Rationalisation and design of molecular recognition: computational and experimental approaches.

A thesis submitted to the University of Manchester

for the degree of PhD

in the Faculty of Medical and Human Sciences

2013

Luz del Socorro Flores Michel

School of Pharmacy and Pharmaceutical Sciences

List of content

Word Count: 64,688

List of figures

of PI3Kγ showing the hydrophobic interactions. Note that the view is rotated to highlight the different protein-ligand interactions. …………………………………

Figure 5.5: Comparison between representative structures from the MD trajectories of the PI3Kγ/*s*-cis inhibitor **2** system (simulation **II**, shown in purple) and the PI3Kγ/*s*-trans inhibitor **2** system (simulation **VI**, shown in cyan). (a) Cartoon representation of the enzymes, showing the ligands as sticks. (b) Active site of PI3Kγ showing the key hydrogen bonding residues. The yellow dashed lines indicate the key hydrogen bond interactions. For clarity, only the interactions of the PI3Kγ/*s*-trans inhibitor **2** system are indicated. (c) Active site of PI3Kγ showing the hydrophobic interactions. Note that the view is rotated to highlight the different protein-ligand interactions. …………………………………

Figure 5.6: Bicyclic moieties of *s*-cis inhibitor **1** (left, in green) and *s*-cis inhibitor **2** (right, in purple) and corresponding protein fragments (containing peptide bond Ile881-Val882). The average distance for the *s*-cis inhibitor **1** quinoxaline N···HN(Val882) hydrogen bond is 3.23 Å, whereas the average distance for the *s*-cis inhibitor **2** 1,3-benzodioxole ring O···HN(Val882) hydrogen bond is 3.65 Å. ………………………………………………………………………

Part II: Synthesis of Novel fluorescent Ca^{2+} indicators

Figure 1.10: Schematic representation of the FRET process between the

donor (D) and acceptor (A) molecules. The result is an increase in the emission
intensity of the acceptor with accompanying loss of the emission intensity of
the donor (dashed lines). The overlap integral between the donor emission and
the acceptor absorption spectra is shown in yellow. ……………………………… 182

List of tables

Part II: Synthesis of Novel fluorescent Ca^{2+} indicators

List of schemes

Part II: Synthesis of Novel fluorescent Ca^{2+} indicators

List of abbreviations

ABD	adaptor-binding domain
ACN	acetonitrile
AM	acetoxymethyl ester
AMBER	assisted model building with energy refinement
ATP	adenosine-5'-triphosphate
BAPTA	bis(o-aminophenoxy)ethane-N,N,N',N'-tetraacetic acid
BFP	blue fluorescent protein
Bn	benzyl
Boc$_2$O	di-$tert$-butyldicarbonate
BODIPY	boron-dipyrromethene
^{t}Bu	$tert$-butyl group
Bz	benzoyl
CFP	cyan fluorescent protein
CHARMm	chemistry at Harvard molecular mechanics
COOT	crystallographic object-oriented toolkit
COSY	two dimensional correlated spectroscopy
DCM	dichloromethane
DEPT	distortionless enhancement through polarisation transfer
DIEA	N,N-diisopropylethylamine
DMF	dimethylformamide
DNA	deoxyribonucleic acid
EGTA	ethyleneglycol bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid
ES	electrospray
Et$_3$N	triethylamine
EtOAc	ethyl acetate
EtOH	ethanol
eq	equivalent(s)
FDPB	finite difference Poisson-Boltzmann
FEP	free energy perturbation
FRET	fluorescence resonance energy transfer
FT-NMR	Fourier transform nuclear magnetic resonance spectroscopy
GAFF	general AMBER force field
GB	generalised Born

GBSA	generalised Born surface area
GFP	green fluorescent protein
GPCRs	G-protein coupled receptors
h	hour(s)
HMQC	heteronuclear multiple-quantum coherence experiment
HOMO	highest occupied molecular orbital
HPLC	high performance liquid chromatography
IR	infrared
K_d	dissociation constant
LDA	lithium diisopropylamide
LIE	linear interaction energy
LUMO	lowest unoccupied molecular orbital
MAD	multiple-wavelength anomalous diffraction method
MD	molecular dynamics
MeOH	methanol
min	minute(s)
MIR	multiple isomorphous replacement method
MM	molecular mechanics
MM-PBSA	molecular mechanics-Poisson Boltzmann surface area
m.p.	melting point
MR	molecular replacement method
mTOR	mammalian target of rapamycin
MW	microwave
m/z	mass/charge ratio
NMR	nuclear magnetic resource
NOESY	nuclear overhauser effect spectroscopy
NVE	constant number of atoms, volume and energy
NPT	constant number of atoms, pressure and temperature
NVT	constant number, volume and temperature
PB	Poisson Boltzmann
PBSA	Poisson Boltzmann surface area
PCT	photoinduced tharge transfer
PDP	protein data bank
PES	potential energy surface
PET	photoinduced electron transfer
PI3K	phosphoinositide 3-kinase

PIP_2	phosphatidylinositol 4,5-bisphosphate
PIP_3	phosphatidylinositol 3,4,5-trisphosphate
PTEN	tensin homologue deleted on chromosome 10
QM	quantum mechanics
PMEMD	particle mesh Ewald molecular dynamics
RBD	Ras-binding domain
RESP	restrained electrostatic potential
R_f	retention factor
RMSD	root mean square deviation
RMSF	root means square fluctuations
r.t.	room temperature
SA	surface area
SAD	single-wavelength anomalous diffraction method
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TI	thermodynamic Integration
t.l.c.	thin layer chromatography
UV	ultraviolet radiation
VMD	visual molecular dynamics
wt	wild type
YFP	yellow fluorescent protein

Abstract

Molecular recognition is essential to all biological interactions and processes. Knowledge of the structural basis of recognition offers a powerful mechanism for understanding, predicting and controlling the behaviour of biological systems. In this thesis, we present, firstly a computational and crystallographic analysis of molecular recognition in protein-ligand systems; and secondly, progress towards the synthesis of a fluorescent probe for calcium ion recognition.

Class I phosphoinositide 3-kinases (PI3Ks), in particular PI3Kγ, have long been considered promising drug targets for the treatment of inflammatory and autoimmune disorders. As a step towards improved understanding of PI3K binding preferences, we examine the basis on which PI3Kγ distinguishes γ-selective inhibitors AS-605240 and AS-604850, with a ~30-fold preference for the former. Interestingly, despite the chemical similarity of the two ligands, the X-ray structures for their PI3Kγ complexes exhibit the molecules in different conformers, *s*-cis for AS-604850 and *s*-trans for AS-605240. Here, we re-examine the PI3Kγ/AS-605240 crystallographic data and find that not only is a *s*-cis conformation possible but in fact it has a much higher occupancy (87%) than the originally modelled *s*-trans isomer (13%). Subsequently, to account for the isomeric complexities presented by the ligands, we perform 140 ns MD simulations of the four PI3Kγ complexes in explicit solvent: this reveals similar conformational flexibility at the active site for all systems. Yet, the conformations sampled by the *s*-cis isomers are more consistent with the conformations reported by the X-ray crystal structures. Subsequent energetic analysis was performed incorporating ensemble-averaging and desolvation effects via the Poisson-Boltzmann continuum solvent model. For both inhibitors the *s*-cis isomers are predicted to be the most favourable conformations. Additionally, the results indicate a preference for AS-605240, as observed experimentally. The molecular basis for this preference is discussed, together with a comparison of molecular mechanical and quantum chemical treatments of the key ligand-Val882 interaction. This study provides structural, dynamical and energetic insights into the subtle basis of molecular recognition by PI3Kγ.

Fluorescent probes have evolved into an extremely useful tool for the detection of calcium in biological systems. Benzothiazole derivatives BTC, and its iminocoumarin analogue BTIC, are two low affinity calcium indicators featuring many desirable properties for cellular calcium measurement. In an effort to produce fluorophores that can be chemically conjugated with a screening protein, such as Green Fluorescent Protein (GFP), we aimed to derivatise BTC and BTIC. Two synthetic approaches towards the synthesis of these potential fluorescent probes are outlined.

Declaration

The author declares that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

i The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions", which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is avail- able in the University IP Policy (see http://www.campus.manchester.ac. uk/medialibrary/policies/intellectual-property.pdf), in any rele- vant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/ library/aboutus/regulations) and in The University's policy on presen- tation of Theses.

For my family.

A mi familia, con amor y gratitud.

Acknowledgements

I would like to express my immense gratitude to a number of people, without whom completing this thesis would not have been possible.

I owe this work to the guidance and patience of my supervisor Dr Richard Bryce. Thank you for your kindness and unending support in this long journey. I also owe a great deal of gratitude to Prof Ken Douglas, my second supervisor. Thank you for your encouragement during the course of this project.

I would like to thank Dr Mark Dunstan for his assistance with the crystallographic study; Dr Willem Nissink, AstraZeneca, for his helpful comments and for providing the homology models; Dr Neil Burton, for providing computational resources; and Dr David Berk for his support and advice.

I would also like to thank Dr Charles Laughton and Dr Sally Freeman, my examiners, for their guidance and comments.

Thanks must be given to Dr Laura Etchells for her never-ending support, patience and friendship. Thanks to all the current and former members of the Computational Biophysics and Drug Design and Action group for teaching me so much. Special thanks go to Dr Neil Bruce. Thanks to Dr Sergio Mares Sámano, Dr Zahra Hamrang and Dr Nicholas Rattray, for their support, scientific discussions and advice.

I am grateful for the generous financial support that has been provided by the Mexican Council for Science and Technology.

Finally, I would like to thank my family, muchas, muchísimas gracias por todo. My deepest thanks go to Tarik Kochi whose patient love and understanding enabled me to complete this work.

Introduction to molecular recognition

Molecular recognition is essential to all biological interactions and processes[1, 2]. Every bimolecular reaction, whether it occurs in the gas phase, dilute solution, or an enzyme's interior, begins with a recognition event[3]. Phenomena such as enzyme-substrate or enzyme-inhibitor interactions, the transmission of cellular signals, antigen-antibody interactions, ion binding by a fluorescent probe, the control of transcription and translation, and the fidelity of DNA replication, are all underpinned by molecular recognition[4-9].

Biological recognition is very specific and selective. It requires size, shape and functional group complementarity[10]; and it depends on non-covalent interactions such as hydrogen bonding, van der Waals forces, π-π stacking, cation-π and hydrophobic interactions[11, 12]. Furthermore, biomolecules such as proteins can be highly flexible and this plasticity can have important consequences for molecular recognition[13, 14].

Knowledge of the structural basis of recognition offers a powerful mechanism for understanding, predicting and controlling the behaviour of biological systems. A detailed understanding of which interactions are responsible for affinity, reactivity and selectivity is essential for rational drug design, as many drugs and biological probes act by binding, and usually inhibiting, a specific macromolecular target.

Molecular recognition is the foundation for the two topics of this thesis: analysis of protein-ligand interactions, and fluorescent indicators design. Each topic will be discussed and presented independently in a stand-alone fashion.

The first part of this thesis is devoted to exploring the structural basis of molecular recognition in protein-ligand complexes. Specifically, we examine the basis on which the γ isoform of phosphoinositide 3-kinase (PI3Kγ) distinguishes selective inhibitors AS-605240 and AS-604850, with a ~30-fold preference for the former[15]. To this end, computational and crystallographic techniques are used.

In the second half of this thesis we consider the application of molecular recognition to fluorescent sensor design. Progress towards the synthesis of potential fluorescent indicators for Ca^{2+} detection that can be chemically conjugated to a Green Fluorescent Protein (GFP) mutant is presented.

1. Schneider, H. J., Introduction to Molecular Recognition Models. In *Protein-Ligand Interactions: From Molecular Recognition to Drug Design*, Böhm, H. J.; Schneider, G., Eds. Wiley Germany, 2003.

2. Kuriyan, J.; Konforti, B.; Wemmer, D., Molecular Recognition: The Thermodynamics of Binding. In *The Molecules of Life: Physical and Chemical Principles*, Garland Science: 2012.

3. Rebek, J., Jr., Introduction to the Molecular Recognition and Self-Assembly Special Feature. *Proceedings of the National Academy of Sciences of the United States of America* **2009,** *106*, 10423-10424.

4. Lehn, J. M., Supramolecular chemistry - scope and perspectives: molecules, super molecules and molecular devices *Angewandte Chemie-International Edition* **1988,** *27*, 89-112.

5. Gilson, M. K.; Zhou, H.-X., Calculation of protein-ligand binding affinities. In *Annual Review of Biophysics and Biomolecular Structure*, 2007; Vol. 36, 21-42.

6. Babine, R. E.; Bender, S. L., Molecular recognition of protein-ligand complexes: Applications to drug design. *Chemical Reviews* **1997,** *97*, 1359-1472.

7. Valeur, B.; Leray, I., Design principles of fluorescent molecular sensors for cation recognition. *Coordination Chemistry Reviews* **2000,** *205*, 3-40.

8. Houk, K. N.; Leach, A. G.; Kim, S. P.; Zhang, X. Y., Binding affinities of host-guest, protein-ligand, and protein-transition-state complexes. *Angewandte Chemie-International Edition* **2003,** *42*, 4872-4897.

9. Zaher, H. S.; Green, R., Fidelity at the Molecular Level: Lessons from Protein Synthesis. *Cell* **2009,** *136*, 746-762.

10. Rebek, J., Model studies in molecular recognition *Science* **1987,** *235*, 1478-1484.

11. Hobza, P.; Müller-Dethlefs, K., *Non-covalent Interactions: Theory and Experiment*. The Royal Society of Chemistry: UK, 2009.

12. Harmat, V.; Naray-Szabo, G., Theoretical Aspects of Molecular Recognition. *Croatica Chemica Acta* **2009,** *82*, 277-282.

13. Demchenko, A. P., Recognition between flexible protein molecules: induced and assisted folding. *Journal of Molecular Recognition* **2001,** *14*, 42-61.

14. Brooijmans, N.; Kuntz, I. D., Molecular recognition and docking algorithms. *Annual Review of Biophysics and Biomolecular Structure* **2003,** *32*, 335-373.

15. Camps, M.; Ruckle, T.; Ji, H.; Ardissone, V.; Rintelen, F.; Shaw, J.; Ferrandi, C.; Chabert, C.; Gillieron, C.; Francon, B.; Martin, T.; Gretener, D.; Perrin, D.; Leroy, D.; Vitte, P. A.; Hirsch, E.; Wymann, M. P.; Cirillo, R.; Schwarz, M. K.; Rommel, C., Blockade of PI3K gamma suppresses joint inflammation and damage in mouse models of rheumatoid arthritis. *Nature Medicine* **2005,** *11*, 936-943.

Part I. Computational and crystallographic analysis of molecular recognition by the γ-isoform of phosphoinositide 3-kinase (PI3Kγ)

1. Introduction

The main aim of this chapter is to introduce the phosphoinositide 3-kinase (PI3K) family of enzymes. Firstly, the family classification will be presented. A brief summary of the different isozymes' relevance to human disease will then be provided. Following this, key structural information that provides a rationale for ATP binding will be discussed. Finally, two PI3Kγ-selective inhibitors, AS-605240 and AS-604850, will be introduced, leading to the aims of this research project.

1.1 Phosphoinositide 3-kinases (PI3Ks)

Phosphoinositide 3-kinases (PI3Ks) are a family of enzymes that regulate numerous cellular processes including proliferation, motility, survival and apoptosis[1-3]. PI3Ks can act both as lipid and as protein kinases[4-6].

Protein kinases catalyse the transfer of a phosphate group from a nucleoside triphosphate (mostly ATP or GTP) to a protein substrate. This is called phosphorylation, and usually takes place on the hydroxyl group of serine, threonine, or tyrosine[7], but there are also kinases that act on other amino acids such as histidine, arginine or lysine[8, 9]. Protein phosphorylation can result in enzyme activation, enzyme inhibition, the creation of recognition sites for recruitment of other proteins, and transitions in protein state from order to disorder or disorder to order[10, 11].

Similarly, lipid kinases catalyse the phosphorylation of lipids within cells. These enzymes, and the resulting phosphorylated lipids, serve a variety of roles in receptor-mediated signal transduction, cytoskeletal remodelling, nuclear events and membrane trafficking[12-14].

Based on their structure and substrate specificity, PI3Ks can be classified into three different families (denoted class I, II and III)[15, 16]. The most studied of these and the focus of this research project is class I PI3Ks. Class I PI3Ks primarily catalyse the phosphorylation of phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5)P_2; also referred to as PIP$_2$; Figure 1.1) at the 3'-hydroxyl group, giving rise to the lipid second

messenger phosphatidylinositol 3,4,5-trisphosphate (PtdIns(3,4,5)P_3; also known as PIP$_3$)[4, 17]. PIP$_3$ acts as a plasma membrane docking site for an assorted group of signalling proteins containing PX domains, pleckstrin homology domains (PH domains), FYVE domains and other phosphoinositide-binding domains. While there are a wide variety of PI3K downstream signalling events, most of them are mediated through AKT (also known as Protein Kinase B, PKB) and mTOR (mammalian target of rapamycin)[17, 18]. The phosphatase PTEN (tensin homologue deleted on chromosome 10) acts as a negative regulator of the PI3K signalling by converting PIP$_3$ to PIP$_2$[19].
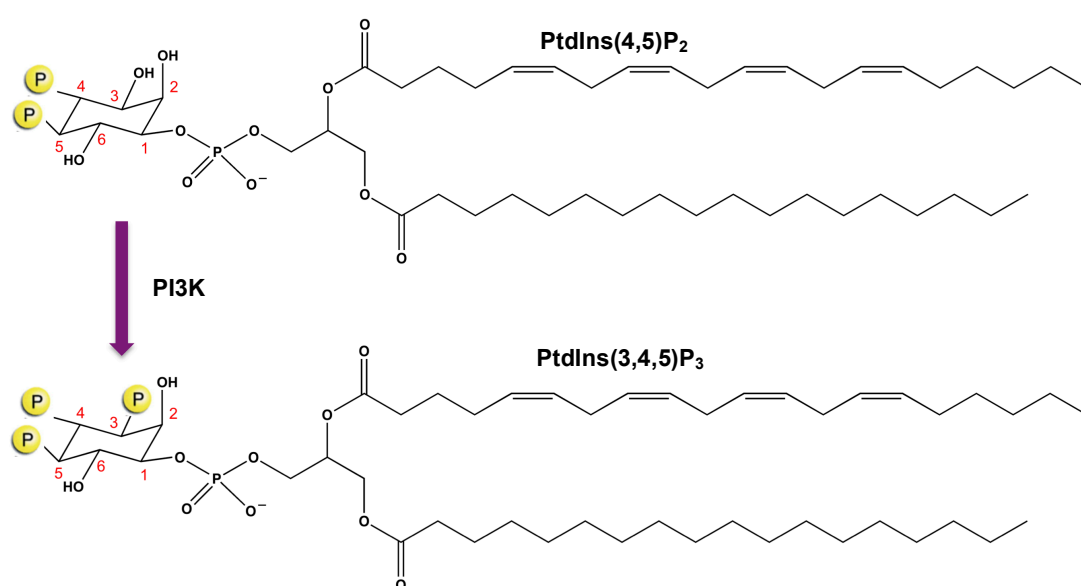
Figure 1.1: Structure of PtdIns(4,5)P_2 and PtdIns(3,4,5)P_3. These lipid second messengers consist of a cytoplasmic inositol head group and a hydrophobic tail formed by two long fatty acid hydrocarbon chains. PI3Ks transfer a phosphate group (represented by a yellow P) from ATP to phosphorylate the 3' position of the inositol ring.

Class I PI3Ks are heterodimeric proteins that contain a larger (110 kDa) catalytic domain and a smaller regulatory subunit, the nature of which determines a further subdivision into class IA and class IB PI3Ks[20, 21]. The catalytic domain can occur in four isoforms differentiated as p110α, p110β, p110γ and p110δ. The class IA catalytic subunits (namely α, β and δ) generally combine with a p85 regulatory subunit, of which five isoforms exist (p85α, p55α, p50α, p85β and p55γ). All p85 isoforms have Src homology 2 (SH2) domains that allow them to dock to phosphorylated tyrosine residues on other proteins[22]. On the other hand, the single class IB isotype (p110γ) associates with p101 or the closely related p84-p87 regulatory subunits[23]. Class IA PI3Ks are activated by receptor tyrosine kinases (RTK), while class IB is activated by

G-protein-coupled receptors (GPCRs)[24]. Additionally, all four isoforms are activated by Ras[25].

Sequence alignment of the four isoforms reveals a high sequence homology, particularly at the active site. PI3Kα and PI3Kβ are ubiquitously expressed, whereas PI3Kγ and PI3Kδ are predominantly found in hematopoietic cells[26]. Although the role of each of the four isoforms is not yet fully understood, studies have linked genetic deregulation (i.e. oncogenic gain-of-function mutations, overexpression) of PI3Kα to many tumour types, including breast, ovarian and colon cancer[27, 28]. PI3Kβ has a role in platelet aggregation and thrombosis[29]; PI3Kγ and PI3Kδ have been associated with chronic inflammation and autoimmune disorders[30-34]. As such, it is possible to appreciate why Class I PI3Ks have been the subject of growing attention over the recent years.

The broad spectrum of biological activities associated with these enzymes has led to an intense search for potent and isotype-specific inhibitors for therapeutic applications. During this process, the X-ray crystal structures of the apo and ATP-cocrystallised holo form of p110γ[35, 36] have been solved. This was followed by the resolution of the apo forms of p110α/p85α[37], p110δ[38] and more recently, p110β/p85β[39]. Additionally, crystallographic structures of both p110γ and p110δ complexed with a myriad of small molecule inhibitors have also been solved. However, the molecular basis of ligand recognition by PI3Ks and the mechanisms underlying isoform selectivity still remain unclear. The catalytic subunit isoform p110γ (henceforth referred to as PI3Kγ) is the focus of this investigation. In the coming sections, to provide a more detailed understanding of the key structure-activity relationships, the emphasis will be on the structural information available.

1.1.1　Structural overview of PI3Kγ

In 1999, Walker *et al.*[35, 36] determined the structure of the catalytic subunit of PI3Kγ for a fragment comprising residues 144-1102 (PDB codes 1E8Y and 1E8X for the human and porcine enzymes, respectively). The protein has a modular structure consisting of four separately identifiable domains. These domains are: the Ras-binding domain (RBD), the C2 domain, the helical domain and the double-lobed catalytic domain (Figure 1.2). The structure, however, did not provide information about the second substrate, PIP$_2$, or about the p101 regulatory subunit.

The missing N-terminal region of PI3Kγ (residues 1-143) constitutes a separate fifth domain called the adaptor-binding domain (ABD). This N-terminus region is crucial for interaction with the p101 regulator subunit[40].

The Ras-binding domain (RBD; residues 220-311) mediates interaction with Ras in a GTP-dependent manner[25, 41]. The RBD comprises a five-stranded mixed β-sheet, flanked by two α-helices. In the absence of Ras, residues 228–230 and 255–267 are disordered. The segments leading from the RBD into the C2 domain and from the C2 domain to the helical domain are not ordered.

The C2 domain (residues 357-522) mainly interacts with the helical domain and exhibits only minor interactions with the C-terminal lobe of the catalytic domain. This domain consists of two four-stranded antiparallel β-sheets that form a β-sandwich. The C2 domain can bind phospholipid in a Ca^{2+}-independent manner and may participate in membrane interactions[35, 42].
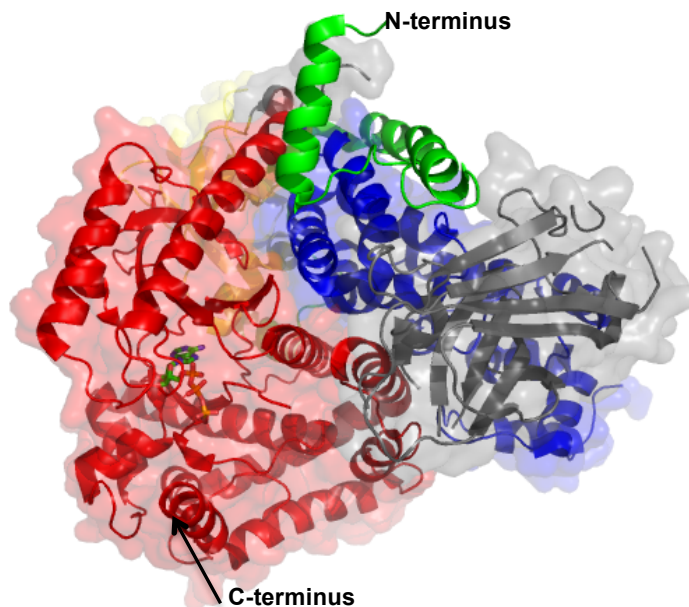
Figure 1.2: Cartoon representation of the structure of PI3Kγ complexed with ATP (using PDB entry 1E8X). The domains are coloured as follows: Ras-binding domain in yellow; C2 domain in grey; helical domain in blue; catalytic domain in red; and the linker regions in green. The ATP is shown as ball-and-stick thereby identifying the location of the active site.

Following the C2 domain is the helical domain (residues 545-725). The helical domain is formed by five A/B pairs of anti-parallel helices. Given the fact that a large portion of the helical domain remains exposed to the solvent, it has been speculated[43] that this surface might be a site of interaction with the regulatory subunits.

Lastly, the catalytic domain (also known as kinase domain) is composed of two lobes, the N-terminal lobe and the C-terminal lobe (Figure 1.3). The N-terminal lobe (residues 726-883) is composed of seven β-sheets, one helical hairpin (kα1-kα2 loop) and three flanking α-helices. On the other hand, the C-terminal lobe (residues 884-1092) primarily consists of α-helices. The ATP binds in the cleft formed by the N- and C-terminal lobes, in a similar fashion to that adopted in protein kinases.

The functionally important regions of the catalytic domain of PI3Kγ are listed in Table 1.1, and will now be briefly described. Firstly, it is interesting to note the extended hairpin between helices kα1 and kα2. This hairpin structurally distinguishes PI3Ks from protein kinases[38]. In contrast to PI3Kα and PI3Kδ, PI3Kγ has two additional residues inserted in this loop region, making it possible for this hairpin to interact with the P-loop below it.
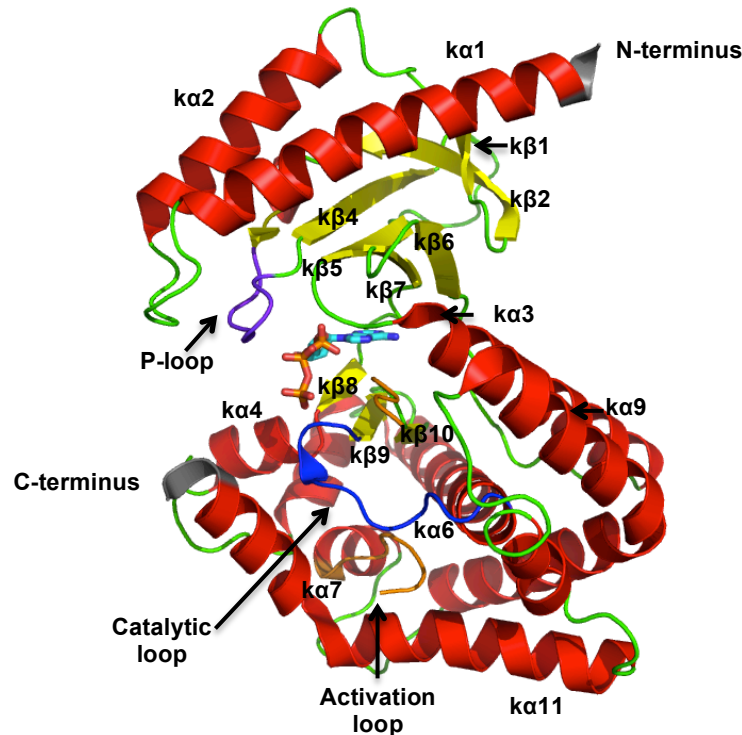
Figure 1.3: Catalytic domain of PI3Kγ complexed with ATP (using PDB entry 1E8X). The structure is coloured by secondary structure with N- and C-termini in grey; the P-loop in purple; the catalytic loop in blue and the activation loop in orange. The ATP is shown as ball-and-stick thereby identifying the location of the active site.

The loop between kβ3 and kβ4 (residues 803-809) corresponds to the phosphate-binding loop (also called P-loop or glycine-rich loop), a motif commonly found in adenine and guanine nucleotide-binding proteins[44]. In contrast to protein kinases, the P-loop in PI3Kγ contains no glycine residues. Instead, contact with the ATP β-

phosphate is preserved through the side chain of Ser806, a residue that is conserved in all PI3Ks[35]. The P-loop has been reported[45] to be a region of high conformational flexibility.

Table 1.1: Definition of functionally important regions of the catalytic domain of PI3Kγ[23].

Region	Residues
P-loop	803-809
Specificity pocket	Met804, Trp812
Helix αC (or kα3)	839-858
Gatekeeper	Ile879
Hinge	Glu880, Val882
Catalytic loop	940-951
Hydrophobic region I	Asp841, Tyr867, Ile879, Ile963
Hydrophobic region II	Ile831, Met953, Phe961
Activation loop	964-989
DFG motif	964-966

The N- and C-terminal lobes are connected through a single polypeptide strand, which lies between kβ7 and kβ8. This loop acts as a hinge, allowing the two domains to rotate with respect to one another upon binding of ATP and/or substrate, without the disruption of the enzyme's secondary structure[46]. The hinge region forms the deepest wall of the ATP-binding pocket and provides a hydrophobic pocket for binding of the adenine moiety of ATP.

A loop directly required for the catalytic mechanism is the catalytic loop, which is located between kα6 and kβ9 (residues 940-951). It has been reported that mutations to residues Asp946, Arg947, Asp950 and Asn951 in the catalytic loop of PI3Kγ lead to elimination of kinase activity[6, 43].

The activation loop (residues 964–988) between kβ10 and kα7 is essential for the substrate specificity of PI3Ks[35, 47]. In the various crystallised forms of PI3Kγ, the activation loop (also known as T-loop) is either missing or only partly observed. This suggests conformational flexibility of the region. Also note that the activation loop is on the surface of the enzyme. It has been argued[48] that the PI3K activation loop can only assume a well-defined structure upon PIP$_2$ binding.

Loop-swapping studies carried out by Pirola *et al.*[48] have demonstrated that two conserved positively charged residues (Lys942 and Arg949, PI3Kα numbering) in the activation loop of PI3Kα are indispensable for phosphatidylinositol 4,5-bisphosphate-specificity[48]. Additionally, the authors suggest that Lys942-p110α interacts with the 5-phosphate and Arg949-p110α with the 4-phosphate group of PtdIns(4,5)P$_2$. Note that these residues correspond to Lys973 and Lys980 in PI3Kγ.

Unfortunately, it has not yet been possible to obtain a crystal structure of PI3K in the presence of phosphoinositides or their head group analogues. Therefore, the mechanisms by which PI3Ks are able to achieve phosphoinositide substrate specificity remain obscure.

The N-terminal end of the activation loop contains the DFG motif (residues 964-966). The presence of this short Asp-Phe-Gly motif is nearly invariant in protein kinases[49]. The DFG aspartate is believed to be crucial for catalysis. In fact, in many protein kinases, a characteristic conformational change (the "DFG flip") connects catalytically active and inactive conformations[11, 50]. In the PI3Kγ/ATP complex, the catalytic aspartate Asp964 is observed to directly coordinate both of the Mg^{2+} ions at the active site (see below).

1.1.2 The ATP-binding site of PI3Kγ

The ATP-binding site of PI3Kγ is located in a cleft between the N- and C-terminal lobes of the catalytic domain. This region is amongst the most conserved regions of PI3Ks[51]. Conceptually, the ATP-binding site can be divided into five different regions. These are: the hinge region, the adenine region, the hydrophobic regions I and II, the ribose region and the phosphate region (Figure 1.4)[23].

