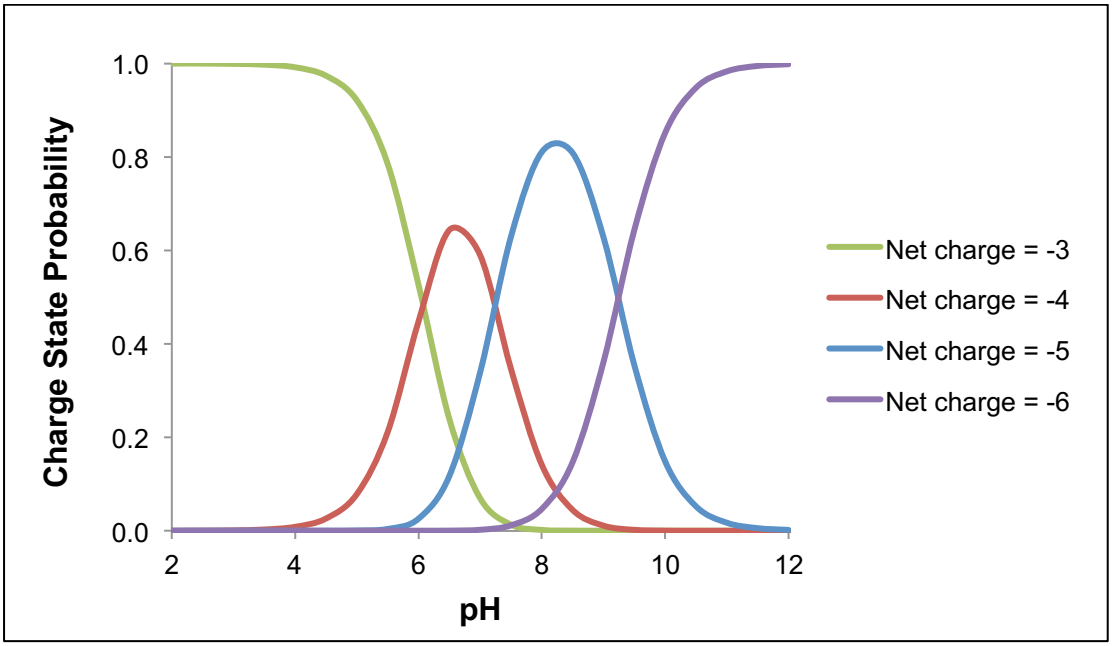


(b)

Figure 22: The eight Ins(1,3,4)P_3 (a) and Ins(1,4,5)P_3 (b) isomers. Diagram adapted from Borkovec and Spiess.[3] The horizontal line indicates the axial hydroxyl group at position 2. For Figures (a) and (b) circles represent phosphate groups 1, 3 and 4, and 1, 4 and 5 respectively. Filled circles represent mono-protonated phosphate groups, and empty circles represent fully deprotonated phosphate groups. The name given to, net charge of, and *in vitro* probability of each isomer are provided.[146]



(a)



(b)

Figure 23: Variation of the charge state probabilities as a function of pH for **(a)** Ins(1,3,4)P_3 and **(b)** Ins(1,4,5)P_3 in Et_4NClO_4 10% D_2O.[145, 146]

5. 5. 1 Validation of Parameters

The parameters assigned to Ins(1,3,4)P_3 and Ins(1,4,5)P_3 were validated by applying them in a molecular mechanics energy minimisation stage, as for Ins(1,3,4,5)P_4 in section 4.6.2. For all of the Ins(1,3,4)P_3 and Ins(1,4,5)P_3 isomers, the energetic orderings from the QM optimisation and MM energy minimisation stages were in full agreement.

5. 6 MD Protocol

5. 6. 1 Initial Configuration

The X-ray crystal structure of the PKB PH domain in a complex with Ins(1,3,4,5)P_4 [from the RCSB Protein Data Bank, accession code 1UNQ] provided the initial co-ordinates.[11] In order to study the PKB PH domain in a complex with Ins(1,3,4)P_3 and Ins(1,4,5)P_3, the PDB was modified by substituting phosphate groups 5 and 3 of Ins(1,3,4,5)P_4 with hydroxyl groups respectively. The acetyl group capping the N-terminus of the PH domain was deleted as it is sufficiently far from the inositol phosphate binding site, and is not present *in vivo.* Crystallographic water molecules present in the PDB were kept. The Leap module of Amber was used to add hydrogen (and missing) atoms to the crystal structure. The systems were charge neutralised with Na^+ counter-ions and solvated with TIP3P water molecules[99] within a truncated octahedral unit cell, with a minimum of 8 Å distance from the edge of the box to any protein or ligand atom.

5. 6. 2 Minimisation, Equilibration and Production Run

The complexes were minimised in two stages. In the first stage, the water molecules and Na^+ counter-ions were energy minimised, whilst harmonic restraints of 500 kcal mol^{-1} $Å^{-2}$ were applied to the protein and ligand; 500 steps of steepest descent, followed by 500 steps of conjugate gradient minimisation were carried out. In the second stage of minimisation, the restraints were removed, and the entire

system was minimised using 1500 steps of steepest descent, followed by 2000 steps of conjugate gradient minimisation.

In all simulations, periodic boundary conditions were applied. A cut-off of 8 Å was used for the explicit calculation of non-bonded interactions. For the calculation of long-range electrostatic interactions, the particle mesh Ewald method was applied with a direct space cut-off of 8 Å.[98] Each minimised system was heated from 0 K to 300 K at constant volume over 50 ps. Following this, 50 ps of equilibration was carried out at constant pressure (1 bar) and temperature (300 K). Production MD simulations of each system were then run for 50 ns at constant pressure (1 bar) and temperature (300 K), with snapshots saved every 10 ps. For each simulation the SHAKE algorithm was applied to constrain all bonds involving hydrogen atoms, allowing a time-step of 2 fs to be used.[102] In order to maintain the temperature during the simulations, Langevin dynamics was used with a collision frequency of 1.0 ps^{-1}.

5. 6. 3 MM-PBSA Calculations

MM-PBSA (Molecular Mechanics–Poisson–Boltzmann Surface Area) calculations were implemented to calculate the free energy of binding for the inositol phosphate-PKB PH domain complexes. These calculations were performed on the conformational ensembles generated by the production MD simulations. The theory and explanation behind this free energy approach is provided in section 3.7.

Structural snapshots of the production MD simulations were saved every 10 ps, giving a total of 5000 snapshots for each 50 ns simulation. Each snapshot was stripped of water and counter-ions. MM contributions were calculated with an interaction cut-off of 999 Å and a dielectric constant of 1.0. Contributions to the solvation energy were calculated using the pbsa module of Amber. An internal dielectric of 1.0, an external dielectric of 80.0, a solvent radius of 1.4 Å and grid spacing of 0.5 Å was applied. Non-polar contributions to the solvation energy were calculated with a solvent surface tension of 0.0072 kcal mol^{-1} $Å^{-2}$. The SASA was calculated with the Molsurf program with a solvent probe radius of 1.4 Å and Bondi atomic radii.[147, 148]

5. 6. 3. 1 Reasoning for the MM-PBSA Method

The main reason the MM-PBSA method was chosen for this work is due to the fact it is computationally efficient compared to many other free energy methods, such as FEP or TI. Further reasoning for applying the MM-PBSA method in this work is its previous success in ranking and/or predicting the binding affinities of numerous ligands to a common target.[115, 149-151] Its recent use in assessing the selectivity of protein kinases is also particularly relevant. [152]

5. 6. 3. 2 Entropy

In this work, the entropic contribution to binding is neglected due to its large computational expense. Whilst the entropic contribution can increase accuracy when calculating free energies of binding, it can also have a detrimental effect.[115] Inclusion of the entropic term in free energy calculations has been shown, in some cases, to worsen the agreement of computational calculations with experimental data. This is because inclusion of inaccurate entropies can cause more harm than simply neglecting this term.[153, 154]

In addition to the above reasoning, the structural similarity of the inositol phosphate ligands investigated in this work means that their entropic contributions to binding would be expected to be comparable.[107, 150 152]

5. 6. 3. 3 Dielectric Constant

The computed free energies of binding in this study are significantly overestimated. Although this can partly be ascribed to the lack of entropic contribution, the extremely large electrostatic nature of the interaction should also be considered.

The MM-PBSA method can often over-estimate the free energy of binding when dealing with a highly charged binding interface.[115] Previous research has shown that altering the solute dielectric constant in the calculation can scale down binding free energies to bring them into experimental range.[115, 153, 155] However the validity of such an approach is still under debate. A significant issue with this approach is how to

choose the solute dielectric constant for each individual system.[156-158] Whilst Hou *et al.* suggest a method to guide the choice of solute dielectric constant based on interface properties, considerable further work in this area is required.[115]

Problems with altering the solute dielectric constant have also been experienced when investigating ligands with different net charges. Rastelli *et al.* found that free energies of binding tended to only scale uniformly for ligands with the same net charge.[153] Thus when a solute dielectric constant other than 1 was applied in calculations for ligands with different net charges, the correlation to experiment was detrimentally affected. As the ligands within this study fall into charge subsets ranging from -3 to -8, an internal dielectric constant of 1 has been used throughout.

5. 6. 3. 4 Weighted Free Energy Calculations

In order to allow a simple comparison of the computational results to *in vitro* studies, the free energies of binding for the different isomers of each inositol phosphate were weighted according to their *in vitro* probabilities, Figures 15, 16, 22 and 23.[3, 145, 146]

5. 6. 4 Hydrogen Bond Interactions

The Ptraj program within Amber 11 was used to analyse the hydrogen bond interactions for each 50 ns production MD simulation. A list of all potential hydrogen bond donors and acceptors was specified, and pair interactions between these were analysed for the duration of each simulation. A cut-off distance of 3.5 Å, and angle cut-off of 120° were specified for the pair interactions.

5. 6. 5 RMS Deviations and Fluctuations

The Ptraj program was used to calculate the RMS deviations and fluctuations of amino acid residues in the PKB PH domain. RMS deviations were calculated for the PKB PH domain backbone atoms (C, Cα, N) over the time of the simulation, with reference to the starting structure in the production MD simulation. RMS fluctuations of the Cα atoms were calculated with reference to the average structure of the PKB

PH domain during the 50 ns production MD simulation.

5. 6. 6 Cluster Analysis

The Ptraj program was used to carry out cluster analysis. Co-ordinate frames from each 50 ns production MD simulation were grouped into five distinct sets. The frames were clustered using the pairwise RMSD of Cα atoms between frames as a metric, and average linkage as the clustering algorithm.

5. 7 Results and Discussion

5. 7. 1 Free Energies of Binding

5. 7. 1. 1 PKB PH Domain with Inositol Phosphates

The free energy of binding for each Ins(1,3,4,5)P_4, Ins(1,3,4)P_3 and Ins(1,4,5)P_3 isomer to the PKB PH domain is provided in Table 8. These results illustrate the significant variation in the free energies of binding for different isomers of the same inositol phosphate. Interestingly isomers with the same net charge but different arrangement of protonation can differ in their computational free energies of binding by as much as 30 kcal mol^{-1}. Additionally inositol phosphate isomers with a greater net charge do not necessarily bind with a larger free energy, highlighting the complexity of their interactions.

Generally, the statistical errors for the individual simulations are low. Exceptions to this are highlighted in Table 8. The larger errors are due to multiple conformational states in these simulations. Whilst these conformational states are similar to each other, the highly charged nature of this system means that small changes in conformation can result in large differences in the free energy of binding. For example, the 3O145H isomer has the highest statistical error of the simulations. Cluster analysis identified three main conformational states for this system. Whilst all three states illustrated extremely similar hydrogen bond interactions *via* phosphate groups 3 and 4, slight variations at phosphate groups 1 and 5 were apparent, Figure 24.

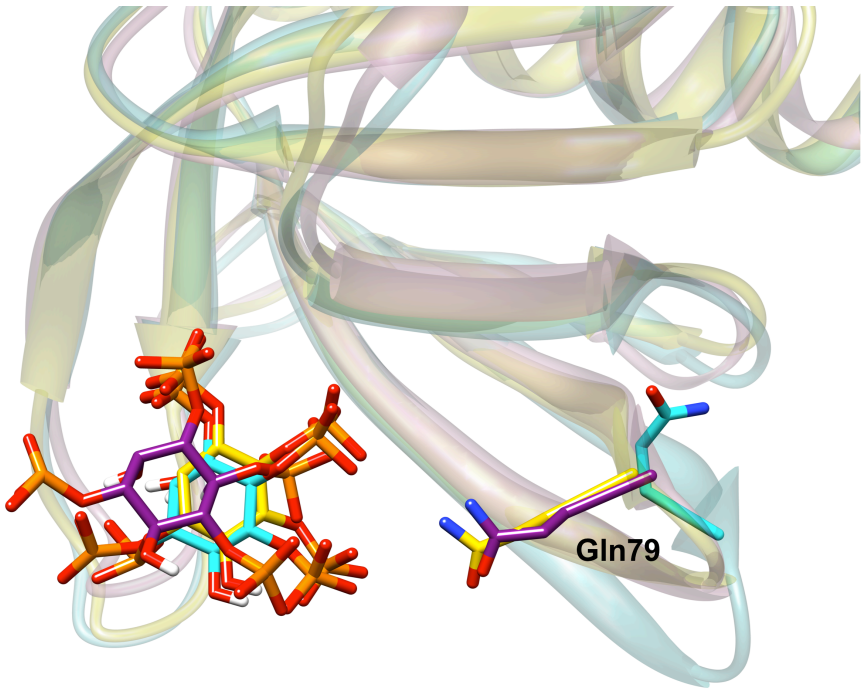


Figure 24: The three conformational states of the 3O145H Ins(1,3,4,5)P_4 isomer in a complex with the PKB PH domain. Cluster 1 shown in cyan is similar to the crystal structure of this complex. Cluster 2 shown in yellow has increased interactions with Gln79 on loop 3 *via* phosphate group 5. Cluster 3 shown in magenta has decreased interactions between Arg23 and phosphate group 1. Most hydrogen atoms have been omitted from this figure for ease of visualisation.

Table 8: Computationally determined binding free energies (denoted ΔG) for the individual inositol phosphate isomer-PKB PH domain simulations. The free energies of binding were calculated using the MM-PBSA method. The statistical error was estimated from the deviation between block averages.

Ligand	Charge	Isomer	ΔG in kcal mol^{-1}
Ins(1,3,4,5)P_4	-4	1345H	-71.86 ± 4.16
	-5	1O345H	-79.43 ± 3.06
		3O145H	-60.33 ± 16.13†
		4O135H	-86.40 ± 5.40
		5O134H	-69.46 ± 1.39
	-6	13O45H	-71.90 ± 9.09
		14O35H	-65.58 ± 6.05
		15O34H	-81.85 ± 5.48
		34O15H	-60.81 ± 10.64†
		35O14H	-76.50 ± 6.99
		45O13H	-96.31 ± 5.92
	-7	134O5H	-64.66 ± 15.77†
		135O4H	-64.64 ± 9.46
		145O3H	-79.41 ± 2.49
Ins(1,3,4)P_3	-3	134H	-70.17 ± 8.55
	-4	1O34H	-57.97 ± 7.53
		3O14H	-72.65 ± 9.63
		4O13H	-62.66 ± 9.93
	-5	13O4H	-65.22 ± 9.55
		14O3H	-63.67 ± 11.64†
		34O1H	-53.48 ± 4.58
	-6	134O	-61.46 ± 9.50
Ins(1,4,5)P_3	-3	145H	-30.01 ± 5.09
	-4	1O45H	-32.56 ± 11.82†
		4O15H	-42.79 ± 3.58
		5O14H	-41.74 ± 5.68
	-5	14O5H	-47.08 ± 7.09
		15O4H	-27.33 ± 8.17
		45O1H	-42.52 ± 10.28†
	-6	145O	-32.83 ± 12.40†

† Larger statistical errors due to multiple conformational states.

In order to allow a simple comparison of the computational results to *in vitro* data, the free energies of binding for the different isomers have been weighted for each inositol phosphate, Table 9. The inositol phosphate isomers were weighted according to their *in vitro* probabilities at pH7.

Table 9: Weighted computationally determined binding free energies (denoted ΔG) and *in vitro* dissociation constants (K_d) for inositol phosphate-PKB PH domain complexes. *In vitro* dissociation constants for the parent phosphoinositide lipid-PKB PH domain complexes are also provided.

Ligand	Computational ΔG (kcal mol^{-1})[a]	*In vitro* K_d for inositol phosphate (μM)[22]	*In vitro* K_d for parent phosphoinositide lipid (μM)[22]
Ins(1,3,4,5)P_4	-71.51 $\pm$ 3.07	1.5 $\pm$ 0.2	0.40 $\pm$ 0.03
Ins(1,3,4)P_3	-65.61 $\pm$ 4.35	ND[b]	0.57 $\pm$ 0.03
Ins(1,4,5)P_3	-40.45 $\pm$ 2.76	1.2 $\pm$ 0.1	2.5 $\pm$ 0.5

[a] The statistical error was estimated from the deviation between block averages for each simulation, and propagated appropriately for each weighted ligand.[115]

[b] Not determined.

As mentioned in section 5.6.3.3 the computed free energies of binding are significantly overestimated due to the highly charged binding interface of this system. Therefore, the comparison between the computational results and experimental data herein is qualitative.

The weighted computational results mirror PKB's widely accepted preference for 3-phosphorylated phosphoinositides, Table 9. The computational results show Ins(1,3,4,5)P_4 and Ins(1,3,4)P_3 to bind to the PKB PH domain with a similar free energy of binding, with a preference for Ins(1,3,4,5)P_4, agreeing with the *in vitro* affinities of their parent lipids. In comparison, Ins(1,4,5)P_3 binds to the PKB PH domain with a significantly lower free energy, agreeing with the low *in vitro* affinity of its parent lipid PtdIns(4,5)P_2.[22, 24, 25] The relatively small statistical errors for the weighted free energies of binding impart a high level of confidence in these results.

Whilst the weighting method used in this work has been justified it is important to note that alternative approaches, such as averaging results for all isomers or taking the result for the strongest binding isomer, also give similar outcomes, Table 10. All approaches illustrate PKB's preference for 3-phosphorylated inositol phosphates, in particular Ins(1,3,4,5)P_4.

Table 10: A comparison of different methods to weight the computationally determined binding free energies (denoted ΔG). The different weighting methods include weighting according to the *in vitro* probabilities of the isomers, averaging results for the isomers and simply taking results for the strongest binding isomer.

Ligand	Weighted computational ΔG (kcal mol^{-1})	Averaged computational ΔG (kcal mol^{-1})	Isomer with largest computational ΔG (kcal mol^{-1})
Ins(1,3,4,5)P_4	-71.51 $\pm$ 3.07	-73.51$\pm$ 7.29	-86.40 $\pm$ 5.40†
Ins(1,3,4)P_3	-65.61 $\pm$ 4.35	-63.41$\pm$ 8.86	-72.65 $\pm$ 9.63
Ins(1,4,5)P_3	-40.45 $\pm$ 2.76	-37.11$\pm$ 8.01	-47.08 $\pm$ 7.09

† Free energy result for 45O13H isomer ignored, as the *in vitro* probability for this isomer within the -6 subset is only 0.5%.

The *in vitro* data of Frech *et al.* suggests that the inositol phosphate head-groups do not follow the same pattern of selectivity as their parent lipids. Rather than illustrating a preference for 3-phosphorylated inositol phosphates, the experimental results show that the PKB PH domain binds to Ins(1,3,4,5)P_4 and Ins(1,4,5)P_3 with a similar affinity.

Visual analysis of the MD trajectories of Ins(1,4,5)P_3 and the PKB PH domain prompted an explanation for this *in vitro* behaviour. When visualised in VMD the Ins(1,4,5)P_3 isomers were seen to rotate 180° along the P1/P4 axis in the binding site, illustrated in Figure 25. A movie of this is provided in the Electronic Supplementary Information, Movie M1. In general, the Ins(1,4,5)P_3 ligands spent the majority of the 50 ns simulations re-orientating, only reaching the fully flipped/inverted conformation towards the end of the simulation.

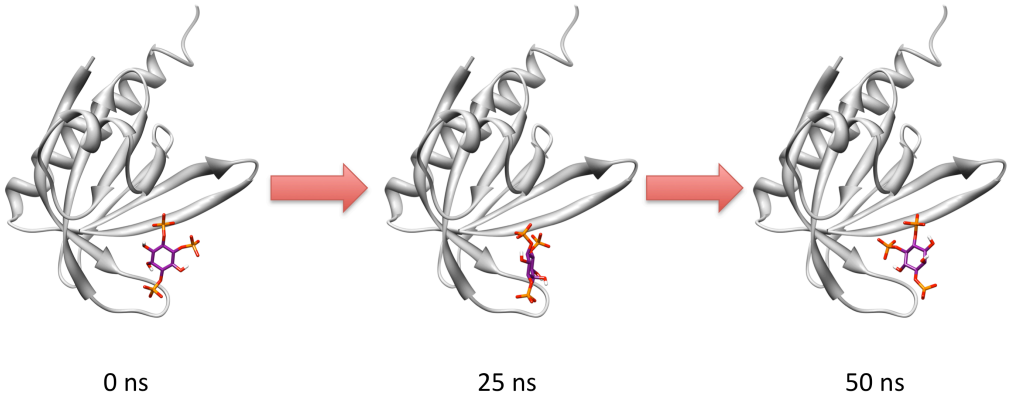


Figure 25: Rotation of Ins(1,4,5)P_3 in the PKB binding site. Images are taken from snapshots at 0, 25 and 50 ns of a simulation. At 0 ns Ins(1,4,5)P_3 is in its initial orientation, with the 5-phosphate group pointing away from binding site. At 25 ns Ins(1,4,5)P_3 has rotated approximately 90° along the P1/P4 axis, with the 5-phosphate group now pointing out of the page. At 50 ns Ins(1,4,5)P_3 has flipped 180° along the P1/P4 axis from its initial orientation, with the 5-phosphate group now pointing towards the PKB binding site.

To confirm the stability of the Ins(1,4,5)P_3 isomers in this rotated/flipped binding mode the simulations were run for a further 50 ns. The Ins(1,4,5)P_3 isomers were shown to maintain this rotated orientation. Furthermore, free energy analysis of the isomers following rotation in the PKB binding site corresponded to a dramatic increase in their free energies of binding, illustrated in Figure 26.

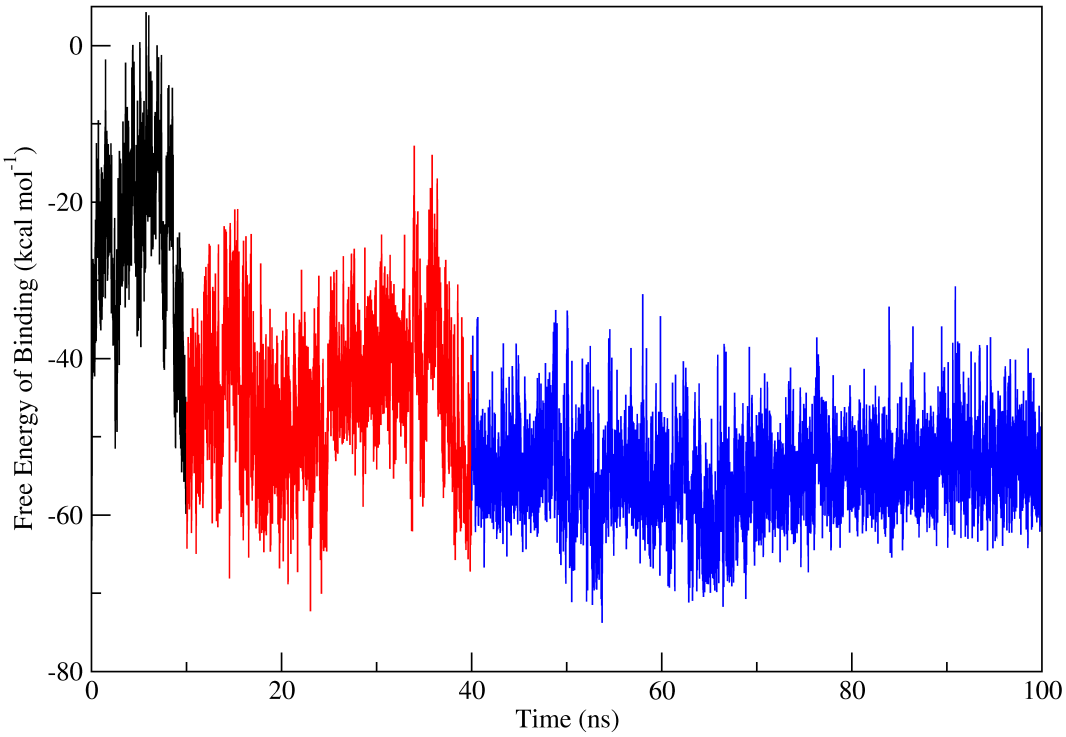


Figure 26: The free energy of binding for Ins(1,4,5)P_3 to the PKB PH domain during an 100 ns MD simulation. The first 10 ns (black) represent the Ins(1,4,5)P_3 ligand in its initial weakly binding orientation in the PKB binding site with its 5-phosphate group solvent exposed. The next 40 ns (red) illustrate Ins(1,4,5)P_3 re-orientating in the binding site. The final 60 ns (blue) illustrate Ins(1,4,5)P_3 in its strongly binding flipped orientation.

Based on these results, we proposed a theory to explain the high *in vitro* affinity of Ins(1,4,5)P_3 for the PKB PH domain. We hypothesised that when not bonded to a phosphatidate tail the Ins(1,4,5)P_3 head-group is no longer restricted to bind in the same orientation as its parent lipid. Instead Ins(1,4,5)P_3 binds to the PKB PH domain in a flipped orientation, in which the inositol ring is inverted 180° along the P1/P4 axis, with respect to the other inositol phosphate ligands, Figure 27. In this flipped orientation phosphate group 5 points towards the PKB binding site, thus mimicking phosphate group 3 of Ins(1,3,4)P_3. Ins(1,4,5)P_3 would therefore act in a similar manner to Ins(1,3,4)P_3, the head-group of PtdIns(3,4)P_2 known to exhibit high affinity for the PKB PH domain.

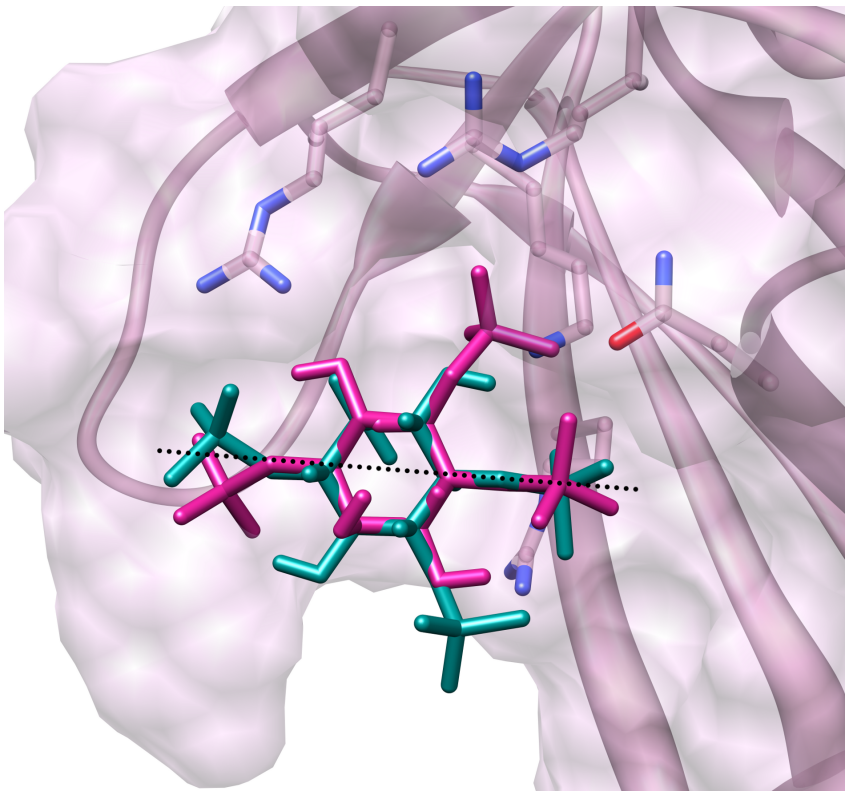


Figure 27: Ins(1,4,5)P_3 in its initial (cyan) and flipped (magenta) orientation in the binding site of the PKB PH domain. A dashed line represents the P1/P4 axis around which the Ins(1,4,5)P_3 ligand was rotated 180°.

5. 7. 1. 2 PKB PH Domain with Flipped Inositol Phosphates

In order to further investigate Ins(1,4,5)P_3's alternative binding mode, MD simulations were performed with the starting orientation of Ins(1,4,5)P_3 flipped/inverted 180° along the P1/P4 axis in the PKB PH domain. Free energy analysis of these simulations (Table 11) showed that in this flipped orientation Ins(1,4,5)P_3 binds with a similar free energy to Ins(1,3,4)P_3 and Ins(1,3,4,5)P_4 in their initial binding mode (Table 9). Furthermore, Ins(1,4,5)P_3 appears stable in this flipped orientation and makes remarkably similar hydrogen bond interactions to the 3-phosphorylated inositol phosphates.

Table 11: Weighted computationally determined free energies of binding (denoted ΔG) for the flipped inositol phosphate-PKB PH domain complexes.

Ligand	Computational ΔG (kcal mol^{-1})[a]
Ins(1,3,4,5)P_4 flip	-60.06 $\pm$ 2.29
Ins(1,3,4)P_3 flip	-42.07 $\pm$ 5.17
Ins(1,4,5)P_3 flip	-60.89 $\pm$ 2.34

[a] The statistical error was estimated from the deviation between block averages for each simulation, and propagated appropriately for each weighted ligand.[115]

To extend this study, simulations of the 3-phosphorylated inositol phosphates flipped 180° along the P1/P4 axis in the PKB PH domain were also performed. Free energy analysis showed that Ins(1,3,4)P_3 has a significantly lower free energy of binding in its flipped mode (Table 11), mirroring that of Ins(1,4,5)P_3 in its un-flipped and weakly binding orientation. Remarkably, towards the end of the simulations the flipped Ins(1,3,4)P_3 isomers rotated back along the P1/P4 axis to their binding mode in which phosphate group 3 points towards the binding site.