The adenine moiety of ATP is placed near the loop between kβ7 and kβ8, as discussed above. In this hinge region, the N1 of ATP acts as a hydrogen bond acceptor with the backbone amide nitrogen of Val882. Moreover, the NH$_2$ of ATP forms a hydrogen bond interaction with the backbone amide carbonyl of Glu880 (Figure 1.5). In addition to these two polar interactions, the adenine moiety also makes extensive non-polar interactions with a set of conserved hydrophobic residues located in both the hydrophobic regions I and II (Table 1.1).
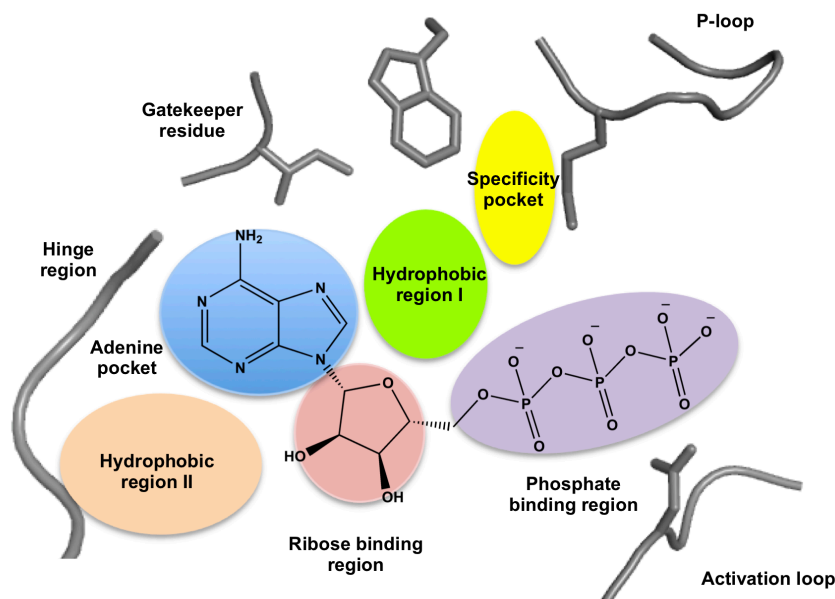
Figure 1.4: Schematic representation of the ATP-binding pocket of PI3K showing its different regions.

In the phosphate binding region, Lys833 (a residue conserved in all PI3Ks[36] and located at the end of kβ5) interacts with the α-phosphate of ATP. Furthermore, contact with the ATP β-phosphate is made through the side chain of Ser806 (part of the P-loop).

The 'gatekeeper' residue is a residue structurally conserved between PI3Ks and protein kinases[52]. This residue, Ile879, forms part of the wall of the small hydrophobic region I, and partially restricts access to this region. It has been documented[52] that mutations to the gatekeeper residue of PI3Kα (from Ile to Gly or Ala) lead to severely impaired enzymatic activity.
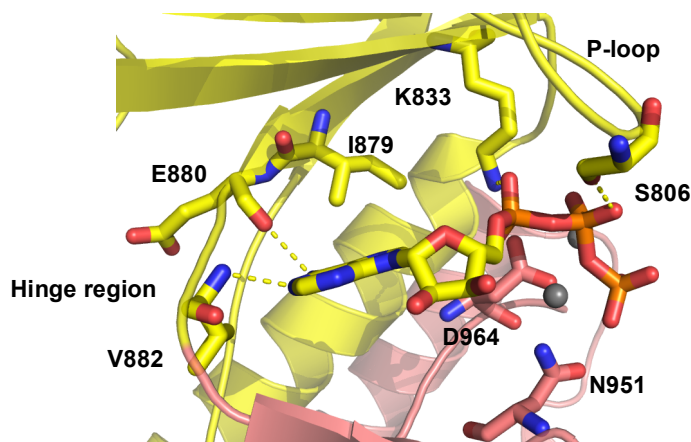
Figure 1.5: Active site of PI3Kγ complexed with ATP (using PDB entry 1E8X). The N-lobe is coloured yellow, while the C-lobe is coloured pink. The metal ions are shown as grey spheres. The yellow dashed lines indicate the key hydrogen bond interactions.

The trisphosphate group is coordinated by two metal ions that are ligated by Asp964 (located on the DFG motif, at the beginning of the activation loop) and Asn951 in the catalytic loop (Figure 1.5). An interesting observation is that the side chain of Asp964 exists in different conformations in the structures of the apo and ATP-cocrystallised holo form of PI3Kγ (Figure 1.6).
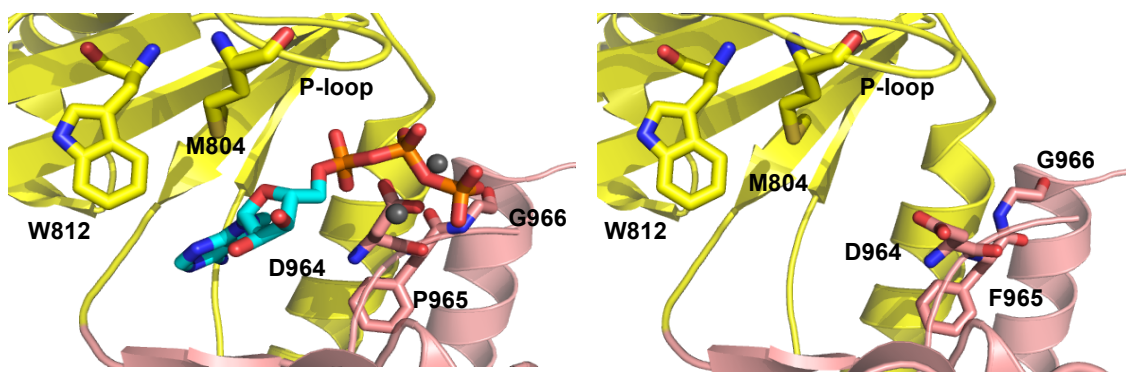
Figure 1.6: View of the active site of PI3Kγ ATP-binding pocket with (left) and without (right) ATP (using PDB entries 1E8X and 1E8Y, respectively). Asp964 swings out to accommodate the trisphosphate moiety of ATP. Residues M804 and W812 from the specificity pocket are also shown.

1.1.3   Selective PI3Kγ inhibitors AS-605240 and AS-604850

Turning now to kinase inhibition, three different types of kinase inhibitors (type I, II and III) have been documented[53]. Type I inhibitors bind in and around the region occupied by the adenine moiety of ATP (adenine region, Figure 1.4), essentially mimicking the ATP-bound conformation. This type of inhibitor also forms hydrogen bonds with the hinge region of the enzyme, and does not require a specific conformation of the DFG motif. Type I inhibitors can bind to kinases in both the active (i.e. phosphorylated) and inactive (i.e. not phosphorylated) conformations[54]. On the other hand, type II inhibitors preferentially bind to the inactive conformation of the enzyme[54], which assumes a DFG-out conformation. In the DFG-out form, the DFG aspartate and phenylalanine side chains swap positions, leaving the aspartate pointing away from the binding site[55]. The flip of the DFG motif involves an approximately 10 Å change in the position of the two side chains[50]. Most importantly, this conformational change creates a hydrophobic pocket (known as allosteric site) available for occupation by type II inhibitors. Lastly, type III inhibitors are non-ATP competitive kinase inhibitors that block kinase activity without displacing ATP[53].

All PI3K inhibitors that have been reported are ATP-competitive type I inhibitors, which occupy the adenine region of the ATP-binding site. To present a detailed

account of all Class I PI3K inhibitors reported to date is beyond the scope of this work (for excellent reviews, see refs[15, 23, 56, 57]). Suffice to say, since the first known inhibitors of PI3Ks, the fungal metabolite wortmannin[58] and the synthetic compound LY294002[59], a multitude of chemically diverse inhibitors have been developed. This wide variety of compounds includes, amongst others, staurosporine, quinazolinone, morpholino, benzoxazine and thiazolidinedione derivatives.

Thiazolidinedione derivatives AS-605240 and AS-604850 (Figure 1.7) are selective, ATP-competitive inhibitors of PI3Kγ developed by Camps *et al.* (Merck-Serono)[30]. These two orally active small molecule inhibitors were developed for the treatment of rheumatoid arthritis and other chronic inflammatory disorders. As seen in Table 1.2, PI3Kγ shows a ~30-fold preference for AS-605240 (henceforth referred to as inhibitor **1**) over AS-604850 (henceforth referred to as inhibitor **2**) (Table 1.2). It is still not clear as to what exactly leads to this particular selectivity profile, what molecular interactions at the binding site confer it and the key residues involved.

AS-605240
s-trans inhibitor 1

AS-604850
s-cis inhibitor 2
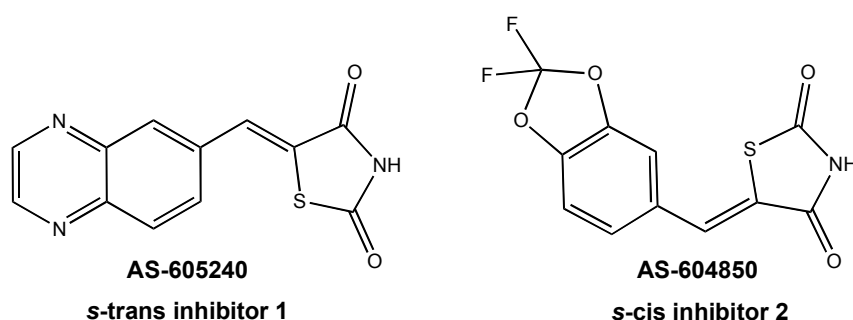
Figure 1.7: Chemical structure of the PI3Kγ-selective *s*-trans AS-605240 (i.e. inhibitor **1**) and *s*-cis AS-604850 (i.e. inhibitor **2**).

Table 1.2: Selectivity profiles (IC$_{50}$ values in μM) for Class I PI3K isoforms[30].

Inhibitor	PI3Kα	PI3Kβ	PI3Kγ	PI3Kδ
1	0.06	0.27	0.008	0.3
2	4.5	>20	0.25	>20

Interestingly, despite the chemical similarity of the two ligands, the publicly available X-ray structures for their PI3Kγ complexes[30] exhibit the molecules in different conformers, *s*-trans for inhibitor **1** and *s*-cis for inhibitor **2** (PDB accession codes 2a5u and 2a4z, respectively; Figure 1.8).
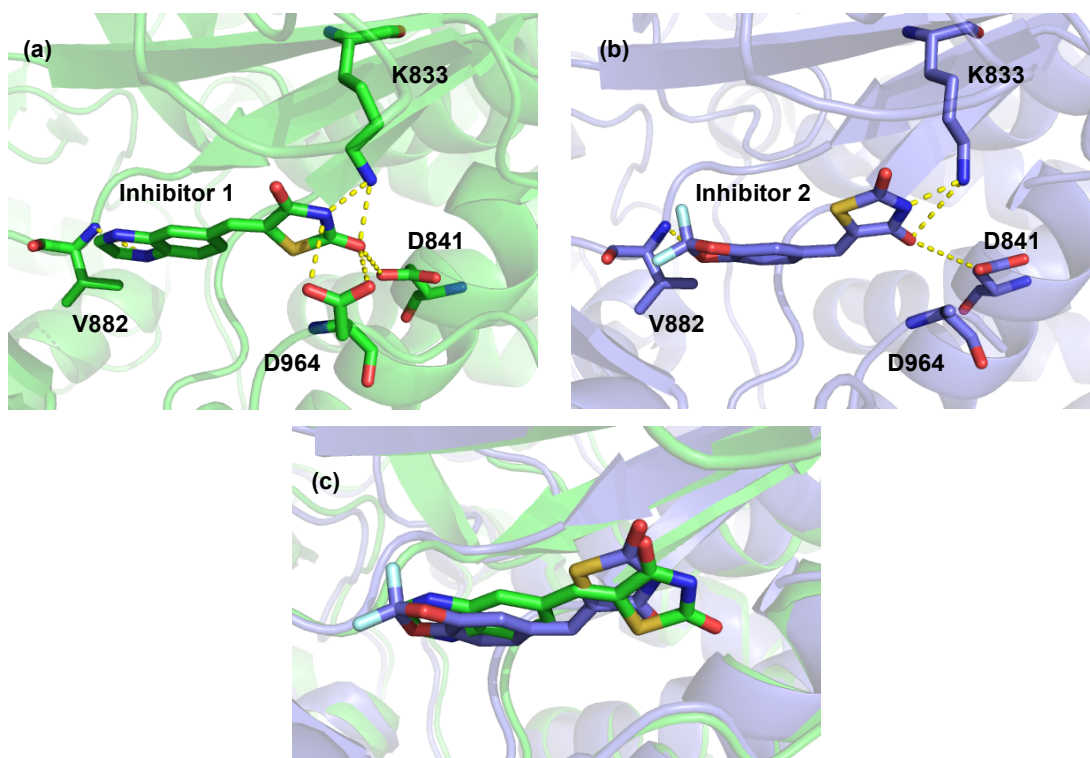
Figure 1.8: Stereodiagram showing some of the key residues of the ATP-binding pocket of PI3Kγ complexed with *s*-trans inhibitor **1** (green; panel a) and *s*-cis inhibitor **2** (purple; panel b). These compounds are tethered by making hydrogen-bond interactions with Val882 and Lys833. (c) Co-crystallised structures of *s*-trans inhibitor **1** (green) and *s*-cis inhibitor **2** (purple) showing dual binding modes.

The X-ray crystal structures of PI3Kγ/*s*-trans inhibitor **1** and PI3Kγ/*s*-cis inhibitor **2** show the thiazolidinedione moiety forming a hydrogen bond with the positively charged side chain of Lys833 and the negatively charged side chains of Asp964 and Asp841 (Figure 1.8). In both cases, the bicyclic moiety, the distinguishing aspect of the two ligands' chemical structure, forges a hydrogen bond interaction with PI3Kγ at the backbone NH of Val882. Furthermore, the quinoxaline moiety of inhibitor **1**, as well as the 1,3-benzodioxole ring of inhibitor **2**, produce favourable interactions with the hydrophobic pocket of PI3Kγ formed by residues Ile831, Ile879, Ile881, Met953 and Phe961. PI3Kγ-selective inhibitors **1** and **2**, together with their PI3Kγ co-crystallised structures, are of particular importance for the present project.

1.2 Aims of this work

Class I PI3Ks, in particular PI3Kγ, have long been considered promising drug targets for the treatment of inflammatory and autoimmune disorders. As a step towards improved understanding of PI3K binding preferences, the aim of this research project is to examine the basis on which PI3Kγ distinguishes γ-selective inhibitors **1** and **2**.

Interestingly, despite the chemical similarity of the two ligands, the X-ray structures for their PI3Kγ complexes exhibit the molecules in different conformers, *s*-trans for inhibitor **1** and *s*-cis for inhibitor **2**. Given the uncertainty regarding the conformation of the ligands, the first objective is to determine the conformational preference of inhibitors **1** and **2**. In order to achieve this, the intrinsic conformational preference of the isolated ligands will be computationally investigated; subsequently, the available crystallographic data will be revisited to determine which isomer is predominantly observed in the protein active site. To provide a dynamical insight into the subtle basis of molecular recognition by PI3Kγ and to identify the key differentiating motifs, fully solvated molecular dynamics simulations, as well as free energy calculations via the Poisson-Boltzmann continuum solvent model, will be employed.

2. Theory

2.1 Protein X-ray crystallography

Three-dimensional structures of biological macromolecules, such as proteins and nucleic acids, are central to our understanding of the processes of life. They allow us to look into biological processes at the most basic level: which molecules interact, how they interact, how enzymes catalyse reactions, and how drugs act. The most common experimental means of obtaining a detailed three-dimensional model of macromolecular complexes is by X-ray crystallography.

The following section provides a simplified overview of the structure determination process by X-ray crystallography, as well as introducing some of the key concepts for structure validation and refinement.

2.1.1   How is a crystal structure determined?

The process of seeing an object involves light rays being reflected by the object and entering the eye through the lens, which in turn reconstructs an image of the object and focuses it on the retina. The same procedure is applied when observing smaller objects under a microscope: electromagnetic radiation (visible light) strikes the object and is diffracted in various directions; the lenses (i.e. the detector) of the microscope then collect the diffracted rays and reassemble them into an image.

The process of protein structure determination by X-ray crystallography follows the same principle (Figure 2.1), but needs additional considerations. Firstly, in all forms of microscopy, the amount of detail (i.e. the resolution) is limited by the wavelength of the electromagnetic radiation used. As the bond length between atoms within a protein is about 0.15 nm or 1.5 Å, to visualise the geometry and structure of protein molecules it is necessary to work with electromagnetic radiation with a wavelength range of 0.1-10 nm or 1-100 Å. Secondly, X-ray scattering from a single protein molecule would be very weak and extremely difficult to detect above noise levels, including scattering from air and water. A way to circumvent this is by using a protein crystal, which consists of a large number of repeating patterns (known as the unit cell) arranged in the same orientation. Each molecule within the crystal therefore diffracts equally, and thus more strongly diffracted X-ray beams can be measured (Figure 2.1.A).
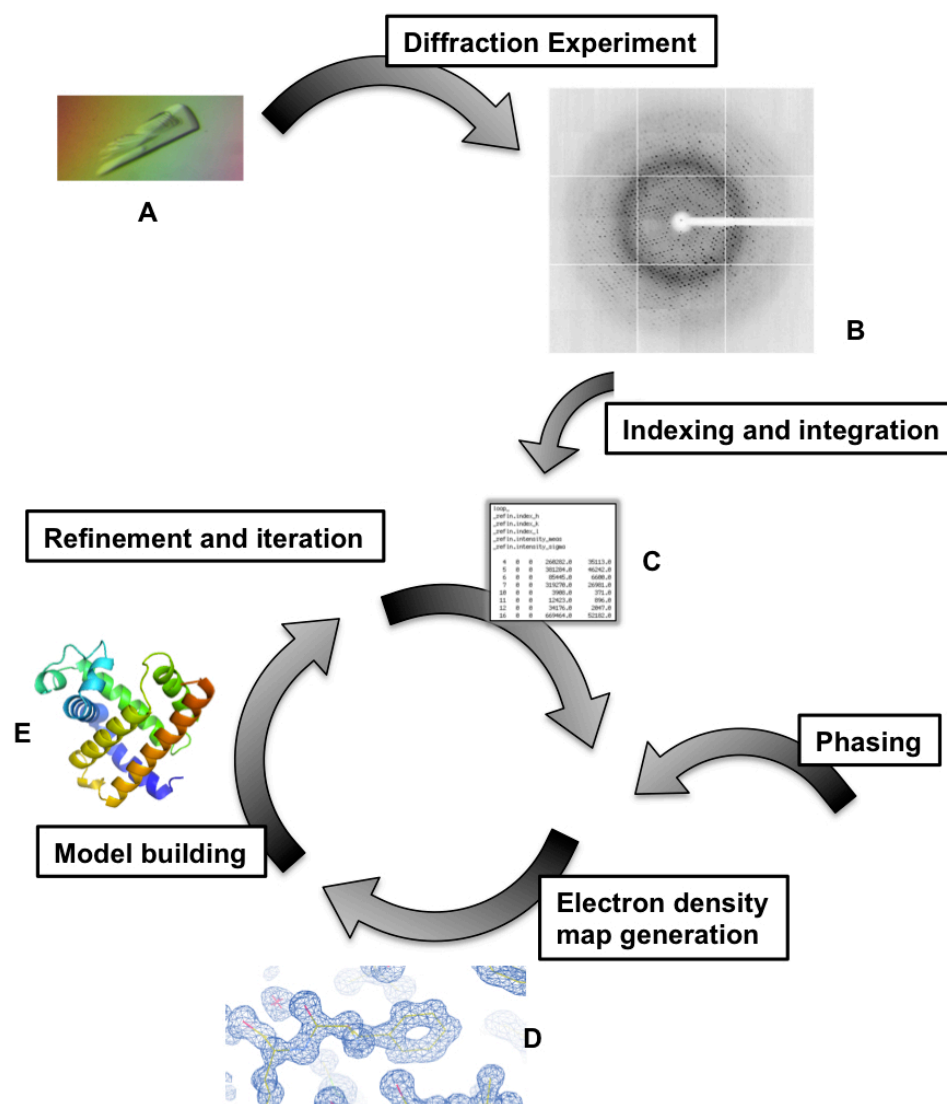
Figure 2.1: Structure determination by X-ray crystallography. A protein crystal (A) is stroked with the X-ray beam. The X-ray beam is diffracted in to smaller X-rays of varying intensity (B) that are recorded by a sensor. The list of reflections is then indexed by their position or Miller indices (*h k l*), intensity estimates and the standard error of those intensities (C). By mathematical means, this data together with phases (which must be derived by other means, see text for details) is converted to an electron density map (D), which is a 3-dimensional representation of the content of the protein crystal. An atomic structure model (E) is built inside the electron density map. The initial model is usually suboptimal hence it needs to be refined. The cycle of electron density map generation, model building, and refinement is iterated until the model is complete and the validation process is satisfactory. The diffraction pattern B was obtained from: http://www.scripps.edu/mb/stout/stout_research.html. The structure factors C, electron density map D, and model E were generated with the graphic programs COOT and PyMOL using PDB entry 1a6m.

When a protein crystal is placed in an intense beam of X-rays of a single wavelength (monochromatic X-rays), the atoms (or more specifically, the clouds of electrons) scatter X-ray waves. Besides, the X-rays striking the electrons produce secondary spherical waves emanating from the electrons. The result is a regular array of

spherical waves, which can either cancel one another out through destructive interference, or they can undergo constructive interference in accordance to Bragg's law.

Bragg's law[60] describes the general relationship between the wavelength of the incident X-rays, angle of incidence and spacing between the crystal lattice planes of atoms. This is expressed as:

$$2dsin\theta = n\lambda \tag{2.1}$$

where d is the spacing between diffracting planes, θ is the incident angle, n is any integer, and λ is the wavelength of the beam. In order for the waves to interfere constructively, the differences in the travel path must be equal to integer multiples of the wavelength (Figure 2.2). When this constructive interference occurs, a diffracted beam of X-rays will leave the crystal at an angle equal to that of the incident beam[61].
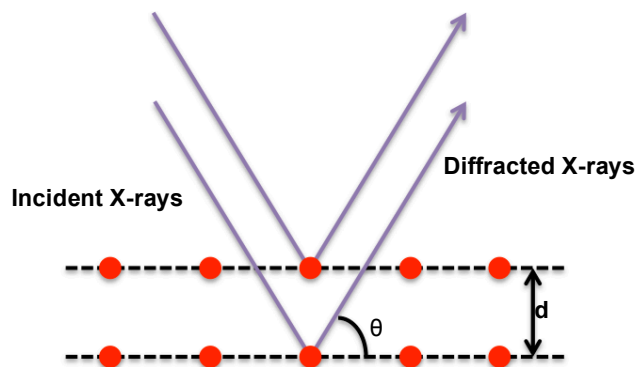
Figure 2.2: Bragg's law. Two beams with identical wavelength and phase approach a crystalline solid and are diffracted by two different atoms within it. The lower beam travels an extra length of $2dsin\theta$. Constructive interference occurs when this length is equal to an integer multiple of the wavelength of the radiation.

As detectors are flat two-dimensional devices, it is only possible to measure the spots in two dimensions. This means the crystal has to be rotated step by step in order to cover all unique orientations of the crystal with respect to the X-ray beam, collecting in this way all the reflections.

Subsequently, the reflections have to be correlated to the orientation of the crystal. This process is called indexing and integration[62]. By identifying the Miller indices *(h k l)* of the crystal plane that give raise to each diffracted beam, the direction of the beam can be specified. From the measured intensity of the reflection, i.e. the intensity of the spots in Figure 2.1.C, its amplitude can be deduced. However, an extra piece of information is required to fully characterise each reflection and that is the phase of the

wave of each diffracted beam. Unfortunately this information is lost during the diffraction experiment (this is known as the phase problem). Consequently the phases need to be calculated in an indirect way. Several methods have been developed in order to derive the phases: the multiple isomorphous replacement method[63,64,65] (MIR), the molecular replacement method[66,67] (MR), or the single[68] or multiple-wavelength anomalous diffraction method[69] (SAD and MAD respectively). Describing the methods of finding the initial phases is beyond the scope of this text, excellent reviews on the subject can be found elsewhere[70,71].

When an X-ray beam strikes a crystal, the origin of the diffraction of the X-rays are the clouds of electrons in the protein molecules of the crystal. Diffraction should therefore reveal the distribution of electrons, namely the electron density, of the molecules within the crystal. Having obtained an initial approximated value for the phases makes it possible to calculate the structure factors **F**(*h k l*), which express both the amplitude and the phase of each reflection. The set of structure factors for all the reflections *hkl* are the primary quantities necessary for the derivation of the three-dimensional distribution of electron density (Figure 2.1.D), which is the image of the crystal structure calculated by Fourier methods.

The fundamental equation for its calculation is:

$$\rho(x\,y\,z) = \frac{1}{V} \sum_h \sum_k \sum_l |F\,(h\,k\,l)| \exp\left[-2\pi i(hx + ky + lz) + i\alpha\,(h\,k\,l)\right] \qquad (2.2)$$

In this equation, the electron density ρ (*x y z*), is expressed as a Fourier transformation of the structure factors **F**(*h k l*). The amplitude of these structure factors, |**F**(*h k l*)|, is obtained from the intensity of the diffracted beam. The phase angles α (*h k l*) cannot be derived in a straightforward manner, but as mentioned earlier, they can be estimated from the current model in an indirect way.

The next step is to interpret the electron density map in terms of the polypeptide chain of the protein, or in other words, to fit the molecular model into the map (Figure 2.1E). However, given the fact that the initial phases are only estimates, the first map is often hard to interpret and may have a considerable number of errors. Nonetheless it serves as a starting point for the iterative process of refinement and model building, which will be discussed on the next section.

2.1.2 The refinement process

The aim of the refinement process is to improve the electron density map as well as the atomic model to achieve the best agreement between the observed reflection intensities and those calculated from the model. To achieve this, the refinement process takes advantage of the following principle: any features that can be reliably determined or inferred from the current map become part of a model for subsequent maps calculation. In this way, based on the atomic coordinates of the already positioned protein, it is possible to calculate how the electron density should look like, using again equation 2.1. The improved electron density maps may then be able to reveal further details, which are subsequently added to the atomic model. Moreover, using the new electron density map, better phases can be calculated by doing a reverse Fourier transform. Using the new phases and atomic coordinates the process is repeated.

It should be noted that the calculated phases usually contain more structural information than the measured amplitudes, thus when recalculating the electron density, it is the phases that dominate the Fourier summation[72]. Consequently, even if the correct amplitudes are known, inaccurate model-based phases generate biased electron density maps. To compensate for the increased impact of the model phases and to ensure the continued influence of the measured amplitudes for improvement of the model, a set of different electron density maps is calculated.

The first electron density map is called a $\mathbf{F}_o$ map because it is calculated from the experimentally observed reflection amplitudes ($\mathbf{F}_o$) and their calculated phases. Another type of electron density map used in the structure determination process is the difference map, which is used to emphasise the errors in the current model. The difference map (F_o - F_c) is calculated with the differences between the observed (F_o) and calculated amplitudes (F_c) using the following equation:

$$\rho(x\,y\,z) = \frac{1}{V} \sum_h \sum_k \sum_l (|F_o| - |F_c|)\exp\left[-2\pi i(hx + ky + lz) + i\alpha\,(h\,k\,l)\right] \quad (2.3)$$

where the Fourier terms can be positive or negative, depending on whether F_o or F_c is larger. As a result, the difference map contains areas of both positive and negative density, which are normally displayed as isosurfaces. The regions of positive density indicate that the contribution of the observed intensities (F_o) to the electron density (ρ) are larger than the contribution of the model (F_c), and consequently that the unit cell (represented by F_o) contains more electron density in this region than implied by the model (represented by F_c). In other words, the model needs to be modified by adding

atoms to this region thus increasing the electron density. Conversely, a region of negative density implies that there is more electron density in the region than the unit cell actually contains. In other words, the electron density does not support the current model and atoms need to be moved away from this region. The difference map is mostly used toward the end of the structure determination process to detect subtle errors after most of the serious errors have been corrected.

Another very insightful map is the $2F_o$ - F_c map, which can be thought as the superposition of the difference map and the F_o map. The $2F_o$ - F_c map is calculated as follows:

$$\rho(x\,y\,z) = \frac{1}{V} \sum_h \sum_k \sum_l (2|F_o| - |F_c|)\exp\left[-2\pi i(hx + ky + lz) + i\alpha\,(h\,k\,l)\right] \quad (2.4)$$

Unless the model contains extremely serious errors, this map displays positive electron density usually resembling a molecular surface. Nevertheless, the $2F_o$ - F_c map (unlike the difference map), still contains undue influence by the current model. As such, it should be treated with caution.

Another map used to increase the quality of the model is the OMIT map. To construct an OMIT map, questionable fragments of the model are eliminated and subsequently new structure factors are calculated[73]. In this way, the possible errors in the suspicious region will not contribute to the phases, thus removing "phase bias" density in the resulting map.

The model can also be refined by manual modification of the atomic coordinates to improve the agreement with the electron density maps while remaining chemically reasonable. For this purpose, geometric and other restraints (bond lengths and bond angles) and constraints are enforced onto the model. To be a conformationally reasonable model, the peptide bonds should be nearly planar, the backbone conformational angles φ and ψ should only fall into allowed ranges and the torsional angles at single bonds in side chains should lie within a few degrees of stable staggered conformations[61]. Additionally, to remain stereochemically reasonable, the model must not have inverted centres of chirality (i.e. no D-amino acids).

Usually, several iterations of density modification and model building are required, and the structural parameters defined above continually improve during the process. Iterative rounds of automated optimisation (e.g. according to least-squares or maximum-likelihood algorithms) are alternated with manual corrections until the validation process is satisfactory.

2.1.3 Validation and structural analysis

The validation process is key to assessing the quality and reliability of protein structures, as well as to identify potential errors within them. In this section, the most commonly used indicators that are utilised to evaluate structural quality are discussed in a simplified manner.

A key criterion when evaluating the accuracy of an experimental protein structure is the resolution of the X-ray data. The resolution at which a structure is determined is a measure of the resolvability of the electron density map and thus dictates the quality that can be achieved. The resolution of the data is expressed in Å, and lower values indicate higher resolution. At higher resolutions, a greater level of detail is obtained, because there is more experimental data available, and hence greater accuracy of the final model. For example, a resolution lower than 1.5 Å implies that at least 95% of the atomic model is a consequence of the experimentally observed data[74], whereas a resolution larger than 2.5 Å means the model is more subjective.

The agreement between the experimental diffraction data and the fitted model is evaluated by the reliability index or crystallographic R-factor, which is defined as follows:

$$R = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|} \tag{2.5}$$

where F_o are the experimental structure factor amplitudes, F_c are the structure factor amplitudes calculated from the model, and the summation extend over all observed reflections. The value of the R-factor lies between zero (perfect agreement between the observed and calculated amplitudes) and 100, its maximum value. It has been suggested that for a reliable model, the R-factor value should not be larger than 25[75]. However, even an inaccurate model can produce a reasonably low R-factor if the data is over-fitted.

This problem can be circumvented by using a better method of cross-validation called free R-factor[76] (R_{free}). The definition of the free R-factor is identical to that of the conventional R-factor defined above, except that the R_{free} is calculated only for a subset of reflections that are never used in the refinement of the model. As such, the R_{free} reliably measures how well the model predicts the experimental observations that are not used to fit the model. It has been suggested[75] that models with R_{free} values larger than 40 should be treated with a high degree of reservation.

When looking at the position of a specific atom or residue, a useful indicator of their reliability is their temperature factor, also known as atomic displacement parameter, or B-factor. The B-factor can be thought as a measure of how much the atom oscillates around the position specified in the model, the oscillations being a consequence of the static and dynamic disorder in the crystal. By using B-factors it is possible to assess which atoms in the model have the most freedom of movement. B-factors are usually expressed in Å^2 and range from ~2 to ~100 Å^2, with the highest values found on the most flexible regions.

A type of validation, which is completely model based, is the Ramachandran plot[77], which displays the backbone conformational angles φ and ψ that fall into allowed and disallowed ranges. These parameters are good indicators of model quality because unlike bond lengths and bond angles, conformational angles are typically not restrained during the refinement process.

In summary, structure validation helps to select the optimal structural model while at the same time estimating its reliability. It should be noted, however, that the uncertainties introduced during the structure determination process are not always appreciated. As previously emphasised, the primary result of an X-ray crystallography experiment is a map of electron density within the crystal. However, the derived atomic model is only an interpretation of the electron density and each step on its determination can introduce ambiguities. This is of course not to say that a structure determined by X-ray crystallography is unreliable, but just to highlight that, as with any other experimental result, it should be treated with caution.

2.2 Molecular modelling

Molecular modelling is concerned with ways to mimic the behaviour of molecules and molecular systems. This may include treating atoms as the smallest individual unit (the molecular mechanics approach), or explicitly modelling electrons of each atom (the quantum chemistry approach). The following section introduces some of elementary concepts of quantum mechanics, and then provides an overview of the theoretical basis for molecular dynamics simulations.