The weighted free energy of binding for the flipped Ins(1,3,4,5)P_4 ligand to the PKB PH domain is approximately 10 kcal mol^{-1} lower than for its un-flipped orientation.

This difference in affinity is interesting, as both binding modes have phosphate groups at the same points along the inositol ring. We therefore hypothesise that this lower affinity is due to the reversed positions of the phosphate groups relative to the plane of the inositol ring, Figure 28, thereby affecting their interactions with amino acids in the binding site.

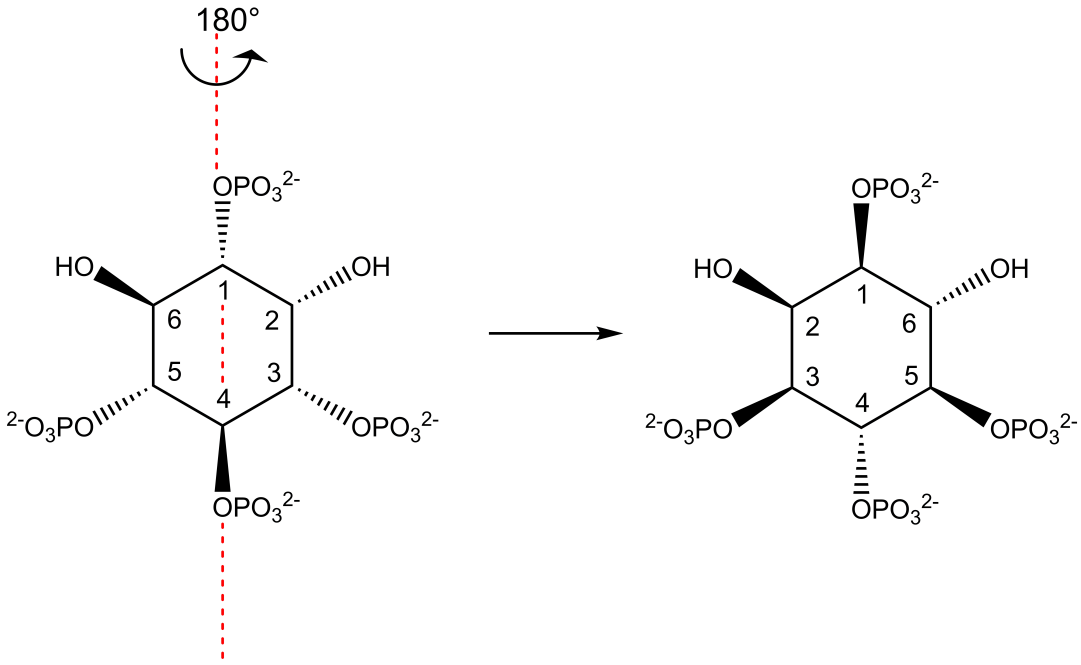


Figure 28: Rotation of the Ins(1,3,4,5)P_4 ligand. Phosphate groups are shown to be fully deprotonated for ease of visualisation.

Inspection of the individual free energies of binding for the Ins(1,3,4,5)P_4 flipped isomers (Table S5 in the Electronic Supplementary Information) highlights the very large variation in their affinity for the PKB PH domain. Interestingly isomers in which phosphate group 4 is singly protonated tend to be unstable in the binding site, displaying large movements and reorientation. Conversely, Ins(1,3,4,5)P_4 isomers in which phosphate group 4 is fully deprotonated generally showed a significantly larger free energy of binding. This may be due to phosphate group 4 forming additional interactions with amino acids in the binding site when fully deprotonated.

Whilst it is feasible that both binding modes are accessible to Ins(1,3,4,5)P_4, results demonstrate that its un-flipped mode of binding is favourable. This is supported by the simulations of the individual complexes, in which each isomer shows a larger free energy of binding in its un-flipped mode compared to its flipped.

5. 7. 1. 3 PKB PH Domain with Phosphoinositide 4, 5-bisphosphate

Whilst we have explained the unexpectedly high *in vitro* binding affinity of Ins(1,4,5)P_3 to the PKB PH domain, one may consider the equally puzzling question of why its parent lipid PtdIns(4,5)P_2 does not exhibit the same relatively high affinity *in vitro* or *in vivo*.[22, 25, 34]

Phosphoinositide lipids have a phosphatidate tail bonded to the inositol phosphate head-group *via* phosphate group 1. It has previously been hypothesised that this tail restricts the binding orientation of the inositol phosphate head-group due to steric hindrance.[143] If PtdIns(4,5)P_2 was restricted to bind with the 5-phosphate group in its solvent exposed position, rather than pointing towards the PKB binding site, this could explain the different behaviour to its head-group.

To investigate this steric hindrance, MD simulations of PtdIns(4,5)P_2 in both its flipped and un-flipped orientations in the PKB PH domain were performed. These simulations were run for 150 ns in order to fully explore the behaviour and movement of the PtdIns(4,5)P_2 ligand. Parameters for PtdIns(4,5)P_2 were generated using the same protocol as for the inositol phosphate isomers. These parameters are provided in section S.6 in the Electronic Supplementary Information.

Visual analysis of the simulations suggests that PtdIns(4,5)P_2's inositol phosphate head-group and phosphatidate tail are locked into a specific conformation by intra-molecular hydrogen bonds. Of particular importance are strong hydrogen bond interactions between the 6-hydroxyl group and the oxygen atoms of phosphate group 1, which are occupied during almost the entire simulation, Figure 29. Interestingly this interaction has also been noted in previous experimental and computational studies.[5, 63] These hydrogen bond interactions encourage phosphate group 1 (and hence the lipid tail to which it is bonded) to tilt towards the 6 hydroxyl group. In its

un-flipped orientation the 6-hydroxyl group of PtdIns(4,5)P_2 is solvent exposed, thus the lipid tail is encouraged to project into solvent. However, in its flipped orientation the 6-hydroxyl group of PtdIns(4,5)P_2 points towards the binding site of PKB and therefore encourages the lipid chain into a position that cannot be accommodated due to steric clashing. As a result, the PtdIns(4,5)P_2 ligand is seen to move away from its flipped binding mode. This movement of PtdIns(4,5)P_2 and the intra-molecular hydrogen bond interactions are illustrated in Movie M2 in the Electronic Supplementary Information.

The simulation of PtdIns(4,5)P_2 in its un-flipped orientation further supports the argument of this lipid's binding mode. Unlike its equivalent inositol phosphate head-group, PtdIns(4,5)P_2 did not rotate in the binding site during the 150 ns simulation.

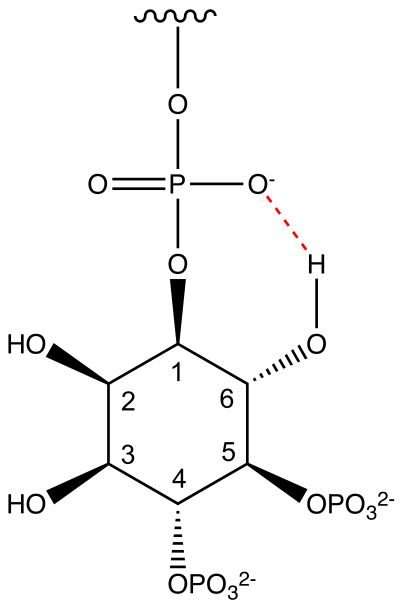


Figure 29: The intra-molecular interaction between the 6-hydroxyl group of PtdIns(4,5)P_2 and phosphate group 1. The remainder of the lipid tail is not illustrated for ease of visualisation, but is attached *via* the 1-phosphate group as indicated.

5. 7. 2 Orientation and Molecular Behaviour

Analysis of the MD simulations showed that the Ins(1,3,4,5)P_4 and Ins(1,3,4)P_3 ligands adopted very similar and stable orientations in the PKB binding site in their initial un-flipped binding mode. This stability is well illustrated by the small RMSD values for the PKB PH domain (and binding site residues) during the simulations. Representative RMSD plots for one simulation are shown in Figures 30 and 31.

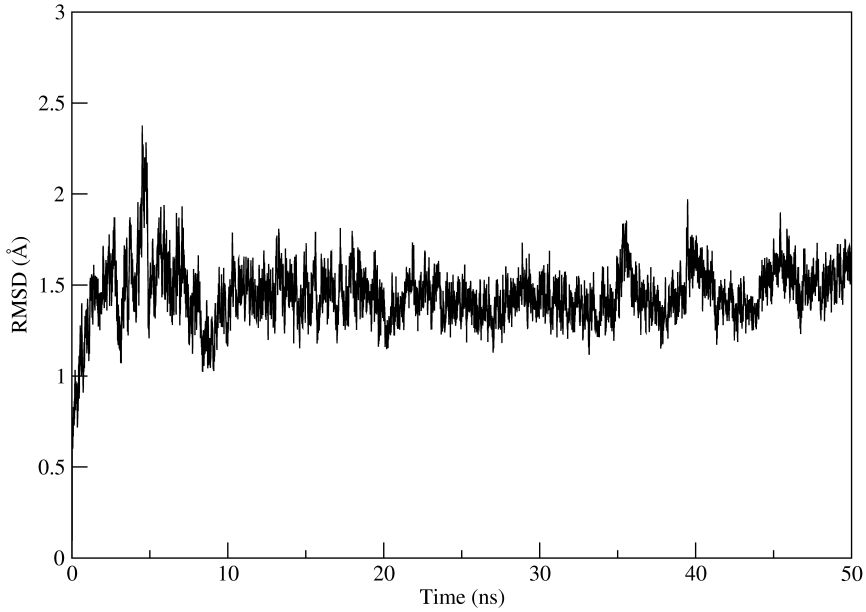


Figure 30: RMS deviation of the backbone atoms of the PKB PH domain when in a complex with Ins(1,3,4,5)P_4 during the 50 ns MD simulation. RMSD values are made in comparison to the starting structure of the production MD simulation.

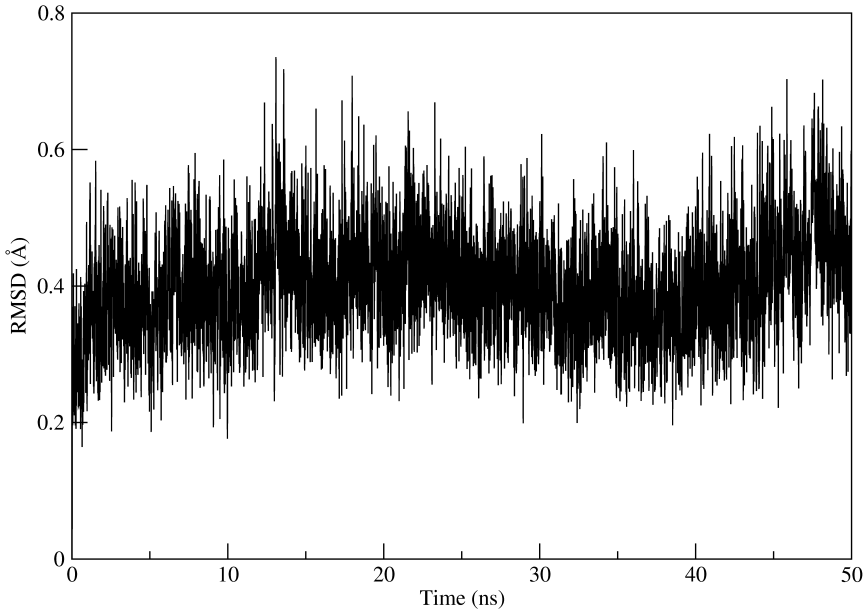


Figure 31: RMS deviation of the backbone atoms of residues in the PKB binding site when in a complex with Ins(1,3,4,5)P_4 during the 50 ns MD simulation. RMSD values are made in comparison to the starting structure of the production MD simulation. Residues include Lys17, Arg23, Arg25, Asn53 and Arg86.

Hydrogen bond analysis of the simulations showed that the 3-phosphorylated inositol phosphates replicated the hydrogen bond interactions apparent from the crystal structure with phosphate groups 3 and 4 forming the major interactions, Table 12.[11] Interestingly in its flipped orientation Ins(1,4,5)P_3 adopted a comparable binding mode, with phosphate group 5 mimicking phosphate group 3 of the 3-phosphorylated inositol phosphates. In contrast, the Ins(1,4,5)P_3 ligand in its initial un-flipped orientation made many fewer H bond interactions and consequently was seen to move away from the PKB binding site, before rotating 180° along the P1 /P4 axis.

Table 12: The major H bond interactions for the PKB PH domain-inositol phosphate complexes. Average occupancies for the hydrogen bond interactions are provided, representing the percentage of time the interactions were formed during the MD simulations. Each phosphate group has three non-ester oxygen atoms that are able to act as hydrogen bond acceptors. The interactions of these oxygen atoms have been combined for each phosphate group, sometimes resulting in occupancies over 100%. The donor atoms of the amino acid side chains are named according to the electronegative atom the hydrogen atoms are attached to. In cases where multiple hydrogen atoms are attached to this electronegative atom, their occupancies were combined. The names of the atoms in the table are taken from Amber.

Acceptor	Donor	Ins(1,3,4,5)P_4	Ins(1,3,4)P_3	Ins(1,4,5)P_3	Ins(1,4,5)P_3 flip
Phosphate group 1	Arg23 (NH2)	121	103	73	83
	Arg23 (NH1)	113	100	57	82
Phosphate group 3	Lys14(NZ)	116	77		
	Arg23 (NH1)	66	26		
	Arg25 (NE)	107	118		
	Arg25 (NH2)	191	141		
	Asn53 (ND2)	110	104		
Phosphate group 4	Lys14 (NZ)	96	90	70	83
	Asn53 (ND2)	105	93	68	84
	Arg86 (NH2)	152	121	83	113
	Arg86 (NE)	122	111	78	111
Phosphate group 5	Lys14(NZ)				96
	Arg23 (NH1)				20
	Arg25 (NE)				105
	Arg25 (NH2)				136
	Asn53 (ND2)				66

Analysis of the MD simulations of the PKB PH domain and Ins(1,3,4,5)P_4 suggested that in addition to the hydrogen bond interactions reported from the crystal structure, interactions *via* phosphate group 5 were also apparent. Whilst phosphate group 5 is considered to be solvent exposed in the crystal structure, significant interactions with Arg86 are apparent in the simulations (with an average occupancy of 59%). Additionally other minor interactions of phosphate group 5 with residue Gln79 positioned on variable loop 3 are also apparent (with an average occupancy of 9%).

The interactions of phosphate group 5 stress the importance of not solely relying on a static crystal structure for binding information. Furthermore, they also highlight the

use of MD simulations in this matter. It is easily understandable how such interactions are not apparent in the crystal structure, as loop 3 (consisting of residues 77-84) is extremely flexible. This flexibility is highlighted by loop 3's high B-factor values, and 7.4 Å difference in position in the *apo*- and complexed crystal structures of the PKB PH domain.[11]

Figure 32 illustrates the RMS fluctuations of the PKB PH domain residues when in a complex with Ins(1,3,4,5)P_4. [Results are displayed for one representative 50 ns simulation]. Fluctuations are calculated relative to the average structure during the simulation. This figure again highlights the high mobility of loop 3 (and particularly residues Gln79 and Trp80).

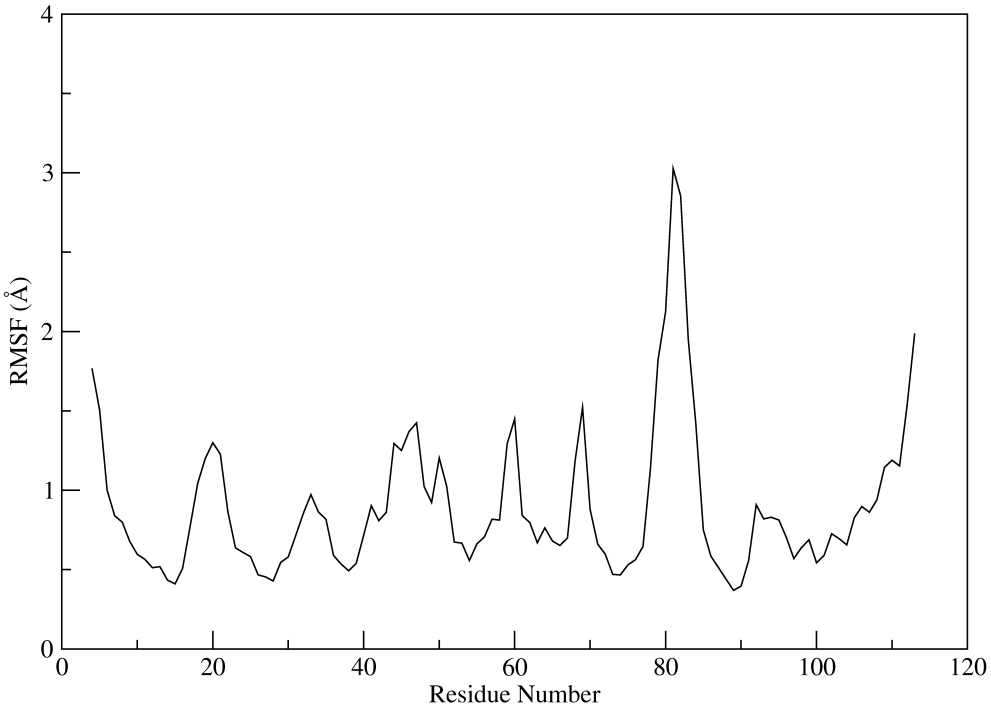


Figure 32: RMS fluctuations by residue of the PKB PH domain when in complex with Ins(1,3,4,5)P_4, during the 50 ns simulation.

5. 8 Conclusion

This chapter describes the use of MD simulations to further understand the interaction of PKB's PH domain with the inositol phosphate head-groups of phosphoinositide lipids. Additionally these simulations act as further validation of the inositol phosphate parameters generated in this work.

The MD simulations successfully mirror PKB's *in vivo* (and *in vitro*) selectivity for 3-phosphorylated phosphoinositides. The hydrogen bond interactions apparent in these simulations also correspond well with that of the crystallographically determined structure of Ins(1,3,4,5)P_4 bound to the PKB PH domain.

In addition to complementing experimental work, MD simulations have also been successfully applied to rationalise and explain experimental data. The atomic level view of the computational simulations has helped to explain the unexpectedly high *in vitro* affinity of Ins(1,4,5)P_3 for the PKB PH domain.[22] Simulations showed that when not bonded to a phosphatidate tail, Ins(1,4,5)P_3 is able to bind to the PKB PH domain in a flipped orientation, with respect to the binding mode of other inositol phosphate ligands and its parent lipid PtdIns(4,5)P_2. This flipped orientation allows phosphate group 5 of Ins(1,4,5)P_3 to form favourable interactions with basic amino acids in the PKB binding site, as opposed to being solvent exposed, thus explaining its higher than expected affinity. A flipped binding mode of inositol-based compounds to the PKB PH domain has previously been noted by Castillo *et al.* for a number of phosphatidylinositol ether lipid analogues with 2-hydroxy substitutions.[159]

Due to Ins(1,4,5)P_3's relatively high affinity for the PKB PH domain, it is important to consider whether this ligand plays a regulatory role *in vivo* by competing with 3-phosphorylated phosphoinositides. Without additional experimental data, it is difficult to hypothesise the physiological relevance of this ligand, as this would depend on a number of factors including its intra-cellular concentration, *in vivo* lifetime, and presence of other Ins(1,4,5)P_3 binding proteins.[143] However, a number of previous reports show many higher order inositol phosphates, including Ins(1,3,4,5)P_4, inhibit the activity of PKB by competing for its PH domain.[160-162] Thus

the regulatory role of Ins(1,4,5)P_3 may be an important consideration for future studies.

Chapter 6 - MD Simulations of the Mutant PKB PH Domain with Inositol Phosphates

6.1 Introduction

In this chapter a PKB mutant that has been reported in human breast, lung, ovarian, and colorectal cancers is investigated using MD simulations.[29, 49-51] Substitution of glutamic acid to lysine at position 17 (E17K) on the PKB PH domain causes PKB to constitutively target the plasma membrane.[29] The mechanism behind this constitutive membrane localisation, and how this leads to oncogenesis, is considered herein.

6. 1. 1 Residue 17

Residue Glu17 is located near to the phosphoinositide binding site of the PKB PH domain. In PKB's native uncomplexed state Glu17 occupies this binding site forming hydrogen bond interactions with Lys14, Figure 33. In contrast, when native PKB is complexed with Ins(1,3,4,5)P_4 Glu17 moves 4.7 Å away from the binding site in order to accommodate the anionic inositol phosphate, Figure 33.

Mutation of the glutamic acid residue at position 17 to lysine (E17K) reverses the charge of this residue and thereby alters the electrostatic interactions in the phosphoinositide binding site. The crystal structure of the E17K PKB PH domain complexed with Ins(1,3,4,5)P_4 shows that Lys17 forms (indirect) hydrogen bond interactions with phosphate group 5 and hydroxyl group 6 of Ins(1,3,4,5)P_4, Figure 33.[29]

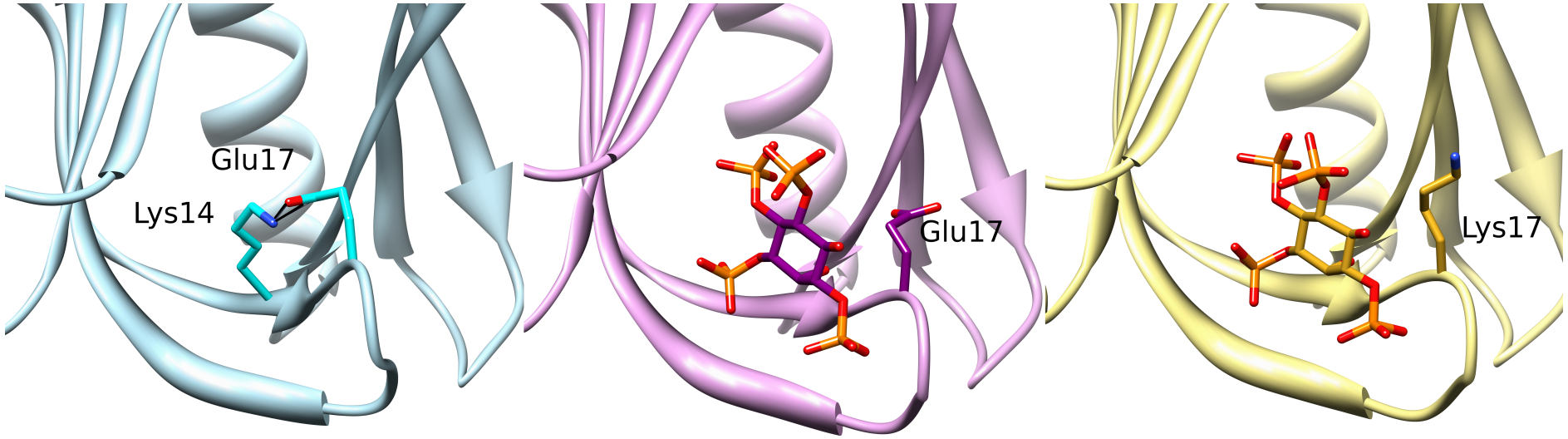


Figure 33: Residue 17 in the native uncomplexed (cyan), complexed (magenta) and E17K mutant (yellow) PKB PH domains. Images are created from the 1UNP, 1UNQ and 2UZS crystal structures respectively.[21, 47] Hydrogen atoms have been omitted for ease of visualisation.

6. 1. 2 Membrane Localisation of Native PKB

Upon cell stimulation, levels of the 3-phosphorylated phosphoinositides in the plasma membrane dramatically increase from their low basal concentrations.[7] These 3-phosphorylated phosphoinositides recruit PKB to the plasma membrane *via* direct interaction of their inositol phosphate head-groups with PKB's PH domain.[11] In contrast PtdIns(4,5)P_2, which is relatively abundant in resting cells, is not known to play a role in this recruitment. At the plasma membrane PKB is phosphorylated, and thereby activated, by PDK1 and MTORC2.[40, 41, 163] Active PKB then dissociates from the plasma membrane and phosphorylates a number of downstream targets. Many of these targets, such as Bad and Caspase-9, regulate the cell's apoptotic machinery.[33, 34] Additionally, many others are associated with cell cycle regulation and progression, including the CDK inhibitors p21 and p27.[36, 37]

6. 1. 3 Membrane Localisation of E17K PKB

The E17K mutation has been found to cause PKB to constitutively target the plasma membrane.[29, 164] This unregulated recruitment of PKB is of particular interest due to its links with cancer.[29, 49, 51] It is understood that constitutive targeting of PKB to the plasma membrane leads to an elevated level of PKB phosphorylation, and thereby activation.[164] As many of PKB's downstream targets are involved in cell proliferation and survival, it is unsurprising that increased levels of active PKB can lead to oncogenesis.

Whilst the crystal structure of the E17K PKB PH domain with Ins(1,3,4,5)P_4 has been solved, the mechanism behind this mutant's constitutive membrane targeting is still under consideration. Two models have previously been proposed for this recruitment.[164] In the first of these models the charge reversal at residue 17 is thought to cause a significant increase in PKB's affinity for 3-phosphorylated phosphoinositides. Therefore even at low basal levels these phosphoinositides would drive PKB recruitment to the membrane. The second model proposes that the E17K mutation alters/broadens PKB's phosphoinositide selectivity, allowing PKB to bind to PtdIns(4,5)P_2, which is relatively abundant in resting cells.

6. 1. 3. 1 Current Understanding

In vitro experiments suggest it is E17K PKB's broadened lipid selectivity that is primarily responsible for its enhanced recruitment to the plasma membrane.[164] In addition, a small increase in the affinity of this mutant for PtdIns(3,4,5)P_3 has also been proposed.[164]

The mechanism behind this broadened selectivity is not well understood. In this chapter, MD simulations of the E17K PKB PH domain with Ins(1,3,4,5)P_4, Ins(1,3,4)P_3 and Ins(1,4,5)P_3 [the head-groups of PtdIns(3,4,5)P_3, PtdIns(3,4)P_2 and PtdIns(4,5)P_2 respectively] have been carried out. It is hoped that the atomic level view from these simulations will help to explain the E17K PKB PH domain's behaviour.

6. 2 Program Details

The Amber11[124] suite of programs implementing the ff99SB force field was used for energy minimisation, equilibration and production MD calculations. Parameter and topology files for the inositol phosphate-E17K PKB PH domain complexes were created using the LEaP module. Minimisation, equilibration and production MD calculations were performed with PMEMD. Analysis of hydrogen bonding was carried out using Ptraj. Free energy calculations, using the MM-PBSA method, were carried out using the Perl scripts in Amber. Graphical representations were produced with VMD and Chimera.[125, 126]

6. 3 MD Protocol

6. 3. 1 Initial Configuration

The X-ray crystal structure of the PKB PH domain in a complex with Ins(1,3,4,5)P_4 [from the RCSB Protein Data Bank accession code 1UNQ][11] provided the initial co-ordinates for the system. Residue Glu17 was modified to Lys by removing all atoms beyond C_δ from the side-chain and altering the name of the residue.

For ligands Ins(1,3,4)P_3 and Ins(1,4,5)P_3, the PDB was modified by substituting phosphate groups 5 and 3 of Ins(1,3,4,5)P_4 with hydroxyl groups respectively. The initial orientation for the inositol phosphates was set to their 'un-flipped' mode, in which the 5-position is solvent exposed, and the 3-position points towards the binding site. As for the native studies in Chapter 5, all physiologically accessible isomers for Ins(1,3,4,5)P_4, Ins(1,3,4)P_3 and Ins(1,4,5)P_3 were considered.

The acetyl group bonded to the N terminus of the PH domain was removed for all simulations. The Leap module of Amber was used to add hydrogen (and missing) atoms to the crystal structure. The systems were charge neutralised with Na^+ counter-ions and solvated with TIP3P water molecules within a truncated octahedral unit cell, with a minimum of 8 Å distance from the edge of the box to any protein or ligand atom.

6. 3. 1. 1 E17K Crystal Structure

The initial orientation for the system was taken from the crystal structure of the native PKB PH domain (PDB accession code 1UNQ), as opposed to the crystal structure of the E17K PKB PH domain (PDB accession code 2UZS).[11, 29]

The reason for using an adapted version of the native PDB in this work was because initially multiple different PKB mutants were investigated for which crystal structures were not available. Therefore, the native crystal structure was adapted for each mutant to maintain consistency. Results from a preliminary investigation of these other mutants are given in the Electronic Supplementary Information (Table S.3). They provide further validation of the molecular model by replicating patterns from experimental data.

A further reason for using the native crystal structure was that the B-factor values for Ins(1,3,4,5)P_4 (and amino acids surrounding it) are significantly larger in the mutant crystal structure: Ins(1,3,4,5)P_4 has an average B-factor value of 21 $Å^2$ in the E17K PKB PH domain, almost three times that of the native crystal structure.

6. 3. 2 Minimisation, Equilibration and Production Run

The complexes were minimised in two stages. In the first stage, the water molecules and Na^+ counter-ions were energy minimised, whilst harmonic restraints of 500 kcal mol^{-1} $Å^{-2}$ were applied to the protein and ligand; 500 steps of steepest descent and 500 steps of conjugate gradient minimisation were carried out. In the second stage of minimisation, the restraints were removed, and the entire system was minimised using 1500 steps of steepest descent and 2000 steps of conjugate gradient minimisation.

In all simulations, periodic boundary conditions were applied. A cut-off of 8 Å was used for the explicit calculation of non-bonded interactions. For the calculation of long-range electrostatic interactions, the particle mesh Ewald method was applied with a direct space cut-off of 8 Å.[98] Each minimised system was heated from 0 K to 300 K at constant volume over 50 ps. Following this, 100.05 ns of equilibration was carried out at constant pressure (1 bar) and temperature (300 K). The equilibration was run for 100.05 ns to allow the PKB PH domain and inositol phosphates to adjust their conformation/orientation due to the hand-altered mutation, and reach equilibrium. Production MD simulations of each system were then run for 50 ns at constant pressure (1 bar) and temperature (300 K), with snapshots saved every 10 ps.

For each simulation the SHAKE algorithm was applied to constrain all bonds involving hydrogen atoms, allowing a time-step of 2 fs to be used.[102] In order to maintain the temperature during the simulations, Langevin dynamics was used with a collision frequency of 1.0 ps^{-1}. MD simulations were performed with the GPU accelerated version of PMEMD in Amber11.[103]

6. 3. 3 Free Energy Calculations

MM-PBSA calculations were implemented to calculate the free energy of binding for the inositol phosphate–E17K PKB PH domain complexes. Structural snapshots from the 50 ns production MD simulations were used for these calculations. Snapshots were saved every 10 ps, giving a total of 5000 snapshots for each simulation. Each

snapshot was stripped of water and counter-ions. MM-PBSA calculations were carried out using the Perl scripts in Amber. For more details about the MM-PBSA method applied, see section 5.6.3.

The necessity for the 100.05 ns equilibration stage is illustrated in Figure 34. This shows that large jumps in the free energy of binding are apparent in the first 50 ns of the simulation, correlating with significant structural movements in the binding site. In the next 50 ns the system is shown to have almost reached an equilibrium state with much smaller jumps being apparent. Finally, the last 50 ns (representing the production MD) display even steadier results. Similar patterns in the free energies of binding were seen for many of the inositol phosphate isomers.

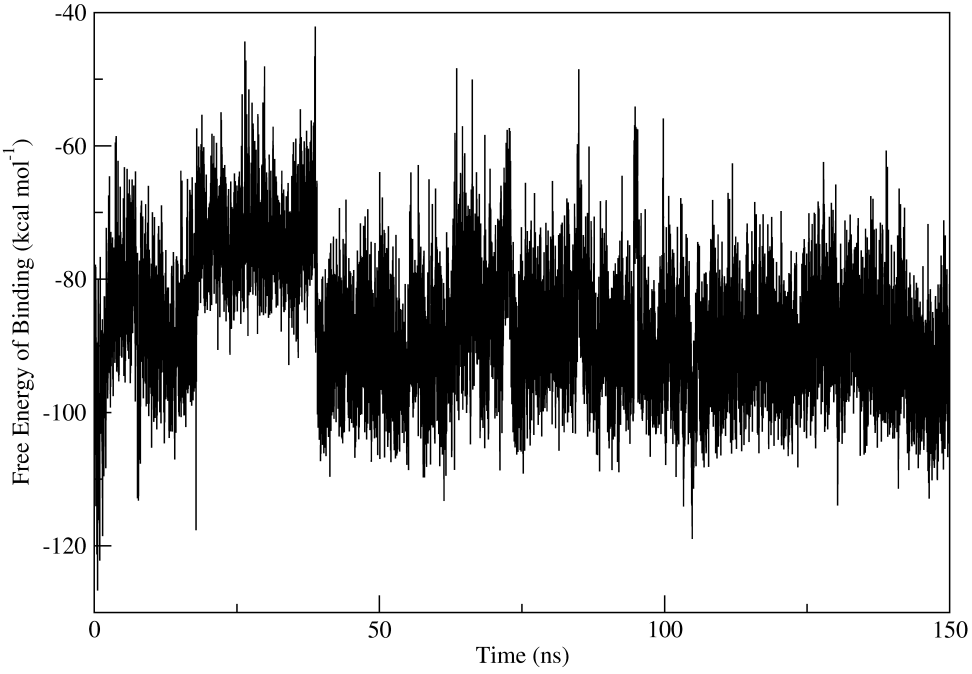


Figure 34: The calculated free energy of binding for a representative E17K PKB PH domain-Ins(1,3,4,5)P_4 complex over the 150 ns simulation.

6. 3. 3. 1 Weighted Free Energy Calculations

In order to allow a simple comparison of the computational results to *in vitro* studies, the free energies of binding for the different isomers of each inositol phosphate were weighted according to their *in vitro* probabilities, Figures 15, 16, 22 and 23 in Chapters 4 and 5 respectively.[3, 145, 146]

6. 3. 4 Hydrogen Bond Interactions

The Ptraj program within Amber 11 was used to analyse the hydrogen bond interactions for each simulation. A list of all potential hydrogen bond donors and acceptors was specified, and pair interactions between these were analysed for the 50 ns production MD simulations. A cut-off distance of 3.5 Å, and angle cut-off of 120 ° were specified for the pair interactions.

6. 4 Results and Discussion

6. 4. 1 E17K PKB PH Domain with Ins(1,4,5)P_3

MD simulations of the E17K PKB PH domain in a complex with Ins(1,4,5)P_3 allow the hypothesis that the E17K mutant has an increased affinity for PtdIns(4,5)P_2 to be investigated.

6. 4. 1. 1 Free Energies of Binding

Results in Table 13 illustrate that the weighted computational free energy of binding of Ins(1,4,5)P_3 to the E17K PH domain is over 8 kcal mol^{-1} greater than to the native PKB PH domain. Results for the individual Ins(1,4,5)P_3 isomers further support this result with each Ins(1,4,5)P_3 isomer binding with greater affinity to the mutant PKB PH domain than the native, illustrated in Tables S.4 and S.6 in the Electronic Supplementary Information.

Table 13: Weighted computationally determined binding free energies (denoted ΔG) of the native and E17K PKB PH domains to inositol phosphates.

Ligand	Computational ΔG for the native PKB PH domain (kcal mol^{-1})[a]	Computational ΔG for the E17K PKB PH domain (kcal mol^{-1})[a]
$Ins(1,3,4,5)P_4$	-71.51 $\pm$ 3.07	-75.67 $\pm$ 2.85
$Ins(1,3,4)P_3$	-65.61 $\pm$ 4.35	-65.04 $\pm$ 2.07
$Ins(1,4,5)P_3$	-40.45 $\pm$ 2.76	-48.24 $\pm$ 2.72

[a] The statistical error was estimated from the deviation between block averages for each simulation, and propagated appropriately for each weighted ligand.[115]

6. 4. 1. 2 Orientation and Molecular Behaviour

The increased affinity of the PKB PH domain for $Ins(1,4,5)P_3$ upon mutation is easily explained by studying this ligand's behaviour in the native and mutant binding sites. $Ins(1,4,5)P_3$ forms many similar interactions in both binding sites, with interactions between P1 and Arg23, and P4 and Lys14, Asn53 and Arg86. However additional interactions between Lys17 and $Ins(1,4,5)P_3$ are apparent in the mutant simulations. Lys17 is shown to form both direct and indirect (i.e. water-mediated) interactions with phosphate groups 4 and 5 of $Ins(1,4,5)P_3$. Direct interactions *via* phosphate group 5 are most apparent with an average occupancy of ~60%, compared to the ~20% occupancy of interactions between Lys17 and phosphate group 4. Additionally indirect interactions between Lys17 and the 6-hydroxyl group of $Ins(1,4,5)P_3$ also contribute to binding.

Visualisation of the simulations shows that, despite $Ins(1,4,5)P_3$'s additional interactions in the mutant binding site, this ligand is still not held tightly in one fixed orientation. This is because many of the interactions are formed *via* water molecules. The indirect nature of these interactions allows flexibility in $Ins(1,4,5)P_3$'s binding orientation.

It is important to note that whilst there is a level of flexibility in Ins(1,4,5)P_3's orientation in the mutant binding site, a fairly stable un-flipped binding mode is still maintained throughout the 150 ns simulations. This is in contrast to the native simulations which show the Ins(1,4,5)P_3 ligand to rotate 180° in the PKB binding site within 50 ns (Chapter 5).

6. 4. 2 E17K PKB PH Domain with 3-phosphorylated Inositol Phosphates

In order to further explore E17K PKB's enhanced affinity for the plasma membrane, the interaction between the E17K PKB PH domain and the 3-phosphorylated inositol phosphates was investigated.

6. 4. 2. 1 E17K PKB PH Domain with Ins(1,3,4,5)P_4

The weighted free energy of binding for Ins(1,3,4,5)P_4 to the E17K PKB PH domain illustrates that this ligand binds with approximately 4 kcal mol^{-1} greater affinity than to the native PKB PH domain, Table 13. Whilst the free energies of binding are close considering the level of error associated with the simulations, they do reflect the behaviour seen *in vitro* for the parent lipids: Landgraf *et al.* illustrated that in addition to the large increase in affinity for PtdIns(4,5)P_2, E17K PKB also showed a smaller increase in affinity for PtdIns(3,4,5)P_3.[164]

Whilst the free energy results for Ins(1,3,4,5)P_4 should be treated with caution, the molecular behaviour of this ligand in the native and mutant binding sites can still be explored. The simulations of Ins(1,3,4,5)P_4 in the mutant binding site illustrate that as well as maintaining the interactions from the native system, additional interactions between Ins(1,3,4,5)P_4 and Lys17 are apparent. The side-chain amine group of Lys17 is seen to form strong direct H-bond interactions with the oxygen atoms of phosphate group 5, with an average occupancy of 96%. These interactions are illustrated in Figure 35. Additional smaller interactions are also apparent between Lys17 and phosphate group 4 with an average occupancy of 31%, as well as indirect interactions between Lys17 and the 6-hydroxyl group.

These interactions between Ins(1,3,4,5)P_4 and Lys17 in the simulations are particularly interesting as many of these are not apparent in the crystal structure.[29] The crystal structure only reports indirect interactions between Lys17 and Ins(1,3,4,5)P_4. Furthermore, the crystal structure does not report any (direct or indirect) interactions between Lys17 and phosphate group 4.

This difference in agreement between the mutant crystal structure and the simulations may just be an artefact of using a static crystal structure for binding information. Alternatively the larger B-factor values for Ins(1,3,4,5)P_4 and its surrounding amino acid residues in the mutant crystal structure could potentially lead to errors in atom placement.[29] However, analysis of the atom placement using the Electron Density Server suggests it is of good quality.[165]

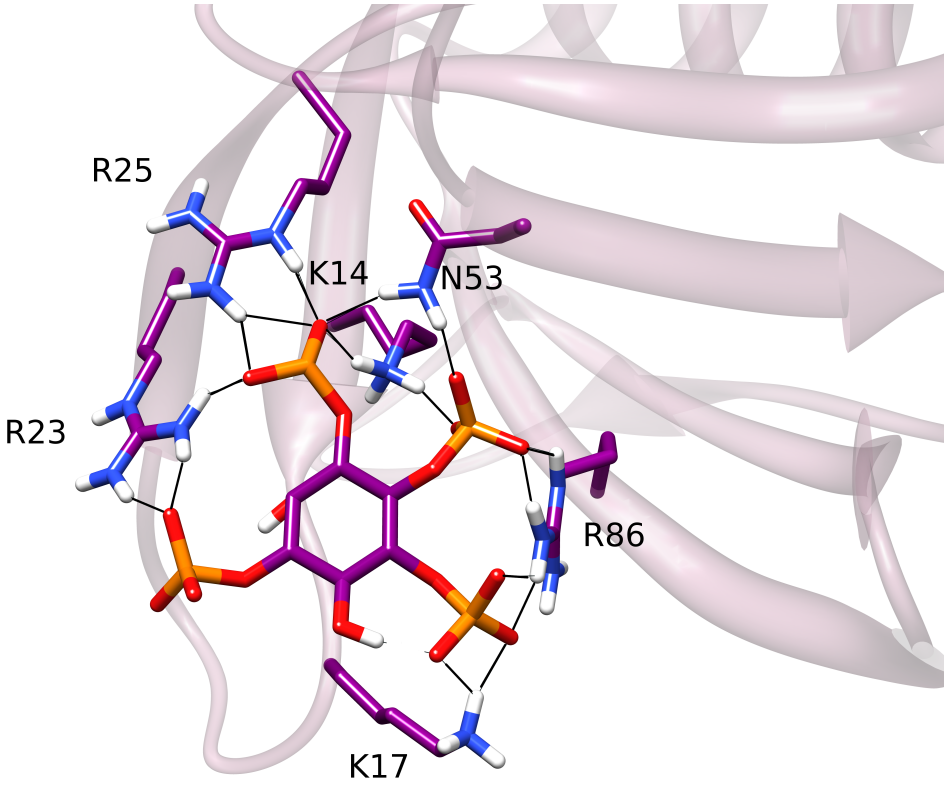


Figure 35: A representative structure of Ins(1,3,4,5)P_4 in the E17K PKB PH domain from the MD simulations. Hydrogen bond interactions are indicated with black lines. Ins(1,3,4,5)P_4 is seen to make similar interactions as in the native PKB binding site. Additional interactions of phosphate group 5 with K17 are also apparent.

6. 4. 2. 2 E17K PKB PH Domain with Ins(1,3,4)P_3

Ins(1,3,4)P_3 has a remarkably similar free energy of binding to both the native and E17K PKB PH domains, with results within 1 kcal mol^{-1} of each other (Table 13). This similar affinity can be attributed to the similar interactions the Ins(1,3,4)P_3 ligand makes with both binding sites.

No significant interactions between Ins(1,3,4)P_3 and Lys17 are apparent in the mutant simulations. This is due to the fact that Ins(1,3,4)P_3 does not contain a phosphate group at position 5. The 5-phosphate group of other inositol phosphates serves to not only form direct interactions with Lys17, but also brings Lys17 into close proximity for other smaller interactions *via* phosphate group 4 and the 6-hydroxyl group to occur.

6. 5 Conclusion

Simulations of the E17K PKB PH domain with Ins(1,3,4,5)P_4, Ins(1,3,4)P_3 and Ins(1,4,5)P_3 have helped explain this mutant's increased affinity for the plasma membrane.

The computational results suggest that the mutant's increased recruitment to the plasma membrane is largely due to a significantly improved affinity for the head-group of PtdIns(4,5)P_2. This result supports, and agrees with, previous experimental work in the field.[29, 164] Additionally, the simulations also capture the somewhat smaller increase in affinity of the E17K PKB PH domain for Ins(1,3,4,5)P_4. This again mirrors experimental data for Ins(1,3,4,5)P_4's parent lipid.[164] However this result should be treated with caution as the difference in Ins(1,3,4,5)P_4's computational (and *in vitro*) affinity for the native and mutant PH domains is fairly small, and the errors associated with the simulations are comparably sized. This highlights a potential limitation of using the MM-PBSA method to deal with this system, as it is not ideal for capturing the small differences in the binding affinities.[166] Additionally, the inability of the MM-PBSA method to account for the molecular

structure of water molecules may also affect contributions from water-mediated interactions.[167]

The simulations in this chapter provide an atomic-level view to help explain the increased affinity of E17K PKB for PtdIns(4,5)P_2 [and PtdIns(3,4,5)P_3]. Simulations show that interactions between Lys17 and phosphate group 5 are the main contribution to this increased affinity. Additionally smaller interactions between Lys17 with phosphate group 4 and hydroxyl group 6 are also favourable for binding. These interactions are particularly interesting as many are not apparent in the crystal structure of the E17K PKB PH domain with Ins(1,3,4,5)P_4.

The interaction of the E17K PKB PH domain with the inositol phosphate head-groups of PtdIns(3,4,5)P_3, PtdIns(3,4)P_2 and PtdIns(4,5)P_2 has been studied herein. Whilst it would naturally be of interest to study the parent lipids, as opposed to the inositol phosphate head-groups, there are a number of difficulties associated with this that are discussed further in Chapter 8. The use of the head-groups to extrapolate results for the phosphoinositide lipids is not unrealistic as it is largely the head-groups that mediate the interactions with the PKB PH domain. Furthermore for the E17K PKB PH domain, all inositol phosphates investigated maintained an orientation that would be accessible to the parent lipid.

Chapter 7 - Inositol Phosphate Derivatives

7. 1 Introduction

The previous chapters provide validation of the inositol phosphate parameters generated in this work. Additionally they also deliver a deeper understanding of the interaction between inositol phosphates and PKB's PH domain.

Interestingly, our simulations suggest that phosphate group 5 of Ins(1,3,4,5)P_4 is capable of interacting with PKB's PH domain. This interaction is particularly exciting as phosphate group 5 has previously been assumed to be solvent exposed when Ins(1,3,4,5)P_4 is bound to PKB.[26] In the simulations significant interactions between phosphate group 5 and Arg86, as well as more minor interactions with Gln79, are apparent.

Phosphate group 5 gained further interest in these studies after investigating the E17K PKB mutant. This mutant is shown to have an increased affinity for Ins(1,4,5)P_3 and Ins(1,3,4,5)P_4 due to the interaction of residue Lys17 with phosphate group 5 of both ligands.

In this chapter the aim is to a design a novel Ins(1,3,4,5)P_4 derivative that binds strongly to the PKB PH domain, preventing its recruitment to the plasma membrane and hence its subsequent phosphorylation and activation.[168] As many of PKB's downstream targets are involved in cell proliferation and survival, such a derivative would prove extremely useful as a potential drug lead and/or as a tool to further probe the intracellular functions of PKB.[60, 169]

The interactions of phosphate group 5 with the PKB PH domain provided an idea for designing an Ins(1,3,4,5)P_4 derivative. Rather than focusing on the 1-, 3- and 4-phosphate positions, which are already well optimised in their interactions with PKB, phosphate group 5 seemed a promising option to modify to improve binding. Additionally modification at this position is likely to be tolerated due to the area surrounding this group being fairly uncongested. Meanwhile phosphate groups 3 and

4 point directly into the tightly packed binding site.

To design a 5-position Ins(1,3,4,5)P_4 derivative with an increased affinity for the PKB PH domain, the aim is to target interactions with residues on loop 3 (i.e. Gln79 and Trp80). These residues serve as a potential source of interactions that have yet to be fully exploited. The desirability to target these residues is enhanced by the role they play in PKB activation. As mentioned in section 1.2.2.1, loop 3 of the PKB PH domain is buried in PKB's kinase domain in its inactive state. Upon binding to 3-phosphorylated phosphoinositides, the PH and kinase domains are thought to separate allowing phosphorylation and activation of PKB.[28, 42] We are encouraged by previous success in targeting Trp80 with the allosteric inhibitor VIII, which has been shown to lock PKB into an inactive conformation, Figure 36.[28, 170, 171]

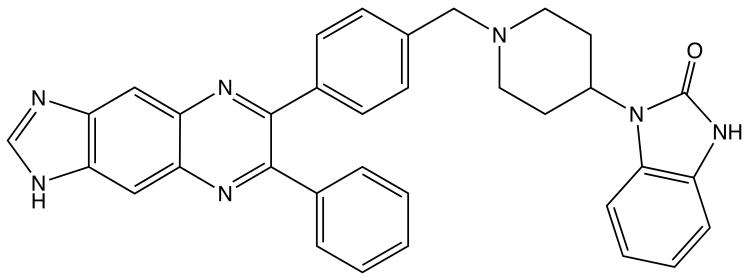


Figure 36: Schematic drawing of inhibitor VIII.

7. 1. 1 Previous Examples of Inositol Phosphate Derivatives

There have been limited attempts at designing inositol phosphate- or phosphoinositide-based derivatives to target the PKB PH domain. However the recently reported inhibitory effects of endogenous $InsP_7$ [5-PP-Ins(1,2,3,4,6)P_5] on PKB suggest this is a promising area of research.[161]

Kozikowski *et al.* have already demonstrated the use of phosphatidylinositol ether lipid analogues as potential therapeutic treatments, Figure 37. These analogues have

been shown to inhibit PKB with low micromolar activity, and additionally reduce the growth of lung and breast cancer cell lines.[159, 172, 173]

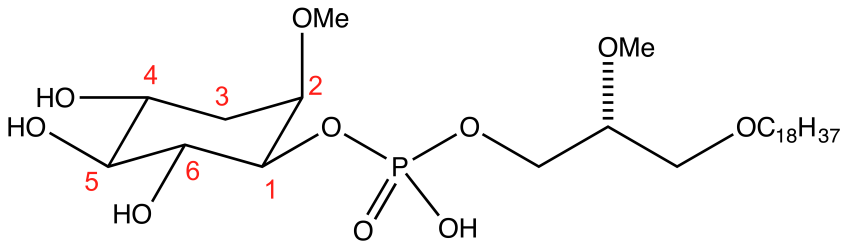
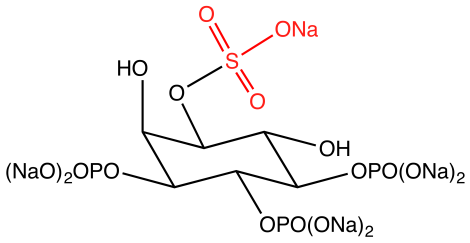
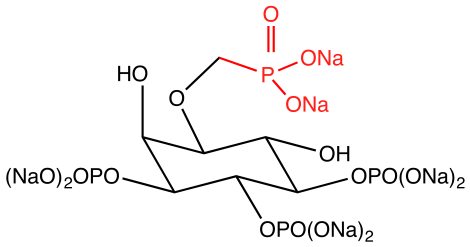
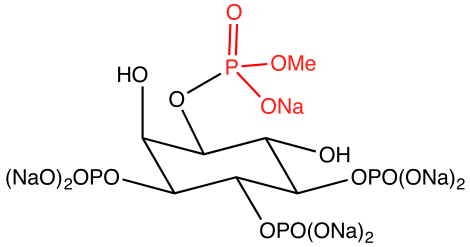
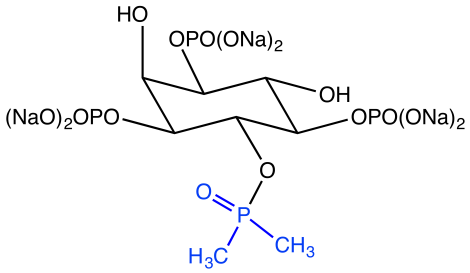
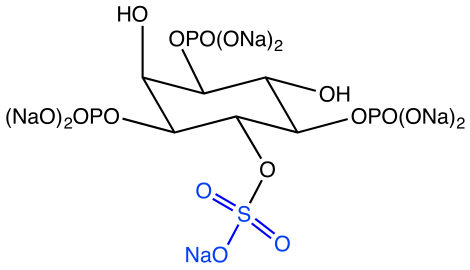
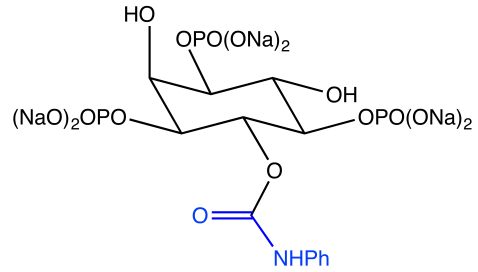
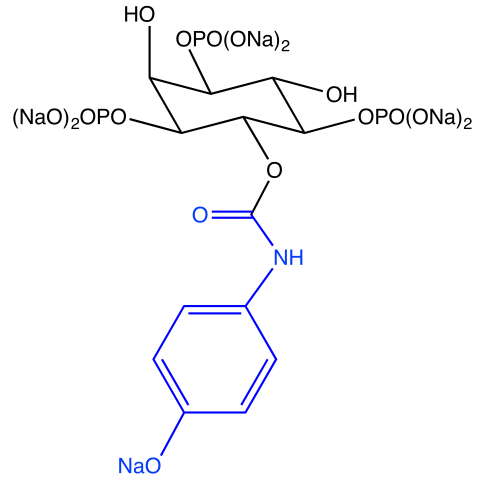
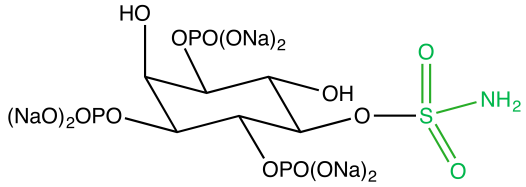
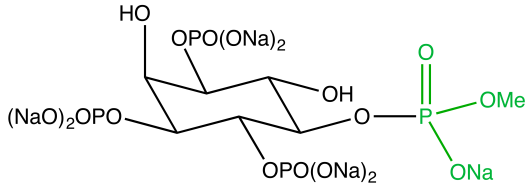
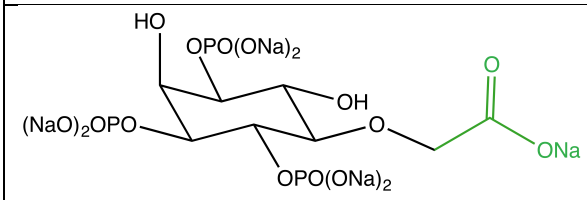


Figure 37: An example of a phosphatidylinositol ether lipid analogue (PIA) designed by Kozikowski *et al.*[159] Positions 4 and 5 are thought to undergo phosphorylation upon entering the cell.

Examples of inositol phosphate-based derivatives largely come from the Conway group at Oxford University. Their work so far has focused on designing inositol phosphate derivatives that have a single substitution of a phosphate group for a phosphate bioisostere. Table 14 illustrates the Ins(1,3,4,5)P_4 derivatives that have been synthesised by PhD members Nemeth and Elliott of the Conway group.[174, 175] The ability of these derivatives to bind to the PKB and GRP1 PH domains is also noted.

Table 14: Summary of the novel Ins(1,3,4,5)P_4 derivatives synthesised by Elliott and Nemeth.[174, 175] Their biological activity against the PKB and GRP1 PH domains is provided, taken from Elliott's thesis.

Position Modified	Derivative	PKB PH Domain Binder	GRP1 PH Domain Binder
1-position modified		No	Yes
		No	Yes
		Yes	Yes
4-position modified		No	No
		No	No

		No	No
		No	No
5-position modified		Yes	No
		ND	No
		ND	No

As illustrated in Table 14, work conducted within the Conway group includes the design and synthesis of 4-position Ins(1,3,4,5)P_4 derivatives by Nemeth.[175] Biological analysis of these derivatives illustrated that they exhibit no binding affinity for the PKB PH domain [No binding was detected even at concentrations of 1mM]. This data therefore supports the argument that a significant modification at phosphate group 4 is unlikely to be tolerated. Building upon Nemeth's work, Elliott broadened the possible modification sites of Ins(1,3,4,5)P_4 to the 1- and 5-phosphate groups.[174] Results from his work showed that modifications at the 1-position also resulted in poor binding to the PKB PH domain. The only exception to this was a 1-position methylphosphate ester.

Whilst only one 5-position derivative was biologically tested against the PKB PH domain, it was promisingly shown to bind therefore supporting the suggestion that modification at the 5-position is tolerated. Additionally, the 5-position Ins(1,3,4,5)P_4 derivatives were shown not to bind to the GRP1 PH domain, suggesting that this position highlights an opportunity for improving PH domain selectivity.

A similar approach to the Conway Group has been carried out by Falasca *et al.,* who have targeted the PH domain of PDK1. They have concentrated on the inhibitory effects of Ins(1,3,4,5,6)P_5 and its derivative 2-*O*-Bn-InsP_5 to disrupt PDK1 translocating to the membrane.[176]

7. 1. 2 A Collaboration with the Conway Group

The work within the following sections was performed in close collaboration with researchers of the Conway group at Oxford University. To clarify, all synthesis and purification of derivatives was performed external to this study by Aine Slowey of the Conway group, unless explicitly stated otherwise. All computational parameterisation and MD simulations performed as part of this study were carried out by myself. Additionally, all *in vitro* biochemical binding and inhibition assays (other than the *in vitro* phosphorylation assay in section 7.4.7) were performed by myself and form part of this study. Any reference to further work by external collaborators is explicitly identified in the text and referenced appropriately.

In collaboration with the Conway group, the aim was to design a 5-position Ins(1,3,4,5)P_4 derivative with an increased affinity for the PKB PH domain, compared to Ins(1,3,4,5)P_4.

The remainder of this chapter is split into three sections:

- Section 7.2 - In the first of these sections preliminary MD simulations were used to screen several 5-position Ins(1,3,4,5)P_4 derivatives. These derivatives were designed using knowledge from our previous MD simulations, in collaboration with the Conway group. The derivatives were screened according to their affinity for, and interactions with, the PKB PH domain. The derivatives were additionally assessed for their synthetic potential. After these considerations were taken into account, the most promising derivative was synthesised by Aine Slowey of the Conway group.

- Section 7.3 - After selecting a 5-position derivative to synthesise, its binding affinity for the PKB PH domain and molecular behaviour was more thoroughly examined by considering all possible ionisation states of the derivative in the MD simulations. Three other derivatives recently synthesised in the Conway group were also investigated using MD simulations. Again all ionisation states were considered.

- Section 7.4 - Once the behaviour of the four Ins(1,3,4,5)P_4 derivatives was predicted from the computational simulations in section 7.3 the binding affinity for, and inhibitory actions of, these derivatives on the PKB PH domain was assessed using *in vitro* techniques.

7. 2 Screening 5-Position Ins(1,3,4,5)P_4 Derivatives

Using knowledge from the previous MD simulations, a number of 5-position derivatives were designed in collaboration with the Conway group. Previous simulations suggested that interactions between the ligand and residues on loop 3 of the PKB PH domain (i.e. Gln79 and Trp80) are a source of interactions that have yet to be fully exploited. Therefore, 5-position Ins(1,3,4,5)P_4 derivatives were designed to target these residues.

The 5-position phosphomonoester group of Ins(1,3,4,5)P_4 was substituted with: an ethylphosphate ester (**5E**), a carboxymethyl ether (**5C**), a pyrophosphate ester (**5D**) a triphosphate ester (**5T**), and an acetamidoethylphosphate ester (**5A**), illustrated in Figure 38.

Preliminary simulations were run of these 5-position Ins(1,3,4,5)P_4 derivatives in a complex with the PKB PH domain. Unfortunately due to time constraints we were unable to investigate all possible isomers of each derivative mentioned. To worsen matters, *in vitro* probability data was not readily available for these derivatives. Therefore in order to decide which isomer to study, numerous assumptions were made. In agreement with the Conway group, all phosphate groups other than the 5-phosphate were assumed singly protonated. This assumption was based on the pK_a values for H_3PO_4 being 2.12, 7.21 and 12.32 for the first, second and third deprotonations respectively.[177] The ethylphosphate, carboxymethyl, pyrophosphate, and triphosphate moieties were all assumed fully deprotonated for ease of consistency, and additionally in the hope of maximising the interactions of these groups with PKB's cationic pocket.