2.2.1 Quantum mechanics background

The electronic structure and properties of any molecule could, in principle, be determined by solving the Schrödinger's (time-independent) equation[78]:

$$\widehat{H}\Psi = E\Psi \qquad (2.6)$$

where $\widehat{H}$ is the Hamiltonian operator for a system of nuclei and electrons, Ψ is the molecular wavefunction, and E is the total molecular energy. For a time-independent multi-electron system the molecular Hamiltonian[79, 80] can be represented as:

$$\widehat{H} = - \sum_{i=1}^{electrons} \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_{\alpha=1}^{nuclei} \frac{\hbar^2}{2m_\alpha}\nabla_\alpha^2 - \sum_{\alpha=1}^{nuclei}\sum_{i=1}^{electrons} \frac{Z_\alpha e^2}{4\pi\varepsilon_0 r_{\alpha i}}$$

$$+ \sum_{\alpha=1}^{nuclei}\sum_{\beta>\alpha}^{nuclei} \frac{Z_\alpha Z_\beta e^2}{4\pi\varepsilon_0 r_{\alpha\beta}} + \sum_{i}^{electrons}\sum_{j>i}^{electrons} \frac{e^2}{4\pi\varepsilon_0 r_{ij}} \qquad (2.7)$$

where $\hbar = h/2\pi$ with h as Planck's constant; m_α is the mass of a nucleus whose charge is Z and m_e is the mass of an electron of charge e. The distances separating pairs of nuclei, pairs of electrons, and electrons from nuclei are $r_{\alpha\beta}$, r_{ij} and $r_{\alpha i}$ respectively. The first two terms in the above equation are summations of the Laplace operator ∇^2 over the entire molecule. Here, this differential operator is defined as:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \qquad (2.8)$$

The first term in equation 2.7 describes the kinetic energy of the electrons (T_{el}) while the second term describes the kinetic energy of the nuclei (T_n). The last three terms correspond to the potential energies arising from Coulombic attraction and repulsion between nuclei and electrons (V_{n-el}), nuclei and nuclei (V_{n-n}), and electrons and electrons (V_{el-el}) respectively. Consequently, the molecular Hamiltonian operator can be understood as the summation:

$$\widehat{H}_{tot} = T_{el}(\boldsymbol{r}) + T_n(\boldsymbol{R}) + V_{n-el}(\boldsymbol{R},\boldsymbol{r}) + V_{n-n}(\boldsymbol{R}) + V_{el-el}(\boldsymbol{r}) \qquad (2.9)$$

where $\boldsymbol{R}$ is the set of nuclear coordinates and $\boldsymbol{r}$ is the set of electronic coordinates.

The Schrödinger equation cannot be solved exactly for systems of multiple nuclei and atoms. Therefore, a number of approximations must be made. A fundamental approximation in quantum chemistry is the Born-Oppenheimer approximation, which leads to the concept of potential energy surfaces.

In the Born-Oppenheimer approximation[81], the fast degrees of freedom are separated from the slow ones. Due to the high mass of nuclei compared to that of the electrons (the resting mass of the lightest nucleus, the proton, is 1836 times heavier than the resting mass of the electron[82]), the former move much slower than the latter. This leads to the approximation that the electrons can adjust almost instantaneously to the position of the nuclei. As the motion of the electrons is decoupled from the motion of the nuclei, the second term (kinetic energy of the nuclei, T_n) in the molecular Hamiltonian operator (equation 2.9) can be neglected. Also, the nuclear repulsion (V_n) term can be considered as a constant, which can be added to the total energy at the end of the calculation. The remaining terms are usually referred to as the electronic Hamiltonian:

$$\widehat{H}_{el} = -\frac{\hbar^2}{2m_e} \sum_{i=1}^{electrons} \nabla_i^2 - \sum_{\alpha=1}^{nuclei} \sum_{i=1}^{electrons} \frac{Z_\alpha e^2}{4\pi\varepsilon_0 r_{\alpha i}} + \sum_{i}^{electrons} \sum_{j>i}^{electrons} \frac{e^2}{4\pi\varepsilon_0 r_{ij}} \quad (2.10)$$

Under the Born-Oppenheimer approximation it is possible to define a purely electronic Schrödinger equation:

$$\widehat{H}_{el}\Psi_{el} = E_{el}\Psi_{el} \quad (2.11)$$

where the electronic wavefunction Ψ_{el} depends explicitly on the electronic coordinates, but only parametrically on the nuclear coordinates (i.e. it only depends on the position of the nuclei, but not on their momentum):

$$\Psi_{el} = \Psi_{el}(\{\boldsymbol{R}\}, \{\boldsymbol{r}\}) \quad (2.12)$$

In this way, the total energy for the fixed nuclei equals the sum of the electronic energy (E_{el}), and the inter-nuclear repulsion energy:

$$E_{tot} = E_{el} + \sum_{\alpha=1}^{nuclei} \sum_{\beta>\alpha}^{nuclei} \frac{Z_\alpha Z_\beta e^2}{4\pi\varepsilon_0 r_{\alpha\beta}} \quad (2.13)$$

Plotting the total energy of the electrons (E_{tot}) as a function of the positions of the nuclei, $\boldsymbol{R}$, gives a potential energy curve, called potential energy surface (PES)[83]. The PES is essentially a plot of molecular energy versus molecular geometry. The motion of the nuclei on the PES can then be solved either classically (Newton) or by quantum (Schrödinger) methods.

For a non-linear molecule, if there are N nuclei, the dimensionality of the PES is $3N$, namely there are $3N$ nuclear coordinates that define the geometry[84]. Of these

coordinates, three describe the overall translation of the molecule, and three describe the overall rotation of the molecule with respect to three axes. Accordingly, the molecule has 3*N*-6 degrees of freedom in Cartesian space. Evidently, the visualisation of the entire (hyper) surface is impossible. However, by restricting the calculations to specific sections of the PES, it is possible to obtain useful information. The most chemically interesting points on the PES are the stationary points, where the gradients with respect to all internal coordinates are zero. These points are the minima and the saddle points. The energy minima correspond to stable structures, while saddle points, linking minima, correspond to transition states. The global energy minimum corresponds to the lowest energy point.

2.2.2 Molecular mechanics: the empirical force field models

Molecular mechanics (MM) techniques are based on a mathematical model of molecules behaving as a collection of balls (corresponding to the atoms) held together by springs (corresponding to the bonds)[85]. These methods rely on simple functional forms and sets of parameters empirically adjusted to reproduce the experimental or quantum chemical properties of molecules. In the MM approach, only the nuclear positions are considered, while the electronic distributions are ignored. Consequently, the MM methods (also referred to as force field methods) are only applicable to situations were there would be no change in the electronic structure.

Molecular mechanics assumes that the potential energy of a molecule arises from a few, specific interactions within a molecule. As such, the force field energy is typically calculated from the sum of bond, angle, dihedral and non-bonded energies:

$$E_{total} = \sum_{bonds} E_{bond} + \sum_{angles} E_{angle} + \sum_{dihedrals} E_{dihedral}$$

$$+ \sum_{atom\,pairs} [E_{eel} + E_{vdW}] \qquad (2.14)$$

The bond, angle and dihedral energy contributions (also known as bonded potentials) describe the interactions between atoms that are directly connected by chemical bonds. Typically, each bond, angle and dihedral energy contribution is calculated individually and then summed. On the other hand, the non-bonded interactions (or non-bonded potentials) represent the influence of non-covalent effects such as electrostatics and van der Waals forces. The non-bonded interaction potentials are pair additive.

Force fields are based on numerous approximations and derived from different types of experimental data. Hence, the exact terms and parameters in the above potential function vary between the different force fields. Examples of commonly used force fields in the area of molecular dynamics of macromolecules include AMBER[86, 87], GROMOS[88] and CHARMM[89]. In this work, the AMBER *ff99SB* force field[86] has been used for the protein, and the general AMBER force field (gaff)[87] for the ligands. The functional form of the AMBER force field is:

$$E_{total}(r^n) = \sum_{bonds} \frac{k_b}{2}(b - b_{eq})^2 + \sum_{angles} \frac{k_\theta}{2}(\theta - \theta_{eq})^2 +$$

$$\sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{atom\ pairs} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\varepsilon_r R_{ij}}\right] \qquad (2.15)$$

where the potential energy E_{total} is expressed as a function of the positions r of n atoms and the following terms are empirically defined parameters: k_b, r_{eq}, k_θ, θ_{eq}, V_n, n, ϕ, γ, A_{ij}, B_{ij}, ε_r, q_i and q_j.

The first and second terms in equation 2.15 correspond to the energy function for stretching a bond, E_{bond}, and to the angle bending function, E_{angle}. The bond and angle contributions are described by harmonic potentials analogous to Hooke's law and account for the interactions between directly bonded (A-B) or directly angled (A-B-C) atoms:

$$E_{bond} = \frac{k_b}{2}(b - b_e)^2 \qquad (2.16)$$

$$E_{angle} = \frac{k_\theta}{2}(\theta - \theta_e)^2 \qquad (2.17)$$

where b corresponds to the bond length and θ to the value of the angle; b_{eq} and θ_{eq} correspond to the associated equilibrium values; and k_b *and* k_θ are the Hooke's law force constant for the bond and angle, respectively. The simple harmonic functions above are generally sufficient for determining most atomic geometries, since it is unusual (and energetically expensive) for a bond length or a bond angle to deviate significantly from equilibrium in a molecular mechanical calculation. This is why the bond stretching and angle bending terms are usually called hard degrees of freedom, since large energies are required to cause significant deviation from the equilibrium geometries.

The dihedral angle term ($E_{dihedral}$), the third term in equation 2.15, describes the change in potential energy associated with rotating a chemical bond around a torsional angle ϕ defined by four consecutively bonded atoms:

$$E_{dihedral} = \frac{\mathcal{V}_n}{2} \left[1 + \cos{(n\phi - \gamma)} \right] \qquad (2.18)$$

where $\mathcal{V}_n$ states the height of the barrier to rotation, n is the periodicity of the torsional barrier (which gives the number of minimum points in the function as the torsion angle changes from 0 to 2π), and γ is the phase factor, which sets the position of the potential minima.

Finally, the non-bonded energy (last term in equation 2.15) is composed of an electrostatic term (E_{eel}) and a Lennard-Jones term (E_{vdW}). The former accounts for the electrostatic interactions between partial charges in a molecule due to polar bonds, while the latter quantifies the van der Waals attractions or repulsions of atoms that come close together. Coulomb's law describes the electrostatic interaction between a pair of charged atoms:

$$E_{eel} = \frac{q_i q_j}{4\pi \varepsilon_0 \varepsilon_r r_{ij}} \qquad (2.19)$$

where ε_0 is the permittivity of vacuum, ε_r is the permittivity of the medium, q_i and q_j are the partial charges of the two atoms, and r_{ij} is the separation between them. The term 'partial atomic charges' indicates that point charges centred on the atoms are being used to represent the charge distribution within the molecule. The partial atomic charges are empirical parameters that must be fitted to match experimental data or quantum mechanical potentials. In this work, the restrained electrostatic potential (RESP) method[90] is used for the derivation of charges.

The van der Waals potential can be represented by the Lennard-Jones 12-6 function[91]:

$$E_{vdW} = \frac{A}{r^{12}} - \frac{B}{r^6} \qquad (2.20)$$

where r is the distance between the two particles; $A = 4\varepsilon\sigma^{12}$ and $B = 4\varepsilon\sigma^6$, with ε as the depth of the well, and σ as the collision distance.

As indicated in equation 2.15, the total energy of the system is taken as the sum over all inter- and intra-molecular terms. However, because the evaluation of the energy between all pairs of atoms can be time consuming in large systems, the

intermolecular terms are usually truncated such that interactions between atoms separated by more than a cut-off distance are ignored. The size of the van der Waals interactions between two atoms decreases rapidly with the distance[83]. Thus, it is possible to truncate the Lennard-Jones potential without introducing significant errors into the calculation. On the other hand, the electrostatic interactions are longer ranged. Hence, truncating them could introduce significant errors into the calculation. In this work, to include all the effects of long-range forces, the particle-mesh Ewald approach (PME)[92] has been utilised, and will be described in section 2.2.5.

2.2.3 Molecular dynamics

Molecular dynamics (MD) simulations describe the time dependent behaviour of a molecular system by numerically solving Newton's equations of motion for all atoms in the system[82, 93]. For this to be possible, MD simulations rely on the following three approximations: firstly, the Born-Oppenheimer approximation, where nuclear and electronic motions are decoupled; secondly, the approximation that nuclei can be treated as classical particles; and thirdly, the use of an empirical force field to describe the potential energy with respect to conformation.

The MD trajectory, which specifies how the positions and velocities of the particles in the system vary with time, is obtained by solving the differential equations embodied in Newton's second law ($F=ma$):

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i} \tag{2.21}$$

The above equation describes the motion of a particle of mass m_i along one coordinate, x_i, with F_{x_i} being the force on the particle in that direction. The force F_{x_i} is evaluated from the first differential of the potential energy expression described by the force field. The acceleration (d^2x_i / dt^2) leads to a change in the atom's velocity and position within a discrete time step δt, chosen sufficiently short such as to capture the fastest motions of the system.

For an efficient integration of Newton's equation of motion, AMBER uses the leap-frog algorithm[94]. This algorithm relies on the following relationships for updating positions and velocities:

$$r_i(t + \delta t) = r_i(t) + \delta t v_i\left(t + \frac{1}{2}\delta t\right)$$ (2.22)

$$v_i\left(t + \frac{1}{2}\delta t\right) = v_i\left(t - \frac{1}{2}\delta t\right) + \delta t a_i(t)$$ (2.23)

where r_i is the position of particle i, v_i is the velocity and a_i the acceleration. To implement the leap-frog algorithm, the velocity $v_i(t + \delta t/2)$ is first calculated, from the velocity at time $t - \delta t/2$ and the acceleration at time t, as described by equation 2.23. Subsequently, using equation 2.22, the position $r_i(t + \delta t)$ is calculated from the position $r_i(t)$ and the velocity just calculated above[82]. The velocity at time t can be calculated from:

$$v_i(t) = \frac{1}{2}\left[v_i\left(t + \frac{1}{2}\delta t\right) + v_i\left(t - \frac{1}{2}\delta t\right)\right]$$ (2.24)

As can be seen from equations 2.22 and 2.23, the position and velocity updates are out of phase by half a time step. First, the velocities leap-frog over the positions and then the other way around. This lack of synchronicity is an obvious disadvantage and means that it is not possible to calculate the kinetic energy contribution to the total energy at the same time as the positions are defined (from which the potential energy is determined). Two advantages of the leap-frog algorithm are its numerical accuracy (which is better than the standard Verlet algorithm), and the fact that the velocities appear directly, which facilitates a coupling to an external heat bath[84].

2.2.4 Solvation models

The solvent has a fundamental influence on the structure, dynamics and thermodynamics of biological systems. For biological systems such as proteins, the solvent of choice is usually water. The methods for evaluating the solvent effect are broadly divided into two types: those describing the individual solvent molecules and those that treat the solvent as a continuous medium. The former are called explicit solvent models, while the latter are termed implicit solvent models.

2.2.4.1 Explicit solvent models

In the explicit solvent models a collection of individual water molecules surrounds the macromolecule. This could cause an enormous increase in the number of interactions present in the system. It is therefore desirable to have as simple a representation of

the water molecule as possible. The simplest water models treat the water molecule as rigid and rely only on non-bonded interactions. The commonly used TIP3P model[95] uses three interaction sites, which corresponds to the three atoms of the water molecule. Each atom gets assigned a point charge, with positive charges on the hydrogen atoms and a negative charge on the oxygen. The oxygen atom also gets the Lennard-Jones parameters. In this work, during the MD simulations, the TIP3P model was used to describe the solvent.

2.2.4.2 Implicit solvent models

Implicit solvent models, also called continuum models, consider the solvent as a uniform polarisable environment with a relative dielectric constant ε. Since there are no discrete water molecules, the solute (i.e. the macromolecule) interacts only with itself, but the electrostatic interactions are modified to account for the solvent.

The solvation free energy (ΔG_{solv}) is the cost in free energy change to transfer a molecule from vacuum to solvent. The solvation free energy can be expressed as:

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{np} \qquad (2.25)$$

where ΔG_{elec} is the electrostatic contribution to solvation, and ΔG_{np} is the non-electrostatic contribution, which in turn has two components: ΔG_{cav} and ΔG_{vdw}. The former is the free energy required to form the solute cavity within the solvent, and the latter is the van der Waals interaction between the solute and the solvent.

The electrostatic component of the solvation free energy, ΔG_{elec}, can be described at different levels of approximation. Among these, the methods based upon the Poisson-Boltzmann (PB) equation[96] are of particular interest to this investigation. These methods treat the protein and ligand as low dielectric regions surrounded by the high dielectric aqueous solvent. For an aqueous solution containing mobile ions, the Poisson-Boltzmann equation can be described as:

$$\nabla \cdot \varepsilon(\boldsymbol{r})\nabla\phi(\boldsymbol{r}) - \kappa' sinh[\phi(\boldsymbol{r})] = -4\pi\rho(\boldsymbol{r}) \qquad (2.26)$$

where ε is the uniform dielectric constant of the solvent, ϕ is the electrostatic potential, and ρ is the charge density, namely the distribution of the charge throughout the system. κ' is related to the Debye-Hückel inverse length κ, which relates to the ionic strength of the solution and measures how far the electrostatic effects extend into the solution:

$$\kappa^2 = \frac{\kappa'^2}{\varepsilon} = \frac{8\pi q^2 I}{\varepsilon k_B T} \tag{2.27}$$

where I is the ionic strength of the solution, q is the magnitude of the charge of the electrolyte ions, T is the temperature and k_B is Boltzmann constant. If the potential is sufficiently small, i.e. if the solute is not strongly charged, the PB equation can be linearised by performing a Taylor expansion of the hyperbolic sine function. Keeping only the first term gives the Linearised Poisson-Boltzmann Equation (LPBE):

$$\nabla \cdot \varepsilon (\boldsymbol{r})\nabla\phi(\boldsymbol{r}) - \kappa'\phi(\boldsymbol{r}) = -4\pi\rho(\boldsymbol{r}) \tag{2.28}$$

The above equation can be solved using the finite difference Poisson-Boltzmann (FD-PB) method[97, 98]. The philosophy behind this method is to map all physical quantities (i.e. atomic charges, electrostatic potentials, ionic strength and dielectric constant) onto a grid, replacing the differential operators in equation 2.28 by grid value differences, hopefully finite (hence the name). Evidently, the accuracy of the PB procedure is determined by the grid spacing. Fine grid resolutions are required for sufficient accuracy, but they also increase calculation time and memory requirements.

In MD applications, the computational costs associated with the numerical solutions of the Poisson-Boltzmann equation are often very high. As such, when many conformations need to be processed, faster generalised Born (GB) models[99] are valuable approximations to the full PB model. These are based on the realisation that dielectric screening of a charge-charge interaction correlates with the degree to which each charge interacts with surrounding water. For a system of particles of radii a_i and charge q_i, the GB method defines the electrostatic component of the solvation free energy as:

$$\Delta G_{elec} = -\left(1 - \frac{1}{\varepsilon}\right) \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{r_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right) \sum_{i=1}^{N} \frac{q_i^2}{a_i} \tag{2.29}$$

where ε is the relative permittivity of the medium and r_{ij} is the distance between the particles. Despite the computational advantages provided by the GB method, in this work, only the more rigorous PB method is used.

Turning now to the non-electrostatic contribution to the solvation free energy, ΔG_{np}, which can be especially significant for solutes that are neither charged nor highly polar. ΔG_{np} decomposes into a repulsive cavity hydration free energy term, ΔG_{cav} and a stabilisation free energy term due to van der Waals interactions between the solute

and solvent, ΔG_{vdw}. Both terms are usually assumed to be proportional to the total solvent accessible surface area (SASA)[100, 101], and are represented using an equation of the following form:

$$\Delta G_{vdw} + \Delta G_{cav} = \gamma A + b \tag{2.30}$$

where *A* is the total solvent accessible area, and γ and *b* are constants usually taken from experimentally determined free energies for the transfer of alkanes from vacuum to water[82]. The effect is to add a positive (unfavourable) solvation energy to conformations with more surface area and thus to favour binding, which reduces surface area.

2.2.5 Periodic boundary conditions

To be computationally feasible, MD simulations in explicit solvent require a much smaller system size than the corresponding macroscopic experiment. Periodic boundary conditions (PBC) enable a simulation to be performed using a relatively small number of particles, but in a way that the particles experience forces as if they were in bulk solvent. To accomplish this, the simulation system is surrounded by periodic images of itself, allowing particles that leave the simulation box (also called unit cell) on one side to instantaneously re-enter it on the opposite side. Periodic boundary conditions are typically used in conjunction with the minimum image convention for short ranged forces, which states that a particle can interact only with the closest image of any other particle.

Unfortunately, the long ranged forces present a more complex scenario. Since they are infinitely ranged, the unit cell and all its periodic images must be considered. In PBC, to include all the effects of long-range forces the particle-mesh Ewald (PME)[92] approach is usually applied. The PME method is a fast implementation of the Ewald summation method. Its fundamental concept is to split the direct summation of interaction energies between two particles into two summations: the real (direct) space sum corresponding to the short ranged interactions, and the reciprocal space sum corresponding to long rage interactions[102]. The short-range force part is calculated from particle-particle interactions, using a modification of Coulomb's Law (equation 2.19). The long-range interactions are calculated using a Fourier transform to build a 'mesh' of charges, interpolated onto a grid.

2.2.6   Temperature and pressure regulation

MD simulations naturally occur in the microcanonical (NVE) ensemble, in which the number of particles, volume and energy remains constant. Algorithms that connect the system to a thermostat or barostat allow the sampling of the canonical (NVT) or the isothermal-isobaric (NPT) ensemble.

The temperature is related to the total kinetic energy of the system; as such it could be kept constant by adjusting the velocities of each particle at each step. This, however, is undesirable because the adjustment of the velocities could cause the trajectories to be no longer Newtonian, and the properties calculated from the trajectories to be less reliable[85]. An alternative way to generate simulations in the canonical ensemble is by using the stochastic collisions methods, such as the Andersen thermostat[103], or Langevin dynamics[104]. Langevin dynamics, the method used in this work, relies on the Langevin equation of motion[105], in which a frictional force added to the conservative force is proportional to the velocity, and it adjusts the kinetic energy of the particles so that the temperature matches the set temperature.

Similar to keeping the temperature of a simulation system constant by the methods described above, a barostat is introduced to keep the pressure of the system constant. In this investigation, the Berendsen barostat[106] has been used. This method controls the pressure by rescaling the volume of the simulation box at each time step so that the pressure oscillated around the desired value.

2.3 Free energy calculations

The free energies of molecular systems describe their tendencies to associate and react[107]. Consequently, being able to predict this quantity gives the ability to predict solvation, binding, stability, phase transitions and many other properties. Computational approaches to evaluate the binding affinity of a receptor-ligand system are generally formulated in terms of estimating the relative free energy differences, ΔG, between two equilibrium states[108]. Thermodynamically rigorous methods for calculating the free energy differences include the free energy perturbation (FEP) methodology[109] and the thermodynamic integration (TI) method[110]. However, both approaches are computationally expensive and time consuming. Faster and simpler approximation methods include the lineal interaction energy (LIE) method[111] and the

molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method[112]. In this work, the MM-PBSA method is used.

2.3.1 MM-PBSA method

The molecular mechanics Poisson-Boltzmann surface area method[112] is based on an analysis of MD trajectories using a continuum solvent approach. In this method the binding free energy of a state is approximated as:

$$\Delta G = \Delta E_{MM} + \Delta G_{PBSA} - T\Delta S \tag{2.31}$$

where ΔE_{MM} is an average molecular mechanical energy, ΔG_{PBSA} is the average solvation energy and the last term, $T\Delta S$, represents the solute entropy. The solute entropy can be estimated by quasi-harmonic analysis of the trajectory or by using normal mode analysis[113].

The molecular mechanical energy ΔE_{MM} is the sum of the bond, angle, torsion, van der Waals and electrostatic terms from a regular force field. Solvation free energies are calculated using a numerical solution of the Poisson-Boltzmann equation for the polar contribution, ΔG_{PB}; and a surface area based estimate of the nonpolar free energy, ΔG_{np}:

$$\Delta G_{PBSA} = \Delta G_{PBSA} + \Delta G_{np} \tag{2.32}$$

The terms in equation 2.31 are evaluated for the complex, receptor and ligand. Subsequently, the binding free energy is estimated as:

$$\Delta G_{bind} = G_{complex} - \left(G_{receptor} + G_{ligand}\right) \tag{2.33}$$

The theory and methods discussed here will be used in the following chapters to provide a structural, dynamic and energetic insight into the subtle basis of molecular recognition by PI3K.

3. Structural refinement

A major aim of this thesis is to examine the basis on which PI3Kγ distinguishes γ-selective inhibitors **1** and **2** (Figure 3.1). However, the ability to do so is hindered by the uncertainty regarding the conformation of the ligands. Despite their chemical similarity, the X-ray structures for their PI3Kγ complexes exhibit the molecules in different conformers, *s*-trans for inhibitor **1** and *s*-cis for inhibitor **2**[30].
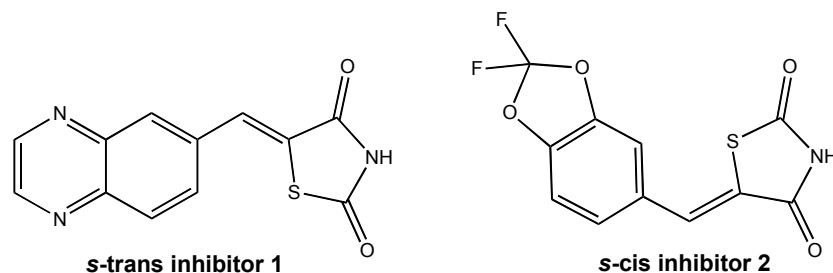
Figure 3.1: Chemical structure of the PI3Kγ-selective *s*-trans inhibitor **1** and *s*-cis inhibitor **2**.

The main goal of this chapter is to determine the conformational preference of inhibitors **1** and **2**. For this purpose, the first objective is to investigate computationally the intrinsic conformational preference of the isolated ligands, and subsequently revisit the available crystallographic data to determine which isomer is predominant in the protein active site.

It should be noted that the structural refinement work presented in this chapter was conducted in collaboration with Dr. Mark Dunstan (Manchester Interdisciplinary Biocentre, University of Manchester).

3.1   Rotational barriers for inhibitors 1 and 2

The stability of inhibitors **1** and **2** was investigated using *ab initio* quantum mechanics (QM) calculations. A relaxed potential energy surface (PES) scan was conducted in order to evaluate the conformational preferences of the compounds.

The relaxed PES scan was performed utilising the Gaussian09 package[114]. For this calculation, the M06L functional of Truhlar *et al.*[115] was used, along with the MG3S basis set[116]. It is worth mentioning that the MG3S basis set is identical to the MG3 basis set[117,118] except that in the former, diffuse functions on hydrogen are absent; and that the MG3 basis set is the same as 6-311++G(3*d2f*,2*df*,2*p*)[119,120] for H–Si systems, but improved[121] for P–Ar systems.

The meta-generalised gradient approximation (GGA) functional M06L was chosen because of its excellent performance, when compared to other density functional theory (DFT) methods such as B3LPY, BLYP and BP86, for calculating non-covalent interactions[115, 122].

A PES scan involves the calculation of potential energy in equispaced steps over a rectangular grid of selected internal coordinates. In this study, the scanning variable is the dihedral angle C1-C2-C3-C4, a conjugated diene system (shown in red, Figure 3.2). The initial value of this variable was set to 0º (s-cis conformation), and it was incremented by 15º a total of 11 times. A constrained geometry optimisation was performed for each dihedral angle value.

s-trans inhibitor 1
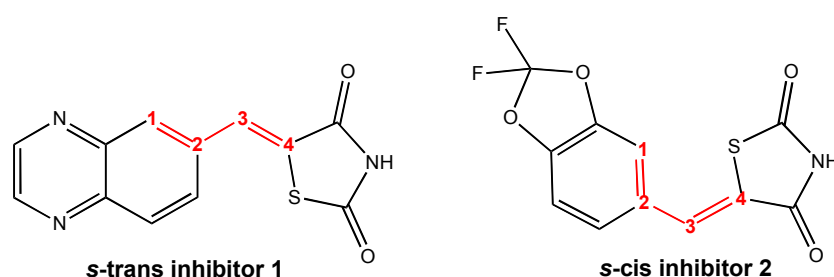
s-cis inhibitor 2

Figure 3.2: Structure of s-trans inhibitor **1** and s-cis inhibitor **2** showing the key dihedral angle C1-C2-C3-C4 in red. The s-cis and s-trans conformations for each structure are related by rotation about the single bond C2-C3.

The resulting potential energy curve (Figure 3.3, where the blue line corresponds to inhibitor **1**, while the red line corresponds to inhibitor **2**) shows that for both inhibitors the s-cis and s-trans conformations are isoenergetic. The relative energies are given in Table 3.1.
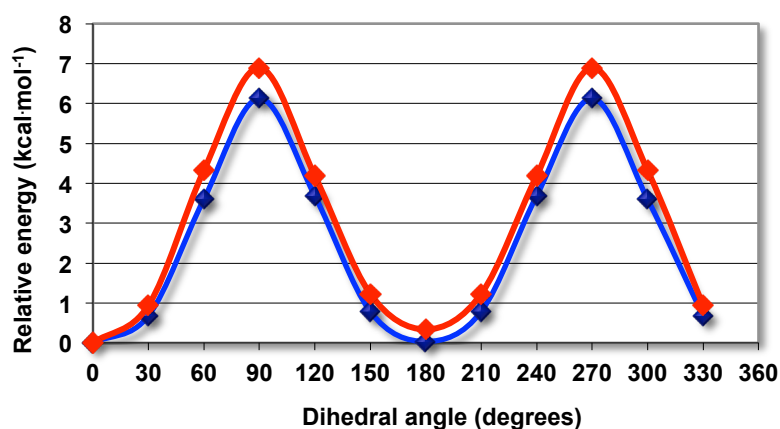
Figure 3.3: Potential energy curve for rotation about the C2-C3 bond of inhibitor **1** (blue line) and inhibitor **2** (red line). The data was determined by *ab initio* calculations at the M06L/MG3S level of theory.

65

The rotational barriers between the *s*-cis and *s*-trans conformation was found to range between 6 and 7 kcal·mol^{-1} (Table 3.1), suggesting reasonably facile interconversion.

Table 3.1: Calculated relative energy and barrier of interconversion for inhibitor **1** and inhibitor **2** as determined by *ab initio* calculations at the M06L/MG3S level of theory.

Structure	E_{rel} (kcal·mol^{-1})	
	Inhibitor 1	Inhibitor 2
s-cis	0.0	0.0
transition state	6.1	6.9
s-trans	0.0	0.3

For the purpose of analysis, the obtained values can be compared to the energy barrier for rotation about carbon-carbon single bonds for small organic molecules presenting rotational isomerism. Some representative examples are listed in Table 3.2.

Table 3.2: Rotational energy barriers for alkanes of the type CH_3-X^a.

Compound	Barrier height (kcal·mol^{-1})
CH_3-CH_3	2.9
CH_3-CH_2CH_3	3.4
CH_3-$CH(CH_3)_2$	3.9
CH_3-$C(CH_3)_3$	4.7

[a] Taken from reference[123]; barriers are those for rotation about the bond indicated in the formula.

It is possible to appreciate how the calculated barriers of 6 - 7 kcal·mol^{-1} are just slightly higher than the typical values for C-C single bonds. This is to be expected due to the presence of the conjugated diene system in inhibitors **1** and **2**. Conversely, the calculated barriers are much lower than the typical values reported for double bonds involving sp^2 carbons where rotation is prevented as a result of the stable π-π bond. For these systems, the high thermodynamic barrier has been estimated to be about 62-65 kcal·mol^{-1} for simple alkenes[124].

In summary, the results of the PES scan, as well as the comparison to previously reported rotational barriers, suggest that from the energetic point of view, for both inhibitors, the *s*-cis and *s*-trans conformer could coexist in dynamic equilibrium in the unbound state.

3.2   Re-examination of the inhibitors crystallographic data

Knowing that the *s*-cis and *s*-trans conformations could coexist, the next step was to re-examine which isomer is predominant in the protein active site. To answer this question, given the differing observed binding mode of the inhibitors and the fairly low resolution of the X-ray crystal structures[30] (2.7 and 2.9 Å for PI3Kγ complexed with inhibitor **1** and **2**, respectively), it was decided to re-evaluate the available crystallographic data.

As discussed in section 2.1, while the electron density maps are the best representation of the crystallographic experiment, an atomic model is just one (partly subjective) interpretation of the data contained in the density maps. Hence, to assess the validity of the current protein-ligand models, an examination of the electron density maps was performed.