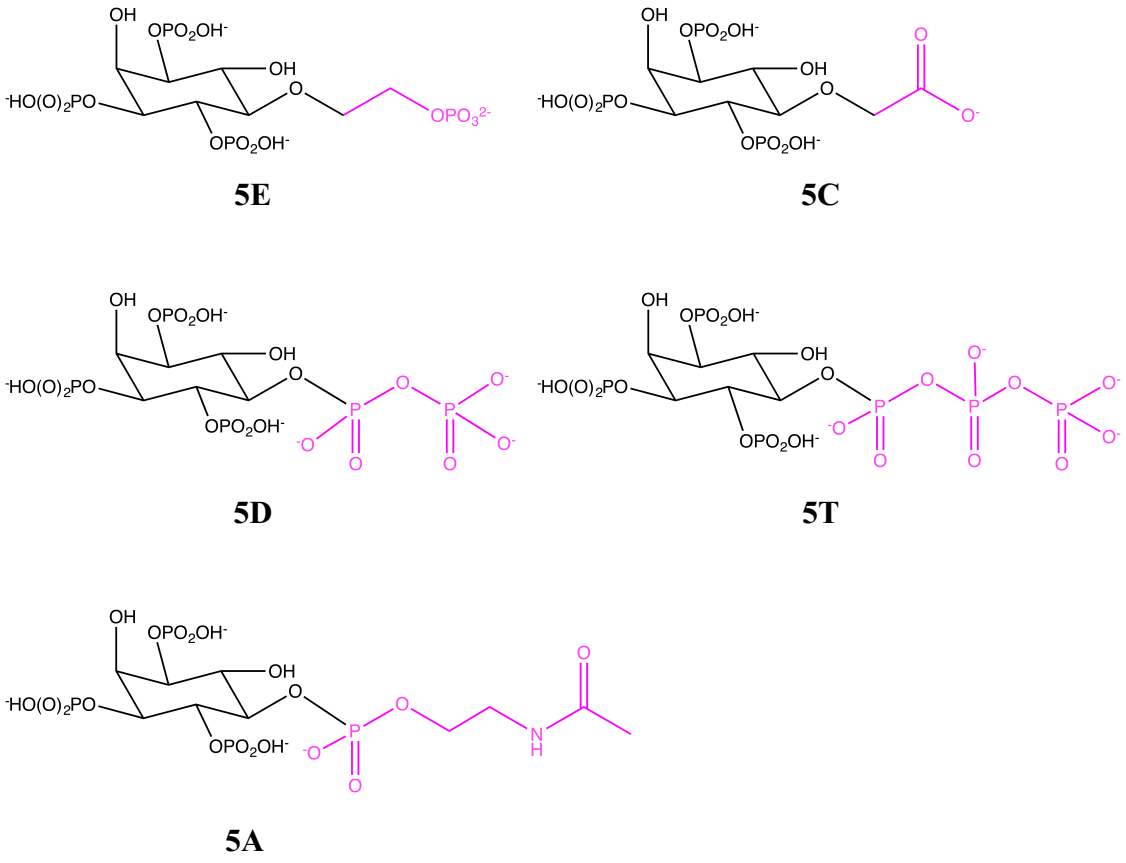


Figure 38: Summary of the novel Ins(1,3,4,5)P_4 derivatives designed in this work. The derivatives are shown in the ionisation state they were investigated in the MD simulations.

7. 2. 1 Program Details

Gaussian 09[122] was used for the geometry optimisation calculations of the Ins(1,3,4,5)P_4 derivatives and to calculate their molecular electrostatic potential (MEP). The RESP program in Amber was used to generate the charge parameters by fitting the QM calculated MEP to atom-centred points. Graphical representations were produced with Gaussview.[123]

The Amber11[124] suite of programs implementing the ff99SB force field was used for energy minimisation, equilibration and production MD calculations. Parameter and topology files for the inositol phosphate derivative-PKB PH domain complexes were creating using the LEaP module. Minimisation, equilibration and production MD calculations were performed with PMEMD. Analysis of hydrogen bonding was carried out using Ptraj. Free energy calculations, using the MM-PBSA method, were carried out using the Perl scripts in Amber. Graphical representations were produced with VMD and Chimera.[125, 126]

7. 2. 2 Parameterisation of 5-position Ins(1,3,4,5)P_4 Derivatives

The derivatives were parameterised using the same protocol as for the inositol phosphates, detailed below. The initial structures for these derivatives were adapted from that of Ins(1,3,4,5)P_4 in Gaussview, by substituting phosphate group 5 with the appropriate functional group.

Each derivative was geometry optimised until it reached a local minimum on the potential energy surface. The derivatives were first optimised using the DFT-B3LYP/3-21g* level of theory and then further optimised using DFT-B3LYP/6-31g**. These calculations were carried out in gas phase. For each optimised structure the MEP was calculated using the HF/6-31g* level of theory. The RESP algorithm was employed to fit the QM calculated MEP to atom-centred points. The RESP generated point charges were assigned as the charge parameters. Atom types were assigned from the ff99SB force field, using analogous atom types to Ins(1,3,4,5)P_4.

7. 2. 3 Minimisation, Equilibration and Production Run

The X-ray crystal structure of the PKB PH domain in a complex with Ins(1,3,4,5)P_4 provided the initial co-ordinates, PDB accession code 1UNQ.[11] The PDB was modified by substituting phosphate group 5 with the appropriate functional group. Crystallographic water molecules present in the PDB were kept. The Leap module of Amber was used to add hydrogen (and missing) atoms to the crystal structure. The systems were charge neutralised with Na^+ counter-ions and solvated with TIP3P water molecules within a truncated octahedral unit cell, with a minimum of 8 Å distance from the edge of the box to any protein or ligand atom.

The complexes were minimised in two stages. In the first stage, the water molecules and Na^+ counter-ions were energy minimised, whilst harmonic restraints of 500 kcal mol^{-1} $Å^{-2}$ were applied to the protein and ligand; 500 steps of steepest descent and 500 steps of conjugate gradient minimisation were carried out. In the second stage of minimisation, the restraints were removed, and the entire system was minimised using 1500 steps of steepest descent and 2000 steps of conjugate gradient minimisation.

In all simulations, periodic boundary conditions were applied. A cut-off of 8 Å was used for the explicit calculation of non-bonded interactions. For the calculation of long-range electrostatic interactions, the particle mesh Ewald method was applied with a direct space cut-off of 8 Å.[98] Each minimised system was heated from 0 K to 300 K at constant volume over 50 ps. Following this, 100.05 ns of equilibration was carried out at constant pressure (1 bar) and temperature (300 K). The equilibration was run for 100.05 ns to allow the PH domain and inositol phosphate derivatives to adjust their conformation/orientation and reach equilibrium. Production MD simulations of each system were then run for 50 ns at constant pressure (1 bar) and temperature (300 K), with snapshots saved every 10 ps.

For each simulation the SHAKE algorithm was applied to constrain all bonds involving hydrogen atoms, allowing a time-step of 2 fs to be used.[102] In order to maintain the temperature during the simulations, Langevin dynamics was used with a

collision frequency of 1.0 ps^{-1}. MD simulations were performed with the GPU accelerated version of PMEMD in Amber11.

7. 2. 4 Free Energy Calculations

MM-PBSA calculations were implemented to calculate the free energy of binding for the inositol phosphate derivative–PKB PH domain complexes. Snapshots from the 50 ns production MD simulations were used in these calculations. Structural snapshots were saved every 10 ps, giving a total of 5000 snapshots for each simulation. Each snapshot was stripped of water and counter-ions. MM contributions were calculated with an interaction cut-off of 999 Å and a dielectric constant of 1.0. Contributions to the solvation energy were calculated using the pbsa module of Amber. An internal dielectric of 1.0, an external dielectric of 80.0, a solvent radius of 1.4 Å and grid spacing of 0.5 Å was applied. Non-polar contributions to the solvation energy were calculated with a solvent surface tension of 0.0072 kcal mol^{-1} $Å^{-2}$. The SASA was calculated with the Molsurf program with a solvent probe radius of 1.4 Å and Bondi atomic radii.[147, 148]

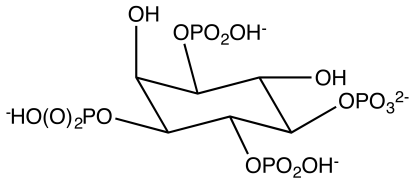
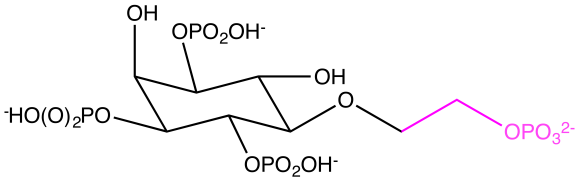
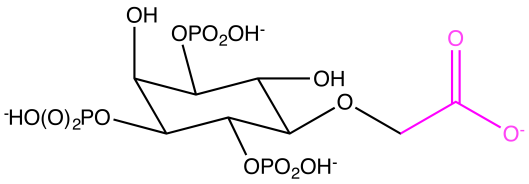
7. 2. 5 Hydrogen Bond Interactions

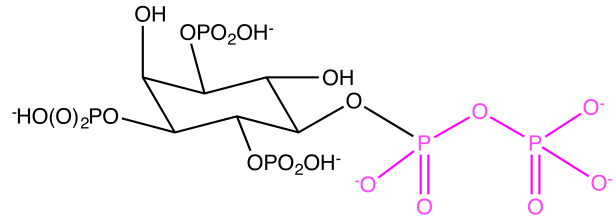
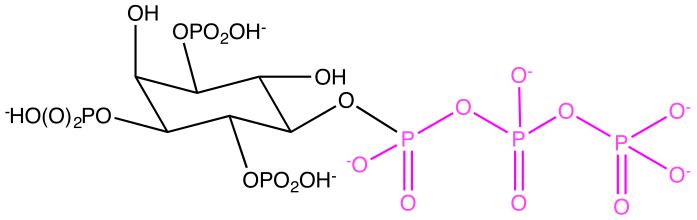
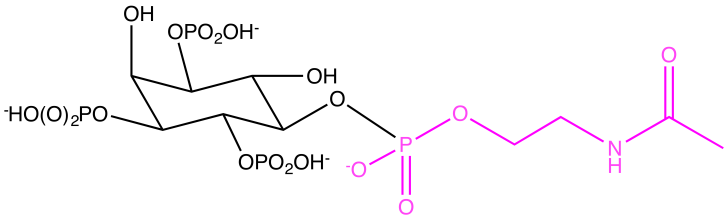
The Ptraj program within Amber 11 was used to analyse the hydrogen bond interactions for each system. A list of all potential hydrogen bond donors and acceptors was specified, and pair interactions between these were analysed for each 50 ns production MD simulation. A cut-off distance of 3.5 Å, and angle cut-off of 120° were specified for the pair interactions.

7. 2. 6 Results and Discussion

MD simulations (and MM-PBSA calculations) were used to assess the affinity of the 5-position Ins(1,3,4,5)P_4 derivatives for the PKB PH domain, Table 15. Additionally, the simulations were also used to investigate the derivatives' binding modes and interactions with the PKB PH domain.

Table 15: Computationally determined free energies of binding for the 5-position Ins(1,3,4,5)P_4 derivatives to the PKB PH domain, from the preliminary simulations. The result for Ins(1,3,4,5)P_4 was taken from previous work in chapter 4.

Derivative	Net charge of derivative	Free energy of binding to the PKB PH domain (kcal mol^{-1}) [a]
	-5	-69.46 ± 1.39
 **5E**	-5	-59.07 ± 6.66
 **5C**	-4	-59.12 ± 5.82

 **5D**	-6	-71.75 ± 5.90
 **5T**	-7	-80.62 ± 2.90
 **5A**	-4	-88.49 ± 5.81

[a] The statistical error was estimated from the deviation between block averages for each simulation.[115]

The initial derivative proposed was the ethylphosphate ester **5E**. It was hoped that by extending the linker between the inositol ring and the 5-phosphate group that interactions between the phosphate moiety and Gln79 would be facilitated. A variation of this derivative was also considered in the form of the carboxymethyl ether **5C.** This derivative (with the same linker length as **5E**) was also considered, as carboxylate moieties are widely employed as phosphate bioisosteres.[178]

Simulations of these derivatives suggest that removal of the 5-phosphate group immediately attached to the inositol ring is detrimental to binding (to the PKB PH domain). Results show that the ethylphosphate ester **5E** and carboxymethyl ether **5C** derivatives bind to the PKB PH domain with a significantly lower affinity than

Ins(1,3,4,5)P_4. This is unsurprising as when the phosphate linker of phosphatidylinositol ether lipid analogues was replaced with a carbonate linker, the inhibitory effects of these compounds was also lost.[159]

The lack of 5-phosphate group immediately attached to the inositol ring means that interactions *via* this site are not possible. The phosphate and carboxylate moieties (on the extended linker) therefore attempt to replicate/replace these interactions with the PKB binding site. However, the orientations the derivatives adopt to maintain these interactions are significantly different to that of Ins(1,3,4,5)P_4, Figure 39. A change in orientation of the ligand is unlikely to be favourable for binding due the already mentioned well-optimised interactions of the 1-, 3- and 4-phosphate groups in the native system. Therefore it is essential to conserve the 5-phosphate group in future derivatives in the hope of stabilising the Ins(1,3,4,5)P_4 ligand in its native conformation, whilst adding further favourable interactions by potentially interacting with loop 3 of the PKB PH domain.

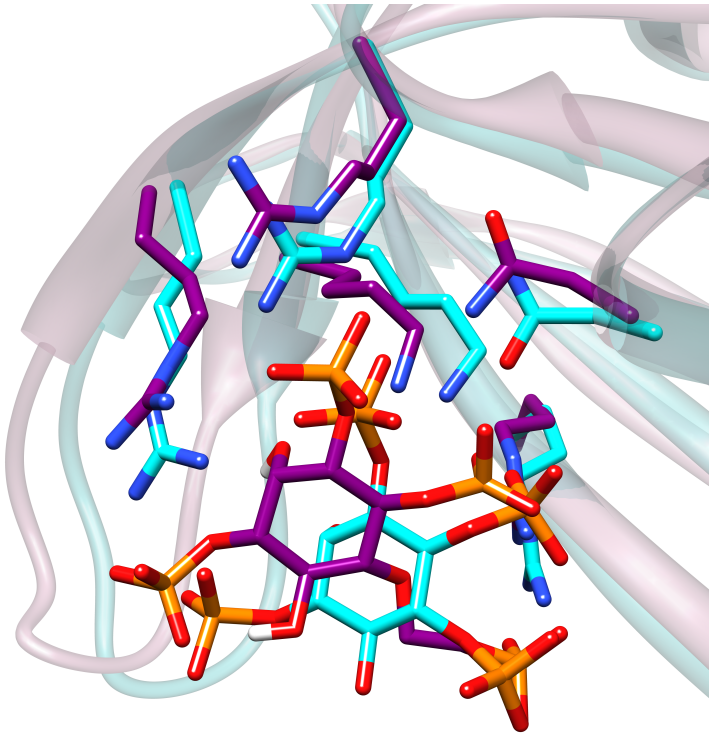


Figure 39: The conformation of the ethylphosphate derivative 5E in the PKB binding site (magenta). The crystal structure of the PKB PH domain and Ins(1,3,4,5)P_4 is shown in cyan for comparison. The altered orientation of the derivative and the binding site residues compared to the crystal structure is illustrated.

A pyrophosphate derivative **5D** was subsequently designed, in the hope of maintaining the native interactions of Ins(1,3,4,5)P_4 whilst extending the phosphate moiety towards loop 3 (and Gln79). Unfortunately this derivative showed a very similar binding affinity to Ins(1,3,4,5)P_4. Analysis of the simulations illustrate that this is due to the fact that both ligands adopted very similar binding modes, and formed similar interactions, in the PKB binding site. The pyrophosphate group did not result in additional interactions with loop 3, but rather maintained similar interactions to that of phosphate group 5 in Ins(1,3,4,5)P_4.

An additional phosphate group was added to the above derivative to form a triphosphate derivative **5T**. Arg86 was shown to form strong hydrogen bond interactions with the triphosphate group (*via* $N\eta^1$ and $N\eta^2$), whilst still maintaining its interactions with phosphate group 4 (*via* $N\varepsilon$ and $N\eta^2$), Figure 40. These hydrogen bond interactions were fully occupied for almost the entire simulation. Furthermore, additional smaller interactions between the triphosphate group and Tyr18 on loop 1 were also apparent. These favourable interactions resulted in an increased free energy of binding for this derivative to the PKB PH domain, compared to Ins(1,3,4,5)P_4.

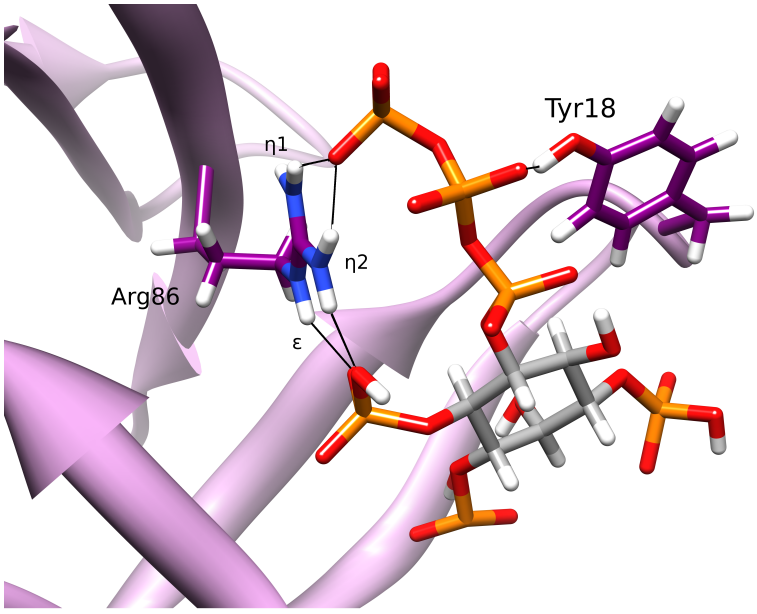


Figure 40: A snapshot of the interactions between the triphosphate derivative 5T and the PKB PH domain. Hydrogen bond interactions are indicated with black lines. Interactions of Arg86 with the triphosphate moiety and phosphate group 4 are illustrated. Additionally interactions between the triphosphate moiety and Tyr18 on loop 1 are also highlighted.

Whilst the triphosphate derivative showed an increased binding affinity for the PKB PH domain, the aim to target loop 3 was still unachieved. In an attempt to interact with loop 3, residue Gln79 on its tip was chosen as a suitable target. An amide-based derivative, **5A**, was therefore designed, in the hope of exploiting glutamine's propensity to form hydrogen bonds.

The simulation of derivative **5A** with the PKB PH domain illustrated that hydrogen bond interactions between the acetamidoethylphosphate ester and Gln79 are able to form *via* both their carbonyl and amine moieties. These interactions were shown to cause loop 3 to move closer towards the phosphoinositide binding site. As a consequence of this movement, a strong intra-molecular hydrogen bond interaction between Asn54 and Trp80 was shown to form, Figure 41. The formation of this hydrogen bond is particularly interesting, as these residues are 14 Å apart in the crystal structure of the PKB PH domain in a complex with Ins(1,3,4,5)P_4.[11] Free energy analysis highlights how the additional interactions mentioned result in a relatively high free energy of binding for this derivative to the PKB PH domain, almost 20 kcal mol^{-1} greater than the equivalent Ins(1,3,4,5)P_4 isomer.

Based on the relatively high binding affinity and molecular behaviour of the acetamidoethylphosphate derivative **5A** it was chosen for synthesis and further evaluation.

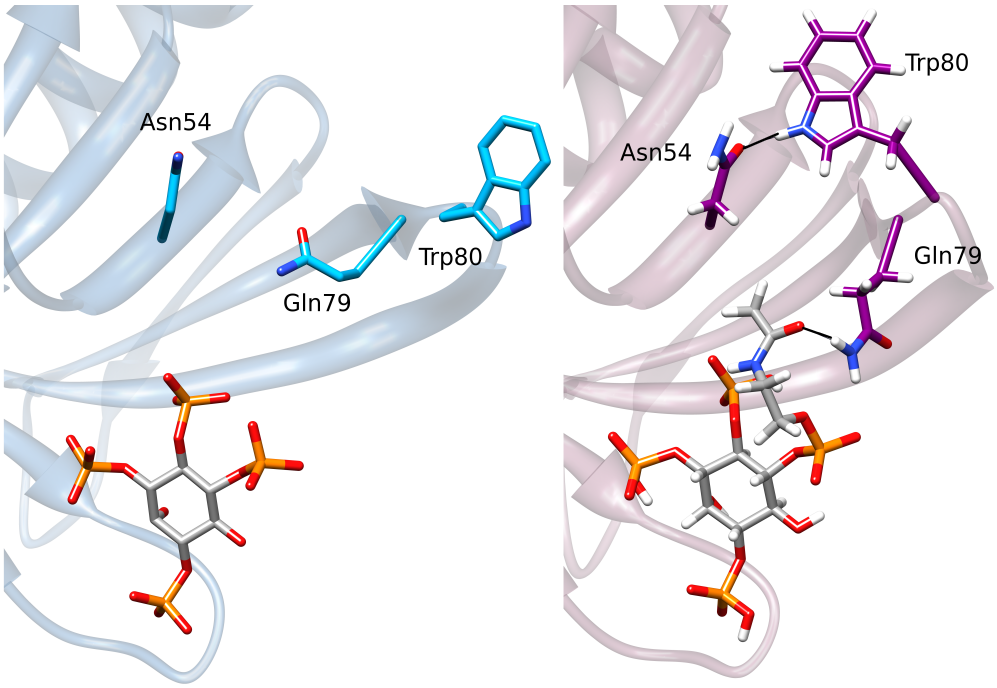
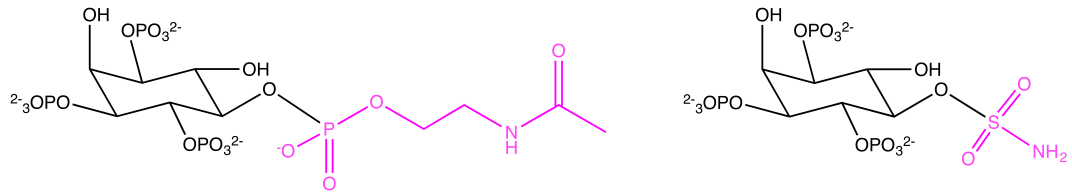


Figure 41: A representative snapshot from the simulation of the PKB PH domain with the Ins(1,3,4)P_3-5-*O*-acetamidoethylphosphate ester (magenta). The crystal structure of the PKB PH domain in a complex with Ins(1,3,4,5)P_4 is shown as a comparison (cyan). The significant distance between Asn54 and Trp80 in the crystal structure is highlighted. In contrast when the Ins(1,3,4)P_3-5-*O*-acetamidoethylphosphate ester is bound to the PKB PH domain, a strong intra-molecular interaction between Asn54 and Trp80 is apparent. Additionally interactions between the acetamidoethylphosphate ester and Gln79 are highlighted.

7. 3 Detailed Computational Investigation of Selected Inositol Phosphate Derivatives

The acetamidoethylphosphate ester **5A** was selected for further analysis using both computational and biochemical techniques. Three other Ins(1,3,4,5)P_4 derivatives recently synthesised in the Conway group were also selected for investigation. These included a 5-position sulfamate-based derivative synthesised by T. Elliott and 3-position derivatives synthesised by A. Slowey, Figure 42.

A thorough computational analysis of these derivatives was carried out prior to their biochemical evaluation. MD simulations allowed the derivatives' binding affinities for the PKB PH domain [compared to Ins(1,3,4,5)P_4] to be predicted, as well as their molecular behaviour to be assessed. In order to thoroughly investigate these derivatives all physiologically accessible ionisation states were considered in the simulations.



5A **5S**

3M **3S**

Figure 42: The 5- and 3-position Ins(1,3,4,5)P_4 derivatives fully investigated using MD simulations. All physiologically accessible isomers were investigated.

7. 3. 1 Parameterisation of Selected Ins(1,3,4,5)P_4 Derivatives

The 5-position acetamidoethylphosphate ester **5A** and 3-position methylphosphate ester **3M** were parameterised using the same protocol as in section 7.2.2 for the other inositol phosphate derivatives. Eight isomers with different levels and arrangements of protonation were investigated for each. The phosphate moieties of the 3-position methylphosphate group and the 5-position acetamidoethylphosphate group were assumed singly deprotonated. The reason for this is that these groups cannot be dianionic due to their substitution, and full protonation is extremely unlikely *in vitro* and *in vivo.*[177]

Each isomer was first optimised using the DFT-B3LYP/3-21g* level of theory and then further optimised using DFT-B3LYP/6-31g** in Gaussian09. These calculations were carried out in gas phase. For each optimised structure the MEP was calculated using the HF/6-31g* level of theory. The RESP algorithm was employed to fit the QM calculated MEP to atom-centred points. The RESP generated point charges were assigned as the charge parameters. Atom types were assigned from the ff99SB force field. Parameters for these derivatives are supplied in the Electronic Supplementary Information, sections S.7 and S.9.

7. 3. 1. 1 Sulfamate-based derivatives

The sulfamate-based derivatives, at both the 3- and 5-positions, were not optimised using the same protocol as the other inositol phosphate derivatives. This was due to complications with the protocol when applying it to these derivatives. When optimised with the B3LYP/3-21g* level of theory followed by B3LYP/6-31g** in gas phase, many of the sulfamate-based isomers failed to reach convergence. A different approach to parameterise these molecules was therefore necessary.

Based on the multiple approaches investigated in Chapter 4, it was decided to carry out the optimisations in implicit solvent instead [i.e. in water (ε=78.3553)]. This adaptation to the protocol allowed all of the sulfamate-based isomers to optimise to low energy conformations. However, upon visualisation of the optimised structures it

was apparent that in many cases a proton from the sulfamate group had migrated onto the neighbouring phosphate group. This structure in which the sulfamate group is deprotonated is extremely unlikely to exist *in vivo*, as the pK_a of $MeSO_2NH_2$ is 17.5.[179]

To overcome this problem the Hartree Fock method was used in place of DFT-B3LYP. The HF method was chosen due to its previous use in the principal study for generating force field parameters for sulfamates.[180] The isomers were each optimised using the HF/6-31g** level of theory in implicit solvent (water). This resulted in sensible structures without proton migration. Therefore subsequent optimisations in lower dielectric solvents were also investigated. Optimisations in octanol (ε=9.8629) were deemed successful, however further optimisations in even lower dielectric constants (e.g. cyclohexane) resulted in the proton migration problem mentioned. Therefore the optimised structures in octanol were used for the following steps.

Following optimisation, the MEP for each isomer was calculated using the HF/6-31g* level of theory in gas phase. The RESP algorithm was employed to fit the QM calculated MEP to atom-centred points. The RESP generated point charges were assigned as the charge parameters. Atom types were assigned from the ff99SB force field. Missing parameters not available from the ff99SB force field, were taken from a study by Huige and Altona based on sulfamates.[180] Any parameters still not accounted for were taken from the GAFF force field. The parameters for these derivatives are provided in the Electronic Supplementary Information, sections S.8, S.10 and S.11.

7. 3. 2 Minimisation, Equilibration and Production Run

The X-ray crystal structure of the PKB PH domain in a complex with Ins(1,3,4,5)P_4 provided the initial co-ordinates, PDB accession code 1UNQ.[11] The PDB was modified by substituting phosphate group 3/5 with the appropriate functional group. Crystallographic water molecules present in the PDB were kept. The Leap module of Amber was used to add hydrogen (and missing) atoms to the crystal structure. The systems were charge neutralised with Na^+ counter-ions and solvated with TIP3P water

molecules within a truncated octahedral unit cell, with a minimum of 8 Å distance from the edge of the box to any protein or ligand atom.

The complexes were minimised in two stages. In the first stage, the water molecules and Na^+ counter-ions were energy minimised, whilst harmonic restraints of 500 kcal mol^{-1} $Å^{-2}$ were applied to the protein and ligand; 500 steps of steepest descent and 500 steps of conjugate gradient minimisation were carried out. In the second stage of minimisation, the restraints were removed, and the entire system was minimised using 1500 steps of steepest descent and 2000 steps of conjugate gradient minimisation.

In all simulations, periodic boundary conditions were applied. A cut-off of 8 Å was used for the explicit calculation of non-bonded interactions. For the calculation of long-range electrostatic interactions, the particle mesh Ewald method was applied with a direct space cut-off of 8 Å.[98] Each minimised system was heated from 0 K to 300 K at constant volume over 50 ps. Following this, 100.05 ns of equilibration was carried out at constant pressure (1 bar) and temperature (300 K). Production MD simulations of each system were then run for 50 ns at constant pressure (1 bar) and temperature (300 K), with snapshots saved every 10 ps.

For each simulation the SHAKE algorithm was applied to constrain all bonds involving hydrogen atoms, allowing a time-step of 2 fs to be used.[102] In order to maintain the temperature during the simulations, Langevin dynamics was used with a collision frequency of 1.0 ps^{-1}. MD simulations were performed with the GPU accelerated version of PMEMD in Amber11.

7. 3. 3 MM-PBSA Calculations

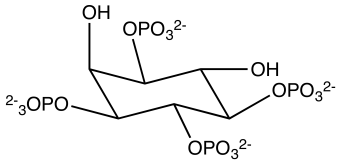
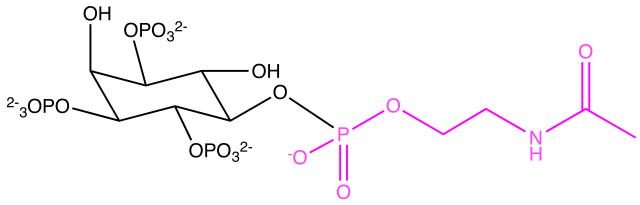
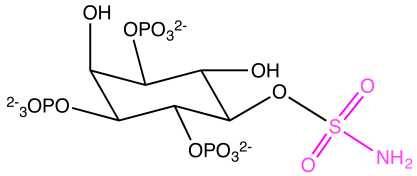
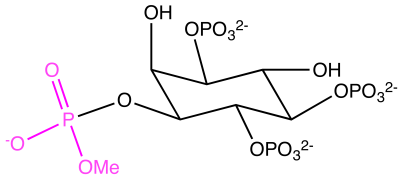
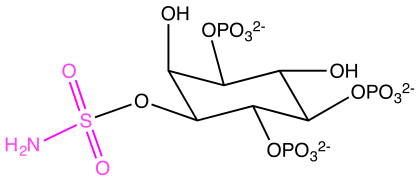
MM-PBSA calculations were implemented to calculate the free energy of binding for the inositol phosphate derivative–PKB PH domain interactions. Snapshots from the 50 ns production MD simulations were used in these calculations. MM-PBSA calculations were carried out using the Perl scripts in Amber, detailed in section 7.2.4.