The $2F_o$ - F_c map, as well as the F_o - F_c difference map, was obtained using the Uppsala Electron Density Server[125] (EDS), utilising the PDB entries 2a5u (inhibitor **1**) and 2a4z (inhibitor **2**)[30]. The graphics program COOT[126,127] was used for visualisation of the maps and manipulation of the atomic model. Unless otherwise stated, all $2F_o$ - F_c and F_o - F_c maps in the following sections have been contoured at the 2.0 and 3.0 σ level, respectively, where σ is the RMSD of all map points from the average value of electron density.

3.2.1   Re-examination of the crystallographic pose of PI3Kγ-bound inhibitor 1

The originally published[30] X-ray crystal structure of PI3Kγ complexed with inhibitor **1** (resolution 2.7 Å) reported the inhibitor in *s*-trans conformation. Interestingly, an analysis of the atomic model and corresponding electron density maps questions its validity.

A superposition of the atomic model and the $2F_o$ - F_c and F_o - F_c maps is shown in Figure 3.4. The blue iso-surface corresponds to the $2F_o$ - F_c map and indicates where the model should lie. Nevertheless, this $2F_o$ - F_c map should be interpreted with caution, because it contains the influence of the assigned structural models. The green and red iso-surfaces correspond to the difference map (F_o - F_c). The green parts are positive density and indicate unsatisfied electron density; in other words, the model needs to be modified by assigning atomic volume to this region thus increasing the electron density. Conversely, the red parts are negative density and imply that the

electron density does not support the occupancy of this region. In this case, the sulfur atom is lying in a negative density area (i.e. red region), suggesting that perhaps it should not be there.
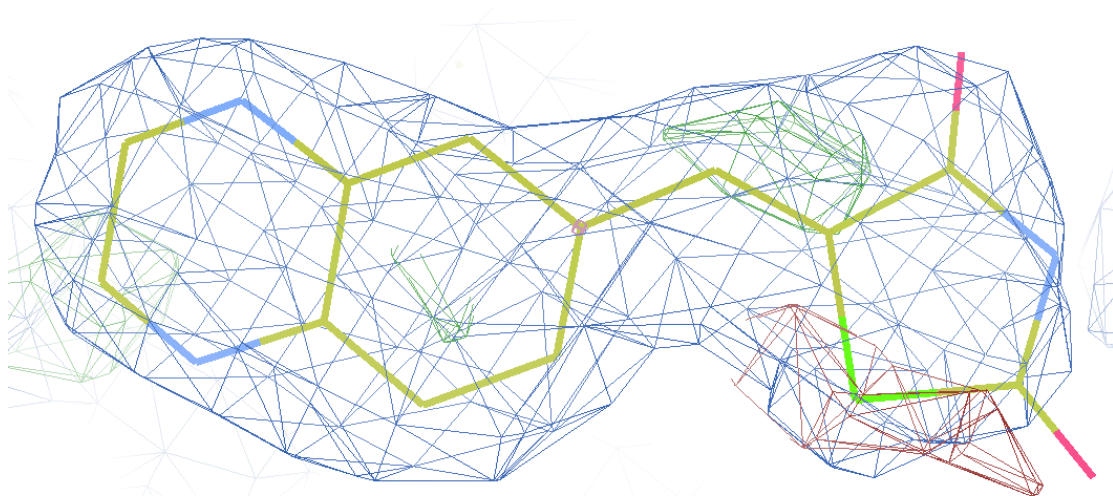
Figure 3.4: Visualisation of the $2F_o$-F_c (blue) and F_o-F_c (green and red) electron density maps for inhibitor **1** in *s*-trans conformation, i.e. as it appears in the originally published X-ray structure.

Upon initial inspection, the existence of an alternative ligand orientation seems plausible. However, if one is to explore different ligand conformations, it is indispensable to be able to rank their validity.

When considering whether the available crystallographic data may be best fitted by an alternative conformation, there are three valuable pieces of information that should be taken into consideration. These are: the quality of the maps, the values of the atomic B-factors, and the occupancy values. Firstly, the quality of the maps is important because it is a direct measurement of how well the current model predicts the observed data. The values of the B-factors are informative because they not only indicate the true static or dynamic mobility of an atom, but they can also reveal where there are errors in model building. Lastly the occupancy values are insightful because it is possible that a mixture of conformations, rather than just one, may best fit the electron density.

The first aim was to improve the quality of the maps. For the purpose of building a new ligand conformation, the most useful electron density maps are the ones that are less influenced by the currently assigned molecular model. The difference map (in red and green in Figure 3.4) fits into this category, emphasising errors in the model. On the other hand, the $2F_o$ - F_c map (in blue in Figure 3.4) contains the influence of the existing model. Therefore, if possible, it is best not to use it. Instead, an OMIT map should be used. As explained in section 2.1.2, an OMIT map is a way to reduce the

model bias in the electron density calculated with model phases. To construct an OMIT map, questionable fragments of the model are eliminated; in this case, the atomic coordinates of the ligand. As a result, when new phases are recalculated for the modified model, the "phase bias" caused by the suspicious region is eliminated. The OMIT map was generated using the program PHENIX[128]. The result is an iso-surface of positive density (in green, Figure 3.5) that indicates where the new ligand model should be built.
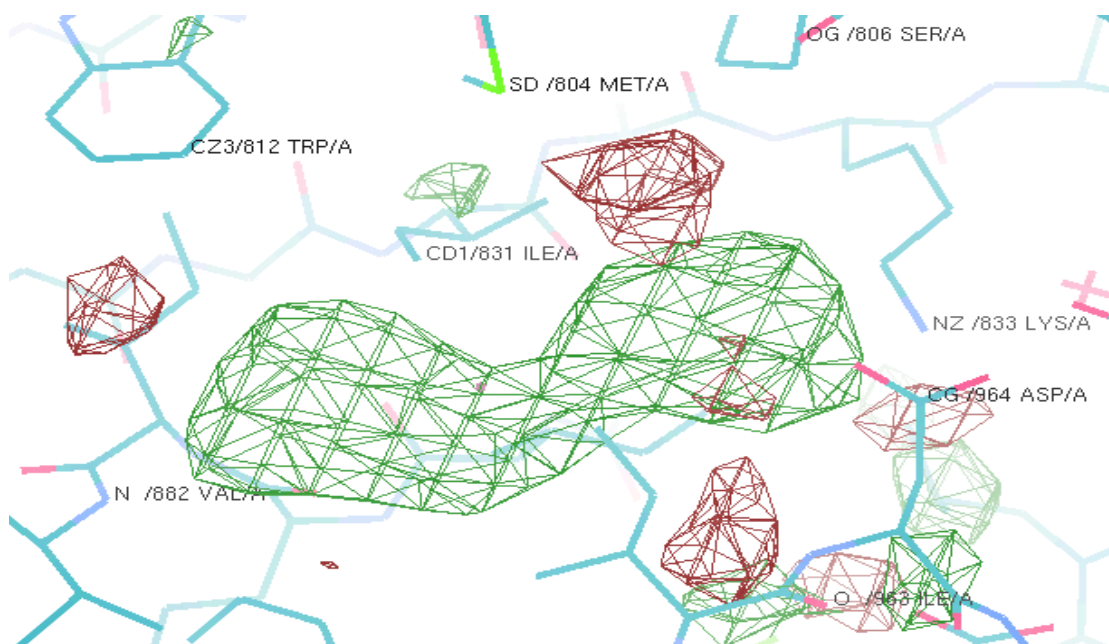
Figure 3.5: Visualisation of the active site of PI3Kγ and OMIT map (contoured at the 3.0 σ level). The region of positive electron density is shown in green and indicates where the ligand model should be built.

Using the information provided by the difference (F_o - F_c) and OMIT maps, the potential alternative conformations for the ligand are constructed.

The *s*-cis conformation for inhibitor **1** can be modelled by rotation about the single bond C2-C3 of the conjugated double bond system (shown in red in Figure 3.2). This alternative model is placed in the green iso-surface of positive density of the OMIT map. The result is shown in Figure 3.6. The next step is to evaluate the potential new model using the F_o - F_c map.
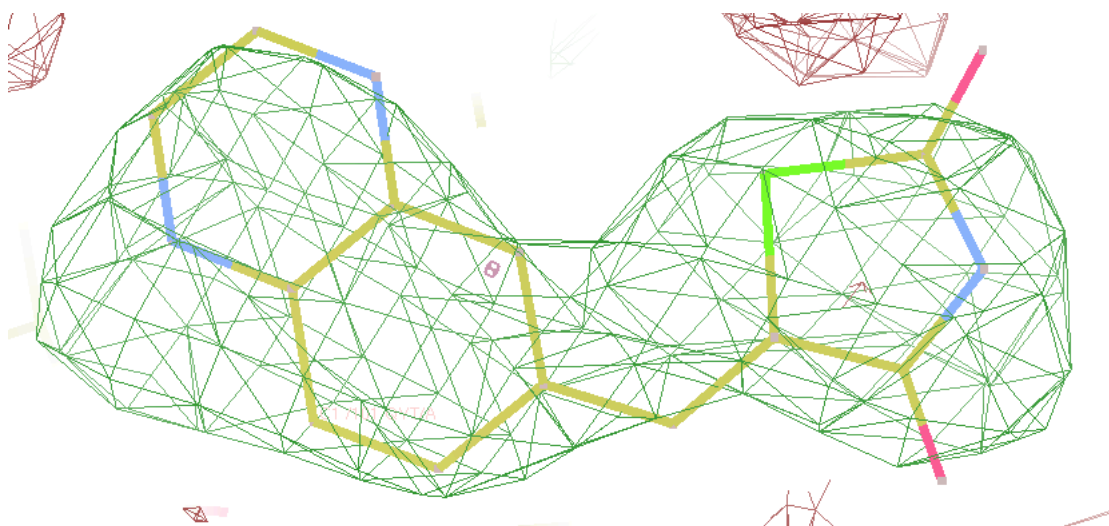
Figure 3.6: Visualisation of the OMIT map (green and red) where inhibitor **1** in *s*-cis conformation has been built. The map was contoured at the 3.0 σ level.

The superposition of *s*-cis inhibitor **1** and the $2F_o$ - F_c and F_o - F_c density maps is shown in Figure 3.7. In this model, the sulfur atom is sitting in the region of unsatisfied electron density (green isosurface). On the other hand, the region of negative density (red part), which previously contained questionable electron density, now hosts an oxygen atom, rather than the more electron rich sulfur atom. This strongly suggests the *s*-cis conformation may be a better fit.
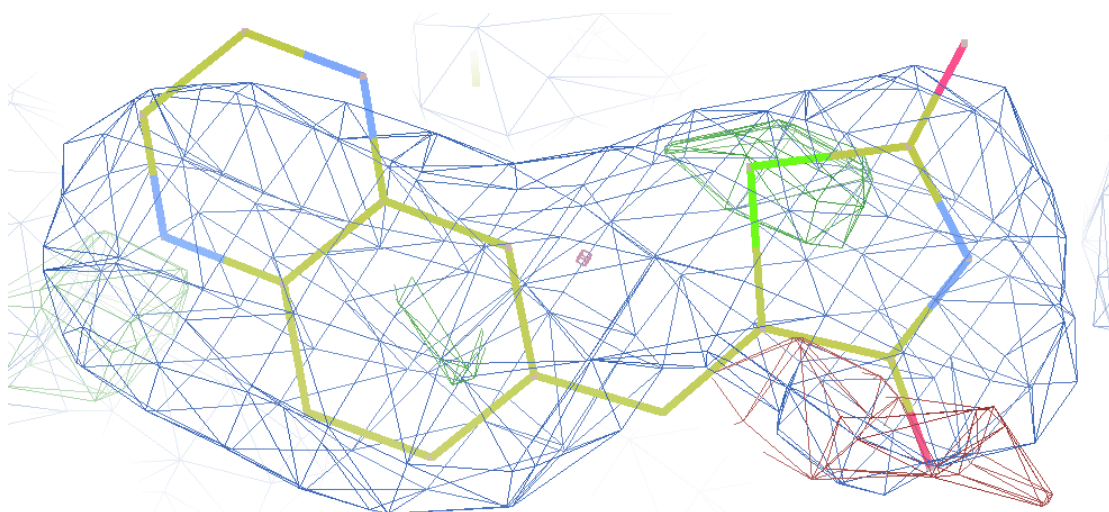
Figure 3.7: Visualisation of the $2F_o$-F_c (blue) and F_o-F_c (green and red) electron density maps for inhibitor **1** in *s*-cis conformation, i.e. after rotation about the single bond from conjugated diene.

Given that the *s*-cis conformation appears plausible for inhibitor **1**, the next stage involved the recalculation of the electron density maps $2F_o$-F_c and F_o-F_c as well as the re-refinement of the atomic model using the *phenix.refine*[129] program from the PHENIX[128] package. The superposition of the re-calculated maps and the re-refined

model is displayed in Figure 3.8. The *s*-cis model exhibits an improved fit over *s*-trans: the regions of unsatisfied electron density, both positive and negative, are now satisfied and thus the iso-surfaces have disappeared.
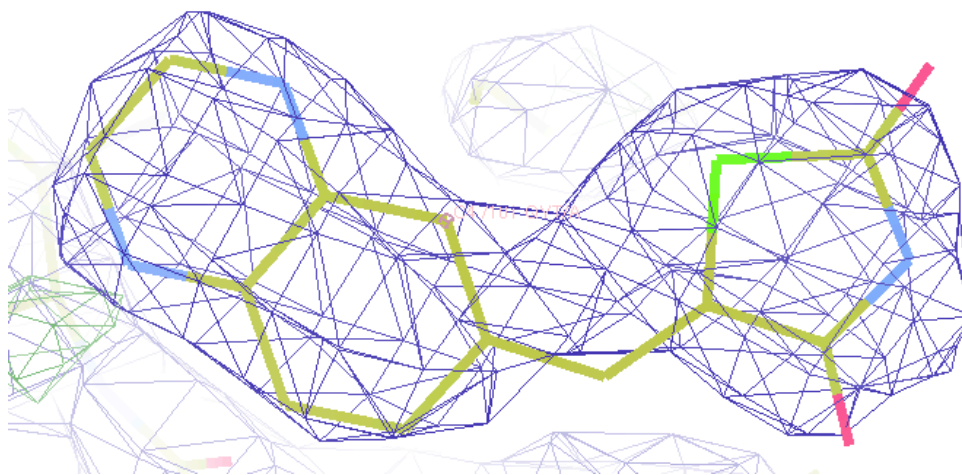
Figure 3.8: Visualisation of the recalculated $2F_o$-F_c (blue) and F_o-F_c (green and red) electron density maps for inhibitor **1** in *s*-cis conformation. The regions of previously unsatisfied electron density have disappeared.

By modelling the *s*-cis conformation of inhibitor **1** into PI3Kγ, the quality of the F_o-F_c map has been improved. Additionally, the B-factors for most of the ligand atoms have decreased (Figure 3.9). The atoms in the thiazolidinedione moiety show the biggest improvement, for example, the B-factor of the sulfur atom decreased from 70 Å^2 to 44 Å^2, while that of the nitrogen atom went from 63 Å^2 to 44 Å^2.
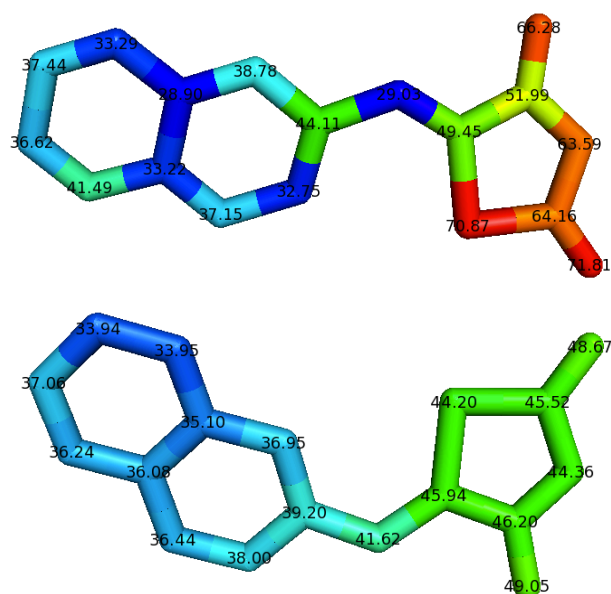
Figure 3.9: Chemical structures of *s*-trans inhibitor **1** (top) and *s*-cis inhibitor **1** (bottom). The molecules are coloured accordingly to their B-factor, from dark blue for low values to red for high B-factor values.

However, at the current resolution of 2.7 Å, B-factor refinement for individual atoms may not be highly informative. When dealing with low resolution structures, it is statistically more informative to use grouped B-factor refinement. The reason for this is the small observation-to-parameter ratio available for the structure model. For example, if an atom is described by its position (x, y, z) and an atomic displacement parameter (i.e. individual B-factor) at a resolution of about 2.8 Å, there is only one observation for each parameter. Consequently, a misleading level of agreement between the model and experiment could be obtained. Nevertheless, in this case the individual B-factor refinement for the ligand atoms was conducted for comparative purposes, but these results should be interpreted with caution.

To explore the possibility that the model might be better fitted by a mixture of both conformations (*s*-cis/*s*-trans), a dual occupancy refinement was performed using again the program PHENIX. In fact, our new *s*-cis conformation was found to have a much higher occupancy (87%) than the reported *s*-trans isomer (13%). This does not necessarily represent the true occupancy in the crystal but it is nonetheless conclusive enough to establish *s*-cis as the preferred conformation for inhibitor **1**.

3.2.2 Re-examination of the crystallographic pose of PI3Kγ-bound inhibitor 2

The originally published[30] X-ray crystal structure of PI3Kγ complexed with inhibitor **2** reported the inhibitor in *s*-cis conformation. In this case, the poorer resolution (2.9 Å) of the publicly available experimental data further complicated the evaluation of alternative conformations, and limited the refinement options.

A superposition of the published atomic model and the $2F_o - F_c$ and $F_o - F_c$ maps is displayed in Figure 3.10. Again, the blue iso-surface represents the $2F_o - F_c$ map and denotes where the model should lie. The green and red iso-surfaces correspond to the difference map ($F_o - F_c$) and indicate errors in the current model. It is possible to appreciate that an oxygen atom from the ligand is sitting on a region of negative density (red part), which implies that the electron density does not support the current model and atoms should be moved away from this region. However, the rest of the atoms do satisfy the electron density. As such, upon initial inspection, the *s*-cis model seems reasonable.
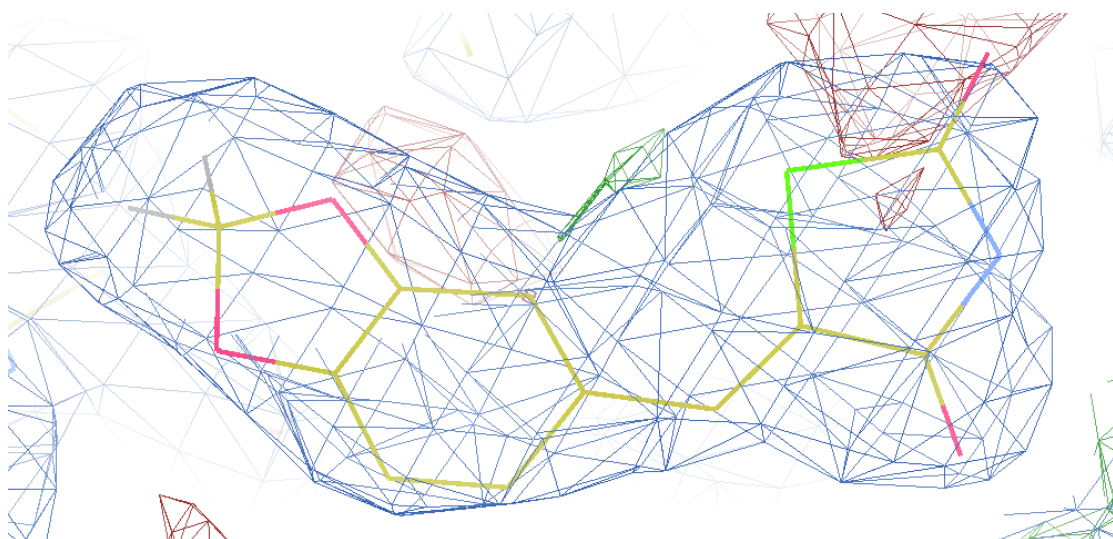
Figure 3.10: Visualisation of the 2F$_o$-F$_c$ (blue) and F$_o$-F$_c$ (green and red) electron density maps for inhibitor **2** in *s*-cis conformation, i.e. as it appears in the originally published X-ray structure.

Interestingly, the inspection of the B-factors for inhibitor **2** in the publicly available X-ray crystal structure showed that most of them were low, ranging from 40 to 52 Å^2 (Figure 3.11). Of particular interest are the atoms in the thiazolidinedione moiety. Some of these atoms displayed similar values to the ones obtained after refinement of inhibitor **1** above, for example, the sulfur atom with 42 Å^2, and the nitrogen atom with 47 Å^2.
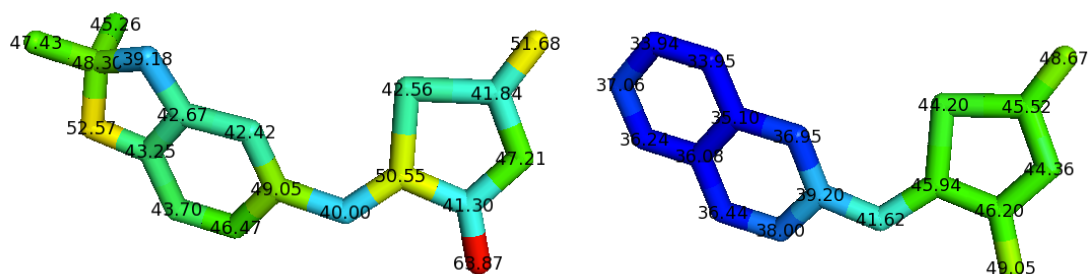
Figure 3.11: Chemical structures of *s*-cis inhibitor **2** (left) and *s*-cis inhibitor **1** (right). The molecules are coloured accordingly to their B-factor, from dark blue for low values to red for high B-factor values.

The low values of the B-factor further support the validity of the *s*-cis model. However, to test whether the quality of the maps could be improved, the *s*-trans conformation for inhibitor **2** was modelled and fitted to the 2F$_o$ - F$_c$ and F$_o$ - F$_c$ maps (Figure 3.12). In this case, due to the lack of evidence provided by the difference map, the placement of the *s*-trans model is more of an estimate. In this new scenario an oxygen atom from the ligand is still sitting on the region of negative density (red area). It can thus be suggested that the alternative *s*-trans conformation does not provide a better fit for the electron density.
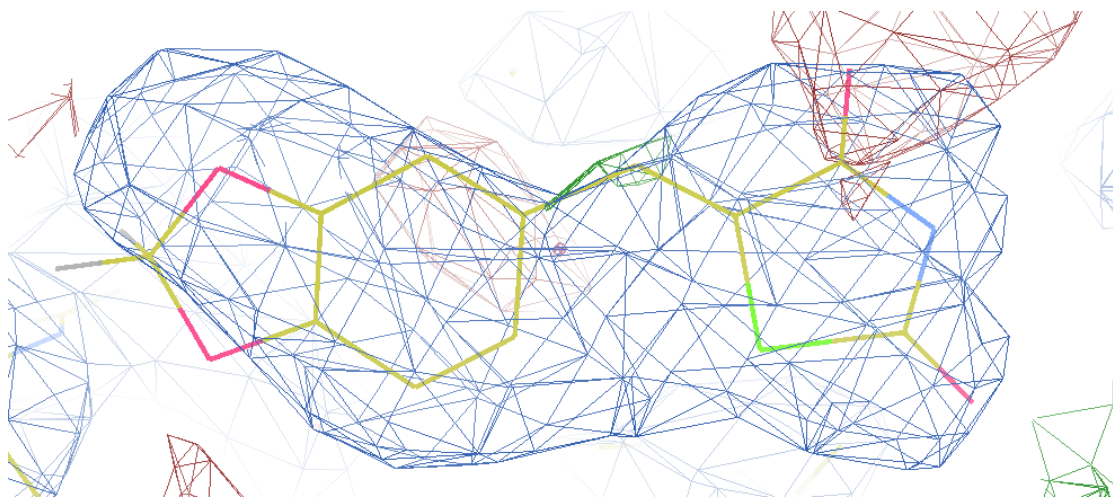
Figure 3.12: Visualisation of the $2F_o$-F_c (blue) and F_o-F_c (green and red) electron density maps superposed to the modelled inhibitor **2** in *s*-trans conformation.

The relatively good fit of the electron density maps with the initially modelled *s*-cis conformation, combined together with its low B-factors values, and the inability to improve the map quality by using the alternative s-trans model, make it possible to suggest the *s*-cis conformation as the preferred one for inhibitor **2**.

In summary, in this section the available crystallographic data was revised and it was found that for inhibitor **1**, not only is a *s*-cis conformation possible but, in fact, it has a much higher estimated occupancy (87%) than the originally modelled *s*-trans isomer (13%). On the basis of the available evidence, for the case of inhibitor **2**, the *s*-cis conformation was found to be the preferred one.

3.3   Re-refinement of the protein structure

In addition to re-examining the crystallographic pose of PI3Kγ-bound inhibitor **1**, the available crystallographic data for the PI3Kγ protein model reported in PDB entry 2a5u was analysed to search for features that required optimisation and re-refinement.

The first step consisted of real space intervention, in other words, map analysis and manual alterations to better fit the model to the observed electron density. To this end, the program COOT[126,127] was used. For the purpose of detecting problem areas in the assigned (published) protein model, the F_o - F_c and $2F_o$ - F_c maps were utilised. The positive and negative density regions indicated by the difference map were of particular interest. When necessary, the encountered anomalies were resolved by

residue repositioning, side-chain refitting, alternate side-chain conformations or peptide plane flipping.

Having performed these alterations to the protein structure, the resulting structure model was refined with the *phenix.refine*[129] program from the PHENIX[128] package using default settings. TLS (Translation Libration Screw-motion) refinement[130] was performed with the whole model as one TLS group. The B-factors were refined grouped by residue.

After the electron density maps were refined, the difference map showed a region of positive density in the proximity of the thiazolidinedione head of the ligand. The atomic distances suggested the presence of a water molecule with hydrogen bonds to the side-chains of residues Asp964 and Asn951. When assigning water molecules, the emphasis is on chemical plausibility; in this case, the requirement was fulfilled. Since only very high resolution X-ray structures (~1.2 Å and higher) can explicitly reveal the position of the hydrogen atoms, water molecules are normally placed in the form of a single oxygen atom. The resulting water placement can be seen in Figure 3.13.
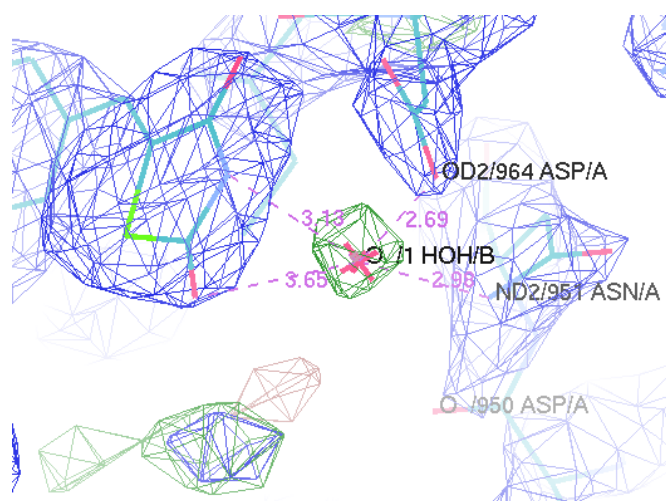
Figure 3.13: Detail of PDB entry 2a5u after refinement showing the thiazolidinedione head of ligand **1**, Asp964 and Asn951 (blue), and the assigned water molecule.

The process of manual rebuilding followed by automated re-refinement resulted in a structure model with improved R and R_{free} factors (Table 3.3). Of particular importance is the R_{free} because, as stated in section 2.1.3, it measures how well the model predicts the experimental observations that are not used to fit the model. After re-refinement, the R_{free} factor improved by 6.9%.

Additionally, the re-refined structure showed geometrical improvement. This is reflected on the increased backbone torsional angles φ and ψ that fall into allowed ranges on the Ramachandran plot (Figure 3.14).

Table 3.3: Quality metrics for PI3Kγ complexed with inhibitor **1**, using PDB entry 2a5u.

Metric	Before re-refinement	After re-refinement
R	27.9	21.1
R_{free}	35.1	28.2
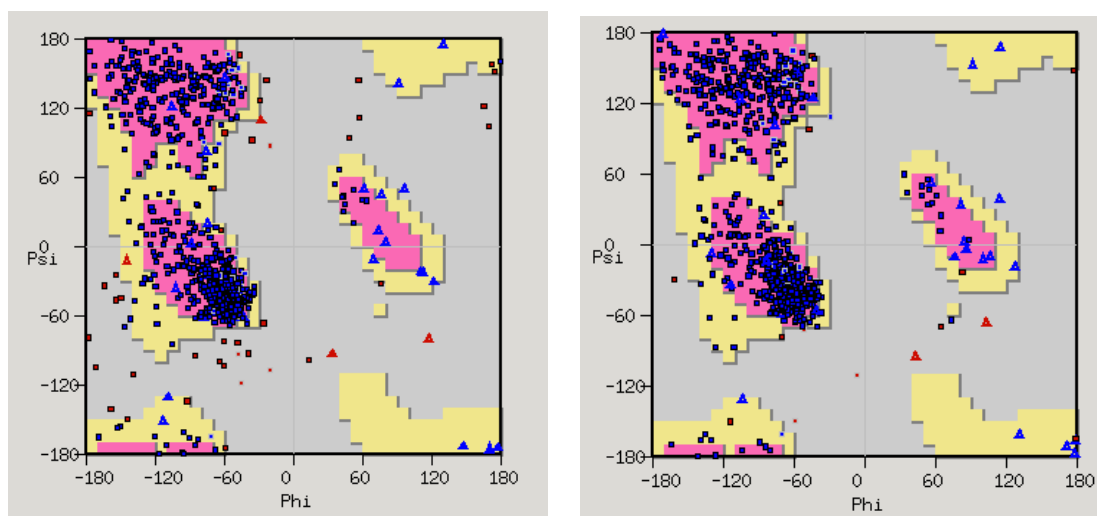
Figure 3.14: Ramachandran plot showing the peptide bonds angles φ and ψ for PI3Kγ complexed to inhibitor **1,** pre-refinement (left) and post refinement (right).

In the Ramachandran plot above, the grey area corresponds to outlier residues whose φ and ψ angles deviate significantly from the standard angles. In other words, these are conformations where atoms in the polypeptide come closer than the sum of their van der Waals radii. Residues in the preferred regions lie in the pink areas and correspond to conformations where there are no steric clashes, and as such the φ and ψ angles best fit the standard torsion angles. The yellow area represents the allowed regions if slightly shorter van der Waals radii are used in the calculation; in this way the atoms are allowed to come a little closer.

In the final re-refined structure, only 1.83% outliers are found (Table 3.4), as opposed to the initially present 5.24%. The decrease in outliers, together with the increase of residues in the preferred regions is an indication of geometrical improvement.

Table 3.4: Residue distribution in the Ramachandran plot (and their percentage of the total number of residues) for PI3Kγ complexed with inhibitor **1**, using PDB entry 2a5u.

Residue distribution	Pre-refinement		Post-refinement	
In preferred regions	702	(85.51%)	740	(90.13%)
In allowed regions	76	(9.26%)	66	(8.04%)
Outliers	43	(5.24%)	15	(1.83%)

It is important to bear in mind that the data for the re-refined structure model contained the ligand as a mixture of *s*-cis and *s*-trans conformations; this in accordance to the results of the occupancy refinement detailed in section 3.2.1.

In summary, the refinement of the PI3Kγ structure determined a protein model with improved quality metrics.

3.4   Conclusions

Returning to the question posed at the beginning of this chapter regarding the conformation of the two PI3Kγ ligands, it can now be stated that for inhibitors **1** and **2** the *s*-cis and *s*-trans conformation are possible from the energetic point of view. The calculated barriers of 6 - 7 kcal·mol^{-1} suggest reasonably facile interconversion between the *s*-cis and *s*-trans forms.

However, after re-examination of the crystallographic data, for inhibitor **1** the *s*-cis conformation possesses a much higher occupancy (87%) than the originally modelled *s*-trans isomer (13%). For inhibitor **2**, the evidence provided by the electron density maps suggests the *s*-cis conformation is the preferred one.

Lastly, the process of manual rebuilding and re-refinement of the PI3Kγ structure resulted in a protein model with improved quality metrics, including R factor, R$_{free}$, and Ramachandran scores.

In this chapter, the focus has been on static models derived from X-ray crystallographic data. These single-structure models are very informative, however they do not adequately account for the inherently dynamic nature of proteins. In the next chapter, molecular dynamics simulations are used to explore the dynamic aspects of PI3Kγ binding.

4. Molecular dynamics

A major aim of this thesis is to examine the basis on which PI3Kγ distinguishes γ-selective inhibitors **1** and **2**, with a ~30-fold preference for the former. However, as discussed in chapter 3, for inhibitors **1** and **2** the *s*-cis and *s*-trans conformations are possible and should be carefully considered.

Modelling protein plasticity during the ligand binding process is essential to understanding the mechanism of the process. Therefore, the main goal of this chapter is to provide a dynamical insight into the subtle basis of molecular recognition by PI3Kγ. For this purpose, the first objective is to use molecular dynamics (MD) simulations to assess the stability of PI3Kγ complexed with the inhibitors. The protein-ligand systems explored are: PI3Kγ complexed with *s*-cis inhibitor **1**; PI3Kγ complexed with *s*-trans inhibitor **1**; PI3Kγ complexed with *s*-cis inhibitor **2**; and PI3Kγ complexed with *s*-trans inhibitor **2**. The method employed to produce the four separate but procedurally identical MD simulation trajectories, along with the evidence that validates the protocol used, is discussed below.

4.1   Methods

4.1.1   Software

The molecular dynamics simulations discussed in this work were performed using the AMBER 11 suite of programs[131] adopting the *ff99SB* force field[86] for the protein and the general AMBER force field (gaff)[87] for the ligands. Partial atomic charges for non-standard units (i.e. ligands) were derived using the *antechamber* module[132]. Gaussian 03[133] was used to generate the electrostatic potential required for the RESP[90, 134] fits.

The simulation models were built using xLEaP; subsequent minimisations were performed with *sander*; and the MD simulations were carried out using the GPU accelerated version of *pmemd*. Trajectory analysis was performed using *ptraj*. Graphical representations were produced using VMD v1.9[135], MOLDEN[136], Grace v5.1.22[137] and PyMOL v1.3[138].

4.1.2 Protein-ligand models

The protein models used throughout this work derive from the X-ray crystal structures of PI3Kγ complexed with inhibitor **1** and inhibitor **2** (PDB code 2a5u and 2a4z, respectively)[30]. As described in the previous chapter, the process of rebuilding and re-refinement of the PI3Kγ structure complexed with inhibitor **1** (from PDB code 2a5u) produced a protein model with improved quality metrics. This re-refined PI3Kγ model was used as a template for the construction of the initial structures for simulation.

As stated in Section 3.2, after re-examination of the crystallographic data, it was concluded that for inhibitor **1**, the *s*-cis conformation possesses a much higher occupancy (87%) than the originally modelled *s*-trans isomer (13%). Similarly, the evidence provided by the electron density maps suggests the *s*-cis conformation is preferred for inhibitor **2**. Given potential population of both conformations, we explore both *s*-cis and *s*-trans bound conformations of each inhibitor. To this end, four different protein/ligand systems needed to be prepared. These are: PI3Kγ complexed with *s*-cis inhibitor **1**; PI3Kγ complexed with *s*-trans inhibitor **1**; PI3Kγ complexed with *s*-cis inhibitor **2**; and PI3Kγ complexed with *s*-trans inhibitor **2**. For the systems involving inhibitor **1**, the structures were based on the re-refined PI3Kγ model described in Chapter 3. For the construction of the systems involving inhibitor **2**, the re-refined PI3Kγ model was superimposed *in silico* on top of the PI3Kγ structure complexed with inhibitor **2** (from PDB code 2a4z) and the necessary changes were made.

4.1.3 Structure preparation

The structure of the catalytic subunit of PI3Kγ (Ser144 to Leu1092) comprises four separately identifiable domains. These domains are: the Ras-binding domain (RBD), the C2 domain, the helical domain and the double-lobed catalytic domain (Figure 4.1). Only the catalytic domain (Thr726 to Leu1092) was used for molecular dynamic simulations. This approach allows consistency with previously published studies by Han and Zhang[45], Berndt *et al.*[38] and D'Abramo *et al.*[139]
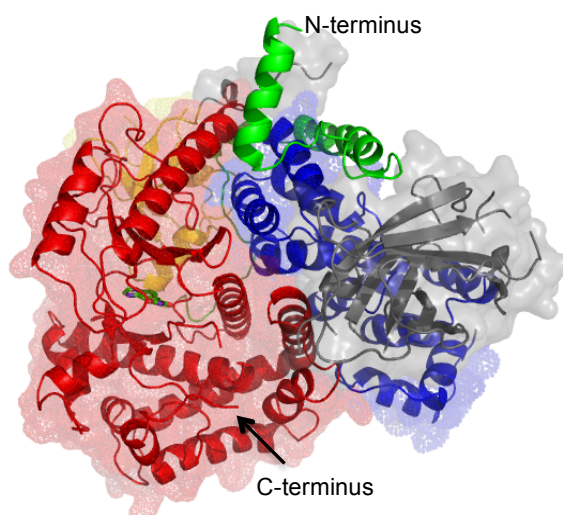
Figure 4.1: Cartoon representation of the structure of PI3Kγ complexed with *s*-cis inhibitor **1**. The domains are coloured as follows: Ras-binding domain in yellow; C2 domain in grey; helical domain in blue; catalytic domain in red; and the linker regions in green. The *s*-cis inhibitor **1** is shown as ball-and-stick thereby identifying the location of the active site.

As a first step in the structure preparation process, the PI3Kγ structure was truncated down to the Thr726 - Leu1092 region. The loop region encompassing residues 968-980 (a part of the activation loop) as well as residue 901 were not solved in the crystal structures. These lost loop regions were added with the software SYBYL 7.3[140] using homology models based on the crystal structure of porcine PI3Kγ (PDB code 1e7u) as templates. Dr. Willem Nissink, AstraZeneca, provided these homology models. It should be noted that the porcine and human enzymes have 95.3% overall sequence identity and complete identity at the ATP binding pocket[36].

The missing hydrogen atoms were then added with the software SYBYL 7.3[140]. Subsequently, to ensure that the added loops would not create instabilities, a BELLY minimisation was performed utilising *sander* module. This included 250 steps of steepest descent followed by 250 steps of conjugate-gradient minimisation where only the added segments were allowed to move. The resulting structures of the PI3Kγ catalytic domain with minimised added loops will be referred to in the following sections as initial PI3Kγ structures.

4.1.3.1 Choice of protonation and tautomeric states

The catalytic domain of PI3Kγ contains ten histidine residues whose protonation state varies according to the surrounding environment. On the basis of an examination of

their hydrogen bonding environment in the crystal structure, histidine residues 730, 834, 909, 948, 962, 967, 1005, 1023 and 1089 were assigned as being protonated on the ε nitrogen (HIE); while histidine residue 1022 was assigned as being protonated on the δ nitrogen (HID). No histidines were assigned as protonated on both nitrogens (HIP), i.e. positively charged. All of the eight cysteine residues (801, 817, 863, 869, 932, 936, 1013, 1075) were assumed to be neutral (CYS).

Similarly, the protonation state of the residues in the active site was determined from inspection of their local hydrogen bonding environment. Lysine 833 was assigned as protonated. Aspartic acid 964 and aspartic acid 841 were assigned as deprotonated. All other ionisable residues were set at their default protonation states at physiological pH.

As illustrated in Figure 4.2, the assigned protonation state of the ligands involved a neutral nitrogen at the 3-position of the thiazolidinedione moiety; and a protonated oxygen in position 2 or 4 of the thiazolidinedione moiety, for the *s*-trans and *s*-cis isomers, respectively. This assignment was based on three crucial considerations. These are: the pK_a values for the thiazolidinedione moiety, the possible tautomeric states presented by the inhibitors and the local hydrogen bonding environment within the active site. We discuss these below.
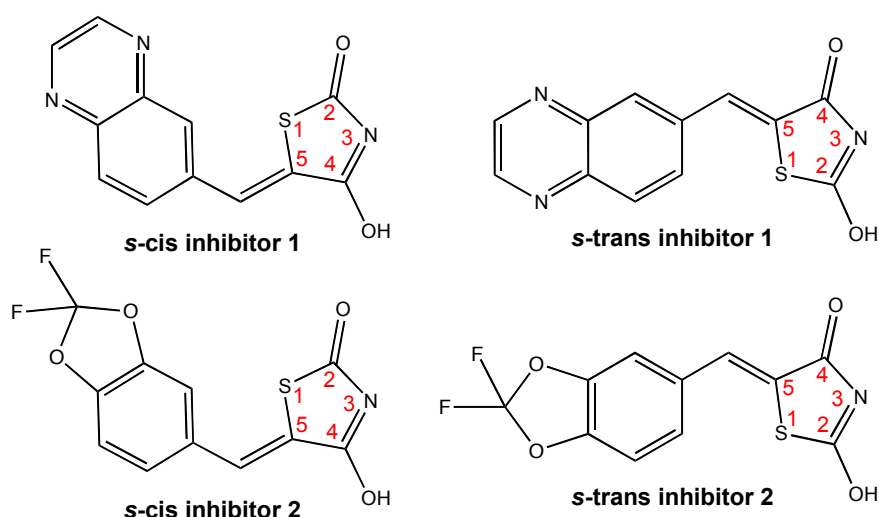
Figure 4.2: Tautomers of the *s*-cis and *s*-trans isomers for inhibitors **1** and **2**.

The first consideration was to estimate the acidity of the inhibitors and the hydrogen bond acceptor potential of their deprotonated forms. The pK_a values of the proton of the 3-N atom of the 2,4-thiazolidinedione moieties in inhibitors **1** and **2** were estimated to be 7.0 and 7.3, respectively. These values were calculated using ACD/I-lab, accessed through the Chemical Database Service[141].

On the other hand, the experimental pK_a (determined by conductivity measurements) of 2,4-thiazolidinedione (R=H, Figure 4.3) has been reported by Kanolt[142] to be 6.74 at 25 ºC. Here, it is interesting to note how the charge formed upon ionisation (in the conjugate base) is stabilised by resonance delocalisation (Figure 4.3).
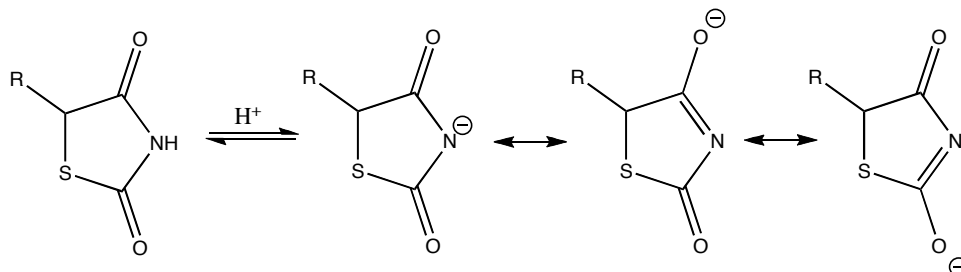
Figure 4.3: 2,4-thiazolidinedione derivatives and resonance stabilisation of the conjugated base.

A previous study by Pinson and co-workers[143] also estimated the pK_a values of the thiazolidinedione group. To this end, the authors evaluated a series of thirty-three thiazolidinedione derivatives and their sulfur analogues, rhodanines. The pK_a of the thiazolidinedione group was predicted to be between 6.42 and 8.44 (average of 7.44) using ACD Chemsketch software. Pinson and co-workers also report that, for their virtual screening study, the ionisation state of the thiazolidinedione moiety had a marked influence on the results. The protonated nitrogen series was more successfully retrieved than the deprotonated nitrogen series.

In a separate study by Pomel and co-workers[144] a series of furan-2-ylmethylene thiazolidinediones was investigated. In this case, the pK_a values of the thiazolidinedione group for ten compounds were calculated experimentally utilising UV spectroscopy. The reported values range between 7.3 and 9.7 (average of 8.7). In this study the authors expected the NH group to be deprotonated.

From the preceding discussion it is possible to appreciate a lack of consensus with regards to the protonation state of the ligand. Additionally, for the successful set-up of our MD simulations, it was also necessary to consider the tautomeric states presented by the inhibitors. Previous investigations, both theoretical[145, 146] and experimental[147, 148], into the structures of 2,4-thiazolidinedione derivatives revealed that these compounds exhibit a number of distinct prototropic tautomeric states. As exemplified in Figure 4.4, prototropic tautomerism[149] exists when the two isomers differ in the position of a proton. This phenomenon is frequently observed under biological conditions and the exact ratio of the tautomers depends on several factors, which

include temperature, solvent and pH[150-152]. Furthermore, the most stable tautomeric state at a given pH might change upon binding.
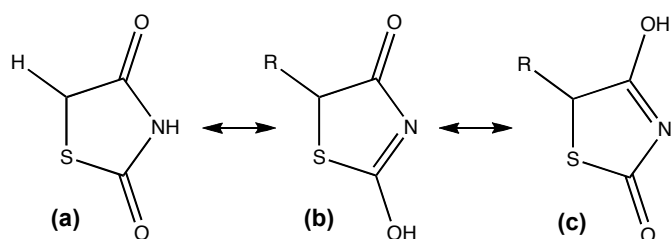
Figure 4.4: Three important tautomers of 2,4-thiazolidinedione derivatives.

It has been documented[152] that the formation of hydrogen bonds can considerably affect both the equilibrium and the rate of the tautomeric process. As such, the key distances and local hydrogen bonding environment within the active site were a decisive element for the determination of the protonation and tautomeric state of the ligands.

Figure 4.5 shows the distances within the active site for the PI3Kγ/*s*-cis inhibitor **1** system and the PI3Kγ/*s*-cis inhibitor **2** system. Notably, the thiazolidinedione nitrogen forms a hydrogen bond with the side chain of Lys833. The thiazolidinedione nitrogen atom was assumed to be deprotonated. Besides, one of the oxygen atoms in the thiazolidinedione head of the inhibitors is in close proximity with the negatively charged carboxylate oxygen of Asp964, the negatively charged carboxylate oxygen of Asp841 and the backbone amide oxygen of Ile963. This suggests two possibilities: that the aminocarbonyl oxygen of the inhibitor is protonated or that a carboxylate oxygen of Asp964 is protonated.

Ideally, it would be desirable to conduct a systematic investigation to evaluate the influence of the ligand/Asp964 protonation states on protein-ligand binding affinities. However, for the purpose of this investigation, it was decided to explore a single state with the protonated aminocarbonyl oxygen of the inhibitor and keeping a negatively charged carboxylate oxygen of Asp964 (i.e. as it appears in Figure 4.5).
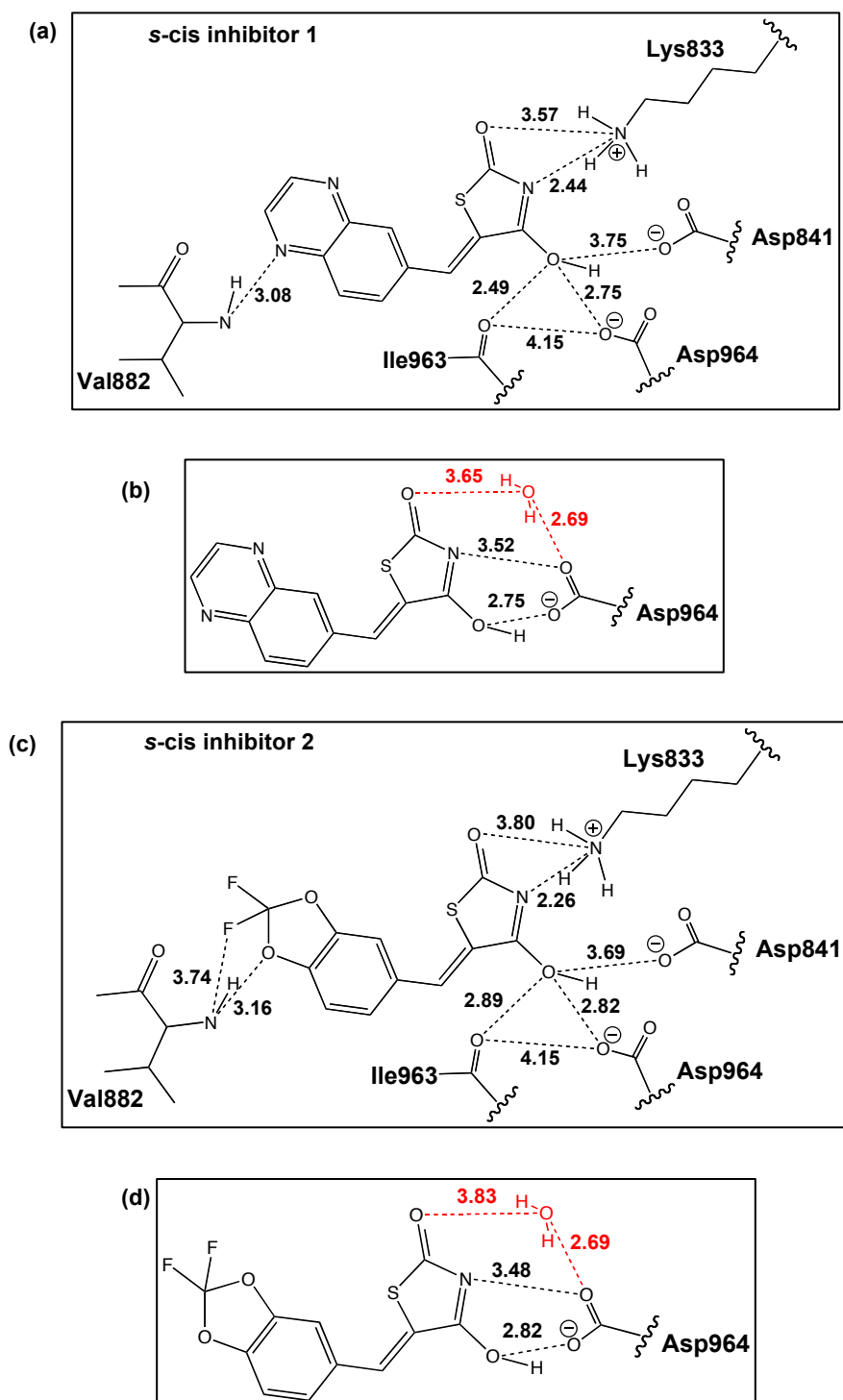
Figure 4.5: Schematic 2D representation of the key active site residues for the PI3Kγ/*s*-cis inhibitor **1** system (a) and (b) and the PI3Kγ/*s*-cis inhibitor **2** system (c) and (d). Panels (b) and (d) refer specifically to a water bridged interaction, which is shown separately for clarity.

The local hydrogen bonding environment for the PI3Kγ systems involving the *s*-trans inhibitors are presented in Figure 4.6. Again, the thiazolidinedione nitrogen was assigned as deprotonated, while the oxygen atom in the proximity of Asp841 and Asp964 was assigned as protonated.
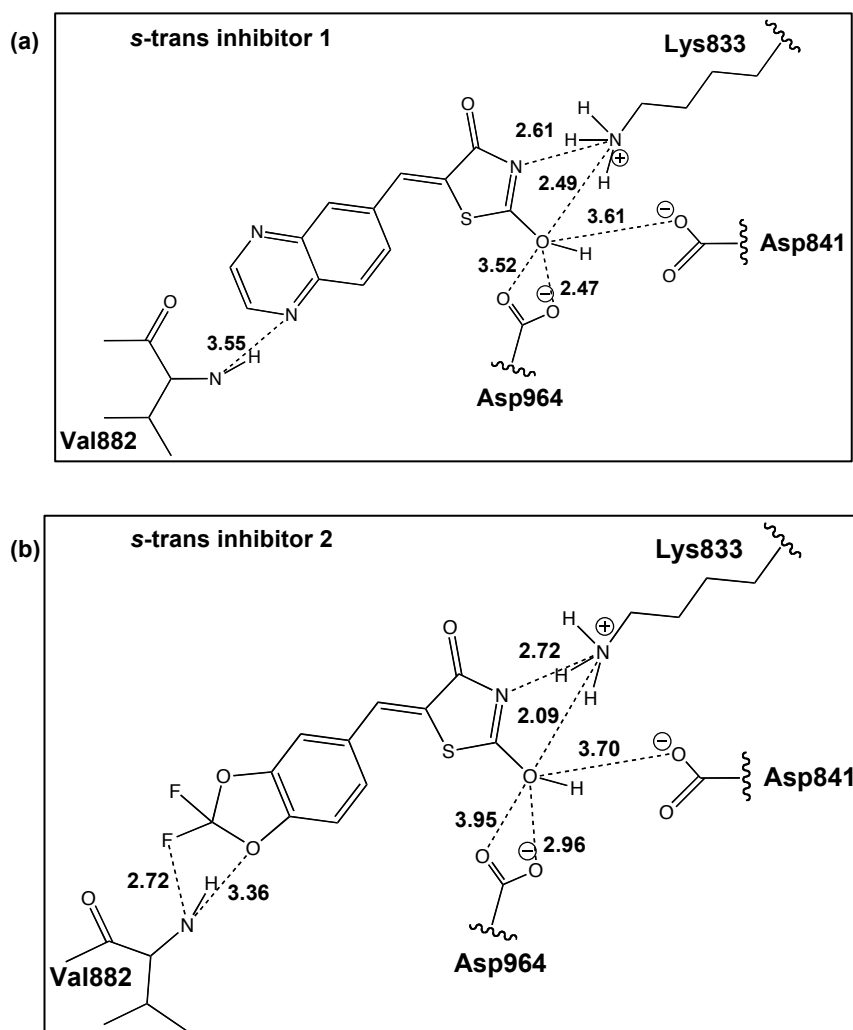
Figure 4.6: Schematic 2D representation of the key active site residues for the PI3Kγ/*s*-trans inhibitor **1** system (a) and the PI3Kγ/*s*-trans inhibitor **2** system (c) and (d).

4.1.3.2 Atom-centred partial point charge derivation for the ligands

Quantum mechanical calculations were performed on both *s*-cis and *s*-trans conformations of inhibitor **1** and **2** to derive the atom-centred partial point charges for use in MD simulations. After geometry optimisation (MP2/6-31G*), the electrostatic potential of each ligand was calculated at the HF/6-31G* level of theory using Gaussian03[133]. The atomic charges were then fitted to the electrostatic potentials using the Restrained Electrostatic Potential (RESP) method[90, 134] as implemented in the AMBER 10.0 *antechamber* module[132]. The calculated values for each inhibitor are shown in Figure 4.7.

s-cis inhibitor **1**
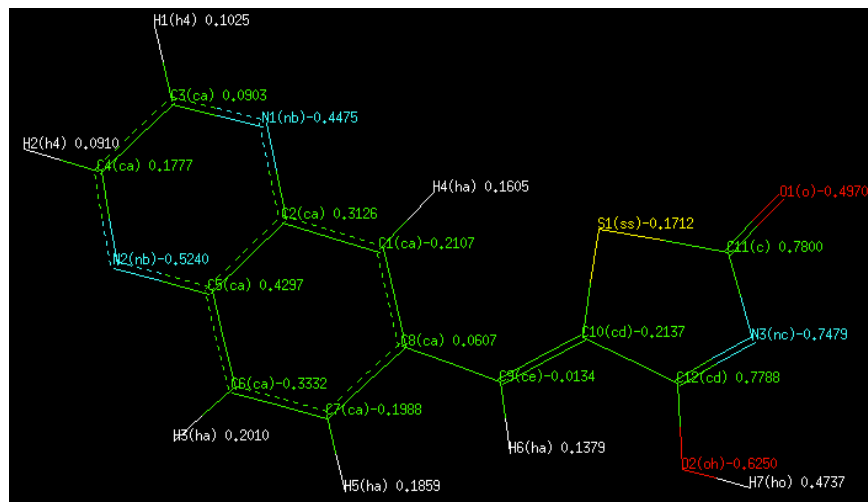
s-cis inhibitor **2**
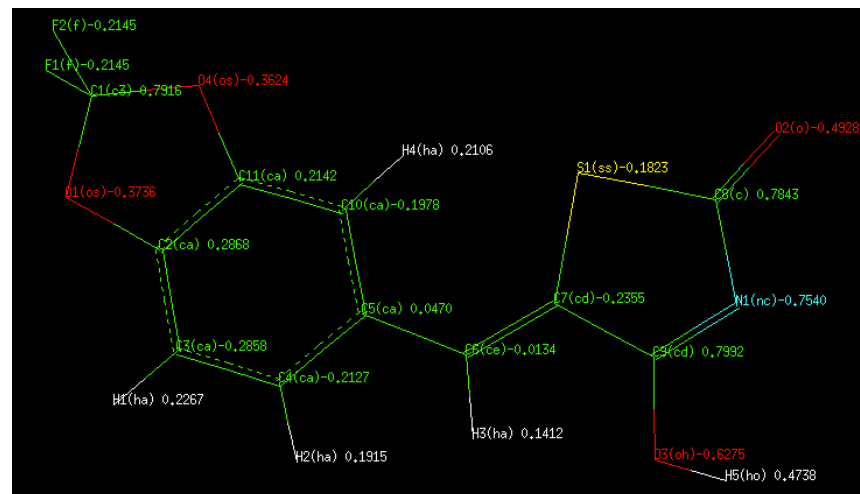
s-trans inhibitor **1**
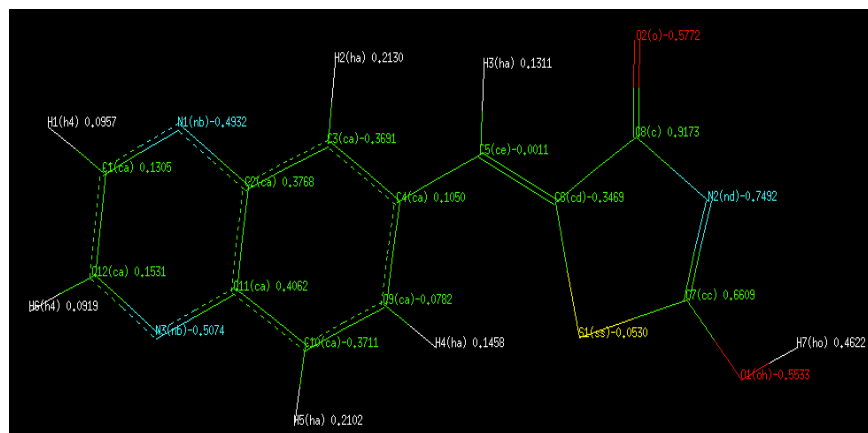
s-trans inhibitor **2**
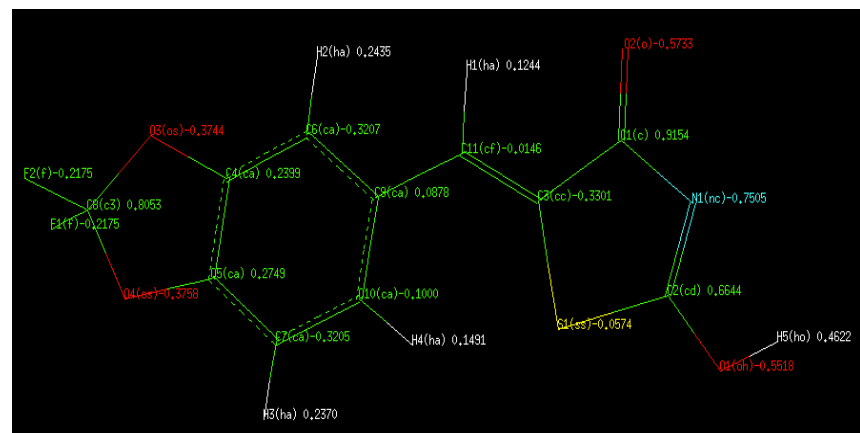
Figure 4.7: Atom names, types and charges for the different inhibitors as assigned by *antechamber*.

In the last stage of the structure preparation process, the systems were neutralised by the addition of five Na^+ ions. Each system was then solvated using the TIP3P water model[95] by placing it in a box with a margin of 8.0 Å along each dimension (Table 4.1).

4.1.4   Molecular dynamics protocol

Four separate but procedurally identical molecular dynamics (MD) simulations were performed on the PI3Kγ systems (Table 4.1) over a period of 160 ns each.

Table 4.1: Composition of each simulation system.

Simulation	Protein model	Added waters	Total number of atoms
I	PI3Kγ/*s*-cis inhibitor **1**	12,346	42,969
II	PI3Kγ/*s*-cis inhibitor **2**	14,108	48,254
III	PI3Kγ/*s*-trans inhibitor **1**	12,498	43,425
IV	PI3Kγ/*s*-trans inhibitor **2**	12,498	43,424

The protocol used for the MD simulations was as follows. Firstly, the system was subjected to minimisation of just the solvent molecules (500 steps of steepest descent followed by 2500 steps of conjugate gradient minimisation). During this restrained minimisation stage, the protein atoms and the ligand were kept fixed by an imposed harmonic force restraint of 500 kcal mol^{-1} Å^{-2}. Afterwards, in the second stage, 1000 steps of steepest descent and 4000 steps of conjugate gradient unrestrained minimisation were carried out to allow the whole system to relax.

During all subsequent simulations, the Particle Mesh Ewald method[92] was applied to calculate long range electrostatic interactions and the non-bonded interactions were truncated at 10.0 Å. All covalent bonds to hydrogen atoms were constrained using the SHAKE algorithm[153] and a simulation time step of 2.0 fs was used. Periodic boundary conditions were applied to all dynamics to avoid edge effects. The simulation temperature was controlled using the Langevin thermostat[104, 154] with a collision frequency of 5.0 ps^{-1}. The system density was controlled using a weak coupling barostat[106].

The prepared minimised structure was then subjected to a gradual equilibration procedure. The system was subjected to 20 ps of slow heating from 0 to 300 K,

followed by 30 ps heating at 300 K. Both stages were conducted in the canonical ensemble (with constant number of atoms, volume and temperature, NVT) and using harmonic restraints of strength 5.0 kcal mol^{-1} Å^{-2} on the protein and inhibitor. Subsequently, the system density was allowed to equilibrate to atmospheric pressure using the isothermal-isobaric ensemble (with constant number of atoms, pressure and temperature, NPT) for 100 ps at the same temperature. The applied harmonic restraints were then gradually reduced by 1 kcal mol^{-1} Å^{-2} every 20 ps until they were completely removed. The free system underwent a further 300 ps of unrestrained dynamics at constant pressure followed by 200 ps dynamics under constant volume conditions (NVT). This final structure (and velocities) was the initial configuration for production dynamics. Finally, a production run for 160 ns was performed in the canonical ensemble. For subsequent analyses, only the last 140 ns were considered.

For the systems involving *s*-cis inhibitor **1** and *s*-cis inhibitor **2**, an important water of crystallisation was identified in the active site as discussed in section 3.3 (see also Figures 4.5b and 4.5d). This crystallographic water, which was involved in a hydrogen bonding network between PI3Kγ residue Asp964 and the ligand, was restrained during the equilibration stage.

4.1.5   Analysis of the results

The generated MD trajectories were analysed qualitatively by visual inspection using the VMD program[135]. The extraction of quantitative information, such as hydrogen bond occupancies and interatomic distances, was performed using the *ptraj* module of AMBER 11[131].

The identification of protein-bound water molecules and of hydrogen-bonded networks in the active site was performed using water density maps created with the grid command of *ptraj*. The resulting grids were visualised as positive density using VMD[135].

4.2   Results and discussion

Following the equilibration protocol outlined above, molecular dynamics simulations in explicit aqueous solutions were run for 160 ns at 300 K. The composition of the different simulation systems can be found in Table 4.1. The stabilities of the different

systems during the production dynamics were confirmed by analysis of the time series of their kinetic, potential and total energies. Additionally, equilibration of each system was indicated by the preservation of approximately constant temperatures (data not shown). For subsequent analyses, only the last 140 ns of the production dynamics were considered.

The root mean square deviations (RMSD) in atomic position from the corresponding initial PI3Kγ structures are given in Figure 4.8. The black lines represent the values of the protein backbone atoms, whereas the red lines correspond to the heavy atoms of the ligands. As can be seen in Figure 4.8, the receptor and ligand in each system appear reasonably stable over the course of the trajectory, although there are gradual fluctuations, which may suggest protein flexibility (see below). The average RMSD values over the production dynamics are summarised Table 4.2, and range from 2.84 to 3.54 Å for the protein backbone atoms; and from 0.29 to 0.43 Å for the heavy atoms of the ligands.