The binding free energy results for the different isomers of each derivative were weighted. The 3-position derivatives were weighted according to the *in vitro* probabilities of the Ins(1,4,5)P_3 isomers and the 5-position derivatives according to the Ins(1,3,4)P_3 isomers.

7. 3. 4 Results and Discussion

The weighted free energies of binding (Table 16) and molecular behaviour of the selected Ins(1,3,4,5)P_4 derivatives provide yet more information about the PKB binding site. Each of these derivatives is discussed in detail in the following sections.

Table 16: Computationally determined free energies of binding for the selected 5- and 3-position Ins(1,3,4,5)P_4 derivatives. All physiologically accessible isomers were considered.

Derivative	Weighted free energy of binding (kcal mol^{-1}) [a]
	-71.51 ± 3.07
 **5A**	-75.04 ± 2.88
 **5S**	-65.20 ± 3.87
 **3M**	-66.37 ± 1.91
 **3S**	-50.44 ± 2.21

[a] The statistical error was estimated from the deviation between block averages for each simulation and propagated appropriately for each weighted ligand.[115]

7. 3. 4. 1 D-*myo*-Ins(1,3,4)P_3-5-*O*-acetamidoethylphosphate ester

As expected, the Ins(1,3,4)P_3-5-*O*-acetamidoethylphosphate ester **5A** is shown to have an increased affinity for the PKB PH domain compared to Ins(1,3,4,5)P_4. However this increased affinity is significantly lower than was suggested by the preliminary simulation of this derivative in section 7.2.6.

In the preliminary simulation, in which only one ionisation state was considered, the Ins(1,3,4)P_3-5-*O*-acetamidoethylphosphate ester was shown to bind to the PKB PH domain with approximately 20 kcal mol^{-1} greater affinity than Ins(1,3,4,5)P_4. The results in Table 16 illustrate a much smaller difference in the free energies of binding for this derivative and Ins(1,3,4,5)P_4 when all isomers are considered (only approximately 3.5 kcal mol^{-1}).

In order to understand why the other isomers/ionisation states of the Ins(1,3,4)P_3-5-*O*-acetamidoethylphosphate ester did not exhibit the same high binding affinity for the PKB PH domain as the initial isomer investigated, their molecular behaviour was investigated. Analysis showed that the inositol ring for each isomer is positioned almost identically to each-other, and to Ins(1,3,4,5)P_4, in the PKB binding site. It is hence the difference in the position and orientation of the acetamide moiety of this derivative, which results in the different free energies of binding. In the preliminary simulation the acetamide moiety was seen to form favourable interactions with Gln79 *via* both its carbonyl and amine groups with a total occupancy of approximately 20%, Figure 41. As a result of these interactions, increased interactions between phosphate group 5 and Gln79 were also apparent (with an occupancy of over 50%). However for the other isomers the interactions of the acetamide moiety and Gln79 are significantly less occupied, with only a 1-2% occupancy. Instead interactions between the carbonyl group of the acetamide moiety and Arg86 are apparent, Figure 43.

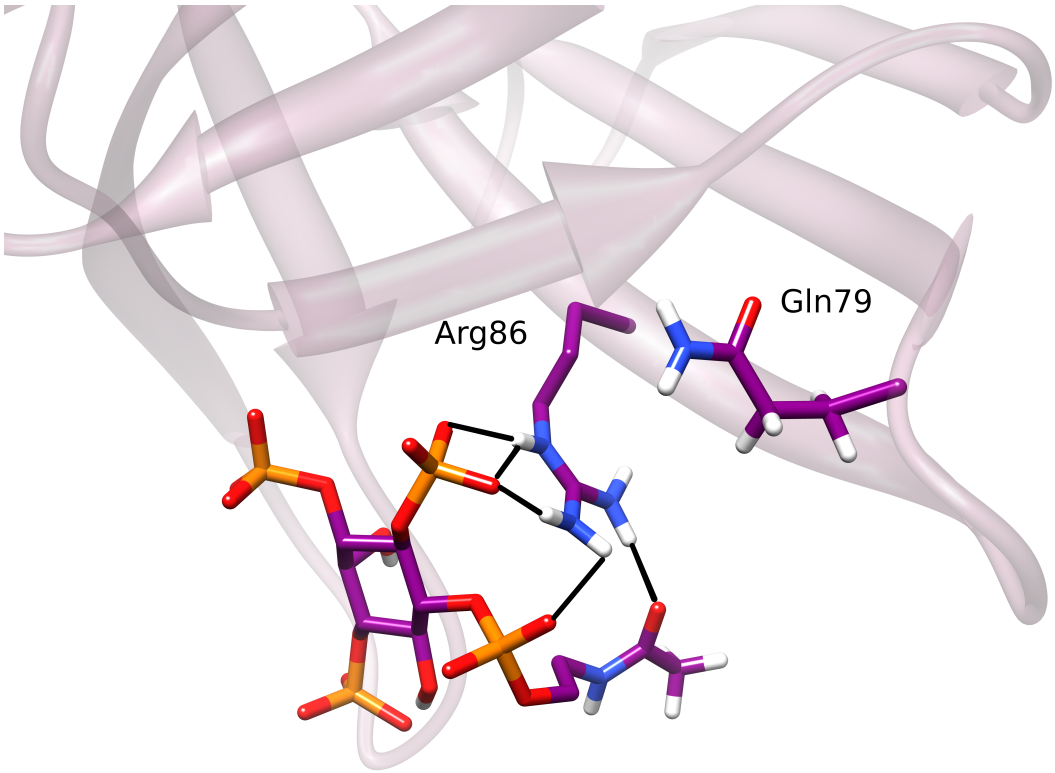


Figure 43: A representative snapshot of the lower affinity binding mode of the Ins(1,3,4)P_3-5-*O*-acetamidoethylphosphate ester in a complex with the PKB PH domain. The hydrogen bond interaction between the acetamide moiety of the derivative and Arg86 is indicated. Additional interactions of Arg86 with phosphate groups 4 and 5 are also shown. The distance of Glu79 from the binding site is highlighted.

7. 3. 4. 2 D-5-*O*-Sulfamoyl-*myo*-Ins(1,3,4)P_3

The D-5-*O*-Sulfamoyl-*myo*-Ins(1,3,4)P_3 derivative **5S** is shown to bind to the PKB PH domain with a similar, but slightly lower, free energy of binding to Ins(1,3,4,5)P_4. This result is easily explained by considering the interactions this derivative makes with the PKB PH domain. Whilst interactions *via* phosphate groups 1, 3 and 4 are almost identical to that of Ins(1,3,4,5)P_4, the 5-position sulfamate group does not replicate the behaviour of a phosphate group. This is unsurprising, as a number of previous studies have shown that sulfamate groups often don't act as suitable phosphate bioisosteres for inositol phosphates.[181, 182]

In contrast to the significant interactions that phosphate group 5 makes with the PKB PH domain (highlighted in Chapter 4), hydrogen bond analysis illustrates that the sulfamate group only makes relatively weak interactions. The lack of interactions *via* the 5-sulfamate group explains why this derivative binds to the PKB PH domain with a remarkably similar affinity (and orientation) to Ins(1,3,4)P_3 in our simulations.

The reason the sulfamate group is unable to replicate the behaviour of a phosphate group is simply due to the fact that sulfamate esters are typically neutral at physiological pH (the pK_a value for methyl sulfamate is 17.5).[179] In comparison phosphomonoesters are negatively charged at physiological pH and adopt either a fully deprotonated or singly deprotonated state. As the sulfamate moiety is neutral, the charges on its oxygen atoms are inherently lower than those for the oxygen atoms of phosphate groups. Our parameters illustrate this, with oxygen atoms of the sulfamate groups assigned partial charges of approximately -0.55 to -0.65, compared to values of around -0.75 to -0.95 for oxygen atoms belonging to the phosphate groups (parameters provided in section S.3 and S.8 of the Electronic Supplementary Information). The lower charge of the oxygen atoms belonging to the sulfamate group causes their electrostatic/hydrogen bond interactions with the PKB PH domain to be dampened.

7. 3. 4. 3 D-*myo*-Ins(1,4,5)P_3-3-*O*-methylphosphate ester

The Ins(1,4,5)P_3-3-*O*-methylphosphate ester **3M** is also shown to bind to the PKB PH domain with a similar, but slightly lower, free energy of binding to Ins(1,3,4,5)P_4. This result is understandable as Elliott's previous work (Table 14) showed that the same derivative at the 1-position also bound to the PKB PH domain with a slightly lower affinity compared to Ins(1,3,4,5)P_4.[174]

The similar binding affinity of Ins(1,3,4,5)P_4 and the Ins(1,4,5)P_3-3-*O*-methylphosphate ester is well explained by studying their molecular behaviour. Both these ligands adopt extremely similar orientations in the PKB binding site, Figure 44. Furthermore, hydrogen bond analysis illustrates that they also make near identical interactions. The only significant difference is that interactions *via* the 3-methylphosphate group are slightly dampened compared to those of the 3-phosphate group. This difference can be attributed to the fact that the oxygen atom of the methoxy moiety has a significantly lower charge than that of a protonated or deprotonated oxygen atom belonging to a phosphate group, with charges of approximately -0.40 to -0.50 and -0.75 to -0.95 respectively. This lower charge on the oxygen of the methoxy moiety therefore results in a lower level of interactions with the PKB PH domain *via* this site.

The orientation of the Ins(1,4,5)P_3-3-*O*-methylphosphate ester in the PKB binding site illustrates how the small methyl moiety is accommodated. For all eight isomers the methyl group is shown to position itself neatly in a pocket between residues Lys14, Arg23 and Arg25, as illustrated in Figure 44.

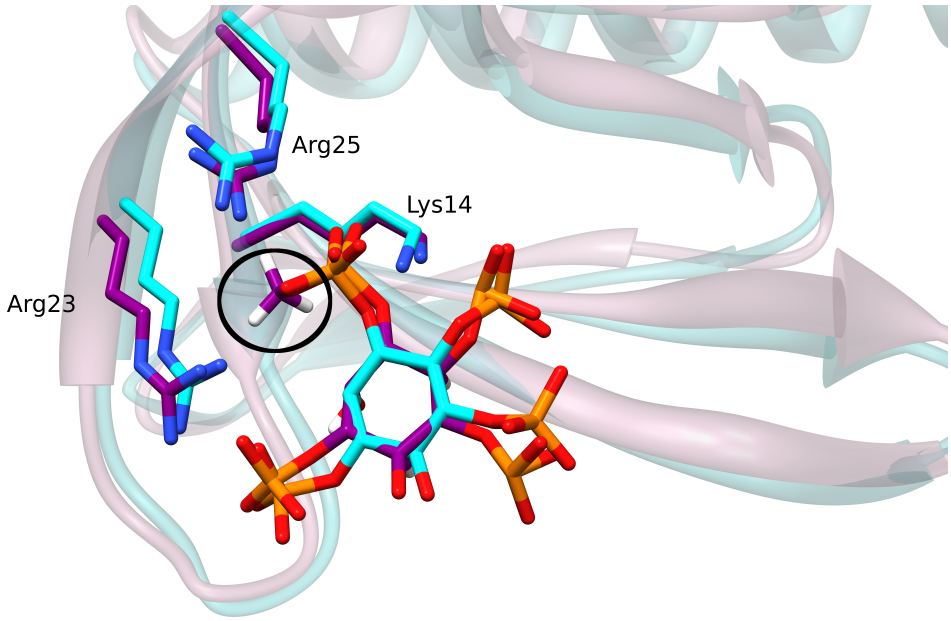


Figure 44: A representative structure of the Ins(1,4,5)P_3-3-*O*-methylphosphate ester 3M in the PKB binding site (magenta). The methyl moiety of the derivative is circled. For comparison, the crystal structure of the PKB PH domain bound to Ins(1,3,4,5)P_4 is shown in cyan. All hydrogen atoms, other than those of the methyl group, have been removed from the image for ease of visualisation.

7. 3. 4. 4 D-3-*O*-Sulfamoyl-myo-Ins(1,4,5)P_3

The 3-*O*-Sulfamoyl-Ins(1,4,5)P_3 derivative **3S** is shown to have a significantly lower affinity for the PKB PH domain compared to Ins(1,3,4,5)P_4 and the other derivatives investigated. The inability of a sulfamate group to act as a phosphate bioisostere for this system has previously been explained in section 7.3.4.2. It is the lower charge on the oxygen atoms of the sulfamate group that causes electrostatic/hydrogen bond interactions with the PKB PH domain to be dampened *via* this site.

As the 3-phosphate group of Ins(1,3,4,5)P_4 is responsible for the critical hydrogen bond/electrostatic interactions with the PKB PH domain, substitution at this site understandably has a large detrimental effect. The effect this substitution has on the free energy of binding is well illustrated in the MM-PBSA results. The electrostatic contribution to binding (i.e. the sum of the MM electrostatic energy and the electrostatic solvation energy) is less favourable for the 3-*O*-Sulfamoyl derivative compared to Ins(1,3,4,5)P_4, with contributions ranging from -1 to -40 kcal mol^{-1} for the 3-*O*-Sulfamoyl-Ins(1,4,5)P_3 isomers and -45 to -70 kcal mol^{-1} for the Ins(1,3,4,5)P_4 isomers.

Whilst the above is true, the significantly higher affinity of the 5-*O*-Sulfamoyl derivative for the PKB PH domain compared to the 3-position derivative leaves many questions. For example, why doesn't the 3-*O*-Sulfamoyl derivative rotate 180° to mimic the 5-position derivative?

In fact the 3-*O*-Sulfamoyl derivative battles between two weakly binding orientations. One of which is indeed the flipped orientation mentioned. This flipped orientation allows the 5-phosphate group to point into the PKB binding site and mimic Ins(1,3,4,5)P_4's 3-phosphate group. However there is a cost associated with reversing the positions of the phosphate groups relative to the plane of the inositol ring. Whilst interactions *via* phosphate group 5 are improved, reduced interactions *via* the 1, 3 and 4 positions are apparent. This cost of rotating the ligand was previously displayed by Ins(1,3,4,5)P_4 in Chapter 4, showing the flipped orientation to be significantly less favourable and stable in the PKB binding site. The other conformation the 3-*O*-Sulfamoyl derivative adopts is similar to that of Ins(1,3,4,5)P_4. Whilst there are

significant interactions *via* the 1, 4 and 5 phosphate groups in this orientation, there is a lack of critical interactions *via* the 3-sulfamate group.

7. 4 *In Vitro* Investigation of Selected Inositol Phosphate Derivatives

In order to validate the computational results for the inositol phosphate derivatives, it was necessary to compare them to experimental data. This section describes the development and application of *in vitro* assays to determine the relative affinity/inhibitory behaviour of the Ins(1,3,4,5)P_4 derivatives for the PKB PH domain. I carried out this work in the Woscholski lab at Imperial College. Thanks go to all members of the group for their assistance, and for the use of their equipment.

There are limited examples of using *in vitro* techniques to investigate the binding of phosphoinositide or inositol phosphate-based ligands to the PKB PH domain. This may partly be due to the micromolar affinity of these complexes, meaning large quantities of PKB and phosphoinositides/inositol phosphates are required for experiments. The lack of experimental data for such complexes is highlighted by the repeated reference to Frech *et al.*'s work in which tryptophan fluorescence was used to determine the binding affinities of a number of inositol phosphates and phosphoinositides to the PKB PH domain.[22] Despite this work being carried out in 1997 there have been few additional investigations in the field, particularly on the native system.

7. 4. 1 Issues with the Inositol Phosphate Derivatives

Deciding an appropriate biological/biochemical assay to test the affinity of the Ins(1,3,4,5)P_4 derivatives for the PKB PH domain was extremely difficult. This was partly due to a number of issues with the derivatives. Due to their complicated synthesis, only relatively small amounts were synthesised by the Conway group. Furthermore, analysis by the Conway group suggested that the derivatives were not stable in solution for extended periods of time. This therefore meant that a stock solution of each ligand could not be made up. Due to the limits in the sensitivity of weighing scales available, this also meant that a very limited number of experiments were possible. As assays generally require multiple parameters to be optimised for the system of interest, Ins(1,3,4,5)P_4 and InsP_6 were used for protocol development in

place of the derivatives. These ligands have complications as well, as Ins(1,3,4,5)P_4 is relatively expensive to purchase and $InsP_6$ has a significantly lower affinity for the PKB PH domain compared to Ins(1,3,4,5)P_4.[22]

7. 4. 2 Expression of PKB

The PKB PH domain was expressed as a GST-fusion protein. The pGEX plasmid expression vector containing the sequence for the PKB PH domain (kindly donated by the Woscholski group) was tranformed into *E.coli* (strain DH5α). The *E.coli* were then grown on LB plates containing ampicillin at 37°C for 16 h. Colonies were transferred to LB media containing ampicillin, and grown for approximately 20 hours at 37°C, until reaching a mid-log phase. Expression of the GST-tagged PKB PH domain was then induced with 1mM IPTG for 24 hours at 23°C. The cells were then harvested and the pellet stored at -20°C. The cells were resuspended in lysis buffer containing 50 mM Tris (pH 7.4), 1% Triton X-100, 2 mM EDTA, 10 mM benzamidine hydrochloride, 100 μg ml^{-1} soybean trypsin inhibitor, 2 mg ml^{-1} lysozyme, 1 mM 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride, and 2 mM DTT. The cell suspension was stirred at 4°C for 1 hour and then sonicated, homogenised and centrifuged at 11500 rpm at 4°C for 1 hour. The supernatant was loaded onto a glutathione sepharose column at 4°C, which was pre-equilibrated with 50mM Tris (pH 7), 140 mM NaCl, 2.7 mM KCl. After loading, the column was washed with:

- 50 mM Tris (pH 7.4), 140 mM NaCl, 2.7 mM KCl, 1% Triton and 2 mM DTT.
- Followed by 50 mM Tris (pH 7.4), 140 mM NaCl, 2.7 mM KCl and 2 mM DTT.
- And finally 50 mM Tris (pH 7.4), 500 mM NaCl, 2.7 mM KCl, and 2 mM DTT.

The GST-tagged protein was eluted with 50 mM Tris (pH 7.4), 20 mM reduced glutathione, 250 mM NaCl and 2 mM DTT. The eluted protein was concentrated at 4°C and 2000 rpm. For a more detailed explanation of the expression protocol please refer to section 9.2 in the Materials and Methods chapter.

Sodium dodecyl sulfate (SDS) – polyacrylamide gel (PAGE) was used to confirm that a protein of the correct size was isolated, Figure 45. Protein concentration was determined using the Bradford assay. Further details of this protocol are provided in section 9.4 in the Materials and Methods chapter.

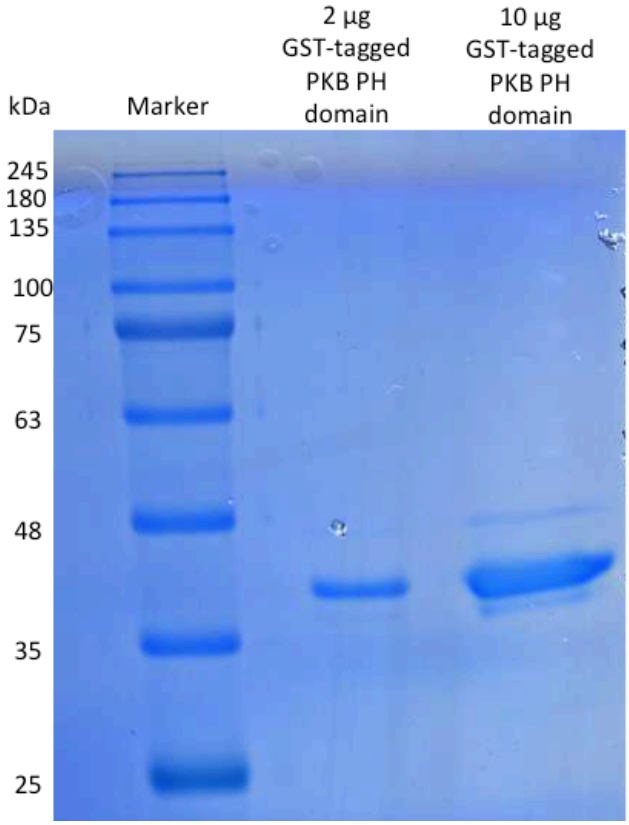


Figure 45: 10% SDS-PAGE gel for the eluted PKB PH domain, visualised by Coomassie Blue staining. Further details are provided in Materials and Methods, section 9.3. A band is clearly shown at around 40 kDa for the GST-tagged PKB PH domain.

7. 4. 3 Protein Lipid Overlay (PLO) Assay

A protein lipid overlay assay (also referred to as a dot blot assay) is a biochemical technique used to identify proteins that bind to lipid ligands. In a PLO assay, lipids are spotted onto a nitrocellulose membrane. This membrane is then incubated with a chosen protein possessing an epitope tag. After an incubation period, the protein bound to the lipid is then detected by immunoblotting with an antibody that recognises the epitope tag.

PLO assays have also been used widely in the Woscholski group to study multiple systems.[183] This technique can provide information about the relative affinity of a protein for various lipids, or various proteins for a particular lipid.[184] The PLO assay can also be adapted for competition-based studies. Here, the ability of a ligand to inhibit binding of the chosen protein to the immobilised lipid is detected. This has previously been applied by Miao *et al.* who have used a competitive PLO assay to illustrate the inhibitory effect of PIT-1 (a non-inositol based inhibitor) on the PKB PH domain binding to PtdIns(3,4,5)P_3.[185]

Advantages of the PLO assay include a) its simplicity and b) that it requires only a small amount of protein and lipid. It is therefore significantly cheaper than many other biochemical techniques. This assay has previously been used to assess the selectivity of different PH domains for phosphoinositides.[23] Additionally it has been applied to determine the relative affinity of different PKB PH domain mutants for PtdIns(3,4,5)P_3 and PtdIns(3,4)P_2.[23, 26] It was therefore hoped that the PLO assay would be suitable for assessing the binding of the Ins(1,3,4,5)P_4 derivatives to the PKB PH domain, by investigating their ability to compete with PtdIns(3,4,5)P_3.

7. 4. 3. 1 General Method

The basic PLO method applied in this work is described below. Any adaptations made to this protocol are detailed where appropriate.

PtdIns(3,4,5)P_3 (di-C16) was diluted in a solution of methanol, chloroform and water (in the ratio 2:1:0.8) to the desired lipid concentration. 2 µl aliquots of the lipid solution were then spotted onto Hybond-C extra membrane and allowed to dry for 1 hour at room temperature. After drying, the membrane was incubated with blocking buffer [50mM Tris-HCl, pH 7.5, 150mM NaCl, 0.1% (v/v) Tween 20 and 2 mg ml^{-1} fatty-acid-free BSA] for 1 hour at room temperature. The membrane was then incubated with the GST-tagged PKB PH domain in blocking buffer for 1 hour. The membrane was washed three times over 45 minutes in TBST [50mM Tris-HCl, pH7.5, 150mM NaCl and 0.1% (v/v) Tween 20], and then incubated with a 1:5000 dilution of anti-GST horseradish-peroxidase (HRP) conjugate in TBST for 1 hour. The membrane was washed as before. For all incubation and washing stages, the

solution was gently agitated. The GST-tagged PKB PH domain bound to the membrane was detected by enhanced chemiluminescence. The following detector solutions were used in combination with the Fujifilm LAS-300 Imaging System: 5ml of detection solution 1 (2.5mM luminol, 400 µM *p*-coumaric acid, 100 mM Tris-HCl, pH 8.5), added to 5ml of detection solution 2 (100mM Tris-HCl, pH 8.5), and 3.1 µl of H_2O_2.

7. 4. 3. 2 Linearity Tests

Before investigating the effect that the inositol phosphate derivatives have on the PKB PH domain binding to PtdIns(3,4,5)P_3, the PLO assay was calibrated. Binding of the GST-tagged PKB PH domain to increasing concentrations of PtdIns(3,4,5)P_3 was investigated. Figure 46 illustrates the binding curve obtained with 10 nM GST-tagged PKB PH domain, and 0-50 µM PtdIns(3,4,5)P_3. From this graph a PtdIns(3,4,5)P_3 concentration of 10 µM was chosen for future experiments as it is within the linear region of the graph and shows a strong signal/intensity.

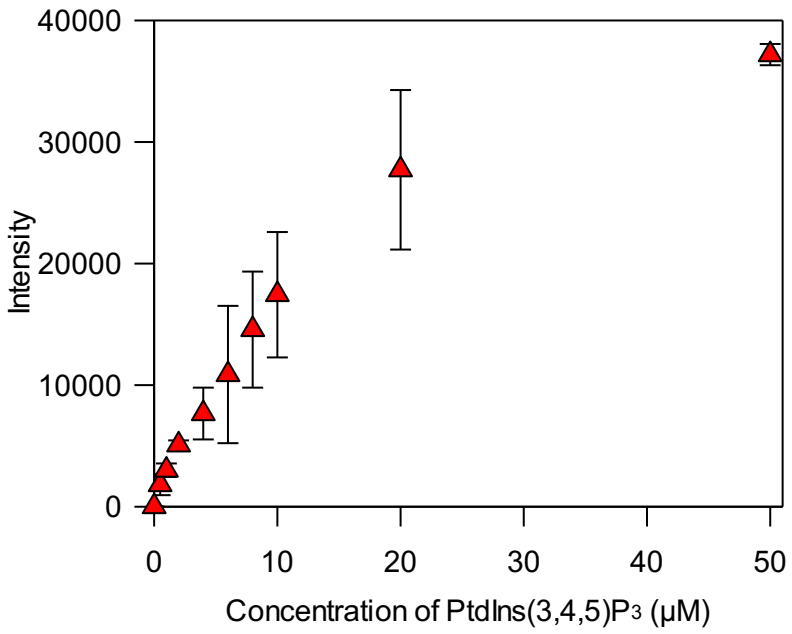


Figure 46: Effect of increasing the concentration of PtdIns(3,4,5)P_3 on the intensity of the chemiluminescent signal in the PLO assay. This chemiluminescent signal represents the amount of GST-tagged PKB PH domain bound to PtdIns(3,4,5)P_3 on the nitrocellulose membrane. The concentration of the PKB PH domain used in this assay is 10 nM. Values are the mean of 3 repeats. Error bars illustrate the standard deviation between these repeats. Graph was produced using GraFit.[186]

Binding of different concentrations of the GST-tagged PKB PH domain to 10 μM PtdIns(3,4,5)P_3 was then tested. Figure 47 illustrates the binding curve obtained with 10 μM PtdIns(3,4,5)P_3, and 0-200 nM GST-tagged PKB PH domain. A concentration of 50 nM for the GST-tagged PKB PH domain was chosen for future experiments, again due to its strong signal within the linear region.

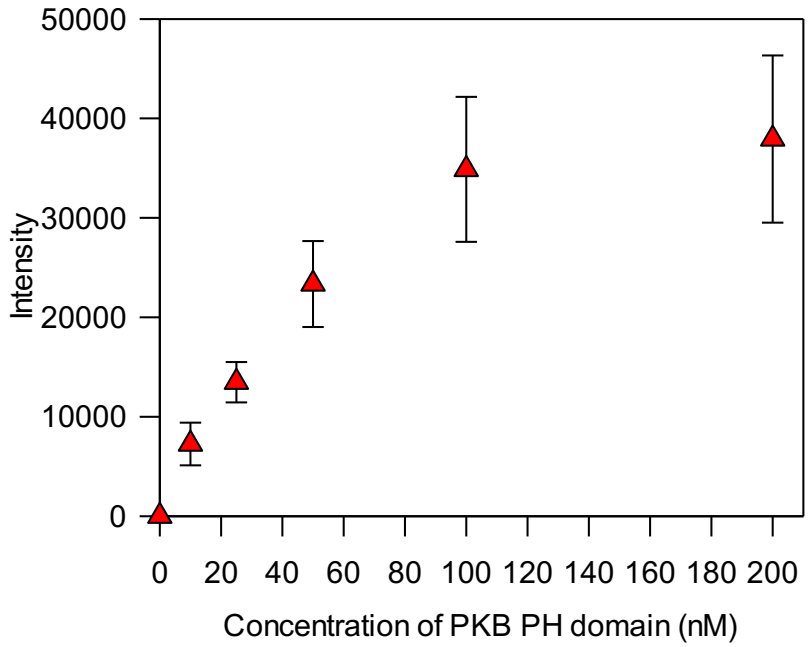


Figure 47: Effect of increasing the concentration of the GST-tagged PKB PH domain on the intensity of the chemiluminescent signal in the PLO assay. This chemiluminescent signal represents the amount of GST-tagged PKB PH domain bound to PtdIns(3,4,5)P_3 on the nitrocellulose membrane. The concentration of PtdIns(3,4,5)P_3 immobilised on the nitrocellulose membrane in this assay is 10 μM. Values are the mean of 3 repeats. Error bars illustrate the standard deviation between these repeats. Graph was produced using GraFit.

7. 4. 3. 3 Competition-based Assays

In order to carry out a competition-based PLO assay, the method in section 7.4.3.1 was adapted slightly: The chosen competitive ligand [i.e. Ins(1,3,4,5)P_4, InsP_6 or an Ins(1,3,4,5)P_4 derivative] was pre-incubated with 50 nM GST-tagged PKB PH domain in blocking buffer for 30 minutes prior to application to the nitrocellulose membrane.

The competition-based assay was originally carried out with InsP_6. Varying concentrations (10^{-3} to 10^{-8} M) of InsP_6 were pre-incubated with the GST-tagged PKB PH domain and the protocol in section 7.4.3.1 carried out. Results from this assay are illustrated in Figure 48. Unexpectedly the chemiluminescent signal, detecting the GST-tagged PKB PH domain bound to PtdIns(3,4,5)P_3, was shown to increase in a dose dependant manner with increasing concentrations of InsP_6. This result was very surprising as it was expected that InsP_6 would bind to the PKB PH domain and thus inhibit/lower the level of protein bound to the immobilised PtdIns(3,4,5)P_3. However, the reverse effect seemed to be occurring. This same 'reverse' effect was also apparent when Ins(1,3,4,5)P_4 and the inositol phosphate derivatives were applied as the competing ligand.