(I)

(II)

(III)

(IV)
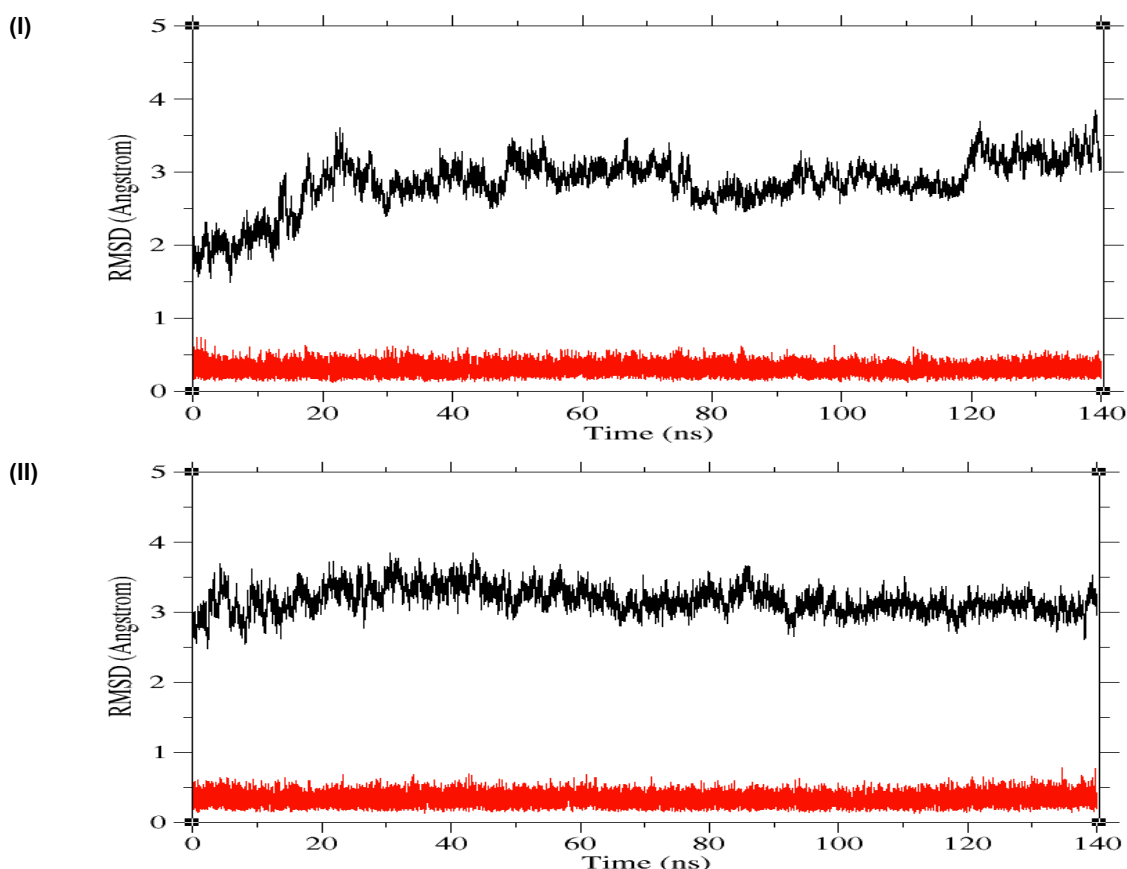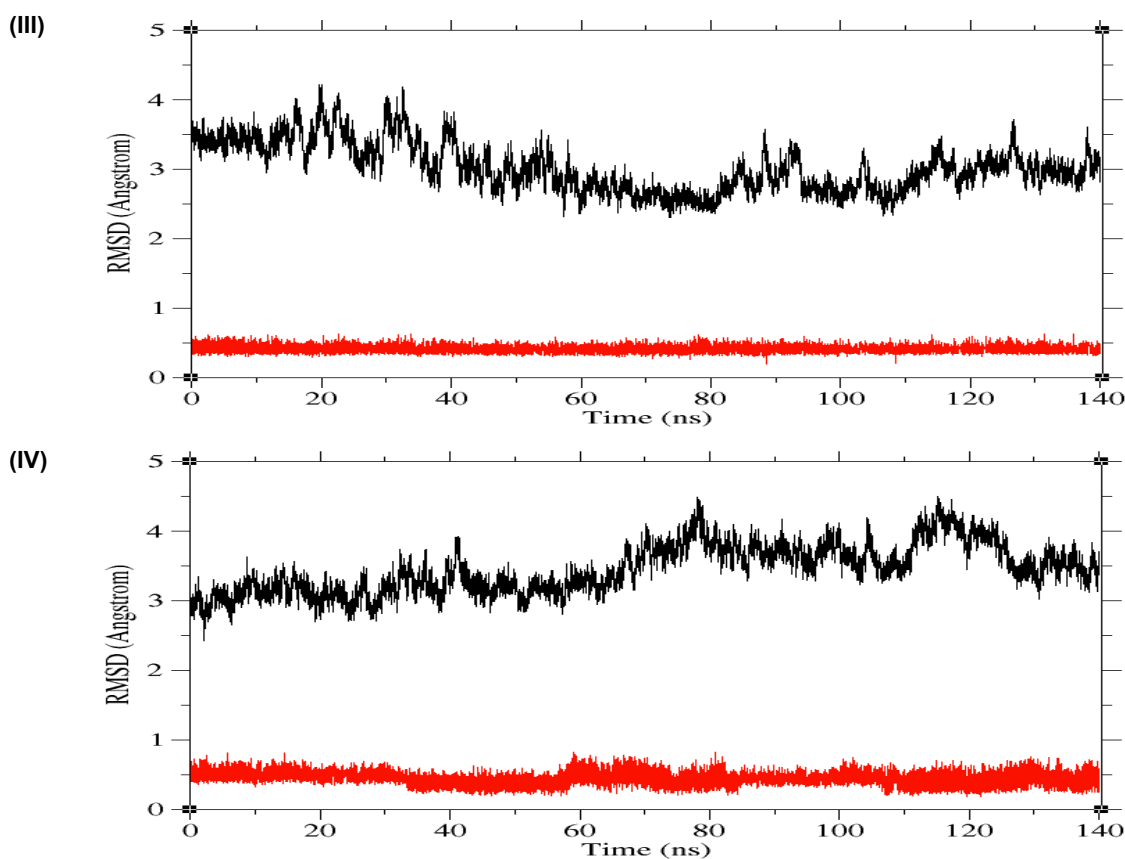
Figure 4.8: Time dependence of the RMSD of protein backbone atoms (black) and the heavy atoms of the ligand (red) throughout the production dynamics for simulations **I** to **IV**.

Interestingly, the systems including inhibitor **1** (i.e. simulations **I** and **III**) show slightly lower average RMSD values for the receptor than the corresponding systems containing inhibitor **2**. On the other hand, the RMSD values for the heavy atoms of the ligand in systems **I** and **VI** appear to be the same within error.

Table 4.2: Average RMSD values throughout the production dynamics for protein backbone atoms and the heavy atoms of the ligand. The standard deviations are in parentheses.

Simulation	Ligand	RMSD (Å)	
		Receptor	Ligand
I	s-cis **1**	2.84 (0.34)	0.29 (0.07)
II	s-cis **2**	3.17 (0.18)	0.30 (0.07)
III	s-trans **1**	3.02 (0.35)	0.39 (0.04)
IV	s-trans **2**	3.45 (0.34)	0.43 (0.08)

4.2.1 Protein flexibility

To assess the dynamic stability of the enzyme, root mean square fluctuations (RMSF) were calculated. The RMSF reflects the mobility of a certain residue around a reference position averaged over a period of time. In this case, the reference positions correspond to the initial PI3Kγ structures. For each system, the mass-weighted RMSF values of each residue averaged over the 140 ns trajectory are given in Figure 4.9.
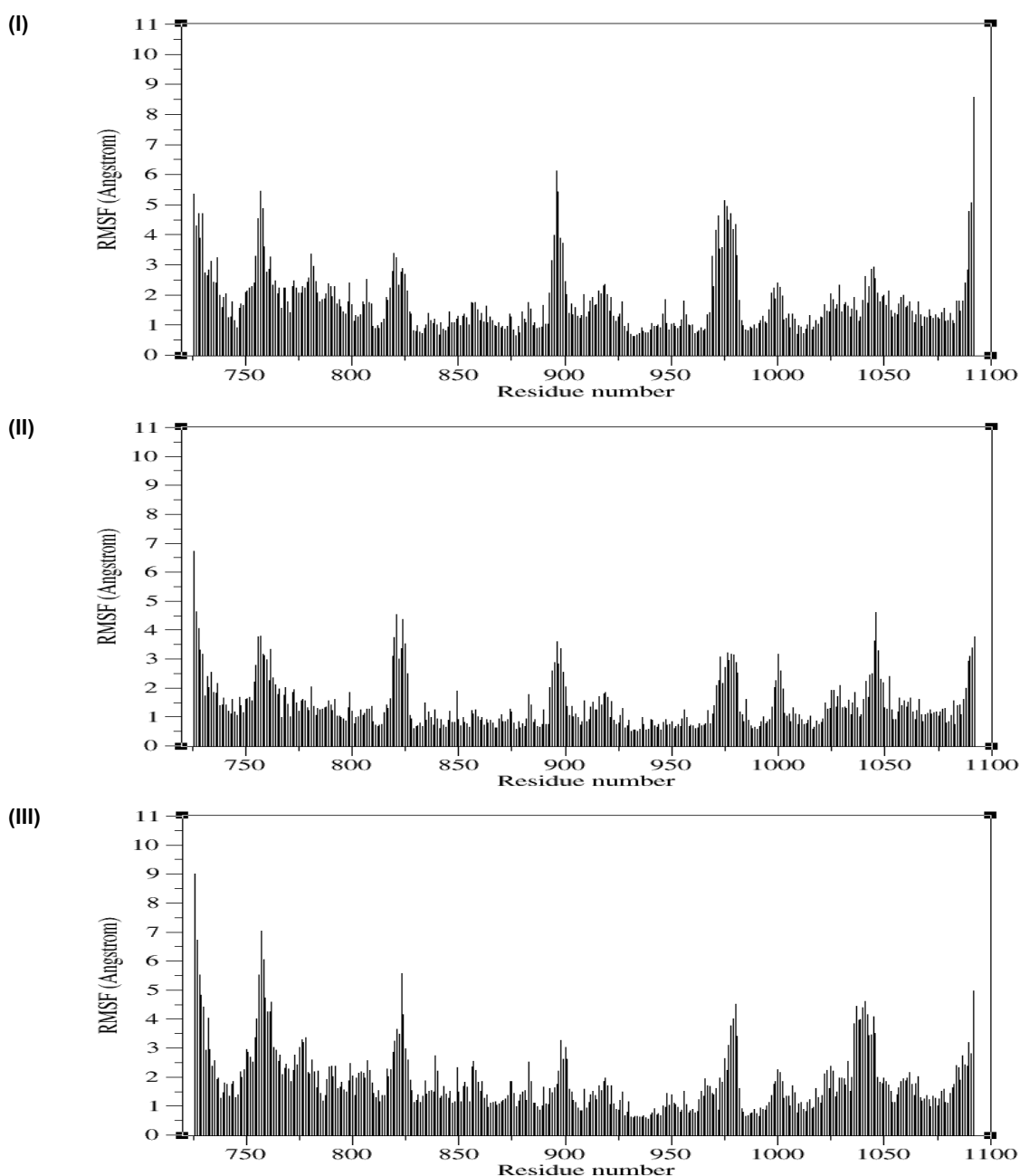
(continued on the next page)
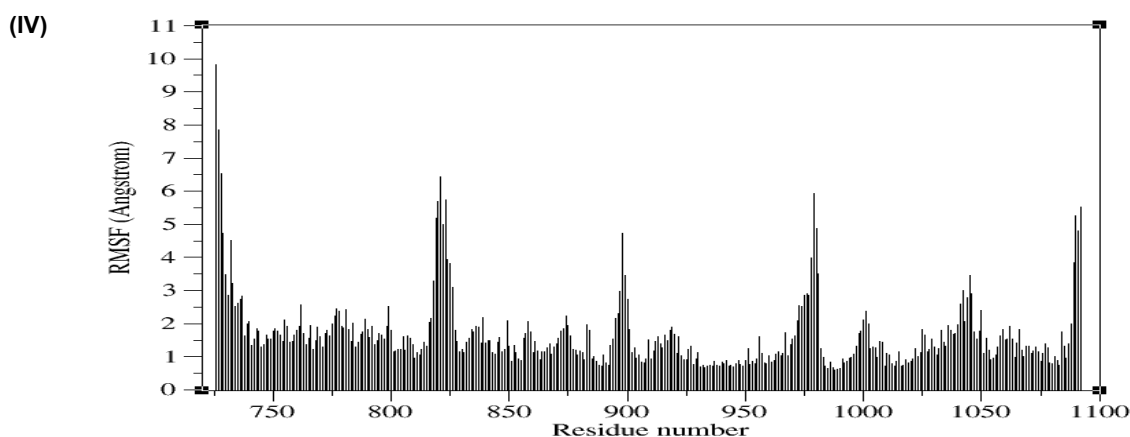
Figure 4.9: Root mean square fluctuations (RMSF) of the protein residues throughout the production dynamics for simulations **I** to **IV**.

Table 4.3: PI3Kγ regions presenting large flexibility during the production dynamics.

Residues	Region
752-760	kα1-kα2 loop
774-781	kα2-kβ1 loop
818-827	kβ4-kβ5 loop
895-904	kα4-kα5 loop
968-982	Part of the activation loop (between kβ10 and kα7)
996-1005	kα8-kα9 loop
1039-1045	kα9-kα10 loop

The results presented in Figure 4.9 indicate that large fluctuations of residues occurred in loops connecting secondary structure elements. Such residues (listed in Table 4.3) belong to the surface regions of the protein, which are exposed to the solvent. Hence, it is not surprising that the regions have a large RMSF value. To better understand the RMSF profiles, the labelled key elements of PI3Kγ secondary structure are presented in Figure 4.11.

A part of the activation loop (residues 968 to 980) as well as residue 901 are missing in the published X-ray crystal structures, which further suggest the conformational flexibility of such regions. In fact, as discussed in Section 1.1.1, in the various crystallised forms of PI3Kγ, the activation loop is either missing or only partially observed[43].

It is worth noting that the RMSF profiles presented in this section (Figure 4.9) are consistent with those reported by D'Abramo *et al.*[139] (Figure 4.10) who used explicitly solvated MD simulations to explore the conformational space of the apo form and the L64-cocrystallised holo form of PI3Kγ (starting from X-ray crystal structures 1E8Y and 3IBE, respectively). Their simulations of the apo form and the L64-cocrystallised holo form of PI3Kγ lasted 1.5 µs and 1 µs, respectively. As can be seen in Figure 4.10, large RMS fluctuations (given here in nm) occurred in loops connecting secondary structure elements, such as the kβ4-kβ5 loop (residues 818-827), the activation loop (residues 964-989) and the kα9-kα10 loop (residues 1039-1045).
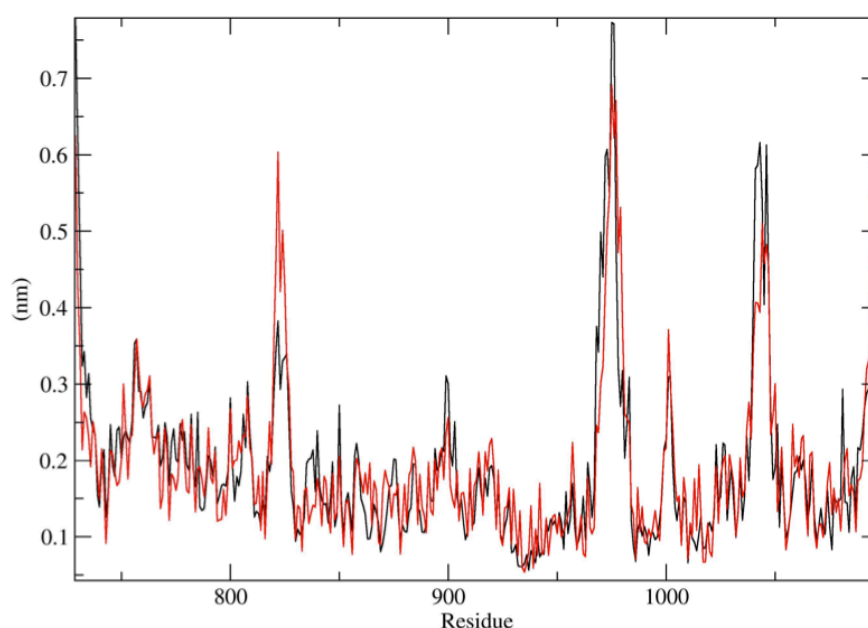
Figure 4.10: Per residue root mean square fluctuations (RMSF in nm) of the apo simulation (shown in black) and the L64-cocrystallised holo simulation (shown in red) of PI3Kγ as reported by D'Abramo and co-workers[139].

To further assess the overall conformational flexibility of the simulation systems, 35 equispaced snapshots were taken from each MD trajectory. The superimposed snapshots for each of the four systems are shown in Figures 4.12 and 4.13. Each representation consists of 35 structures produced at 4 ns intervals over the course of the 140 ns production runs. The functionally important regions of the enzyme, which are listed in Table 4.4, are colour-coded to allow structural differences to be more noticeable.
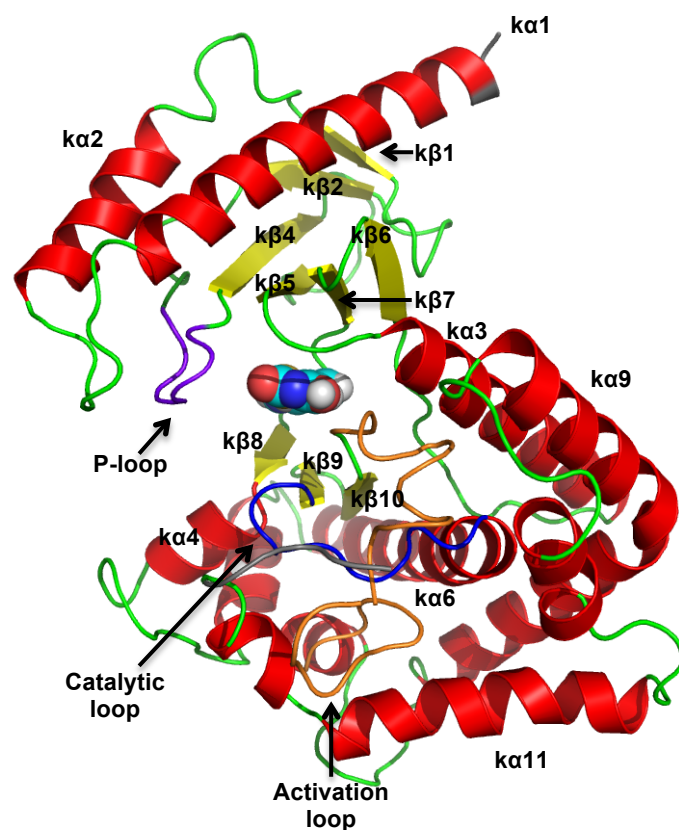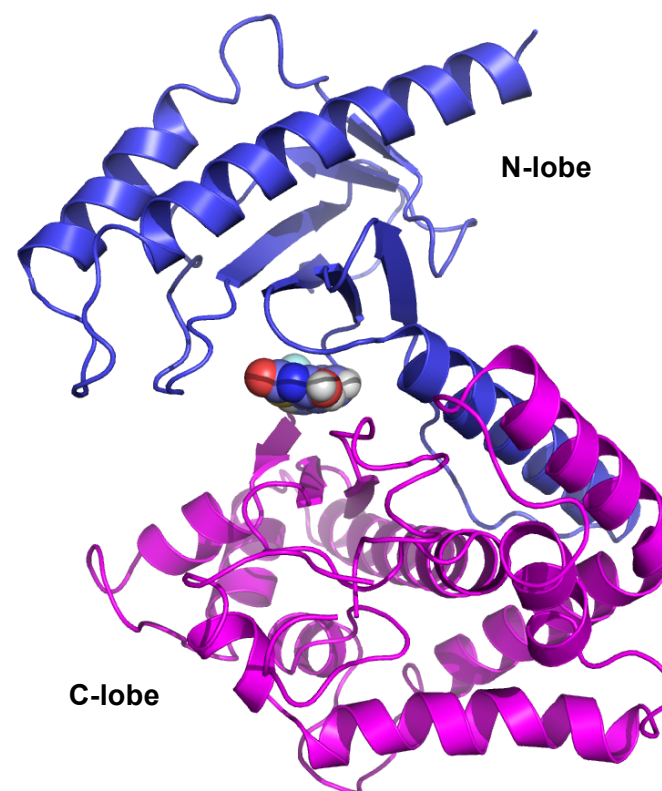
Figure 4.11: Cartoon representation of the structures of PI3Kγ complexed with *s*-cis inhibitor **1** (left) and *s*-cis inhibitor **2** (right). The structure on the left is coloured by secondary structure with N- and C-termini in grey; the P-loop in purple; the catalytic loop in blue and the activation loop in orange. The structure on the right is coloured by lobe, with the N-lobe (residues 726-883) in purple and the C-lobe (residues 884-1092) in magenta. The ligands are shown as spheres thereby identifying the location of the active site.
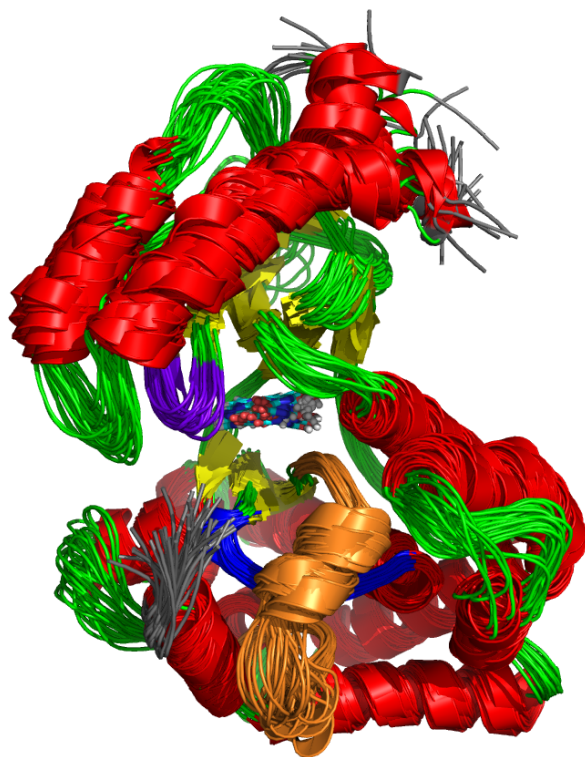
Simulation I: PI3Kγ/*s*-cis inhibitor **1**   **Simulation II:** PI3Kγ/*s*-cis inhibitor **2**
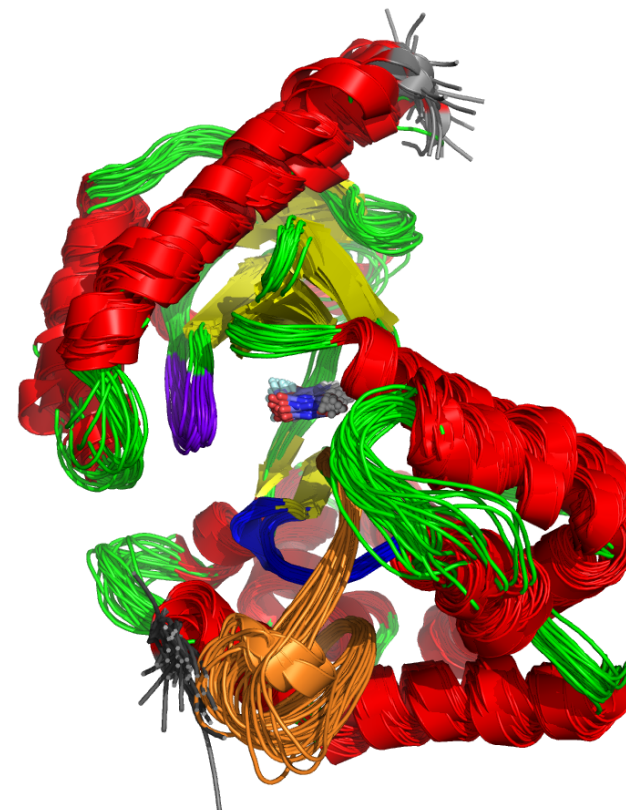
Figure 4.12: Snapshots of the 35 equispaced MD conformations for simulations **I** (left) and **II** (right). The cartoon representations are coloured by secondary structure with N- and C-termini in grey; the P-loop in purple; the catalytic loop in blue; and the activation loop in orange. The ligands are shown as sticks coloured by atom type.

Simulation III: PI3Kγ/*s*-trans inhibitor **1**
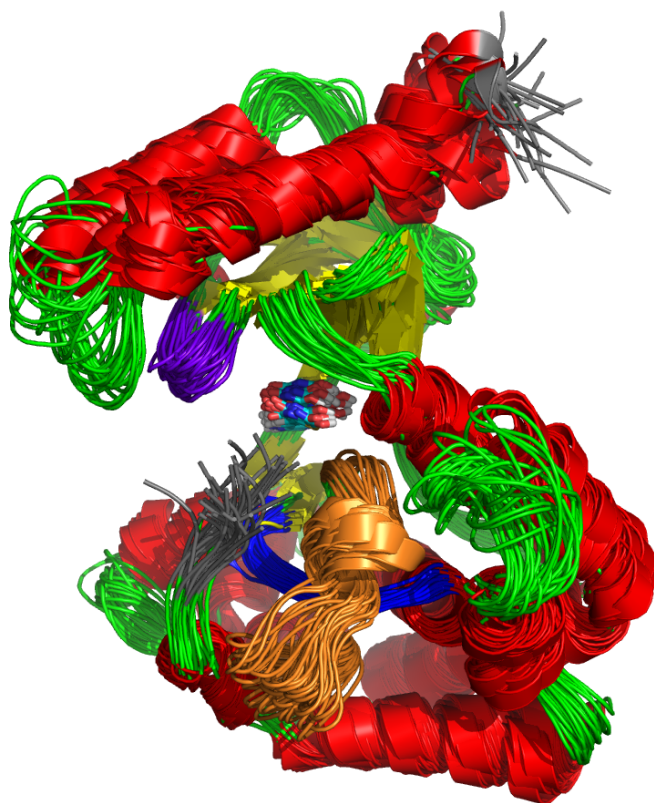
Simulation IV: PI3Kγ/*s*-trans inhibitor **2**
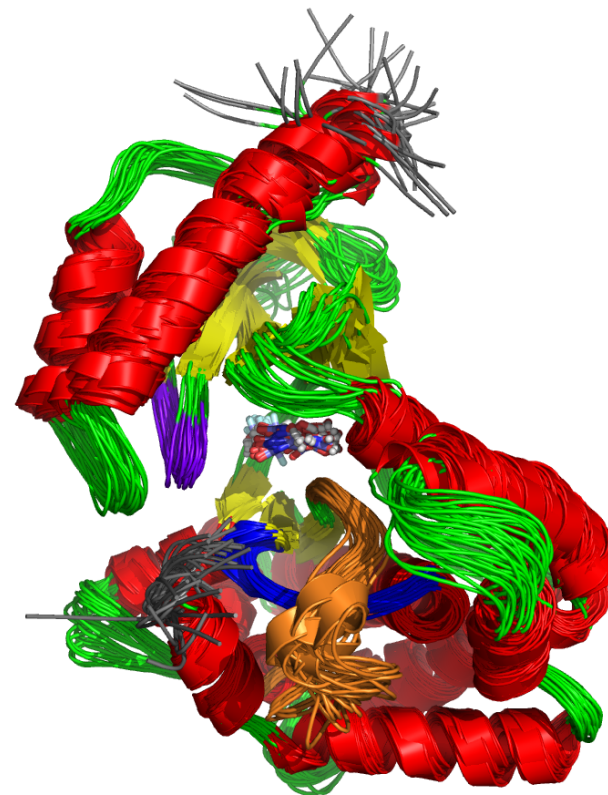
Figure 4.13: Snapshots of the 35 equispaced MD conformations for simulations **III** (left) and **IV** (right). The cartoon representations are coloured by secondary structure with N- and C-termini in grey; the P-loop in purple; the catalytic loop in blue; and the activation loop in orange. The ligands are shown as sticks coloured by atom type.
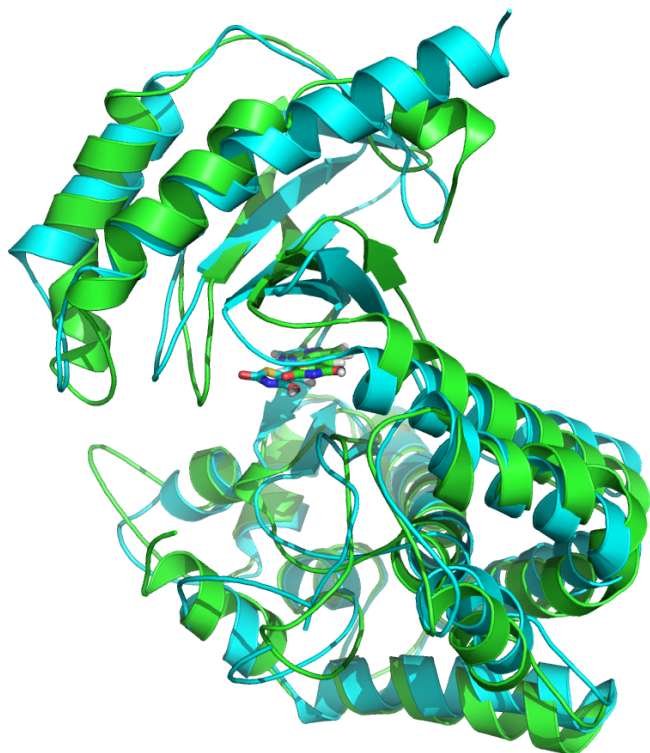
Simulation I: PI3Kγ/*s*-cis inhibitor **1**   **Simulation II:** PI3Kγ/*s*-cis inhibitor **2**
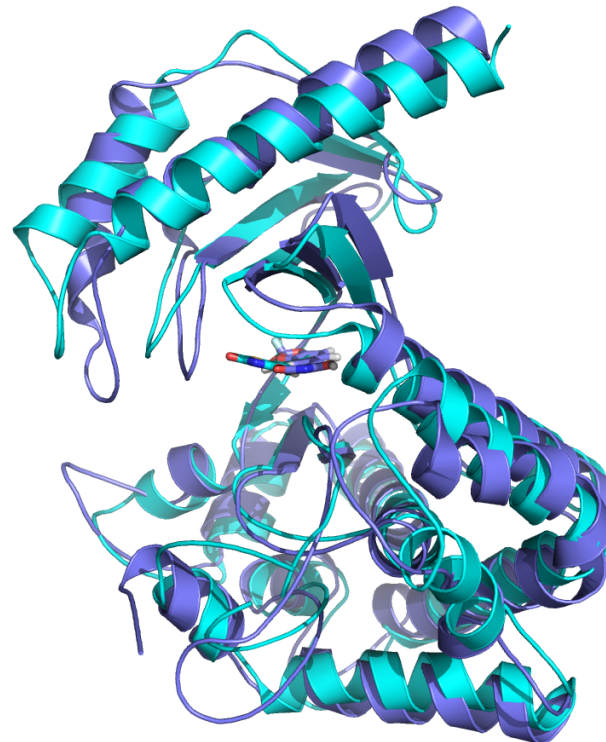
Figure 4.14: Representative structures from the MD trajectories superimposed to their corresponding starting structures. Left: PI3Kγ/*s*-cis inhibitor **1** system, the starting structure is shown in cyan and the representative structure from simulation **I** is shown in green. Right: PI3Kγ/*s*-cis inhibitor **2** system, the starting structure is shown in cyan and the representative structure from simulation **II** is shown in purple.
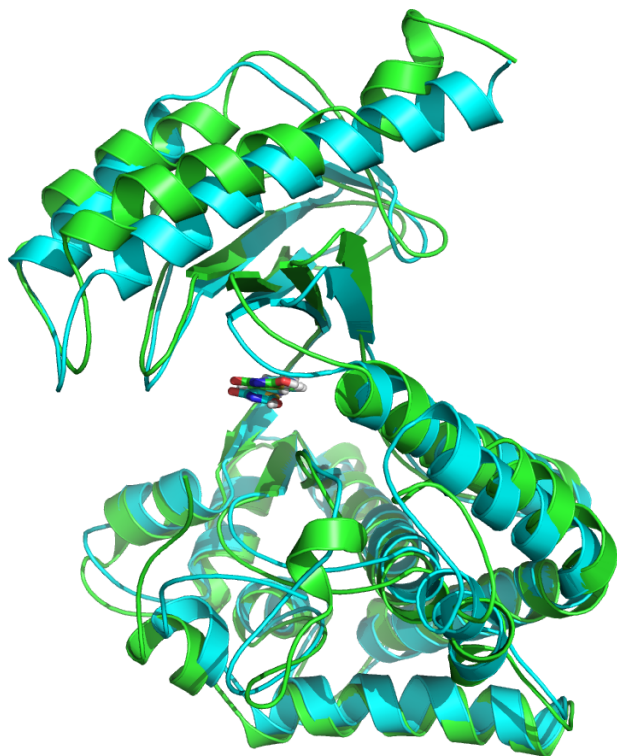
Simulation III: PI3Kγ/*s*-trans inhibitor **1**   **Simulation IV:** PI3Kγ/*s*-trans inhibitor **2**
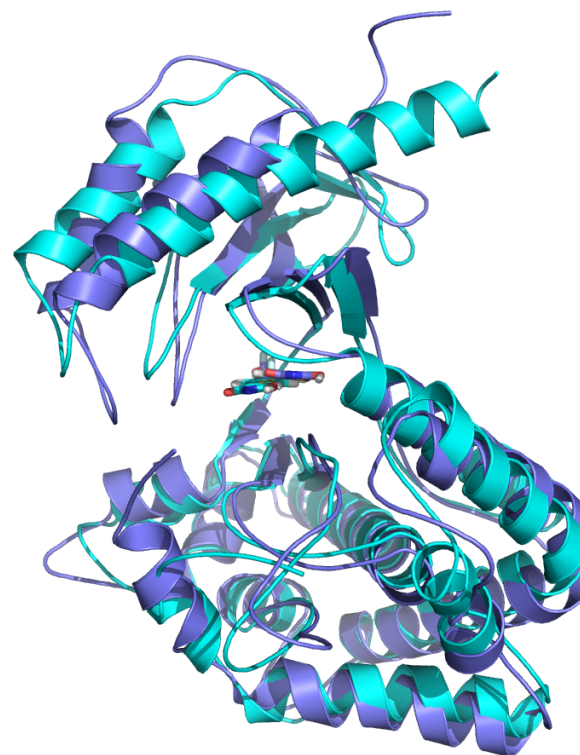
Figure 4.15: Representative structures from the MD trajectories superimposed to their corresponding starting structures. Left: PI3Kγ/*s*-trans inhibitor **1** system, the starting structure is shown in cyan and the representative structure from simulation **III** is shown in green. Right: PI3Kγ/*s*-trans inhibitor **2** system, the starting structure is shown in cyan and the representative structure from simulation **IV** is shown in purple.

Table 4.4: Definition of functionally important regions of PI3Kγ.

Region	Residues
P-loop	803-809
Helix αC (or kα3)	839-858
Catalytic loop	940-951
Activation loop	964-989
DFG motif	964-966

The superimposed snapshots confirm the mobility of the loops connecting secondary structure elements as well as the flexibility displayed by the N- and C- termini. This is in accordance with the calculated RMSF values presented in Figure 4.9. An interesting observation is that the protein structures of all systems share similar RMSF distribution and similar trends of dynamic features. This may suggest all four inhibitors/conformations have similar modes of interaction with PI3Kγ.

Lastly, to assess the degree of conformational relaxation from the initial crystal structures, Figures 4.14 and 4.15 show a superimposition of the starting structures with a representative structure from the simulation trajectory. Here, the representative structure corresponds to the snapshot closest to the average structure over all the MD configurations. Again, it is possible to appreciate the mobility of the loops connecting secondary structure elements.

4.2.2   Structural changes in the PI3Kγ active site

The re-refined X-ray structures of the four inhibitors and PI3Kγ revealed a similar binding mode (see Figures 4.16 to 4.19, panel a). The inhibitors are stabilised in a cavity formed by residues Met804, Ile831, Lys833, Asp841, Tyr867, Ile879, Glu880, Ile881, Val882, Met953, Phe961, Ile963, Asp964, Phe965 and Gly966. In the hinge region of the active site, the inhibitors form hydrogen bonds with the backbone NH of Val882. Additionally, the ligand thiazolidinedione moiety (also referred here as the head of the inhibitor) forms hydrogen bond interactions with the positively charged side chain of Lys833 and the negatively charged side chains of Asp964 and Asp841.

Hydrophobic interactions were formed between the phenyl and furan rings of the inhibitors and several hydrophobic residues located in the hydrophobic pockets I and II. The hydrophobic region I (also known as affinity pocket) is surrounded by residues

Tyr867, Ile879, Ile963 and Asp841; and the hydrophobic region II (or adenine pocket) comprises residues Ile831, Met953 and Phe961[23]. The gatekeeper residue, Ile879, controls the access to the hydrophobic pocket I. The different regions of the ATP-binding pocket are summarised in Table 4.5.

Table 4.5: Definition of the different regions of the active site of PI3Kγ.

Region	Residues
Specificity pocket	Met804, Trp812
Hinge	Val882, Glu880
Hydrophobic region I	Tyr867, Ile879, Ile963, Asp841
Gatekeeper	Ile879
Hydrophobic region II	Ile831, Met953, Phe961
DFG motif	Asp964, Phe965, Gly966

To examine structural changes at the active site, representative snapshots taken from each simulation trajectory were compared with their corresponding initial structure before MD. The aligned structures are presented in Figures 4.16 to 4.19. Note that the representative structure corresponds to the snapshot closest to the average structure over all the MD configurations.

For the four protein-inhibitor systems, it can be seen that overall the MD snapshots are in reasonable agreement with the starting structures (Figures 4.16 to 4.19). The hydrogen bond interactions to Lys833, Asp841 and Val882 were maintained very well. Moreover, the cavity lined by the hydrophobic residues was preserved during the trajectories. This suggests the high RMSD values observed earlier are due to loop flexibility, but the active site itself is on the whole well reproduced by MD. The hydrogen bond occupancies and average distances for each simulation model will be discussed in detail in section 4.2.3. Additionally, the water-mediated interactions with the protein will be discussed in section 4.2.4.
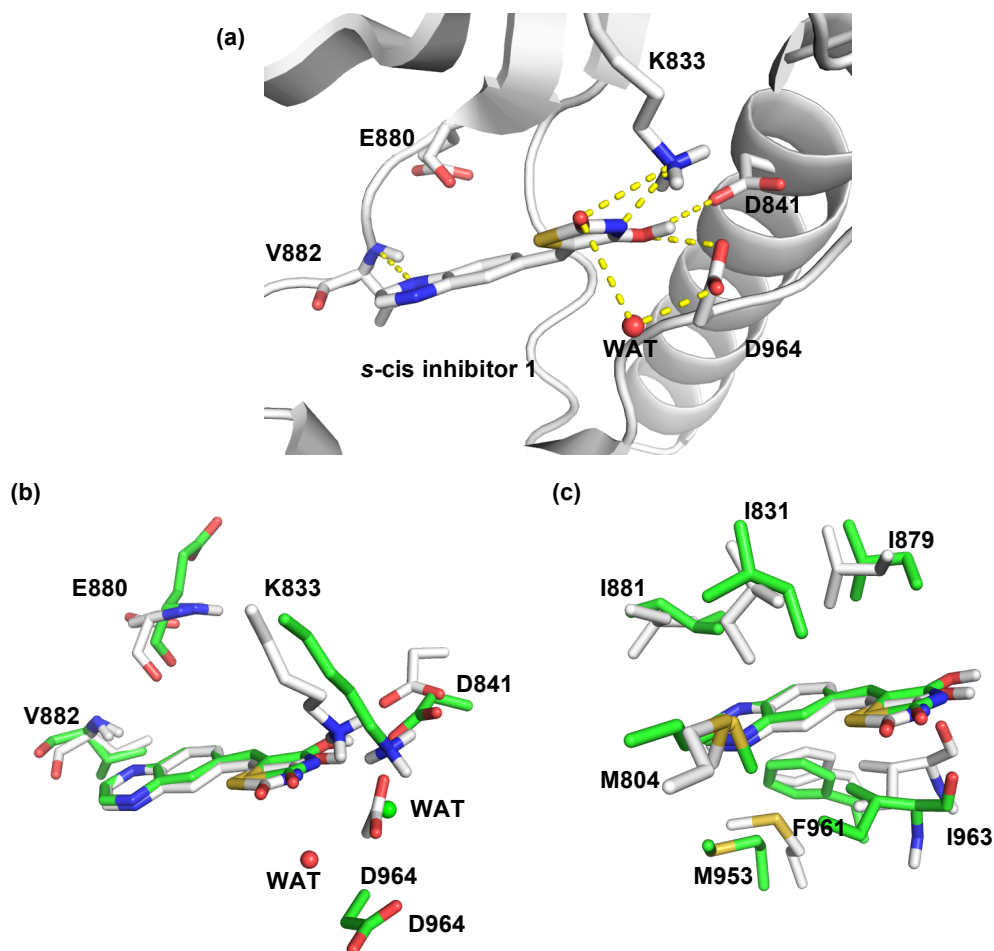
Figure 4.16: Representation of the active site of PI3Kγ complexed with *s*-cis inhibitor **1.** (a) Initial structure. The water molecule is shown as a red sphere. The yellow dashed lines indicate the key hydrogen bond interactions. (b) and (c) Structure comparison between initial (white) and representative snapshot from simulation **I** (green). (b) Shows the hydrogen bonding residues, while (c) shows the hydrophobic interactions.
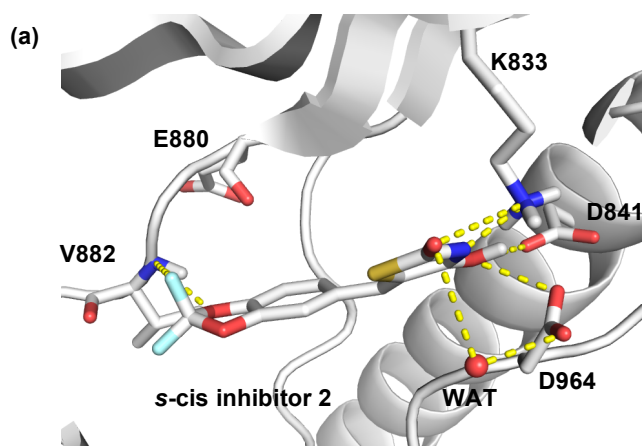
(continued on the next page)
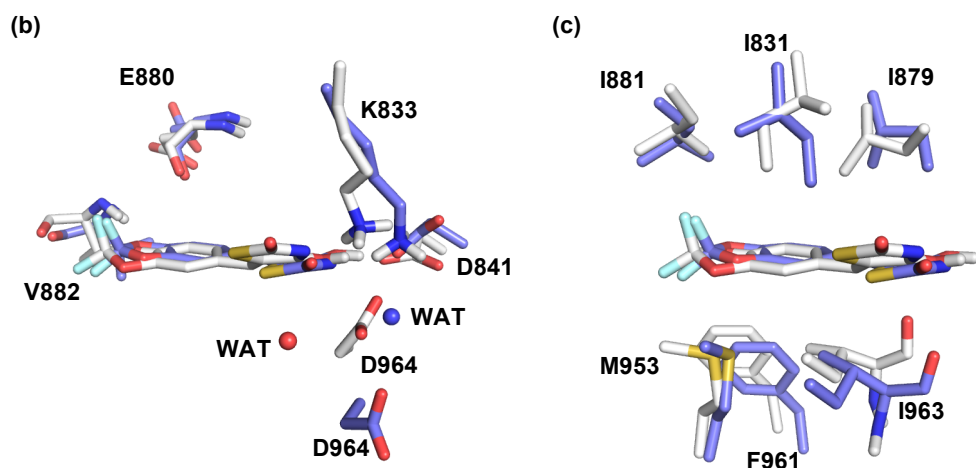
Figure 4.17: Representation of the active site of PI3Kγ complexed with *s*-cis inhibitor **2.** (a) Initial structure. The water molecule is shown as a red sphere. The yellow dashed lines indicate the key hydrogen bond interactions. (b) and (c) Structure comparison between initial (white) and representative snapshot from simulation **II** (purple). (b) Shows the hydrogen bonding residues, while (c) shows the hydrophobic interactions.
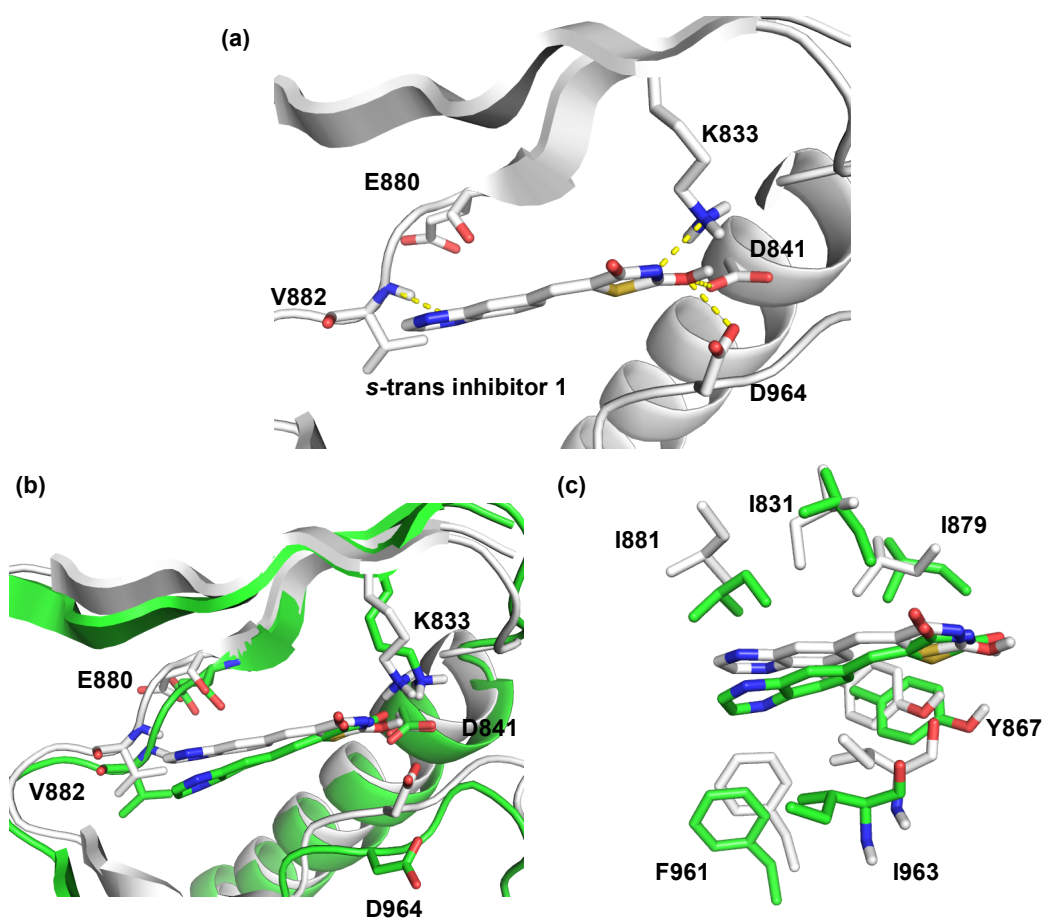
Figure 4.18: Representation of the active site of PI3Kγ complexed with *s*-trans inhibitor **1.** (a) Initial structure. The yellow dashed lines indicate the key hydrogen bond interactions. (b) and (c) Structure comparison between initial (white) and representative snapshot from simulation **III** (green). (b) Shows the hydrogen bonding residues, while (c) shows the hydrophobic interactions.
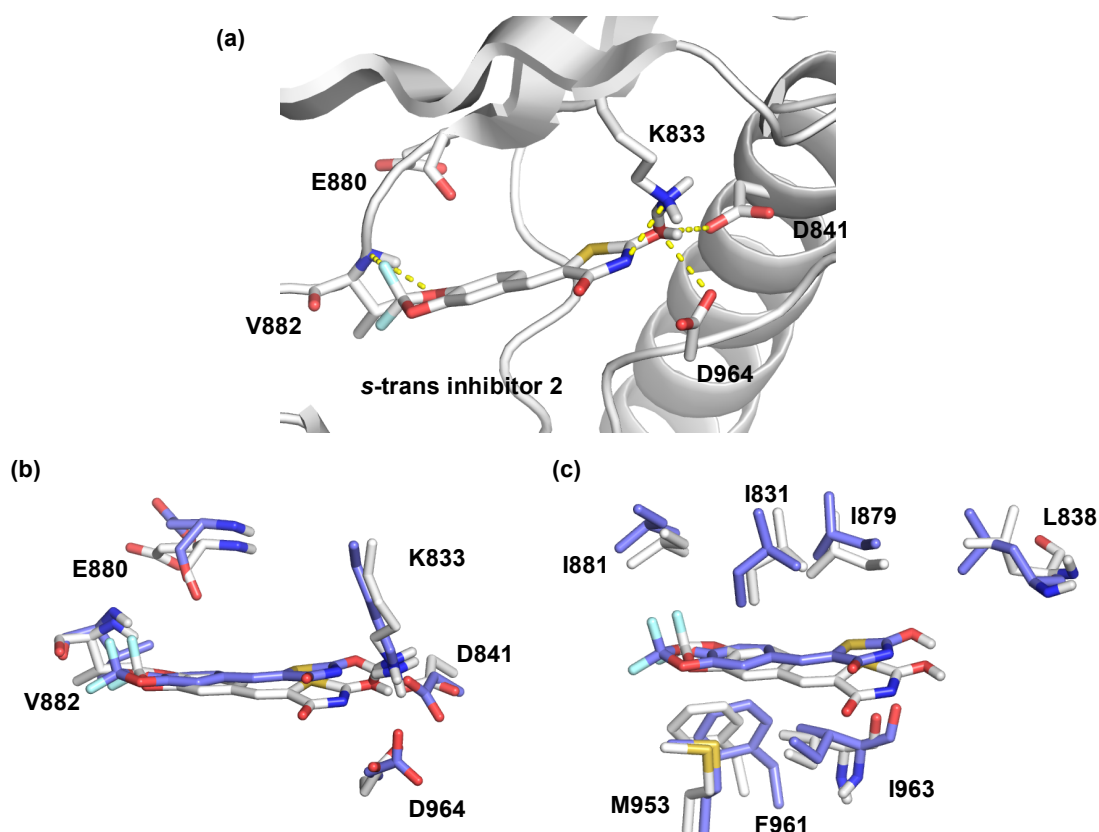
Figure 4.19: Representation of the active site of PI3Kγ complexed with *s*-trans inhibitor **2.** (a) Initial structure. The yellow dashed lines indicate the key hydrogen bond interactions. (b) and (c) Structure comparison between initial (white) and representative snapshot from simulation **IV** (purple). (b) shows the hydrogen bonding residues, while (c) shows the hydrophobic interactions.

A distinct difference between the representative MD snapshots and their corresponding starting structures is the location of residue Asp964. In the starting structure, the carboxylate oxygen atom of Asp964 forms a hydrogen bond with the protonated oxygen of the ligand thiazolidinedione moiety. However, during the MD trajectories, Asp964 swings away from the ligand. It is important to note that Asp964 forms part of the DFG motif, which lies at the beginning of the activation loop (Section 1.1.2). Moreover, in the X-ray structure of PI3Kγ complexed with ATP (PDB code 1e8x), the role of Asp964 is to coordinate two ATP-bound magnesium ions (refer to Figure 1.6).

To better understand the mobility of Asp964, the initial structures of each protein/ligand complex were compared to three snapshots taken from each MD trajectory at 40 ns, 80 ns and 120 ns. Each representation shows the side chains of the DFG motif residues, the activation loop and the αC helix. These are presented in Figures 4.20 to 4.23, panel a. To further assess the motions in the binding pocket, the distance between the protonated oxygen atom of the inhibitors thiazolidinedione

moiety and the center of mass of the side chain atoms of Asp964 was monitored and is also presented in Figures 4.20 to 4.23. Furthermore, the values of the Asp964 dihedral angles φ, ψ and χ_1 were monitored during the simulation (Figure 4.24).
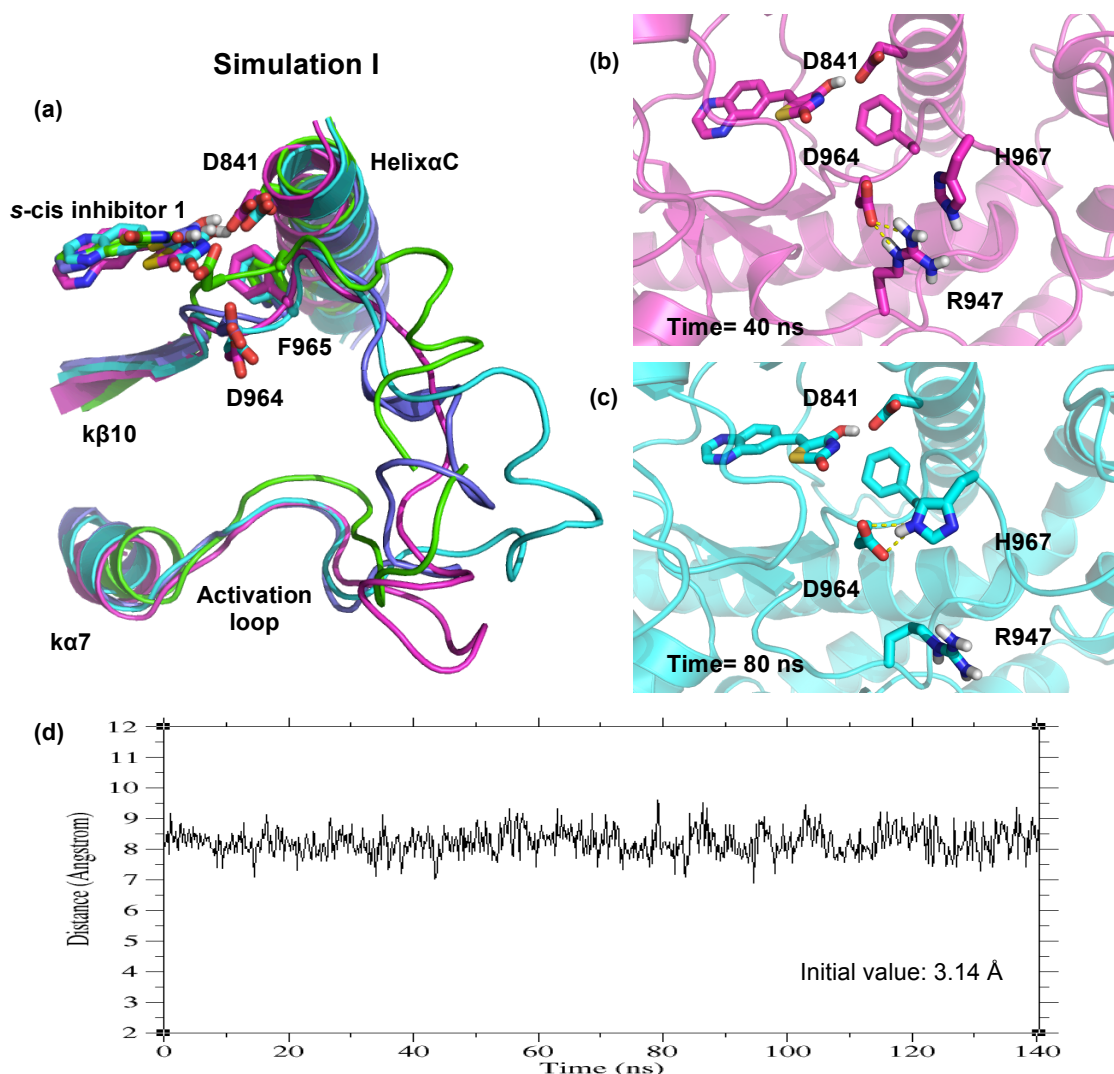
Figure 4.20: Asp964 movement for simulation **I**. (a) View of the DFG-motif and the activation loop. Representative structures of the initial structure (shown in green) and MD snapshots taken at 40 ns (magenta), 80 ns (cyan) and 120 ns (purple). (b) View of the active site at time 40 ns showing the hydrogen bond interactions between Asp964 and Arg947. (c) View of the active site at time 80 ns showing the hydrogen bond interactions between Asp964 and His967. (d) Time dependence of the distance between the protonated oxygen atom of the thiazolidinedione moiety of the ligand and the center of mass of the side chain of Asp964.
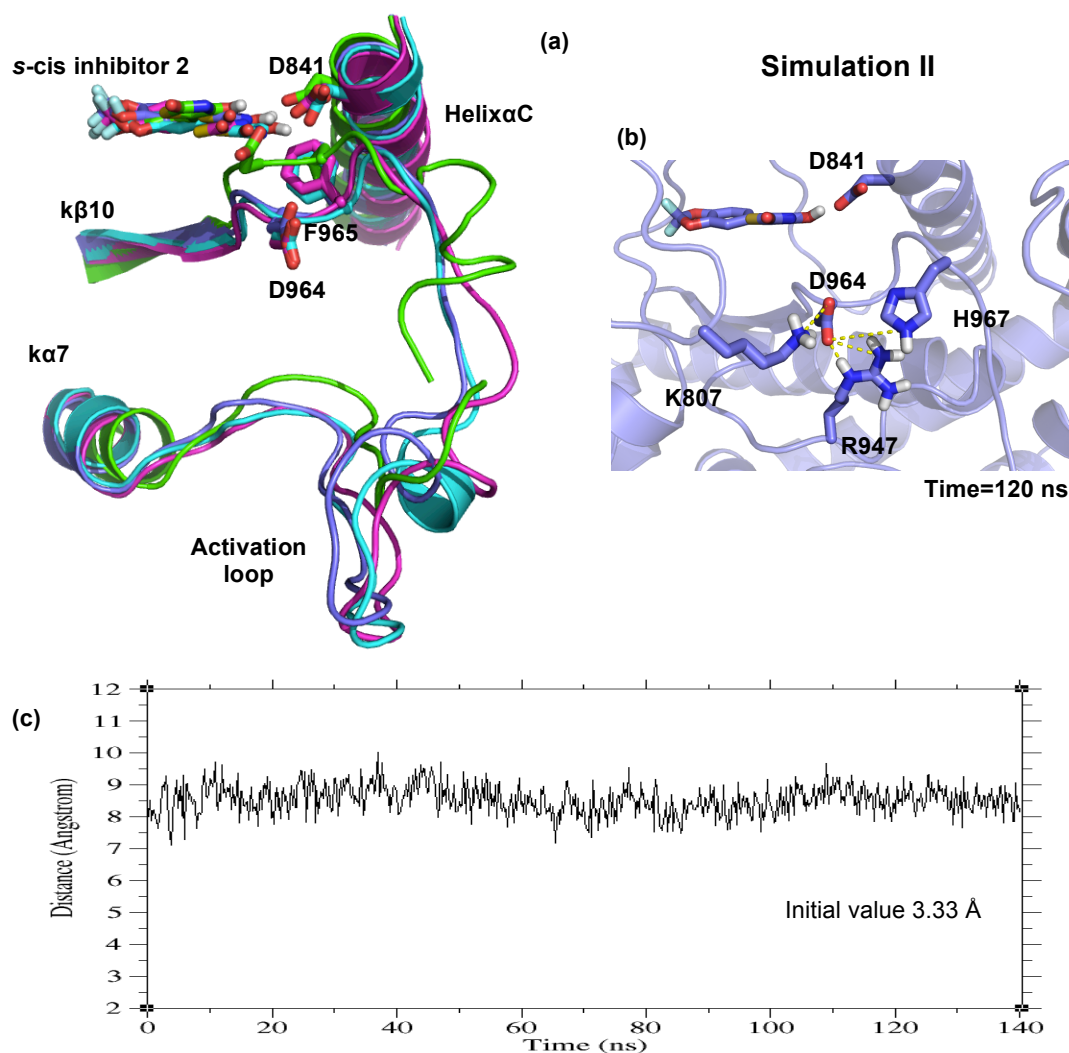
Figure 4.21: Asp964 movement for simulation **II**. (a) View of the DFG-motif and the activation loop. Representative structures of the initial structure (shown in green) and MD snapshots taken at 40 ns (magenta), 80 ns (cyan) and 120 ns (purple). (b) View of the active site at time 120 ns showing the hydrogen bond interactions between Asp964 and: Lys807, Arg947 and His967. (c) Time dependence of the distance between the protonated oxygen atom of the thiazolidinedione moiety of the ligand and the center of mass of the side chain of Asp964.

Whilst one possible explanation for these results may be the incorrect placement of the side chain of Asp964 in the initial atomic model, it is also possible that Asp964 might adopt more than one configuration. As discussed in Section 1.1.2, the side chain of Asp964, the DFG aspartate, exists in different conformations in the structures of the apo and ATP-cocrystallised holo form of PI3Kγ (refer to Figure 1.6). Inspection of the B-factors for the side chain of Asp964 in the publicly available X-ray crystal structure of PI3Kγ complexed with inhibitor **1** (PDB code 2a5u) shows that most of them are high, ranging from 55.5 to 66.8 Å^2 (Table 4.6). Furthermore, in the X-ray crystal structure of PI3Kγ complexed with inhibitor **2** (PDB code 2a4z), the atoms

belonging to the side chain of Asp964 are not resolved; these observations further suggest the high flexibility of the residue.
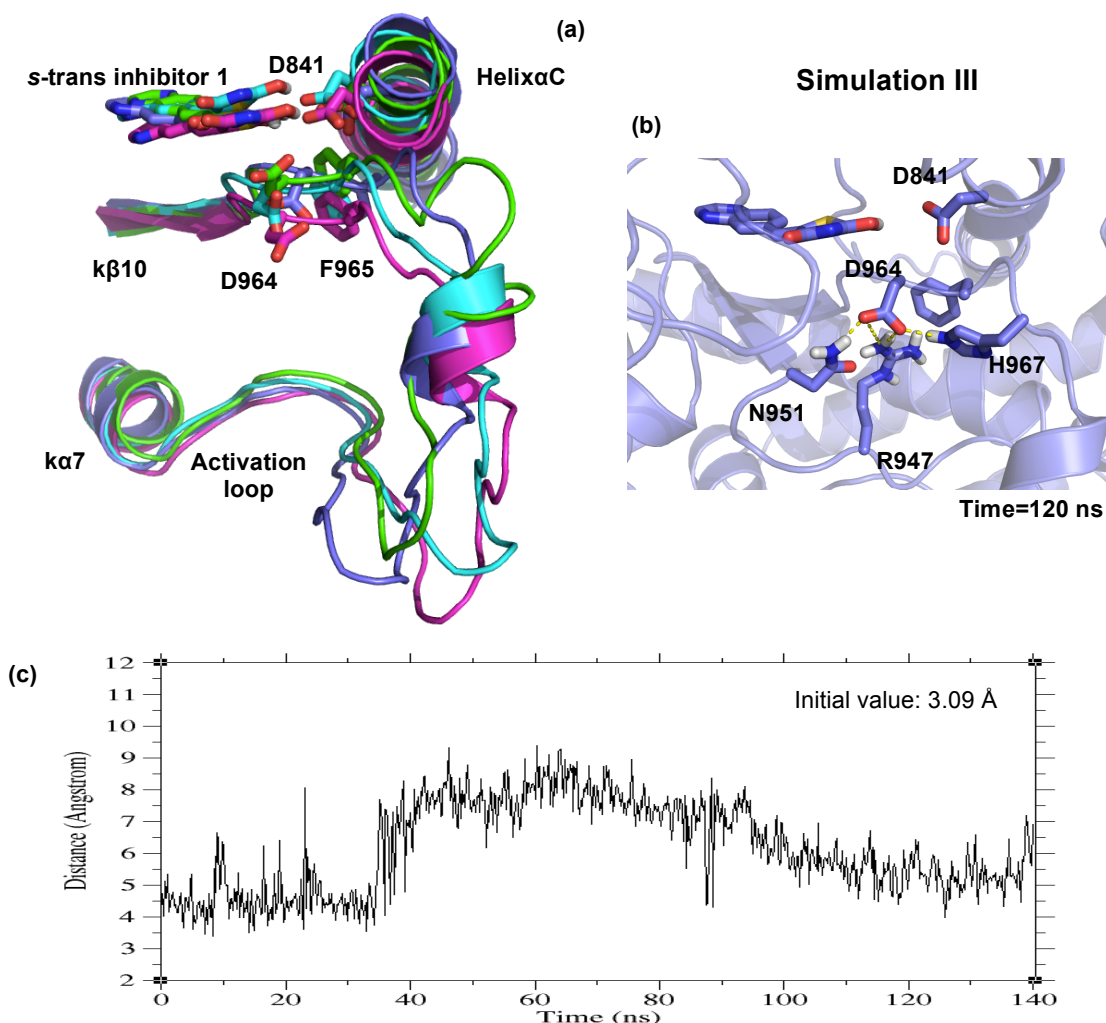
Figure 4.22: Asp964 movement for simulation **III**. (a) View of the DFG-motif and the activation loop. Representative structures of the initial structure (shown in green) and MD snapshots taken at 40 ns (magenta), 80 ns (cyan) and 120 ns (purple). (b) View of the active site at time 120 ns showing the hydrogen bond interactions between Asp964 and: Lys807, Arg947 and His967. (c) Time dependence of the distance between the protonated oxygen atom of the thiazolidinedione moiety of the ligand and the center of mass of the side chain of Asp964.