To detect whether this effect was also apparent with other PH domains, the GST-tagged GRP1 and PLC δ1 PH domains were also tested. [PtdIns(4,5)P_2 instead of PtdIns(3,4,5)P_3 was spotted for the PLCδ1 assay]. However these PH domains illustrated the expected decrease in the chemiluminescent signal when Ins(1,3,4,5)P_4 and Ins(1,4,5)P_3 were pre-incubated with the GRP1 and PLCδ1 PH domains respectively.

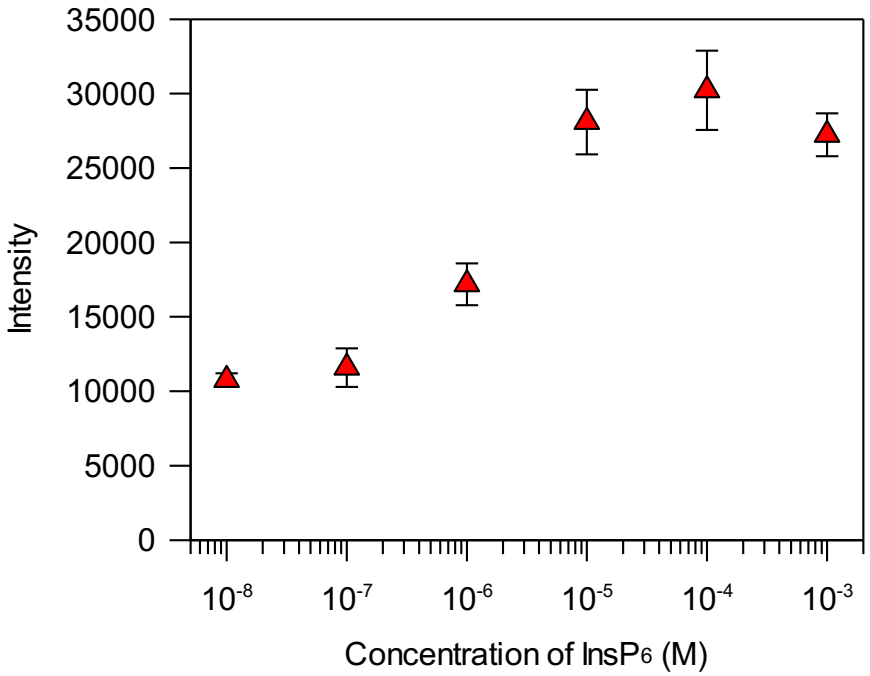


Figure 48: Effect of increasing levels of $InsP_6$ pre-incubated with the GST-tagged PKB PH domain on the intensity of the chemiluminescent signal in the PLO assay. This chemiluminescent signal represents the amount of GST-tagged PKB PH domain bound to PtdIns(3,4,5)P_3 on the nitrocellulose membrane. The concentration of PtdIns(3,4,5)P_3 immobilised on the nitrocellulose membrane in this assay is 10 μM. The concentration of the PKB PH domain used in this assay is 50 nM. Values are the mean of 3 repeats. Error bars illustrate the standard deviation. Graph was produced using GraFit.

An argument for the unusual behaviour of the PKB PH domain is that the inositol phosphate-based compounds are causing it to dimerise. Early experiments investigating PKB's activation mechanism suggested that dimerisation of PKB's PH domain was required for PKB's autophosphorylation and self-activation.[25, 187, 188] As PDK1 and mTORC2 are now considered responsible for PKB's phospho-activation there is currently a lack of evidence for the role of PKB dimerisation. Nevertheless recent experiments using surface plasmon resonance spectroscopy further support the argument for the dimerisation of the PKB PH domain, suggesting the PH domain-PH domain interaction to have a nanomolar affinity.[54]

PKB's speculated dimerisation could increase its avidity for phosphoinositides, as previously reported for numerous other PH domains.[189] It is possible that such

dimerisation could be caused/enhanced by non-specific interactions between the PKB PH domains and an inositol phosphate ligand at an alternative binding site to PtdIns(3,4,5)P_3. This corresponds well with the low affinity inositol phosphate binding site proposed by Auguin *et al*. for the closely related PKBβ: The authors note that inositol phosphates might bind to an extended surface of the molecule in a non-specific manner.[190] Dimerisation of PKB's PH domain has also been reported to be encouraged by the TCL-1 protein (T-cell leukemia 1).[191]

7. 4. 3. 4 Adaptations to the Competitive PLO Assay

In order to further explore if the PKB PH domain is binding to inositol phosphates in a non-specific manner, numerous adaptations to the PLO assay were attempted. These included varying the Tween and salt concentrations in the washing buffer, adding cellular levels of $MgCl_2$ to the PKB PH domain solution, and using a higher level of fatty-acid free BSA in the blocking buffer. However these all illustrated the same increase in PKB PH domain bound to the immobilised PtdIns(3,4,5)P_3 in the presence of $InsP_6$ or Ins(1,3,4,5)P_4.

The *in vitro* behaviour of PKB binding to phosphoinositides has previously been shown to be strongly dependent on the buffer used.[24] Although TBS buffer was previously applied in a competitive PLO assay of PIT-1 and the PKB PH domain, it is important to note that this ligand is charge neutral.[185] The use of TBS, which is a cationic buffer, may not be suitable for studying the highly anionic inositol phosphate-based ligands. Therefore PBS buffer (phosphate buffered saline), which is anionic in nature, was investigated. PBS was also the buffer of choice for Kozikowski *et al.* for *in vitro* assays investigating the binding of the PKB PH domain to phosphatidylinositol analogues.[55]

When PBS buffer was used in place of TBS in the PLO assay, pre-incubating the PKB PH domain with $InsP_6$ or Ins(1,3,4,5)P_4 no longer caused an increase in the chemiluminescent signal. However, the expected decrease in signal also wasn't apparent even at very high concentrations (100 μM) of $InsP_6$ and Ins(1,3,4,5)P_4. The reason for these results, and the action of the PBS buffer on the assay, is still under debate. It is possible that the PBS buffer is hindering the dimerisation of the GST-

tagged PKB PH domain caused by non-specific interactions of inositol phosphate ligands.

Although the use of PBS buffer in the PLO assay appeared promising, the lack of inhibitory effect of the inositol phosphate ligands still needs to be considered. High concentrations of phosphoinositide lipids have previously been shown, in cases, to form non-specific interactions with the PKB PH domain which could not be competed for with an excess of inositol phosphates.[192] Whilst numerous precautions to prevent non-specific interactions (such as increased salt and Tween in washes, increased fatty-acid free BSA in the blocking buffer and cellular levels of Mg^{2+} in the protein solution) were attempted, this effect may still be occurring. This theory is supported by the fact that pre-incubation of the GST-tagged PKB PH domain with PtdIns(3,4,5)P_3 does illustrate the expected decrease in binding to the immobilised PtdIns(3,4,5)P_3, Figure 49.

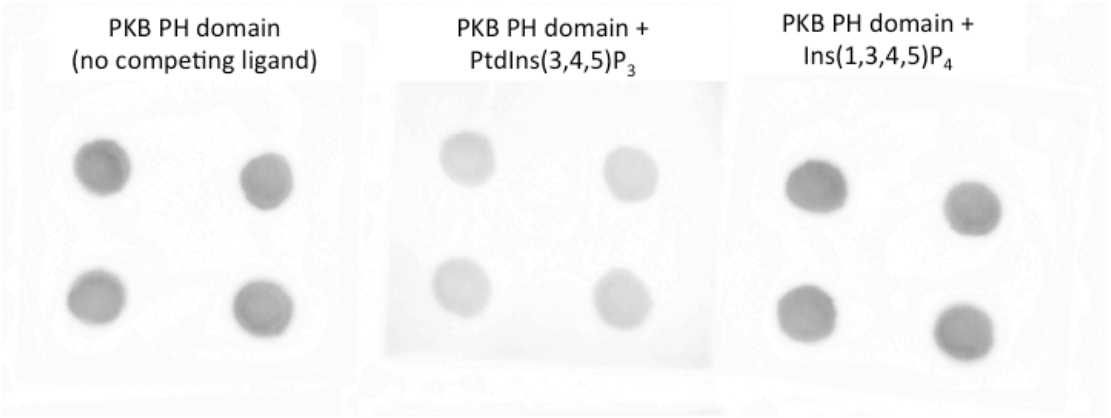


Figure 49: Image of the chemiluminescent signal on the nitrocellulose membrane from the competitive PLO assay. The PKB PH domain with no competing ligand, with 100 µM PtdIns(3,4,5)P_3, and 100 µM Ins(1,3,4,5)P_4 are shown.

In the following sections, other biochemical techniques are considered to investigate the interaction between the PKB PH domain and the Ins(1,3,4,5)P_4 derivatives.

7. 4. 4 Isothermal Titration Calorimetry (ITC)

Isothermal titration calorimetry (ITC) offers an alternative technique for measuring the affinity of the different Ins(1,3,4,5)P_4 derivatives for the PKB PH domain. This technique is not competition-based, and therefore eliminates any problems that may occur due to non-specific binding of phosphoinositides that cannot be competed for.

ITC is a biochemical technique that can be used to determine the binding affinity and stoichiometry of a complex. Additionally it can be used to determine the thermodynamic properties of an interaction, providing the enthalpy (ΔH) and entropy (ΔS) change and hence the change in Gibbs free energy (ΔG).[193]

An isothermal titration calorimeter consists of two cells (a reference cell and a sample cell) surrounded by an adiabatic jacket. The reference cell is filled with buffer or water and the sample cell with the desired protein. Both cells are maintained at a constant temperature by thermopile/thermocouple circuits. When the ligand is titrated into the sample cell in precise aliquots, a temperature change is detected. This temperature change is measured by the time-dependant input of power required to maintain the temperature equilibrium between the reference and sample cells.[193] Origin software is used to analyse this data and produce a binding curve.

ITC has previously been used to investigate the binding patterns of several PH domains.[23, 27] However only two accounts of using ITC to study the PKB PH domain exist.[194, 195] Both these studies noted difficulties in performing such experiments due to the relatively high concentrations of PKB PH domain required.

The calculation used to assess the concentration of protein required for an ITC experiment is provided in equation 23. This equation relies strongly on the predicted binding affinity (K_d) of the complex. As the binding affinities of inositol phosphates for the PKB PH domain are only micromolar, a very high concentration of the PKB PH domain is thus required.[22]

$$c = ([M_{tot}]n)/K_d$$

Equation 23: Equation used to calculate the amount of protein $[M_{tot}]$ required for an ITC experiment. $[M_{tot}]$ is measured in M. *c* is a unitless constant. The recommended value for *c* is between 10 and 100.[196] The value of *n* (the binding stoichiometry) is presumed to be 1. K_d is the dissociation constant in M.

7. 4. 4. 1 Method

For initial experiments $InsP_6$ was chosen as the ligand of interest due it being readily available and inexpensive compared to Ins(1,3,4,5)P_4 and the Ins(1,3,4,5)P_4 derivatives.

Using equation 23 a concentration of 6 mg ml^{-1} for the PKB PH domain was decided upon. This concentration is equivalent to a value of 17.2 for *c* using the dissociation constant of 8.7 μM from Frech *et al.*'s work.[22] This concentration was significantly higher than that applied in a previous ITC experiment for the PKB PH domain and Ins(1,3,4,5)P_4 in which a concentration of 230 μg ml^{-1} PKB PH domain was used.[194]

It was necessary that an even higher concentration (than 6 mg ml^{-1}) of the PKB PH domain was obtained in the protein expression stage. This is because the PKB PH domain is typically stored in 50% glycerol to prevent it denaturing during the freeze-thaw cycle. Additionally for the ITC experiment the PH domain should be diluted to give as low a volume of glycerol as possible. This is because glycerol is viscous in nature and hence can affect mixing.[197, 198]

In order to obtain a high concentration of the PKB PH domain, a new aliquot was expressed. The same protocol in section 7.4.2 was carried out with an additional stage of buffer exchange and concentration. The GST-tagged PKB PH domain was transferred from the Tris pH 7.4 buffer into 25 mM MOPS, pH 6.8, 100 mM NaCl and 1 mM DTT using a spin concentrator. Unfortunately in concentrating the protein, insolubility issues arose. A significant amount of the expressed protein was seen to

crash out of the solution. However a sizeable amount was retrieved, and shown to be active using PLO assays.

Titrations were performed at 25°C using the *MicroCal* VP-ITC MicroCalorimeter. For the ITC experiment the PKB PH domain was diluted to a concentration of 6 mg ml^{-1} in 25 mM MOPS, pH 6.8, 100 mM NaCl, 1 mM DTT and placed in the calorimeter cell. Sequential 10 μl aliquots of $InsP_6$ were injected under computer control to cover a range of $InsP_6$:PKB PH domain molar ratios of 0:1 to 5:1. A control experiment for the heat of dilution of $InsP_6$ into buffer was performed separately and the data corrected accordingly. Data was analysed using the Origin software (OriginLab, Northampton, MA). [Note, the first few points of Figure 50 may be ignored, as they are known to give spurious results due to mixing.[199]]

7. 4. 4. 2 Results and Discussion

Although a high concentration of the PKB PH domain was used in the experiments the signal due to binding was still weak, Figure 50. This could be due to multiple reasons, a few of which include: the concentration of the PKB PH domain still being too low (or a higher proportion was denatured than expected), the predicted binding affinity for $InsP_6$ being incorrect, or the temperature the experiment was carried out at being too high.

As there are numerous potential reasons this experiment was unsuccessful, and multiple parameters that could be adapted, it was decided that ITC was not a suitable technique to use in this work. This is mainly due to the fact that for each experiment (at the current concentration of PKB PH domain used) approximately 12 mg of GST-tagged PKB PH domain is required. Therefore to assess the numerous parameters mentioned, vast amounts of protein would be required.

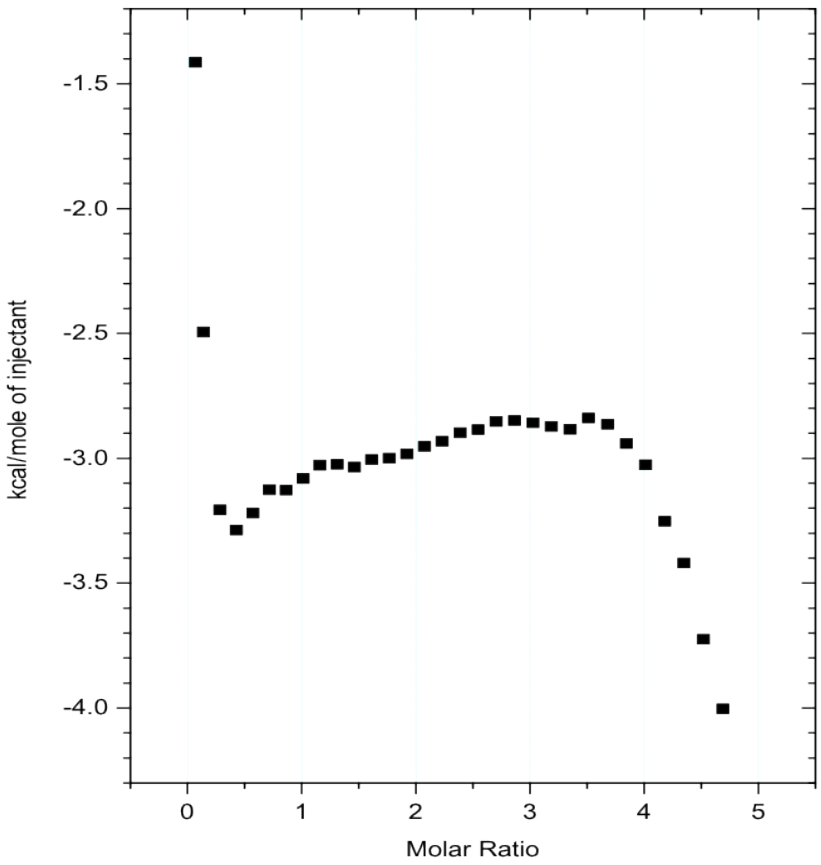


Figure 50: $InsP_6$ injected into 25 mM MOPS, pH 6.8, 100 mM NaCl, 1 mM DTT buffer containing GST-tagged PKB PH domain. Initial concentration of the PKB PH domain was 6 mg ml^{-1}. $InsP_6$ was injected in sequential aliquots (10 µl) to cover a range of $InsP_6$:PKB PH domain molar ratios from 0:1 to 5:1.

7. 4. 5 Intrinsic Tryptophan Fluorescence Spectroscopy

Intrinsic tryptophan fluorescence spectroscopy provides another option to assess the binding affinity of inositol phosphate derivatives for the PKB PH domain. This technique was used in the frequently referenced work of Frech *et al.* in determining the binding affinities of a number of inositol phosphates and phosphoinositide ligands for the PKB PH domain.[22]

Tryptophan's intrinsic fluorescence allows it to act as a probe of the conformational state of a protein. The binding affinity of a ligand for a protein can therefore be calculated for systems in which the protein: a) changes conformation upon binding the ligand and, b) that a tryptophan residue is suitably placed in the protein so that it also experiences a shift in environment.

The technique essentially works by exciting the tryptophan residue/s at a wavelength in the range of 290-305 nm and measuring the resulting fluorescence at around 345 nm using a fluorescence spectrometer. As tryptophan's fluorescence parameters are particularly sensitive to its surrounding environment, a change in fluorescence intensity is detected.

In this work a GST-tagged PKB PH domain was expressed. The presence of multiple tryptophan residues in the GST tag could result in the binding data being impaired. For this reason other biochemical techniques were first considered, as removal of the GST tag has proved difficult in regards to the stability of the PKB PH domain.[194]

7. 4. 6 Enzyme-Linked ImmunoSorbant Assay (ELISA)

A competitive ELISA is typically used to determine the competitive binding of antigens to an antibody. This protocol has been adapted slightly herein to investigate the competitive binding of PtdIns(3,4,5)P_3 and inositol phosphate derivatives to the GST-tagged PKB PH domain.

In this protocol PtdIns(3,4,5)P_3 is coated onto a well plate. Meanwhile the GST-tagged PKB PH domain is pre-incubated with an inositol phosphate derivative. When the pre-incubated solution is added to the well plate, the GST-tagged PKB PH domain is forced to compete between the inositol phosphate derivative and PtdIns(3,4,5)P_3. The amount of GST-tagged PKB PH domain bound to PtdIns(3,4,5)P_3 on the well plate is detected using an HRP-conjugated GST antibody and a TMB (3,3′,5,5′-Tetramethylbenzidine)-based detector solution. TMB acts as a substrate for HRP, causing a colour change in the well by producing a bright blue signal. This is illustrated in Figure 51.

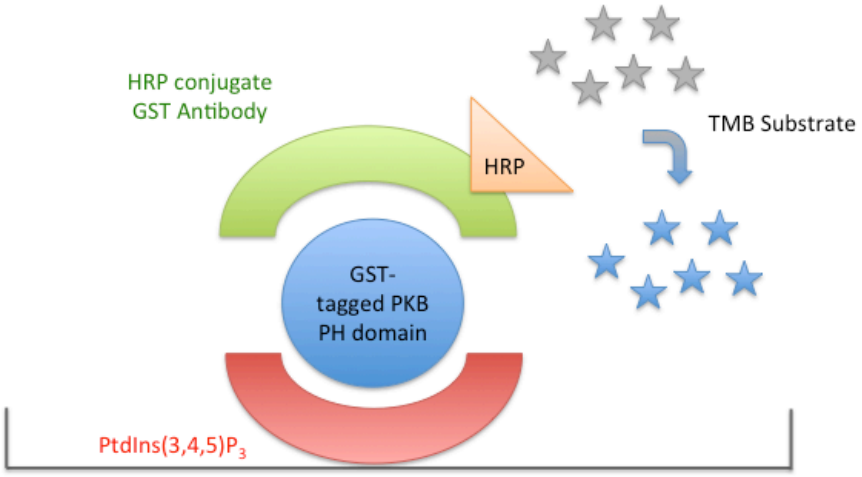


Figure 51: A schematic representation of the ELISA assay carried out in this work. For ease of visualisation the competing inositol phosphate derivative is not included in this diagram.

7. 4. 6. 1 Linearity Tests

Before investigating the effect the inositol phosphate derivatives have on the PKB PH domain binding to PtdIns(3,4,5)P_3, the assay was first calibrated to determine optimum concentrations of the PKB PH domain and PtdIns(3,4,5)P_3.

7. 4. 6. 1. 1 Method

The desired concentration/s of PtdIns(3,4,5)P_3 (diC16) was added to each well of a 96-well Thermo Scientific Nunc MultiSorp^{TM} plate and incubated overnight at 4°C. After incubation, the wells were washed with PBS 3 times over 15 minutes. The well plate was then blocked with 3% (w/v) fatty-acid-free BSA in PBS for 1.5 hours at 37°C and washed as before. The GST-tagged PKB PH domain (at the desired concentration/s) was added to the well plate for 1 hour at room temperature. The well plate was washed as before and then incubated with an HRP-conjugated GST antibody at a 1:2000 dilution for 1 hour. The plate was washed 3 times over 15 minutes with PBST (0.1% Tween 20) and subsequently 3 times over 15 minutes with PBS. Detergent was only included after incubation of the final reactant as recommended by Esser.[200] Following these washes a TMB-based substrate from Frey *et al.* was added to the wells for detection, and allowed to develop for 30 minutes in darkness.[201] After 30 minutes, a stop solution of 1M H_2SO_4 was added to the wells and the plate was read at an absorbance of 450 nm.

7. 4. 6. 1. 2 Results

Binding of the GST-tagged PKB PH domain to increasing concentrations of PtdIns(3,4,5)P_3 was investigated. Figure 52 illustrates the binding curve obtained with 2 nM GST-tagged PKB PH domain, and 0-5 μM PtdIns(3,4,5)P_3. From this graph a PtdIns(3,4,5)P_3 concentration of 1 μM was chosen for future experiments as it is within the linear region.

Different concentrations of the GST-tagged PKB PH domain were then investigated binding to 1 μM PtdIns(3,4,5)P_3. Figure 53 illustrates the binding curve obtained with

1 μM PtdIns(3,4,5)P_3, and 0-32 nM GST-tagged PKB PH domain. A concentration of 3 nM for the GST-tagged PKB PH domain was chosen for future experiments.

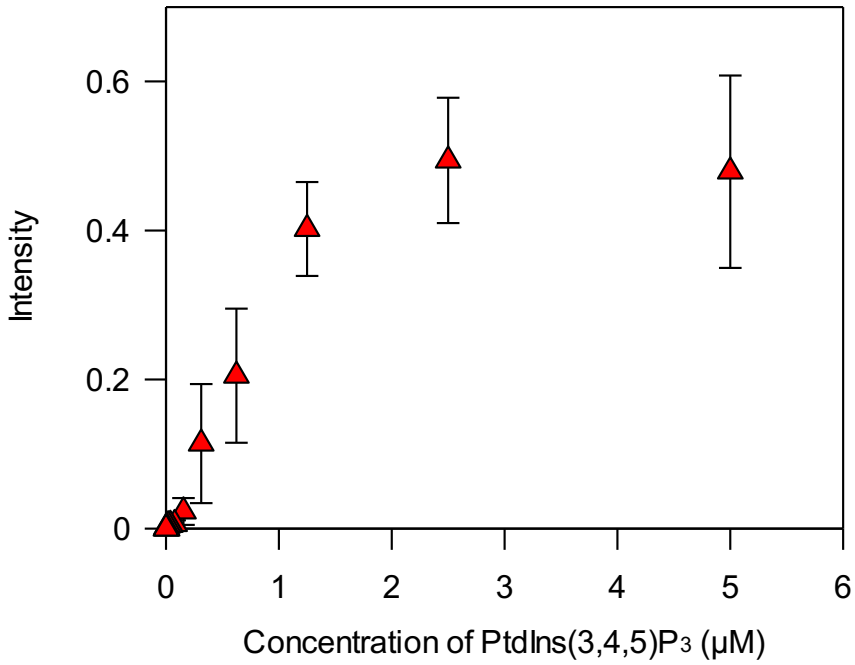


Figure 52: Effect of increasing the concentration of PtdIns(3,4,5)P_3 on the intensity of the detected signal in the ELISA assay. This signal represents the amount of GST-tagged PKB PH domain bound to PtdIns(3,4,5)P_3 on the well plate. The concentration of the PKB PH domain used in this assay is 2 nM. Values are the mean of 3 repeats. Error bars illustrate the standard deviation. Graph was produced using GraFit.

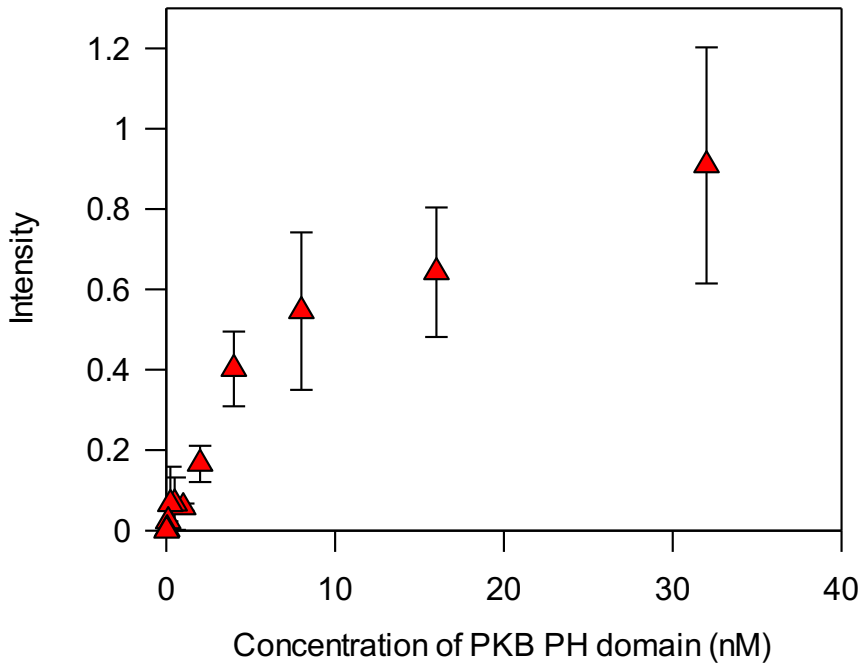


Figure 53: Effect of increasing the concentration of the GST-tagged PKB PH domain on the intensity of the signal in the ELISA assay. This signal represents the amount of GST-tagged PKB PH domain bound to PtdIns(3,4,5)P_3 on the well plate. The concentration of PtdIns(3,4,5)P_3 coated on the well plate in this assay is 1 μM. Values are the mean of 3 repeats. Error bars illustrate the standard deviation. Graph was produced using GraFit.

7. 4. 6. 2 Optimisation of the ELISA assay

Multiple considerations were made in the ELISA assay in order to optimise the signal-to-noise ratio. These included varying blocking times and washes for the well plate, determining the optimum dilution of the HRP-conjugated GST antibody and considering different TMB-based detector solutions. The final assay decided upon in this work for the competition-based assays is provided in section 7.4.6.3.1.

7. 4. 6. 3 Competitive ELISA assays with the Ins(1,3,4,5)P_4 Derivatives

7. 4. 6. 3. 1 Method

A 50 µl solution of 1 µM PtdIns(3,4,5)P_3 (diC16) was added to each well of a 96-well Thermo Scientific Nunc MultiSorp^{TM} plate and incubated overnight at 4°C. After incubation, the wells were washed with PBS 3 times over 15 minutes. The well plate was then blocked with 3% (w/v) fatty acid-free BSA in PBS for 1.5 hours at 37°C and washed as before. The GST-tagged PKB PH domain (3nM) was pre-incubated in PBS with the desired Ins(1,3,4,5)P_4 derivative for 30 minutes, and then added to the well plate for 1 hour at room temperature. Seven concentrations (ranging from 0 µM to 100 µM) of each derivative were each pre-incubated with the GST-tagged PKB PH domain. The well plate was washed as before and then incubated with an HRP-conjugated GST antibody at a 1:2000 dilution for 1 hour. The plate was washed 3 times over 15 minutes with PBST (0.1% Tween 20) and subsequently 3 times over 15 minutes with PBS. Detergent was only included after incubation of the final reactant as recommended by Esser.[200] Following these washes a TMB-based substrate from Frey *et al.* was added to the wells for detection, and allowed to develop for 30 minutes in darkness.[201] After 30 minutes, a stop solution of 1M H_2SO_4 was added to the wells and the plate was read at an absorbance of 450 nm.

7. 4. 6. 3. 2 Results

After carrying out the linearity tests, preliminary competition-based assays with Ins(1,3,4,5)P_4 and InsP_6 were attempted. In these preliminary assays the ligands were shown to compete with the immobilised PtdIns(3,4,5)P_3 for the GST-tagged PKB PH domain in a dose-dependant manner. It was therefore decided that this assay would be suitable to test the relative affinities of the Ins(1,3,4,5)P_4 derivatives.

The protocol detailed in section 7.4.6.3.1 was carried out using seven different concentrations of Ins(1,3,4,5)P_4 and the Ins(1,3,4,5)P_4 derivatives. The seven concentrations (ranging from 0 µM to 100 µM) of the derivatives were each

investigated in three independent experiments. Within each independent experiment, three repeats were carried out.

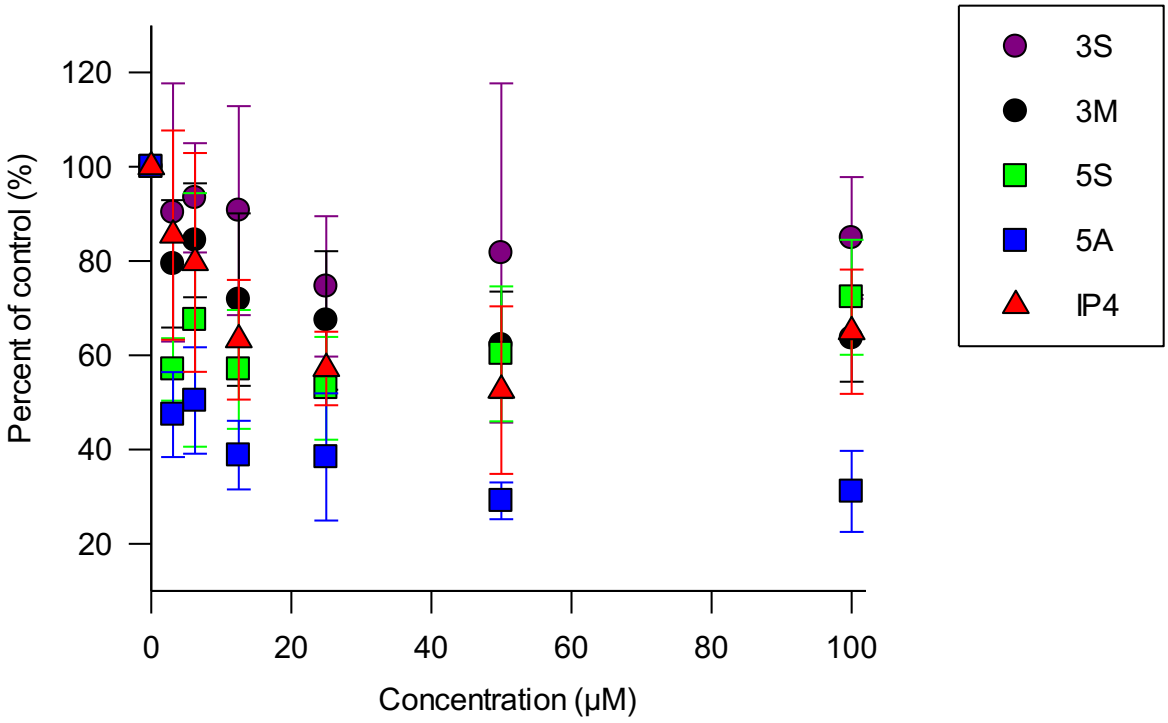


Figure 54: ELISA competitive binding assay for the Ins(1,3,4,5)P_4 derivatives. Percent of control represents the amount of PKB PH domain bound to PtdIns(3,4,5)P_3 on the well plate for the given sample as a percentage of the amount of PKB PH domain bound to PtdIns(3,4,5)P_3 on the well plate for the control (i.e. with no competing ligand/derivative). Competition binding curves for compounds D-3-*O*-Sulfamoyl-*myo*-Ins(1,4,5)P_3 (3S), D-*myo*-Ins(1,4,5)P_3-3-*O*-methylphosphate ester (3M), D-5-*O*-Sulfamoyl-*myo*-Ins(1,3,4)P_3 (5S), D-*myo*-Ins(1,3,4)P_3-5-*O*-acetamidoethylphosphate ester (5A) and D-*myo*-Ins(1,3,4,5)P_4 (IP4). Values are the mean of 3 independent experiments, results for which are provided in section S.12 in the Electronic Supplementary Information. Error bars illustrate the standard deviation between the 3 independent experiments.

Results for these experiments are shown in Figure 54. Unfortunately due to the significant variability in the results (discussed shortly), only qualitative analysis of the data has been carried out.

Figure 54 illustrates a pattern in the relative affinities/inhibitory behaviour of the Ins(1,3,4,5)P_4 derivatives for the PKB PH domain. Reassuringly, this same pattern is apparent at almost all concentrations. The 3-position sulfamate ester **3S** is shown to be the weakest binding derivative at all concentrations. Conversely the 5-position acetamidoethylphosphate ester **5A** is shown to be the strongest. Between these, the 3-position methylphosphate ester **3M**, the 5-position sulfamate ester **5S** and Ins(1,3,4,5)P_4 all illustrate similar affinity for the PKB PH domain. In general, the 3-position methylphosphate ester binds to the PKB PH domain with a slightly weaker affinity than Ins(1,3,4,5)P_4, and the 5-sulfamate ester a slightly stronger affinity. However, variation in these results makes it difficult to draw a clear conclusion here.

Many of the derivatives [and Ins(1,3,4,5)P_4] appear to show a slight decrease in affinity/inhibition at a concentration of 100 μM. Interestingly this effect is also apparent in Mahadevan *et al.*'s results when using a competitive ELISA assay to study the binding of non-inositol based compounds to the PKB PH domain.[202] The reason for this effect is uncertain, and may just be an artefact of the assay. Alternatively precipitation of the ligand at high concentrations could be occurring.

As mentioned, unfortunately the results illustrate a significant amount of variation. This can partly be explained by the fact that stock solutions were not made up of the derivatives, due to the reported instabilities of these ligands in solution for long periods of time by the Conway group. Additionally, the signal for the uninhibited PKB PH domain in this assay (after subtracting background noise) is not as high as one may hope (approximately 0.2 au). Therefore a small amount of variation in the background noise can have an effect on the data. Attempts to maximise the signal, whilst maintaining a low background noise, were made by investigating various blocking procedures, washes, well plates and detector solutions. Whilst the assay has been partially optimised, the use of other buffers, Mg^{2+}, and different blocking agents would also be useful considerations for future work. Due to time limitations, these parameters were not taken into account.

7. 4. 7 *In Vitro* Phosphorylation Assay

In addition to the *in vitro* assays detailed, additional assays were subsequently carried out by Mingxuan Wu and Dorothea Fiedler at Princeton University. The inhibitory behaviour of the Ins(1,3,4,5)P_4 derivatives [and Ins(1,3,4,5)P_4] was investigated by assessing their ability to inhibit phosphorylation of PKB on Thr308. It was hypothesised that the derivatives would bind to PKB and hence stabilise it in an inactive conformation, thus preventing its phosphorylation at Thr308 by the upstream kinase PDK1.[203]

7. 4. 7. 1 Method

Inactive human PKB was incubated with the Ins(1,3,4,5)P_4 derivatives at two concentrations (625 nM and 10 μM) for 20 minutes. Following this, activated PDK1 was then added. After a further 30 minutes, SDS loading buffer was used to quench the reaction. The samples were boiled for 5 minutes at 95°C and resolved by SDS-PAGE. After transferring onto a nitrocellulose membrane, PKB and phosphorylated PKB were detected by western blot analysis with PKB or phospho-PKB (Thr308) antibodies. Bands were quantified using ImageJ software and data were analysed using GraphPad Prism 5 software. For further details about the protocol used by the group, please refer to reference 203.[203]

7. 4. 7. 2 Results

The results of Wu and Fiedler gratifyingly illustrate the same inhibition pattern as our computational and *in vitro* results, Figure 55. Unfortunately their results for Ins(1,3,4,5)P_4 are void, as the authors noted that their batch of this ligand was insoluble. We also experienced the same problem with the batch of Ins(1,3,4,5)P_4 synthesised by the Conway group, and therefore purchased this ligand from Sigma Aldrich for use in our assays.

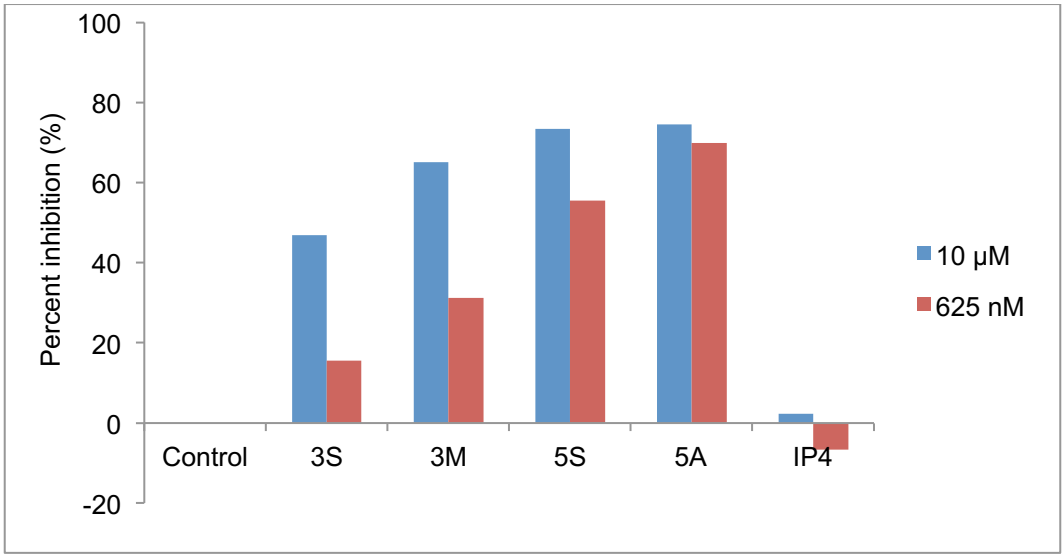


Figure 55: Inhibition of PKB phosphorylation at Thr308 by inositol phosphate derivatives. Inhibition was calculated as the percent decrease in the proportion of phosphorylated PKB in the samples incubated with the inositol phosphate derivatives compared to the proportion in the control (in which PKB was incubated with water). The compounds tested include D-3-*O*-Sulfamoyl-*myo*-Ins(1,4,5)P_3 (3S), D-*myo*-Ins(1,4,5)P_3-3-*O*-methylphosphate ester (3M), D-5-*O*-Sulfamoyl-*myo*-Ins(1,3,4)P_3 (5S), D-*myo*-Ins(1,3,4)P_3-5-*O*-acetamidoethylphosphate ester (5A) and D-*myo*-Ins(1,3,4,5)P_4 (IP4).

In agreement with our previous findings, Wu and Fiedler's assay illustrates that the 5-position acetamidoethylphosphate ester **5A** has the greatest inhibitory effect on the phosphorylation of PKB at Thr308, with approximately 70% inhibition at 625 nM. Additionally they also show the 3-position sulfamate-based derivative **3S** to have the poorest inhibitory effect, with only 15% inhibition at 625 nM.

Wu and Fiedler's results further support the idea that modification at the 5-position of Ins(1,3,4,5)P_4 is suitable and promising for the design of high affinity Ins(1,3,4,5)P_4 derivatives. As well as the strong inhibitory effect of the 5-position acetamidoethylphosphate ester **5A** on PKB phosphorylation, the 5-position sulfamate ester **5S** also illustrates good inhibition, with 55% inhibition at 625 nM.

As suggested from our previous results, the 3-position methylphosphate ester **3M** shows less inhibitory activity than the 5-position derivatives, with approximately 30% inhibition at 625 nM, but has a greater effect than the 3-position sulfamate derivative.

7. 5 Comparison of *In Vitro* and Computational Results

The computational results and the two *in vitro* assays all illustrate a very good agreement. All three show that the 5-position acetamidoethylphosphate ester **5A** binds with the strongest affinity/has the greatest inhibitory effect on PKB, and the 3-position sulfamate ester **3S** the weakest. Results for the remaining derivatives are also in very good agreement, with the 5-position sulfamate derivative **5S** showing good inhibitory behaviour, and the 3-postion methylphosphate derivative **3M** behaving very similarly to Ins(1,3,4,5)P_4.

7. 6 Conclusion

In this chapter numerous Ins(1,3,4,5)P_4 derivatives were assessed for their ability to bind to/inhibit PKB using both computational and experimental techniques. All approaches showed a good agreement. This agreement not only provides further confidence in our computational model, but additionally provides important information about PKB's binding interface.

Of particular importance are the results for the 5-position derivatives. Both *in vitro* assays, as well as the computational data, show that the 5-position acetamidoethylphosphate ester **5A** has the strongest affinity for/greatest inhibitory effect on PKB, of the ligands investigated. This derivative could thus potentially serve as a starting point for second-generation designs. Additionally, results for the 5-position sulfamate derivative further support the argument that modification at this site on Ins(1,3,4,5)P_4 is suitable and tolerated by the PKB PH domain.

The 3-position methylphosphate ester illustrates a similar affinity to Ins(1,3,4,5)P_4 in both the computational and experimental studies. This result informs us that adding a methyl moiety to the 3-position phosphate group appears to be accommodated. The MD simulations suggest that the small methyl group is able to fit into a pocket in the PKB binding site and that the only significant effect to binding is a slightly dampened interaction *via* the methoxy group. The accommodation of the methyl moiety by the PKB PH domain is also supported by work carried out by Elliott, which shows that this modification is also tolerated at the 1-position.[174]

Results herein suggest that a sulfamate moiety is not the optimum phosphate bioisostere for this system if hoping to improve binding affinity. This is due to the fact that this system's binding interface is highly dependant on electrostatic and hydrogen bond interactions. Whilst simulations show that the sulfamate group is still capable of forming interactions with the PKB PH domain, they are significantly dampened compared to those of a phosphate group. Despite this, sulfonamide-based compounds have previously been reported as PKB inhibitors. However these

compounds are not inositol-based and were shown to form numerous other interactions in addition to those *via* the sulfonamide moiety.[57]

The computational work carried out in this chapter prompted additional questions about the protocol used to study this system. Firstly, the parameterisation of the sulfamate-based derivatives highlighted that different levels of theory may be suitable for the optimisation of different ligands. The idea of optimising ligands in solvent was also re-introduced, and should be a consideration for future studies.

The simulations also highlighted the issue of sampling phase space. For the 5-postion acetamidoethylphosphate ester, one isomer was shown to exhibit a significantly higher binding affinity for the PKB PH domain compared to the other isomers. Additionally, different inter- and intra-molecular interactions were also apparent for this high affinity isomer. The different behaviour of this isomer could in fact be a true insight into its actions in the PKB binding site. However it is also possible that the system is sampling a rare event in phase space and is essentially trapped in a different local minimum to the other systems. Alternatively the other systems may have not reached an equilibrium state. To test this theory the systems were run for a further 100 ns, however no significant changes in the free energies of binding or molecular behaviour were apparent. Additional computational and experimental investigations could assist in deciding if the binding mode illustrated for this high affinity isomer is true, or an artefact of the simulation.

Overall this chapter has provided a large amount of information about the binding interface of the PKB PH domain, using a combination of computational and experimental techniques. The route to obtain this information has itself been useful, providing invaluable information about suitable computational and biochemical techniques to study this unique and complicated system.

Chapter 8 – Summary and Future Work

This thesis reports the first full set of parameters for Ins(1,3,4,5)P_4, Ins(1,3,4)P_3 and Ins(1,4,5)P_3, generated using a clearly defined algorithmic approach that is in-keeping with the Amber force field. Previous attempts at generating parameters for these highly charged ligands are sparse. Moreover the examples that do exist typically only consider one ionisation state for these molecules.[62, 63, 65, 68] The Charmm force field has recently been extended, providing parameters for mono- and di-anionic phosphate groups attached to carbohydrates.[61] However the use of these parameters in simulations of Ins(1,3,4,5)P_4 with the PKB PH domain failed to capture some critical experimentally determined interactions, placing doubt on the quality of the parameters.

Aside from the Charmm study mentioned above, the only other published example of investigating the interaction of the PKB PH domain with phosphoinositides/inositol phosphates using MD simulations is by Rong *et al*.[68] Limited information is available from this study due to the lack of crystal structure available at the time, and the short time-scale of the simulations. The work herein therefore provides the first in depth investigation into the interaction of inositol phosphates with the PKB PH domain using MD simulations.

We are gratified to note that our simulations support PKB's widely accepted preference for 3-phosphorylated phosphoinositides.[22-25] In addition to supporting experimental findings, our MD simulations capture subtle structural movements and behaviour that are not apparent from the crystal structure. Perhaps most notably, the simulations help to rationalise the relatively high affinity of Ins(1,4,5)P_3 for the PKB PH domain by illustrating a different binding mode to that previously presumed.[22]

MD simulations of the E17K PKB PH domain help explain this mutant's increased affinity for the plasma membrane, and hence its affiliation with a number of human cancers.[29] The simulations illustrate that this mutant has an increased affinity for the head-group of PtdIns(4,5)P_2 [and PtdIns(3,4,5)P_3], as has been hypothesised from

experimental work.[164] The atomic level view provides reasoning for this, indicating interactions of Lys17 with phosphate group 5 to be mainly responsible.

Information from the simulations mentioned was used to assist in the design of inositol phosphate derivatives to target the PKB PH domain. We aimed to design a high affinity derivative that would prevent PKB's recruitment to the plasma membrane and hence its subsequent phosphorylation and activation. Whilst a similar approach has been attempted by others, there has been limited success in this area.[54, 159] Herein MD simulations have allowed a number of derivatives to be rapidly screened, without requiring the time-consuming and expensive synthesis of such compounds. Both computational and experimental data suggest a 5-position derivative with an increased affinity/inhibitory behaviour for the PKB PH domain, compared to Ins(1,3,4,5)P_4, has been successfully designed.

Additional considerations for adapting and building upon this thesis are discussed in the following sections.

8. 1 Inositol Phosphate Parameters

The parameters developed herein provide a basis to study the broad range of interactions that inositol phosphates are involved in. The atomic level view that MD simulations afford can help us understand these interactions in greater detail, providing assistance to experimental investigations. Particularly relevant applications include studying the interaction of these ligands with PDK1, which is also involved in the PI3K pathway. PDK1 has a lysine residue at a similar position to that of the E17K PKB mutant.[204] It would therefore be interesting to determine if PDK1's increased affinity for PtdIns(3,4,5)P_3 compared to PtdIns(3,4)P_2 is due to this residue.[29]

Whilst not detailed herein, we have also applied the inositol phosphate parameters to study their interactions with InsP_3R (inositol tris-phosphate receptor). InsP_3R acts as Ca^{2+} release channel activated by Ins(1,4,5)P_3. Results from these simulations are in good agreement with what is experimentally known, providing further validation of the inositol phosphate parameters.

8. 2 Parameterisation of Phosphoinositide Lipids

Inositol phosphate parameters are useful for understanding the interactions of their parent phosphoinositide lipids. Whilst interactions of phosphoinositide lipids are primarily mediated *via* their head-groups, contacts *via* their lipid tails are also of importance. This has previously been illustrated by the different binding modes of PtdIns(4,5)P_2 and its head-group Ins(1,4,5)P_3 to the PKB PH domain. Other lipids within a bilayer are also believed to provide additional contacts with many PH domains, in order to supplement their interaction with phosphoinositide lipids.[65, 66]

To investigate lipid-binding proteins in a native membrane environment, parameters for lipids need to be generated for the Amber force field. During my PhD, I also attempted to generate charge parameters for the phosphoinositide lipids PtdIns(3,4,5)P_3, PtdIns(3,4)P_2 and PtdIns(4,5)P_2. Again all physiological ionisation states were considered. Unfortunately there were a number of issues with optimising the geometry of these compounds using the DFT-B3LYP method. The HF level of theory was therefore applied. The HF level of theory has also been used in previous attempts to generate charge parameters for PtdIns(4,5)P_2.[62, 63] As these calculations were extremely time consuming, a complete set of optimised geometries and generated parameters were not obtained for these lipids. However, this work will be continued within the Gould group.

During a two month visit to Ross Walker's group in San Diego, I also helped in the development of other lipid parameters.[205] Members of both the Gould and Walker groups are now focusing on parameterising an Amber lipid force field.[205, 206] These parameters will not only allow the investigation of membrane docking proteins, but additionally the behavior of membrane proteins, mixed lipid bilayers and the possible existence of lipid rafts, amongst many others applications.

8. 3 Full PKB Simulations

Whilst carrying out my PhD, PKB's PH and kinase domains were co-crystallised for the first time.[28] Although this crystal structure was not applied in our simulations, it would be of great interest for future studies. Given the current advances in computational speed, with GPU accelerated calculations, it is a perfect time to study a system of this size (approximately 430 amino acids, and 40,000 atoms when solvated with 8 Å of explicit water). The simulation of both domains could provide critical information about their binding interface and mechanism of action. Additionally the study of both domains may also be of interest for investigating the E17K PKB mutant, as Glu17 has been proposed to interact with residues on PKB's kinase domain in its inactive conformation.[28]

8. 4 Selectivity of Ins(1,3,4,5)P_4 Derivatives

In Chapter 7, a number of Ins(1,3,4,5)P_4 derivatives were screened for their ability to bind to the PKB PH domain using both experimental and computational techniques. These approaches led to the discovery of an Ins(1,3,4,5)P_4 derivative with a proposed increased affinity for the PKB PH domain compared to Ins(1,3,4,5)P_4. It would therefore be of great interest to test these ligands against the E17K PKB PH domain, as well as other closely related kinases to determine its selectivity.

8. 5 Conclusion

The work described in this thesis provides a platform upon which other studies can now be based. The information obtained about the PKB PH domain-inositol phosphate binding interface will aid in future attempts to target this system. The importance of such information should not be understated, as targeted PKB therapeutics are currently of great interest. [54, 57, 159, 202] PKB plays a key role in one of the most frequently activated proliferation pathways in cancer (the PI3K pathway), yet currently no PKB-targeted drugs have reached the market.

In addition to providing information about the PKB PH domain and its interactions, this thesis also provides parameters for members of the inositol phosphate family. This family of ligands are involved in a number of critical cellular activities.[207] Whilst their interaction with the PKB PH domain has been assessed herein, these parameters also provide the opportunity to obtain an atomic level view of other important interactions that have yet to be explored.

Chapter 9 – Materials and Methods

9. 1 General

Reagents were obtained from GE Healthcare® unless stated otherwise. Ultrapure laboratory water was obtained from MilliQ® Millipore water purification systems. Ins(1,3,4,5)P_4, Ins(1,4,5)P_3 and InsP_6 were purchased from Echelon® and PtdIns(3,4,5)P_3, PtdIns(3,4)P_3 and PtdIns(4,5)P_3 from Sigma Aldrich® (C16, sodium salt).

9. 2 Protein Expression

DH5α *E.coli* cells were removed from the -80°C freezer and put immediately on ice. 0.5 µl of the pGEX plasmid expression vector containing the sequence for the PKB PH domain (kindly donated from the Woscholski group) was added to the cells, and incubated on ice for 30 minutes. After 30 minutes, the cells were heat-shocked for 90 seconds at 42°C and then put back on ice for two minutes. Under sterile conditions, the *E.coli* cells were spread onto an agar plate containing ampicillin and incubated at 37°C for 16 hours.

The plate was removed from the incubator and stored at 4°C. Under a sterile environment, 10 ml of lysogeny broth (LB) media was added to a sterile 50 ml falcon tube. 10 µl of ampicillin was added to the 10 ml of LB media. A colony from the agar plate was picked with a P2 pipette tip and released into the falcon. Cells were grown at 37°C for 16 hours, with shaking.

Under a sterile environment, 1 ml of ampicillin was added to 1 l of LB media in an Erlenmeyer flask. The 10 ml of culture grown for 16 hours was added to the 1 l of LB media and grown for a further 3-5 hours at 37°C, with shaking. The OD (optical density) was checked regularly. At an OD between 0.6 and 0.7 (mid-log phase of bacterial growth), protein expression was induced with IPTG (1mM final concentration) and fresh ampicillin (1 ml), for 24 hours at 23°C.

Cells were harvested by centrifugation at 4600 rpm at 4°C for 15 minutes. The cells were resuspended in 60 ml of lysis buffer containing 50 mM Tris (pH 7.4), 1% Triton X-100, 2 mM EDTA, 10 mM benzamidine hydrochloride, 100 μg ml^{-1} soybean trypsin inhibitor, 2 mg ml^{-1} lysozyme, 1 mM 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride, and 2 mM DTT. The cell suspension was stirred at 4°C for 1 hour and then sonicated (5 x 1 minute). The cell suspension was then manually homogenised until no longer viscous, followed by centrifugation at 11500 rpm at 4°C for 1 hour, ensuring the supernatant was clear.

The glutathione sepharose column was pre-equilibrated with 50mM Tris (pH 7.4), 140 mM NaCl, 2.7 mM KCl. The column was washed twice with this equilibrating buffer, with each wash totalling a volume of 30 ml. The supernatant was then filtered through a 0.45 μm syringe filter and loaded onto the column. After loading, the column was washed:

- Twice with 50 mM Tris (pH 7.4), 140 mM NaCl, 2.7 mM KCl, 1% Triton and 2 mM DTT (each wash equaling a total volume of 20 ml)
- Twice with 50 mM Tris (pH 7.4), 140 mM NaCl, 2.7 mM KCl and 2 mM DTT (each wash equaling a total volume of 20 ml)
- Twice with 50 mM Tris (pH 7.4), 500 mM NaCl, 2.7 mM KCl, and 2 mM DTT (each wash equaling a total volume of 20 ml)

After washing, 10 ml of elution buffer [50 mM Tris (pH 7.4), 20 mM reduced glutathione, 250 mM NaCl and 2 mM DTT] was added to the beads and incubated overnight. The eluted protein was collected in small fractions and concentrated at 4°C and 2000 rpm.

9. 3 SDS PAGE

SDS PAGE gels were cast with the following compositions:

Table 17: Components for 10% running gel

Component	Volume
Distilled H_2O	8.2 ml
30% Acrylamide/Bis acrylamide	6.6 ml
1.5 M Tris-HCl pH 8.8	5 ml
10% w/v Sodium dodecyl sulfate (SDS)	200 µl
10% w/v Ammonium persulfate (APS)	100 µl
Tetramethylethylenediamine (TEMED)	10 µl

Table 18: Components for 5% stacking gel

Component	Volume
Distilled H_2O	5.7 ml
30% v/v Acrylamide/Bis	1.7 ml
0.5 M Tris-HCl pH 6.8	2.5 ml
10% w/v Sodium dodecyl sulfate (SDS)	100 µl
10% w/v Ammonium persulfate (APS)	50 µl
Tetramethylethylenediamine (TEMED)	10 µl

The GST-tagged PKB PH domain was diluted in distilled water to concentrations of 0.5 and 1.0 mg ml^{-1}. 4 µl of 0.5 mg ml^{-1} GST-tagged PKB PH domain was added to 4 µl SDS loading buffer (total 2 µg of protein). 10 µl of 1.0 mg ml^{-1} GST-tagged PKB PH domain was added to 10 µl SDS loading buffer (total 10 µg of protein). Both aliquots of the GST-tagged PKB PH domain in SDS loading buffer were incubated at 95°C for 5 minutes, and then on ice for 5 minutes. Each aliquot was then loaded into separate wells of the gel. Molecular weight calibration was achieved by loading 2.5 µl of Prism Ultra Protein Ladder (10-245 kDa) from abcam®. Gels were run at 150 V in electrophoresis running buffer. The components of the SDS loading buffer, and electrophoresis running buffer are detailed below. Gels were visualised using Coomassie Blue Staining. The gel was stained with Coomassie blue solution overnight with agitation and then destained for an hour with distilled water.

Table 19: SDS loading buffer and SDS electrophoresis running buffer compositions.

SDS loading buffer	62.5 mM Tris-HCl, pH 6.8, 2% w/v SDS, 10% v/v glycerol, 50 mM DTT, 0.01% w/v bromophenol blue.
SDS electrophoresis running buffer	25 mM Tris, 192 mM glycine, and 0.1% SDS.

9. 4 Bradford Assay

The Bradford assay was used to determine protein concentration. 200 µl of Bradford reagent was added to 20 µl of eluted protein (at varying dilutions) on a 96 well-plate. The absorbance of each well at 595 nm was read using a spectrophotometer. An ideal absorbance is between 0.5 and 1, as the assay is only linear over a short range of concentrations (typically 2-1500 µg ml^{-1} of protein). Measurements were made at varying dilutions of the protein to allow the concentration of the eluted protein to be accurately determined. The concentration of the eluted protein was calculated from the following equation:

$$X = \frac{Y - 0.0012}{1.196}$$

Equation 24: Calculating the concentration of protein using the Bradford assay. The concentration of protein within each well X is calculated using Y. Y is the difference between the absorbance of the sample well and of a blank cell (i.e. 20 µl of distilled water in 200 µl Bradford reagent) at 595 nm. The concentration of the eluted protein (before dilution) can then be calculated by multiplying concentration X by the dilution factor.

9. 5 Buffer Solutions

PBS	137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4.2H_2O$ (sodium phosphate dibasic), 2 mM KH_2PO_4 (monopotassium phosphate), pH 7.4
TBS	50 mM Tris-Cl, 150 mM NaCl, pH 7.4

References

1. D. Lietha, *Phosphoinositides – The Seven Species: Conversion and Cellular Roles. In: Encyclopedia of Life Sciences*, John Wiley and Sons, Ltd, 2011.
2. H. J. McCrea and P. De Camilli, *Physiology*, 2009, **24**, 8-16.
3. M. Borkovec and B. Spiess, *Physical Chemistry Chemical Physics*, 2004, **6**, 1144-1151.
4. P. Guedat, G. Schlewer, E. Krempp, A. M. Riley, B. V. L. Potter and B. Spiess, *Chemical Communications*, 1997, **6**, 625-626.
5. E. E. Kooijman, K. E. King, M. Gangoda and A. Gericke, *Biochemistry*, 2009, **48**, 9360-9371.
6. W. Kumler and J. J. Eiler, *J. Am. Chem. Soc.*, 1943, **65**, 2355-2361.
7. M. A. Lemmon, *Nature Reviews Molecular Cell Biology*, 2008, **9**, 99-111.
8. S. Corvera and M. P. Czech, *Trends in Cell Biology*, 1998, **8**, 442-446.
9. M. A. Lemmon, K. M. Ferguson and C. S. Abrams, *Febs Letters*, 2002, **513**, 71-76.
10. T. Itoh and T. Takenawa, *Cellular Signalling*, 2002, **14**, 733-743.
11. C. C. Milburn, M. Deak, S. M. Kelly, N. C. Price, D. R. Alessi and D. M. F. Van Aalten, *Biochemical Journal*, 2003, **375**, 531-538.
12. S. Cockcroft, *Biology of Phosphoinositides*, Oxford University Press, 2000.
13. M. J. Berridge, *Biochemical Journal*, 1984, **220**, 345-360.
14. L. Stephens, A. McGregor and P. Hawkins, *Biology of Phosphoinositides*, 2000, **27**, 32-108.
15. Z. Z. Yang, O. Tschopp, A. Baudry, B. Dummler, D. Hynx and B. A. Hemmings, *Biochemical Society Transactions*, 2004, **32**, 350-354.
16. E. M. Sale, C. P. Hodgkinson, N. P. Jones and G. J. Sale, *Biochemistry*, 2006, **45**, 213-223.
17. M. Sun, G. Wang, J. E. Paciga, R. I. Feldman, Z. Q. Yuan, X. L. Ma, S. A. Shelley, R. Jove, P. N. Tsichlis, S. V. Nicosia and J. Q. Cheng, *American Journal of Pathology*, 2001, **159**, 431-437.
18. M. A. Lemmon, *Cell Biology of Inositol Lipids and Phosphates*, 2007, **74**, 81-93.
19. K. M. Ferguson, M. A. Lemmon, J. Schlessinger and P. B. Sigler, *Cell*, 1994, **79**, 199-209.
20. J. H. Hurley and S. Misra, *Annual Review of Biophysics and Biomolecular Structure*, 2000, **29**, 49-79.
21. L. E. Rameh, A. K. Arvidsson, K. L. Carraway, A. D. Couvillon, G. Rathbun, A. Crompton, B. VanRenterghem, M. P. Czech, K. S. Ravichandran, S. J. Burakoff, D. S. Wang, C. S. Chen and L. C. Cantley, *Journal of Biological Chemistry*, 1997, **272**, 22059-22066.
22. M. Frech, M. Andjelkovic, E. Ingley, K. K. Reddy, J. R. Falck and B. A. Hemmings, *Journal of Biological Chemistry*, 1997, **272**, 8474-8481.
23. J. M. Kavran, D. E. Klein, A. Lee, M. Falasca, S. J. Isakoff, E. Y. Skolnik and M. A. Lemmon, *Journal of Biological Chemistry*, 1998, **273**, 30497-30508.
24. S. R. James, C. P. Downes, R. Gigg, S. J. A. Grove, A. B. Holmes and D. R. Alessi, *Biochemical Journal*, 1996, **315**, 709-713.
25. T. F. Franke, D. R. Kaplan, L. C. Cantley and A. Toker, *Science*, 1997, **275**, 665-668.

26. C. C. Thomas, M. Deak, D. R. Alessi and D. M. F. van Aalten, *Current Biology*, 2002, **12**, 1256-1262.
27. M. A. Lemmon, K. M. Ferguson, R. Obrien, P. B. Sigler and J. Schlessinger, *Proceedings of the National Academy of Sciences of the United States of America*, 1995, **92**, 10472-10476.
28. W. I. Wu, W. C. Voegtli, H. L. Sturgis, F. P. Dizon, G. P. A. Vigers and B. J. Brandhuber, *Plos One*, 2010, **5**, e12913.
29. J. D. Carpten, A. L. Faber, C. Horn, G. P. Donoho, S. L. Briggs, C. M. Robbins, G. Hostetter, S. Boguslawski, T. Y. Moses, S. Savage, M. Uhlik, A. M. Lin, J. Du, Y. W. Qian, D. J. Zeckner, G. Tucker-Kellogg, J. Touchman, K. Patel, S. Mousses, M. Bittner, R. Schevitz, M. H. T. Lai, K. L. Blanchard and J. E. Thomas, *Nature*, 2007, **448**, 439-444.
30. S. J. Mills, D. Komander, M. N. Trusselle, S. T. Safrany, D. M. F. van Aalten and B. V. L. Potter, *Acs Chemical Biology*, 2007, **2**, 242-246.
31. B. D. Manning and L. C. Cantley, *Cell*, 2007, **129**, 1261-1274.
32. I. Vivanco and C. L. Sawyers, *Nature Reviews Cancer*, 2002, **2**, 489-501.
33. S. R. Datta, H. Dudek, X. Tao, S. Masters, H. A. Fu, Y. Gotoh and M. E. Greenberg, *Cell*, 1997, **91**, 231-241.
34. M. H. Cardone, N. Roy, H. R. Stennicke, G. S. Salvesen, T. F. Franke, E. Stanbridge, S. Frisch and J. C. Reed, *Science*, 1998, **282**, 1318-1321.
35. A. Brunet, A. Bonni, M. J. Zigmond, M. Z. Lin, P. Juo, L. S. Hu, M. J. Anderson, K. C. Arden, J. Blenis and M. E. Greenberg, *Cell*, 1999, **96**, 857-868.
36. B. H. P. Zhou and M. C. Hung, *Seminars in Oncology*, 2002, **29**, 62-70.
37. N. Fujita, S. Sato, K. Katayama and T. Tsuruo, *Journal of Biological Chemistry*, 2002, **277**, 28706-28713.
38. J. A. Diehl, M. G. Cheng, M. F. Roussel and C. J. Sherr, *Genes & Development*, 1998, **12**, 3499-3511.
39. D. R. Alessi and P. Cohen, *Current Opinion in Genetics & Development*, 1998, **8**, 55-62.
40. D. D. Sarbassov, D. A. Guertin, S. M. Ali and D. M. Sabatini, *Science*, 2005, **307**, 1098-1101.
41. L. Bozulic and B. A. Hemmings, *Current Opinion in Cell Biology*, 2009, **21**, 256-261.
42. V. Calleja, M. Laguerre and B. Larijani, *J Chem Biol*, 2009, **2**, 11-25.
43. P. X. Liu, H. L. Cheng, T. M. Roberts and J. J. Zhao, *Nature Reviews Drug Discovery*, 2009, **8**, 627-644.
44. C. Le Page, I. H. Koumakpayi, M. Alam-Fahmy, A. M. Mes-Masson and F. Saad, *British Journal of Cancer*, 2006, **94**, 1906-1912.
45. G. Ayala, T. Thompson, G. Yang, A. Frolov, R. L. Li, P. Scardino, M. Ohori, T. Wheeler and W. Harper, *Clinical Cancer Research*, 2004, **10**, 6572-6578.
46. J. I. Kreisberg, S. N. Malik, T. J. Prihoda, R. G. Bedolla, D. A. Troyer, S. Kreisberg and P. M. Ghosh, *Cancer Research*, 2004, **64**, 5232-5236.
47. O. Stal, G. Perez-Tenorio, L. Akerberg, B. Olsson, B. Nordenskjold, L. Skoog and L. E. Rutqvist, *Breast Cancer Research*, 2003, **5**, R37-R44.
48. E. M. Sale and G. J. Sale, *Cellular and Molecular Life Sciences*, 2008, **65**, 113-127.
49. D. Malanga, M. Scrima, C. De Marco, F. Fabiani, N. De Rosa, S. De Gisi, N. Malara, R. Savino, G. Rocco, G. Chiappetta, R. Franco, V. Tirino, G. Pirozzi and G. Viglietto, *Cell Cycle*, 2008, **7**, 665-669.

50. M. S. Kim, E. G. Jeong, N. J. Yoo and S. H. Lee, *British Journal of Cancer*, 2008, **98**, 1533-1535.
51. F. E. Bleeker, L. Felicioni, F. Buttitta, S. Lamba, L. Cardone, M. Rodolfo, A. Scarpa, S. Leenstra, M. Frattini, M. Barbareschi, M. Del Grammastro, M. G. Sciarrotta, C. Zanon, A. Marchetti and A. Bardelli, *Oncogene*, 2008, **27**, 5648-5650.
52. J. K. Morrow, L. Du-Cuny, L. Chen, E. J. Meuillet, E. A. Mash, G. Powis and S. X. Zhang, *Recent Patents on Anti-Cancer Drug Discovery*, 2011, **6**, 146-159.
53. C. W. Lindsley, *Current Topics in Medicinal Chemistry*, 2010, **10**, 458-477.
54. E. J. Meuillet, *Current Medicinal Chemistry*, 2011, **18**, 2727-2742.
55. E. J. Meuillet, D. Mahadevan, H. Vankayalapati, M. Berggren, R. Williams, A. Coon, A. P. Kozikowski and G. Powis, *Molecular Cancer Therapeutics*, 2003, **2**, 389-399.
56. G. A. Ruiter, S. F. Zerp, H. Bartelink, W. J. van Blitterswijk and M. Verheij, *Anti-Cancer Drugs*, 2003, **14**, 167-173.
57. S. A. Moses, M. A. Ali, Z. H. Song, D. C. Lei, L. L. Zhou, R. Lemos, N. Ihle, A. G. Skillman, S. X. Zhang, E. A. Mash, G. Powis and E. J. Meuillet, *Cancer Research*, 2009, **69**, 5073-5081.
58. N. Berndt, H. Yang, B. Trinczek, S. Betzi, Z. Zhang, B. Wu, N. J. Lawrence, M. Pellecchia, E. Schonbrunn, J. Q. Cheng and S. M. Sebti, *Cell Death and Differentiation*, 2010, **17**, 1795-1804.
59. S. B. Kondapaka, S. S. Singh, G. P. Dasmahapatra, E. A. Sausville and K. K. Roy, *Molecular Cancer Therapeutics*, 2003, **2**, 1093-1103.
60. S. J. Conway and G. J. Miller, *Natural Product Reports*, 2007, **24**, 687-707.
61. S. S. Mallajosyula, O. Guvench, E. Hatcher and A. D. MacKerell, *Journal of Chemical Theory and Computation*, 2012, **8**, 759-776.
62. I. Liepina, C. Czaplewski, P. Janmey and A. Liwo, *Biopolymers*, 2003, **71**, 49-70.
63. D. Lupyan, M. Mezei, D. E. Logothetis and R. Osman, *Biophysical Journal*, 2010, **98**, 240-247.
64. Z. Li, R. M. Venable, L. A. Rogers, D. Murray and R. W. Pastor, *Biophysical Journal*, 2009, **97**, 155-163.
65. E. Psachoulia and M. S. P. Sansom, *Biochemistry*, 2008, **47**, 4211-4220.
66. C. N. Lumb, J. He, Y. Xue, P. J. Stansfeld, R. V. Stahelin, T. G. Kutateladze and M. S. P. Sansom, *Structure*, 2011, **19**, 1338-1346.
67. C. N. Lumb and M. S. P. Sansom, *Plos Computational Biology*, 2012, **8**, e1002617.
68. S. B. Rong, Y. H. Hu, I. Enyedy, G. Powis, E. J. Meuillet, X. W. Wu, R. X. Wang, S. M. Wang and A. P. Kozikowski, *Journal of Medicinal Chemistry*, 2001, **44**, 898-908.
69. M. Planck, *Annalen der Physik*, 1900, **306**, 719-737.
70. A. Einstein, *Annalen der Physik*, 1905, **322**, 132-148.
71. E. Schrodinger, *Annals of Physics*, 1926, **79**, 361-376.
72. A. R. Leach, *Molecular Modelling, Principles and Applications*, 2nd edn., Pearson Education, 2001.
73. D. R. Hartree, *Proceedings of the Cambridge Philosophical Society*, 1928, **24**, 89-132.
74. N. J. B. Green, *Quantum Mechanics 1: Foundations*, Oxford University Press, 1997.

75. P. W. Atkins, *Physical Chemistry*, Oxford University Press, 1994.
76. J. B. Foresman and E. Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian, Inc., 1996.
77. R. C. Walker, PhD Thesis, Imperial College London, 2004.
78. P. Hohenberg and W. Kohn, *Physical Review*, 1964, **B136**, 864-871.
79. W. Kohn and L. J. Sham, *Physical Review*, 1965, **A140**, 1133-1138.
80. A. D. Becke, *Journal of Chemical Physics*, 1993, **98**, 1372-1377.
81. C. T. Lee, W. T. Yang and R. G. Parr, *Physical Review B*, 1988, **37**, 785-789.
82. A. E. Mattsson, P. A. Schultz, M. P. Desjarlais, T. R. Mattsson and K. Leung, *Modelling and Simulation in Materials Science and Engineering*, 2005, **13**, R1-R31.
83. J. C. Slater, *Physical Review*, 1930, **36**, 57-64.
84. S. F. Boys, *Proceedings of the Royal Society (London)*, 1950, **A200**, 542-554.
85. J. S. Binkley, J. A. Pople and W. J. Hehre, *Journal of the American Chemical Society*, 1980, **102**, 939-947.
86. P. C. Hariharan and J. A. Pople, *Theoretica Chimica Acta*, 1973, **28**, 213-222.
87. M. Born and R. Oppenheimer, *Annalen der Physik*, 1927, **84**, 457-484.
88. R. Salomon-Ferrer, D. A. Case and R. C. Walker, *WIRES Comput Mol Sci*, In press.
89. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *Journal of Computational Chemistry*, 2009, **30**, 1545-1614.
90. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *Journal of Chemical Theory and Computation*, 2008, **4**, 435-447.
91. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins-Structure Function and Bioinformatics*, 2006, **65**, 712-725.
92. J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *Journal of Computational Chemistry*, 2004, **25**, 1157-1174.
93. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *Journal of the American Chemical Society*, 1995, **117**, 5179-5197.
94. C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *Journal of Physical Chemistry*, 1993, **97**, 10269-10280.
95. P. Cieplak, W. D. Cornell, C. Bayly and P. A. Kollman, *Journal of Computational Chemistry*, 1995, **16**, 1357-1377.
96. W. D. Cornell, P. Cieplak, C. I. Bayly and P. A. Kollman, *Journal of the American Chemical Society*, 1993, **115**, 9620-9631.
97. B. H. Besler, K. M. Merz and P. A. Kollman, *Journal of Computational Chemistry*, 1990, **11**, 431-439.
98. T. Darden, D. York and L. Pedersen, *Journal of Chemical Physics*, 1993, **98**, 10089-10092.
99. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *Journal of Chemical Physics*, 1983, **79**, 926-935.

100. F. Jensen, *Introduction to Computational Chemistry*, John Wiley and Sons, Ltd, 1999.
101. D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Elsevier, 2002.
102. J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *Journal of Computational Physics*, 1977, **23**, 327-341.
103. A. W. Gotz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, *Journal of Chemical Theory and Computation*, 2012, **8**, 1542-1555.
104. J. Roy and C. A. Laughton, *Biophysical Journal*, 2010, **99**, 218-226.
105. A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, III, C. A. Laughton and M. Orozco, *Biophysical Journal*, 2007, **92**, 3817-3829.
106. T. Werner, M. B. Morris, S. Dastmalchi and W. B. Church, *Advanced Drug Delivery Reviews*, 2012, **64**, 323-343.
107. P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, *Accounts of Chemical Research*, 2000, **33**, 889-897.
108. J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman and D. A. Case, *Journal of the American Chemical Society*, 1998, **120**, 9401-9409.
109. M. K. Gilson and H. X. Zhou, *Annual Review of Biophysics and Biomolecular Structure*, 2007, **36**, 21-42.
110. H. Gouda, I. D. Kuntz, D. A. Case and P. A. Kollman, *Biopolymers*, 2003, **68**, 16-34.
111. I. Miller, B. R., J. McGee, D., J. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *Journal of Chemical Theory and Computation*, 2012, **8**, 3314-3321.
112. A. Onufriev, D. A. Case and D. Bashford, *Journal of Computational Chemistry*, 2002, **23**, 1297-1304.
113. M. Feig, A. Onufriev, M. S. Lee, W. Im, D. A. Case and C. L. Brooks, *Journal of Computational Chemistry*, 2004, **25**, 265-284.
114. H. Gohlke, C. Kiel and D. A. Case, *Journal of Molecular Biology*, 2003, **330**, 891-913.
115. T. J. Hou, J. M. Wang, Y. Y. Li and W. Wang, *Journal of Chemical Information and Modeling*, 2011, **51**, 69-82.
116. S. A. J. Rosen, P. R. J. Gaffney and I. R. Gould, *Physical Chemistry Chemical Physics*, 2011, **13**, 1070-1081.
117. S. Brauchi, G. Orta, C. Mascayano, M. Salazar, N. Raddatz, H. Urbina, E. Rosenmann, F. Gonzalez-Nilo and R. Latorre, *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**, 10246-10251.
118. S. McLaughlin, J. Y. Wang, A. Gambhir and D. Murray, *Annual Review of Biophysics and Biomolecular Structure*, 2002, **31**, 151-175.
119. M. Toner, G. Vaio, A. McLaughlin and S. McLaughlin, *Biochemistry*, 1988, **27**, 7435-7443.
120. D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin

and P. A. Kollman, *AMBER 10*, University of California, San Francisco, 2008.

121. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 03, Revision C. 02*, Gaussian, Inc.,Wallingford CT, 2004.
122. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. Montgomery, J. A., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09, Revision A.1*, Gaussian, Inc., Wallingford CT, 2009.
123. R. Dennington, T. Keith and J. Millam, *GaussView, Version 5*, Semichem Inc., Shawnee Mission KS, 2009.
124. D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, A. Gusarov, A. Kovalenko and P. A. Kollman, *AMBER 11*, University of California, San Francisco, 2010.
125. W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics*, 1996, **14**, 33-37.
126. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *Journal of Computational Chemistry*, 2004, **25**, 1605-1612.
127. J. C. Lindon, D. J. Baker, J. M. Williams and R. F. Irvine, *Biochemical Journal*, 1987, **244**, 591-595.
128. V. V. Murashov and J. Leszczynski, *Journal of Molecular Structure-Theochem*, 2000, **529**, 1-14.

129. P. Hansia, N. Guruprasad and S. Vishveshwara, *Biophysical Chemistry*, 2006, **119**, 127-136.
130. P. Guilbaud and G. Wipff, *Journal of Physical Chemistry*, 1993, **97**, 5685-5692.
131. M. Lee, *Amber Mailing List*, http://archive.ambermd.org/.
132. R. Walker, *AMBER mailing list*, http://archive.ambermd.org/.
133. D. Manna, N. Bhardwaj, M. S. Vora, R. V. Stahelin, H. Lu and W. H. Cho, *Journal of Biological Chemistry*, 2008, **283**, 26047-26058.
134. M. Caricato, G. Scalmani and M. J. Frisch, *Continuum Solvation Models in Chemical Physics: Theory and Applications*, John Wiley and Sons, Ltd, 2007.
135. J. Tomasi, B. Mennucci and E. Cances, *Journal of Molecular Structure-Theochem*, 1999, **464**, 211-226.
136. J. L. Pascual-Ahuir, E. Silla and I. Tuñón, *Journal of Computational Chemistry*, 1994, **15**, 1127-1138.
137. G. D. Hawkins, C. J. Cramer and D. G. Truhlar, *Chemical Physics Letters*, 1995, **246**, 122-129.
138. G. D. Hawkins, C. J. Cramer and D. G. Truhlar, *Journal of Physical Chemistry*, 1996, **100**, 19824-19839.
139. V. Tsui and D. A. Case, *Biopolymers*, 2000, **56**, 275-291.
140. J. Tomasi, B. Mennucci and R. Cammi, *Chemical Reviews*, 2005, **105**, 2999-3093.
141. S. Miertus, E. Scrocco and J. Tomasi, *Chemical Physics*, 1981, **55**, 117-129.
142. S. A. J. Rosen, P. R. J. Gaffney, B. Spiess and I. R. Gould, *Physical Chemistry Chemical Physics*, 2012, **14**, 929-936.
143. H. Takeuchi, T. Kanematsu, Y. Misumi, F. Sakane, H. Konishi, U. Kikkawa, Y. Watanabe, M. Katan and M. Hirata, *Biochimica Et Biophysica Acta-Molecular Cell Research*, 1997, **1359**, 275-285.
144. B. Vanhaesebroeck, S. J. Leevers, K. Ahmadi, J. Timms, R. Katso, P. C. Driscoll, R. Woscholski, P. J. Parker and M. D. Waterfield, *Annual Review of Biochemistry*, 2001, **70**, 535-602.
145. M. Felemez, P. Bernard, G. Schlewer and B. Spiess, *Journal of the American Chemical Society*, 2000, **122**, 3156-3165.
146. B. Spiess and M. Borkovec, *Personal Communication*, 2011.
147. A. Bondi, *Journal of Physical Chemistry*, 1964, **68**, 441-451.
148. W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, *Journal of the American Chemical Society*, 1990, **112**, 6127-6129.
149. B. Kuhn and P. A. Kollman, *Journal of Medicinal Chemistry*, 2000, **43**, 3786-3791.
150. J. M. Wang, P. Morin, W. Wang and P. A. Kollman, *Journal of the American Chemical Society*, 2001, **123**, 5221-5230.
151. S. H. Huo, J. M. Wang, P. Cieplak, P. A. Kollman and I. D. Kuntz, *Journal of Medicinal Chemistry*, 2002, **45**, 1412-1419.
152. E. Muzzioli, A. Del Rio and G. Rastelli, *Chemical Biology & Drug Design*, 2011, **78**, 252-259.
153. G. Rastelli, A. Del Rio, G. Degliesposti and M. Sgobba, *Journal of Computational Chemistry*, 2010, **31**, 797-810.
154. S. Genheden and U. Ryde, *Physical Chemistry Chemical Physics*, 2012, **14**, 8662-8677.

155. T. Y. Yang, J. C. Wu, C. L. Yan, Y. F. Wang, R. Luo, M. B. Gonzales, K. N. Dalby and P. Y. Ren, *Proteins-Structure Function and Bioinformatics*, 2011, **79**, 1940-1951.
156. G. Archontis, T. Simonson and M. Karplus, *Journal of Molecular Biology*, 2001, **306**, 307-327.
157. C. N. Schutz and A. Warshel, *Proteins-Structure Function and Genetics*, 2001, **44**, 400-417.
158. A. Aleksandrov, D. Thompson and T. Simonson, *Journal of Molecular Recognition*, 2010, **23**, 117-127.
159. S. S. Castillo, J. Brognard, P. A. Petukhov, C. Y. Zhang, J. Tsurutani, C. A. Granville, M. Li, M. Jung, K. A. West, J. G. Gills, A. P. Kozikowski and P. A. Dennis, *Cancer Research*, 2004, **64**, 2782-2792.
160. Y. H. Jia, K. K. Subramanian, C. Erneux, V. Pouillon, H. Hattori, H. Y. Jo, J. You, D. Zhu, S. Schurmans and H. B. R. Luo, *Immunity*, 2007, **27**, 453-467.
161. A. Chakraborty, M. A. Koldobskiy, N. T. Bello, M. Maxwell, J. J. Potter, K. R. Juluri, D. Maag, S. Kim, A. S. Huang, M. J. Dailey, M. Saleh, A. M. Snowman, T. H. Moran, E. Mezey and S. H. Snyder, *Cell*, 2010, **143**, 897-910.
162. H. B. R. Luo, Y. E. Huang, J. M. C. Chen, A. Saiardi, M. Iijima, K. Q. Ye, Y. F. Huang, E. Nagata, P. Devreotes and S. H. Snyder, *Cell*, 2003, **114**, 559-572.
163. D. R. Alessi, S. R. James, C. P. Downes, A. B. Holmes, P. R. J. Gaffney, C. B. Reese and P. Cohen, *Current Biology*, 1997, **7**, 261-269.
164. K. E. Landgraf, C. Pilling and J. J. Falke, *Biochemistry*, 2008, **47**, 12260-12269.
165. G. J. Kleywegt, M. R. Harris, J. Y. Zou, T. C. Taylor, A. Wahlby and T. A. Jones, *Acta Crystallographica Section D-Biological Crystallography*, 2004, **60**, 2240-2249.
166. N. Homeyer and H. Gohlke, *Molecular Informatics*, 2012, **31**, 114-122.
167. J. M. Hayes, V. T. Skamnaki, G. Archontis, C. Lamprakis, J. Sarrou, N. Bischler, A. L. Skaltsounis, S. E. Zographos and N. G. Oikonomakos, *Proteins-Structure Function and Bioinformatics*, 2011, **79**, 703-719.
168. C. P. Berrie and M. Falasca, *Faseb Journal*, 2000, **14**, 2618-2622.
169. S. J. Conway, J. Gardiner, S. J. A. Grove, M. K. Johns, Z.-Y. Lim, G. F. Painter, D. E. J. E. Robinson, C. Schieber, J. W. Thuring, L. S. M. Wong, M.-X. Yin, A. W. Burgess, B. Catimel, P. T. Hawkins, N. T. Ktistakis, L. R. Stephens and A. B. Holmes, *Org Biomol Chem*, 2010, **8**, 66-76.
170. V. Calleja, M. Laguerre, P. J. Parker and B. Larijani, *Plos Biology*, 2009, **7**, 189-200.
171. C. W. Lindsley, Z. J. Zhao, W. H. Leister, R. G. Robinson, S. F. Barnett, D. Defeo-Jones, R. E. Jones, G. D. Hartman, J. R. Huff, H. E. Huber and M. E. Duggan, *Bioorganic & Medicinal Chemistry Letters*, 2005, **15**, 761-764.
172. Y. H. Hu, L. X. Qiao, S. M. Wang, S. B. Rong, E. J. Meuillet, M. Berggren, A. Gallegos, G. Powis and A. P. Kozikowski, *Journal of Medicinal Chemistry*, 2000, **43**, 3045-3051.
173. A. P. Kozikowski, H. Y. Sun, J. Brognard and P. A. Dennis, *Journal of the American Chemical Society*, 2003, **125**, 1144-1145.
174. T. Elliott, PhD Thesis, University of St Andrews, 2010.
175. J. Nemeth, PhD Thesis, University of St Andrews, 2008.

176. M. Falasca, D. Chiozzotto, H. Y. Godage, M. Mazzoletti, A. M. Riley, S. Previdi, B. V. L. Potter, M. Broggini and T. Maffucci, *British Journal of Cancer*, 2010, **102**, 104-114.
177. R. M. C. Dawson, D. C. Elliott, W. H. Elliott and K. M. Jones, *Data for Biochemical Research,*, 3rd edn., Oxford University Press, 1986.
178. T. S. Elliott, A. Slowey, Y. L. Ye and S. J. Conway, *Medicinal Chemistry Communications*, 2012, **3**, 735-751.
179. F. G. Bordwell and D. Algrim, *Journal of Organic Chemistry*, 1976, **41**, 2507-2508.
180. C. J. M. Huige and C. Altona, *Journal of Computational Chemistry*, 1995, **16**, 56-79.
181. G. D. Prestwich, *Chemistry & Biology*, 2004, **11**, 619-637.
182. L. Y. Zhang, W. Huang, A. Tanimura, T. Morita, S. Harihar, D. B. DeWald and G. D. Prestwich, *Chemmedchem*, 2007, **2**, 1281-1289.
183. L. H. Mak, S. N. Georgiades, E. Rosivatz, G. F. Whyte, M. Mirabelli, R. Vilar and R. Woscholski, *Acs Chemical Biology*, 2011, **6**, 1382-1390.
184. S. Dowler, G. Kular and D. R. Alessi, *Sci STKE*, 2002, **129**, 1-10.
185. B. C. Miao, I. Skidan, J. S. Yang, A. Lugovskoy, M. Reibarkh, K. Long, T. Brazell, K. A. Durugkar, J. Maki, C. V. Ramana, B. Schaffhausen, G. Wagner, V. Torchilin, J. Y. Yuan and A. Degterev, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 20126-20131.
186. R. J. Leatherbarrow, *GraFit Version 7*, Erithacus Software Ltd., Horley, U.K., 2009.
187. K. Datta, T. F. Franke, T. O. Chan, A. Makris, S. I. Yang, D. R. Kaplan, D. K. Morrison, E. A. Golemis and P. N. Tsichlis, *Molecular and Cellular Biology*, 1995, **15**, 2304-2310.
188. T. O. Chan, S. E. Rittenhouse and P. N. Tsichlis, *Annual Review of Biochemistry*, 1999, **68**, 965-1014.
189. M. A. Lemmon and K. M. Ferguson, *Biochemical Journal*, 2000, **350**, 1-18.
190. D. Auguin, P. Barthe, M. T. Auge-Senegas, M. H. Stern, M. Noguchi and C. Roumestand, *Journal of Biomolecular Nmr*, 2004, **28**, 137-155.
191. M. A. Teitell, *Nature Reviews Cancer*, 2005, **5**, 640-648.
192. M. Frech and B. A. Hemmings, *Methods in molecular biology*, 1998, **88**, 197-210.
193. J. E. Ladbury, *Biotechniques*, 2004, **37**, 885-887.
194. S. C. Cooper, PhD Thesis, Imperial College London, 2010.
195. D. Keleti, *Annual Report: Investigation of Novel Molecular Targets for Pleckstrin Homology (PH) Domains Found in Oncogenes Implicated in Breast Cancer*, University of Pennsylvania, 2006.
196. J. A. Thomson and J. E. Ladbury, *Isothermal titration calorimetry: a tutorial*, John Wiley and Sons, Ltd, 2004.
197. S. Milev, H. R. Bosshard and I. Jelesarov, *Biochemistry*, 2005, **44**, 285-293.
198. N. E. Lewis, A. D. Kight, A. Daily, T. K. S. Kumar, R. L. Henry and R. L. Goforth, *Journal of Biological Chemistry*, 2010, **285**, LE26-LE28.
199. M. R. Duff Jr., J. Grubbs and E. E. Howell, *J. Vis. Exp.*, 2011, **55**, e2796.
200. P. Esser, *Thermo Scientific Nunc Bulletin No. 09*, 2010.
201. A. Frey, B. Meckelein, D. Externest and M. A. Schmidt, *Journal of Immunological Methods*, 2000, **233**, 47-56.

202. D. Mahadevan, G. Powis, E. A. Mash, B. George, V. M. Gokhale, S. X. Zhang, K. Shakalya, L. Du-Cuny, M. Berggren, M. A. Ali, U. Jana, N. Ihle, S. Moses, C. Franklin, S. Narayan, N. Shirahatti and E. J. Meuillet, *Molecular Cancer Therapeutics*, 2008, **7**, 2621-2632.
203. M. X. Wu, B. E. Dul, A. J. Trevisan and D. Fiedler, *Chemical Science*, 2013, **4**, 405-410.
204. D. Komander, A. Fairservice, M. Deak, G. S. Kular, A. R. Prescott, C. P. Downes, S. T. Safrany, D. R. Alessi and D. M. F. van Aalten, *Embo Journal*, 2004, **23**, 3918-3928.
205. A. A. Skjevik, B. D. Madej, R. C. Walker and K. Teigen, *Journal of Physical Chemistry B*, 2012, **116**, 11124-11136.
206. C. J. Dickson, L. Rosso, R. M. Betz, R. C. Walker and I. R. Gould, *Soft Matter*, 2012, **8**, 9617-9627.
207. R. F. Irvine and M. J. Schell, *Nature Reviews Molecular Cell Biology*, 2001, **2**, 327-338.