There is, however, another possible explanation. This can stand alongside local flexibility. The observed Asp964 displacement may be due to a more comprehensive effect involving the rearrangement of the activation loop. The activation loop is an unresolved region of the PI3Kγ structure. In the X-ray structure of PI3Kγ complexed with inhibitor **1** (PDB code 2a5u), residues 968 to 980 are missing, and residues 981 to 985 have B-factors larger than 80 Å^2. Similarly, residues 966 to 980 are missing from the X-ray structure of PI3Kγ complexed with inhibitor **2** (PDB code 2a4z), and residues 981 to 982 have B-factors larger than 100 Å^2.
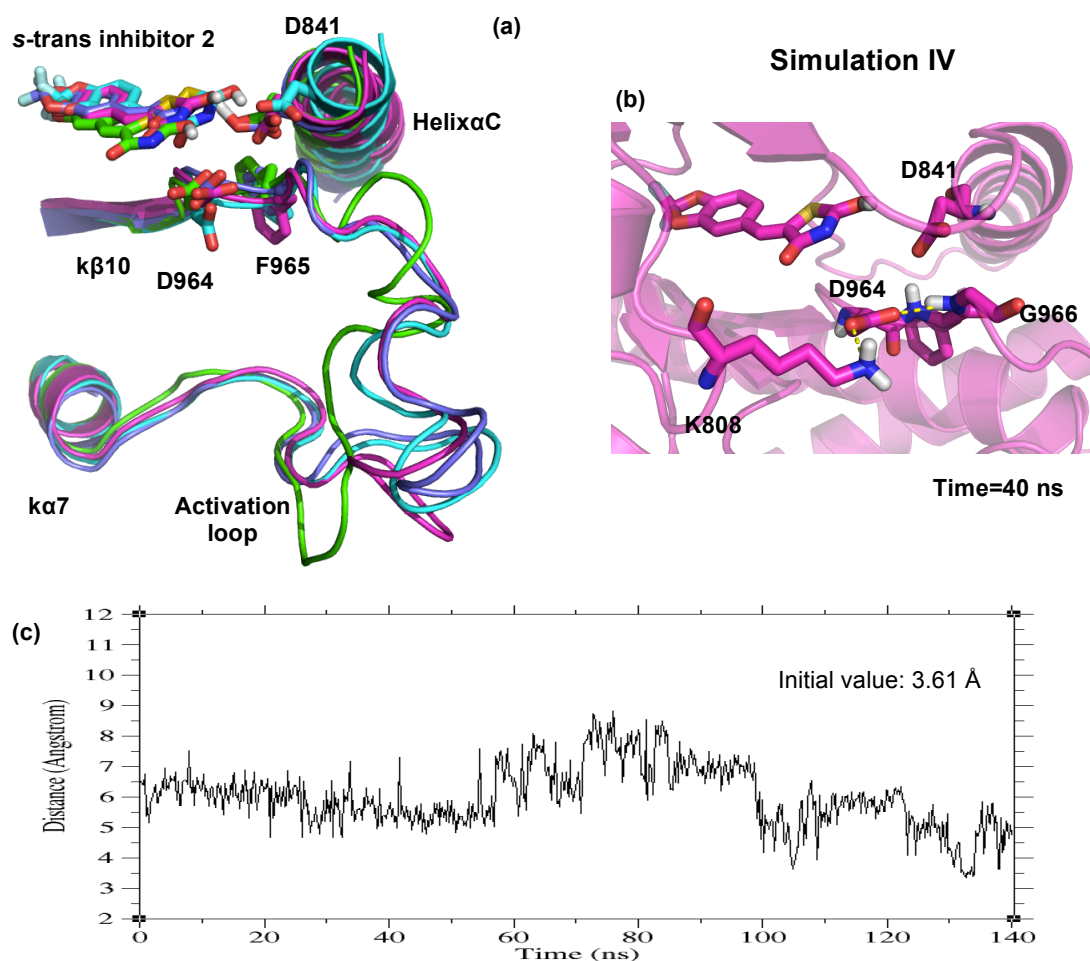
Figure 4.23: Asp964 movement for simulation **IV**. (a) View of the DFG-motif and the activation loop. Representative structures of the initial structure (shown in green) and MD snapshots taken at 40 ns (magenta), 80 ns (cyan) and 120 ns (purple). (b) View of the active site at time 40 ns showing the hydrogen bond interactions between Asp964 and: Lys808 and Gly966. (c) Time dependence of the distance between the protonated oxygen atom of the thiazolidinedione moiety of the ligand and the center of mass of the side chain of Asp964.

As can be seen from Figure 4.20 and 4.21, the Asp964 displacement for the systems involving the *s*-cis inhibitors (i.e. simulations **I** and **II**) is very similar. For both systems, the distance between the protonated oxygen atom of the ligand thiazolidinedione moiety and the center of mass of the side chain of Asp964 remains fairly constant throughout the MD simulation, with average values of 8.23 ± 0.44 Å for simulation **I** and 8.53 ± 0.45 Å for simulation **II**. Also, the dihedral angles for both simulations present similar values during the simulation time (Figure 4.24). For simulation **I**, the carboxylate oxygens of Asp964 hydrogen bond with the side chains of residues His947 and Arg967 (Figure 4.20b and 4.20c). Meanwhile, for simulation **II**, the side chain of Asp964 forms hydrogen bond interactions with His947 and Arg967, and also Lys807 (Figure 4.21b).

On the other hand, the Asp964 side chain displacement for the systems involving the *s*-trans isomers (i.e. simulations **III** and **IV**) has larger fluctuations. In other words, the distance between the protonated oxygen atom of the ligand thiazolidinedione moiety and the center of mass of the side chain of Asp964 does not remain constant during the MD simulation (Figure 4.22c and 4.23c). Notably, the PI3Kγ/*s*-trans inhibitor **1** system (i.e. simulation **III**) is the only system presenting a large variation in the dihedral angle values. From the plot presented in Figure 4.24, it can be seen that the value of the dihedral angle χ_1 angle considerably changes after ~92 ns. In simulation **III**, the carboxylate oxygens of Asp964 hydrogen bond with the side chains of residues Arg967 and His947 and Asn951 (Figure 4.22b). However, in simulation **IV** the side chain of Asp964 forms hydrogen bond interactions with the side chain of Lys803 and backbone NH of Gly966 (Figure 4.23b).

Table 4.6: B-factors (in Å^2) for residues Asp964 and Asp841.

Residue	Atom	X-ray structure	
		2a5u	**2a4z**
Asp964	CB	55.51	*
	CG	62.82	*
	OD1	66.86	*
	OD2	64.42	*
Asp841	CB	50.90	35.29
	CG	54.23	67.39
	OD1	52.52	61.95
	OD2	43.52	76.75

The star (*) indicates that the residue coordinates are missing.
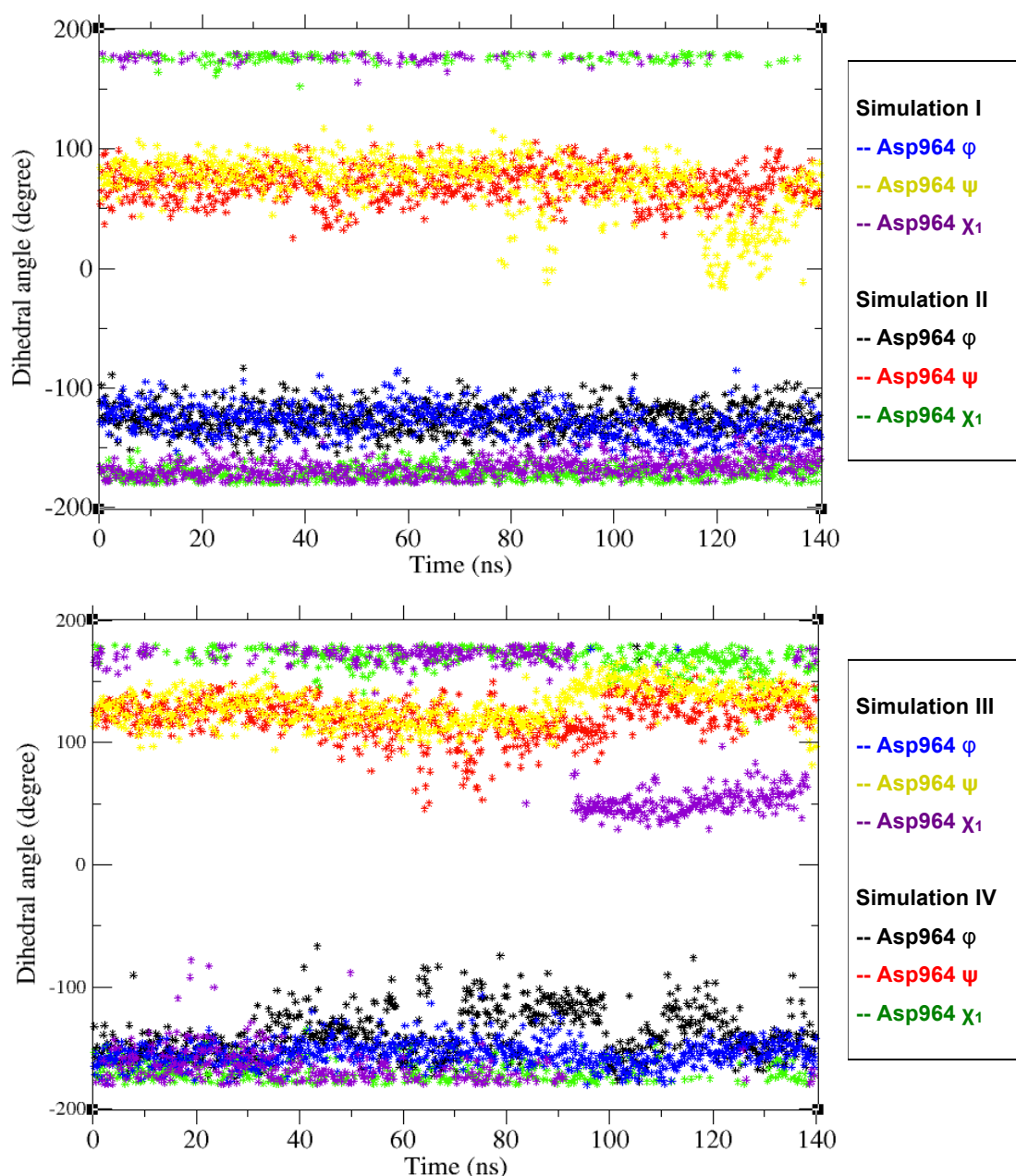
Figure 4.24: Key dihedral angles for Asp964. The data for simulations **I** and **II** is shown in the top panel. The lower panel presents the data for simulations **III** and **IV**.

Prior studies carried out by Berndt *et al.*[38] have also found conformational flexibility of the DFG aspartate. Their published X-ray structures of PI3Kδ in complex with inhibitors SW13 and SW14 (PDB access code 2wxg and 2wxh, respectively) show the catalytic DFG aspartate (Asp911 for PI3Kδ) assuming two alternative conformations (Figure 4.25). Interestingly, the chemical structures of inhibitors SW13 and SW14 are remarkably similar (Figure 4.25e). In the PI3Kδ/SW13 structure (in yellow, Figure 4.25), the DFG aspartate is found in the 'in' conformation, which coincides with its putative ATP/Mg^{2+}-binding position (based on the PI3Kγ/ATP complex). However, in

the PI3Kδ/SW14 structure (in magenta, Figure 4.25), Asp911 is found in the 'out' conformation. Given the fact that in protein kinases, a shift of the DFG aspartate from the active 'in' conformation (ATP-bound) to the inactive 'out' conformation is characteristic of the catalytic cycle, Berndt and co-workers[38] suggest that, by analogy, inhibitors like SW14 might be inducing conformations characteristic of the PI3K catalytic cycle.
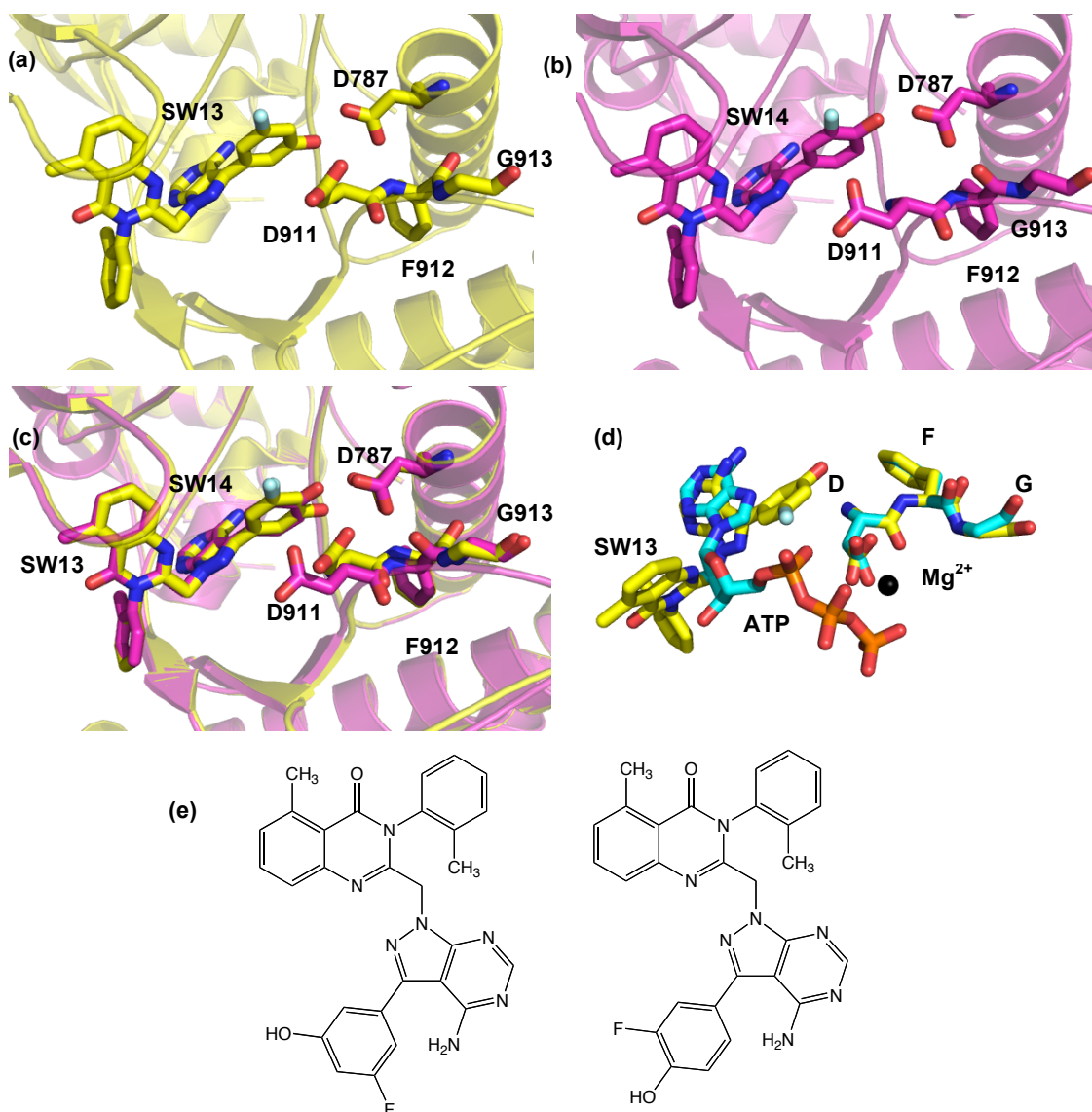
Figure 4.25: Propeller-shaped PI3Kδ-selective inhibitors SW13 and SW14. (a) Structure of PI3Kδ/SW13 (in yellow); the DFG motif is in the 'in' conformation. (b) Structure of PI3Kδ/SW14 (in magenta), the DFG motif is in the 'out' conformation. (c) Structural comparison of PI3Kδ/SW13 (in yellow) and PI3Kδ/SW14 (in magenta). (d) PI3Kδ/SW13 (in yellow) superposed to the PI3Kγ/ATP complex (based on X-ray crystal structure 1E8X, shown in cyan) presenting the DFG motif. The catalytic Mg^{2+} ion is shown as a grey sphere. (e) Chemical structures of inhibitors SW13 and SW14.

4.2.3   Protein-ligand interactions

The hydrogen bonds in the ATP-binding pocket play a crucial role in kinase function and drug inhibition. The conserved hydrogen bonds with the ligand are essential to stabilising the bound conformations. Hence, to compare the binding affinity of the different inhibitors with PI3Kγ, the hydrogen bonding patterns for the four different protein/ligand systems were examined from the 140 ns MD trajectories. The criteria for the formation of a hydrogen bond was taken as a donor-acceptor heavy atom distance ≤ 4.0 Å and with an angle cutoff of 120°.

Tables 4.7 to 4.10 summarise the average hydrogen bonding distances and occupancies between the key PI3Kγ residues Lys833, Asp841 and Val882 and the four different inhibitors over the MD simulations. Additionally, the hydrogen bonding distances for these interactions were followed as a function of the simulation time. The resulting plots are shown in Figures 4.26 to 4.29. To further assess the conformational mobility of the ligand and key residues, a set of equispaced snapshots were taken from each MD trajectory. Each representation consists of 35 superimposed structures taken at 4 ns intervals over the course of the 140 ns production runs. These are also presented in Figures 4.26 to 4.29 (note that for simplicity only polar hydrogen atoms are shown).

For simulation **I**, namely the PI3Kγ/*s*-cis inhibitor **1** system, it can be seen that four hydrogen bonds are formed between the receptor and the ligand (Table 4.7). Firstly, in the hinge region, a hydrogen bond is formed between the ligand quinoxaline nitrogen and the backbone amide nitrogen of Val882 with an occupancy of over 99%. In the phosphate-binding region of the active site, a strong hydrogen bond (average hydrogen bond distance of 2.62 ± 0.11 Å) formed between the carboxylate oxygen of Asp841 and the protonated oxygen of the ligand thiazolidinedione moiety. This is occupied 100% of the simulation time. Also, the positively charged side chain of Lys833 makes two hydrogen bonds with *s*-cis inhibitor **1**, one to the thiazolidinedione nitrogen (3.07 ± 0.16 Å, 97.6% occupancy) and another one to the thiazolidinedione oxygen (3.35 ± 0.23 Å, 50% occupancy). The 35 superimposed structures (Figure 4.26a) show the low conformational mobility of the ligand and key residues, further suggesting the high stability of the interactions throughout the MD trajectory.

Table 4.7: Hydrogen bond occupancies (%) and average distances (Å) of key protein-ligand interactions for simulation **I**.
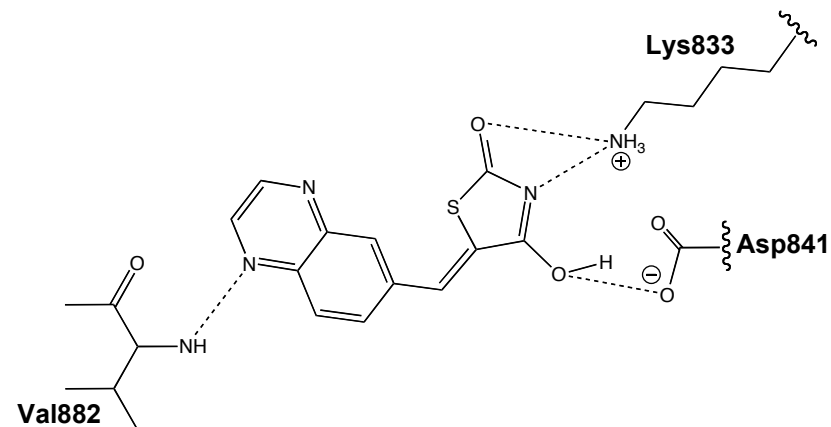
Ligand	Protein	D_{XR} (Å)	Simulation I		
			%occ	D_{HB} (Å)	D_{MD} (Å)
N (tail)	Val882	3.08	99.4	3.21 (0.21)	3.23 (0.23)
OH (head)	Asp841 Oδ1	5.32	0.0	-	4.56 (0.18)
OH (head)	Asp841 Oδ2	3.75	100.0	2.62 (0.11)	2.62 (0.10)
N (head)	Lys833	2.44	97.6	3.07 (0.16)	3.08 (0.17)
O (head)	Lys883	3.57	50.0	3.35 (0.23)	3.40 (0.24)

D_{XR} distances in the starting structure (i.e. re-refined crystal structure). D_{HB} average distances when in a hydrogen bond (Å); and percentage of occupancy (%occ). D_{MD} average distances throughout the MD simulation. The analysis of hydrogen bonds is taken as donor-acceptor heavy-atom distance ≤ 4.0 Å and with an angle cutoff of 120°. The standard deviations (Å) are in parentheses.

Table 4.8: Hydrogen bond occupancies (%) and average distances (Å) of key protein-ligand interactions for simulation **II**.
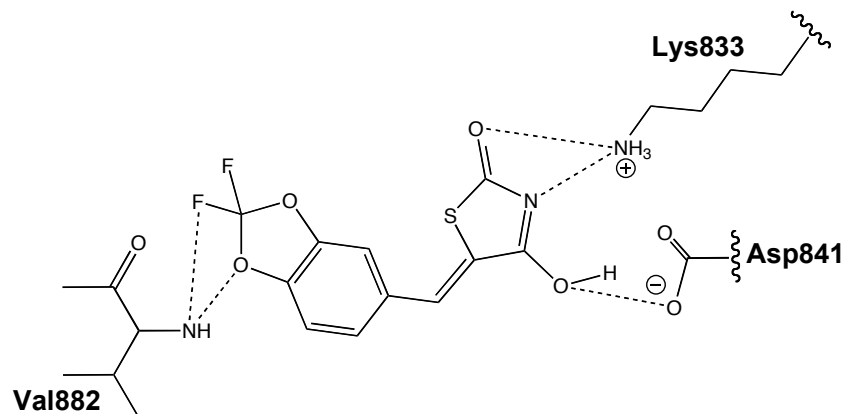
Ligand	Protein	D_{XR} (Å)	Simulation II		
			%occ	D_{HB} (Å)	D_{MD} (Å)
O (tail)	Val882	3.16	81.7	3.50 (0.25)	3.65 (0.41)
F (tail)	Val882	3.79	12.7	3.79 (0.17)	4.82 (0.47)
OH (head)	Asp841 Oδ1	5.24	0.0	-	4.52 (0.18)
OH (head)	Asp841 Oδ2	3.69	100.0	2.62 (0.11)	2.62 (0.10)
N (head)	Lys833	2.26	98.8	3.08 (0.17)	3.08 (0.17)
O (head)	Lys883	3.80	58.7	3.38 (0.23)	3.41 (0.25)

D_{XR} distances in the starting structure (i.e. re-refined crystal structure). D_{HB} average distances when in a hydrogen bond (Å); and percentage of occupancy (%occ). D_{MD} average distances throughout the MD simulation. The analysis of hydrogen bonds is taken as donor-acceptor heavy-atom distance ≤ 4.0 Å and with an angle cutoff of 120°. The standard deviations (Å) are in parentheses.

Table 4.9: Hydrogen bond occupancies (%) and average distances (Å) of key protein-ligand interactions for simulation **III**.
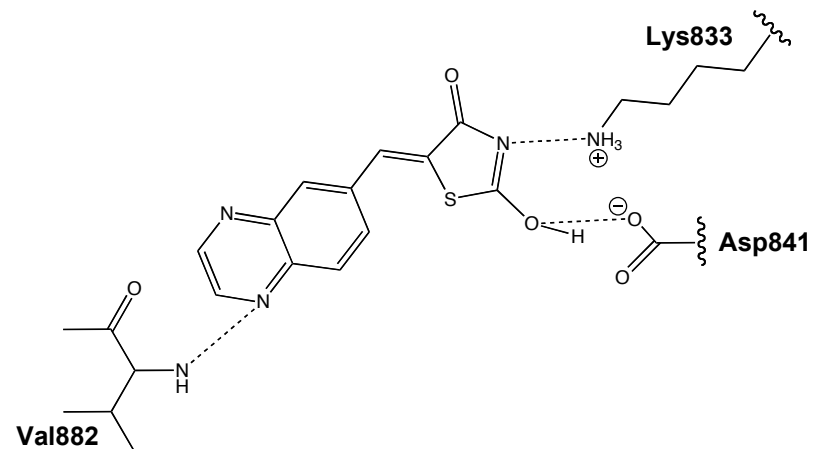
Ligand	Protein	D_{XR} (Å)	Simulation III		
			%occ	D_{HB} (Å)	D_{MD} (Å)
N (tail)	Val882	3.55	95.8	3.36 (0.23)	3.40 (0.29)
OH (head)	Asp841 Oδ1	4.84	36.0	3.11 (0.42)	3.52 (0.81)
OH (head)	Asp841 Oδ2	3.61	71.4	2.64 (0.22)	3.14 (0.96)
N (head)	Lys833	2.61	63.2	3.11 (0.24)	3.50 (0.86)
O (head)	Lys883	4.34	0.0	-	4.17 (1.08)

D_{XR} distances in the starting structure (i.e. re-refined crystal structure). D_{HB} average distances when in a hydrogen bond (Å); and percentage of occupancy (%occ). D_{MD} average distances throughout the MD simulation. The analysis of hydrogen bonds is taken as donor-acceptor heavy-atom distance ≤ 4.0 Å and with an angle cutoff of 120°. The standard deviations (Å) are in parentheses.

Table 4.10: Hydrogen bond occupancies (%) and average distances (Å) of key protein-ligand interactions for simulation **IV**.
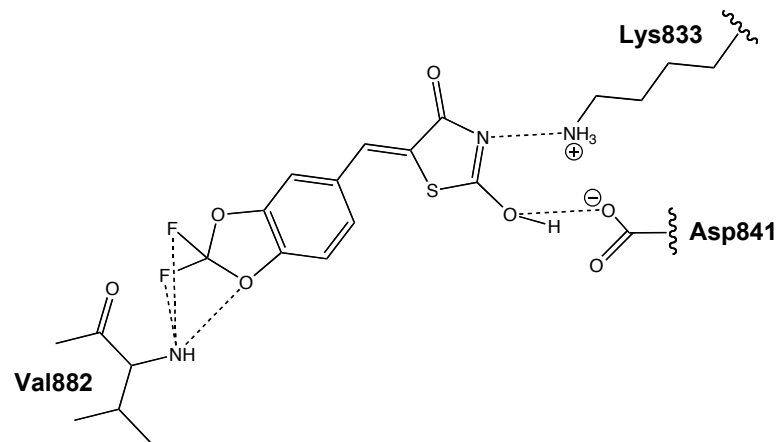
Ligand	Protein	D_{XR} (Å)	Simulation IV		
			%occ	D_{HB} (Å)	D_{MD} (Å)
O (tail)	Val882	3.36	90.9	3.20 (0.26)	3.25 (0.29)
F1 (tail)	Val882	2.72	75.6	3.33 (0.31)	3.44 (0.40)
F2 (tail)	Val882	3.19	22.0	3.82 (0.14)	4.19 (0.29)
OH (head)	Asp841 Oδ1	4.90	29.3	2.90 (0.49)	3.80 (0.78)
OH (head)	Asp841 Oδ2	3.70	60.3	2.63 (0.24)	3.20 (0.81)
N (head)	Lys833	2.72	54.8	3.09 (0.23)	3.94 (1.12)
O (head)	Lys883	4.73	17.6	3.47 (0.33)	4.89 (1.17)

D_{XR} distances in the starting structure (i.e. re-refined crystal structure). D_{HB} average distances when in a hydrogen bond (Å); and percentage of occupancy (%occ). D_{MD} average distances throughout the MD simulation. The analysis of hydrogen bonds is taken as donor-acceptor heavy-atom distance ≤ 4.0 Å and with an angle cutoff of 120°. The standard deviations (Å) are in parentheses.
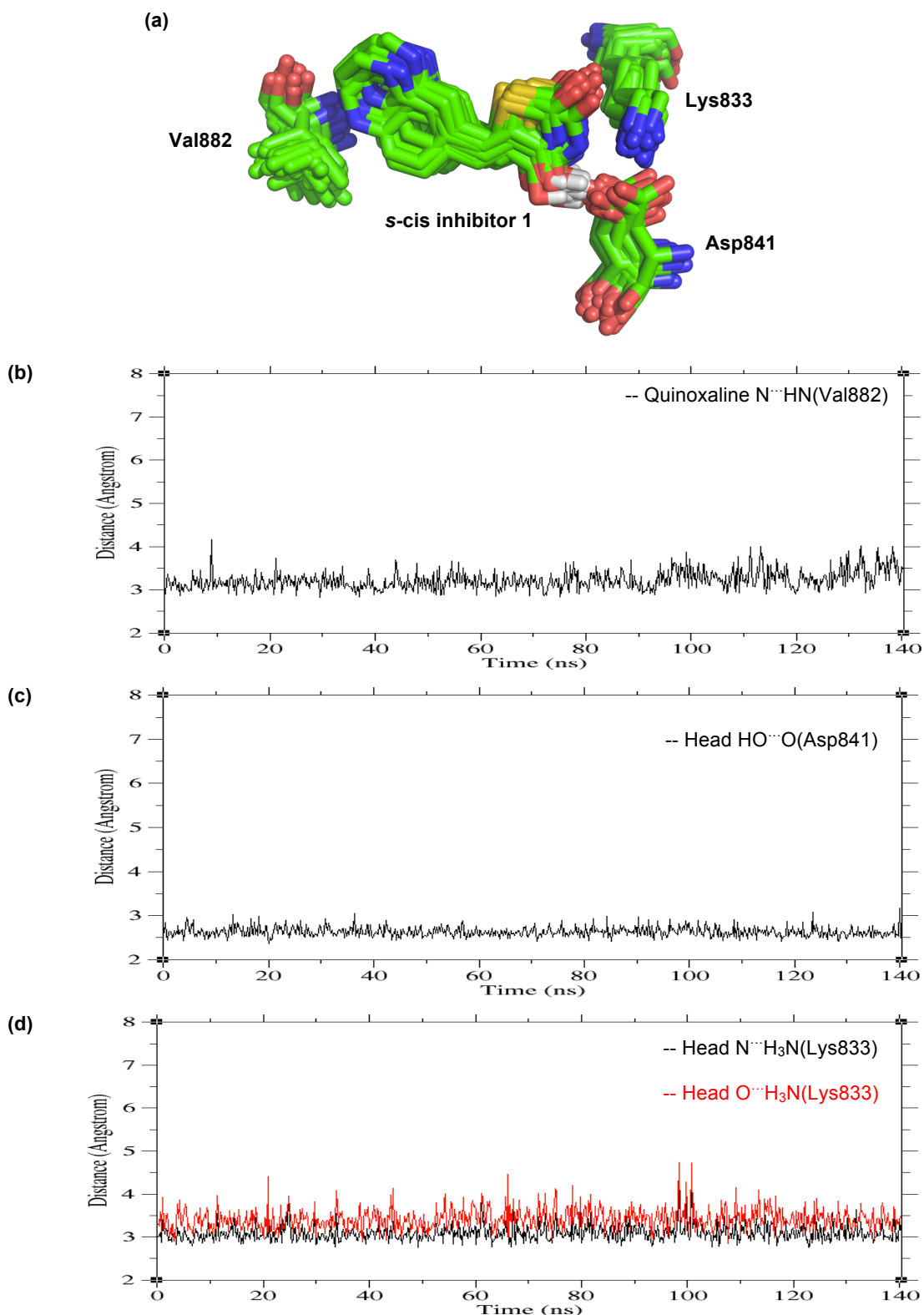
Figure 4.26: Protein-ligand interactions for simulation **I**. (a) Snapshots of the 35 equispaced MD conformations showing the *s*-cis inhibitor **1** and PI3Kγ residues Lys833, Asp841 and Val882. (b) Hydrogen bonding distances between the backbone amide nitrogen of Val882 and the ligand quinoxaline nitrogen. (c) Hydrogen bonding distances between Asp841 Oδ2 and the ligand protonated oxygen. (d) Time dependence of the interactions between Lys833 and: the ligand protonated oxygen (red line), and the ligand head nitrogen (black line).
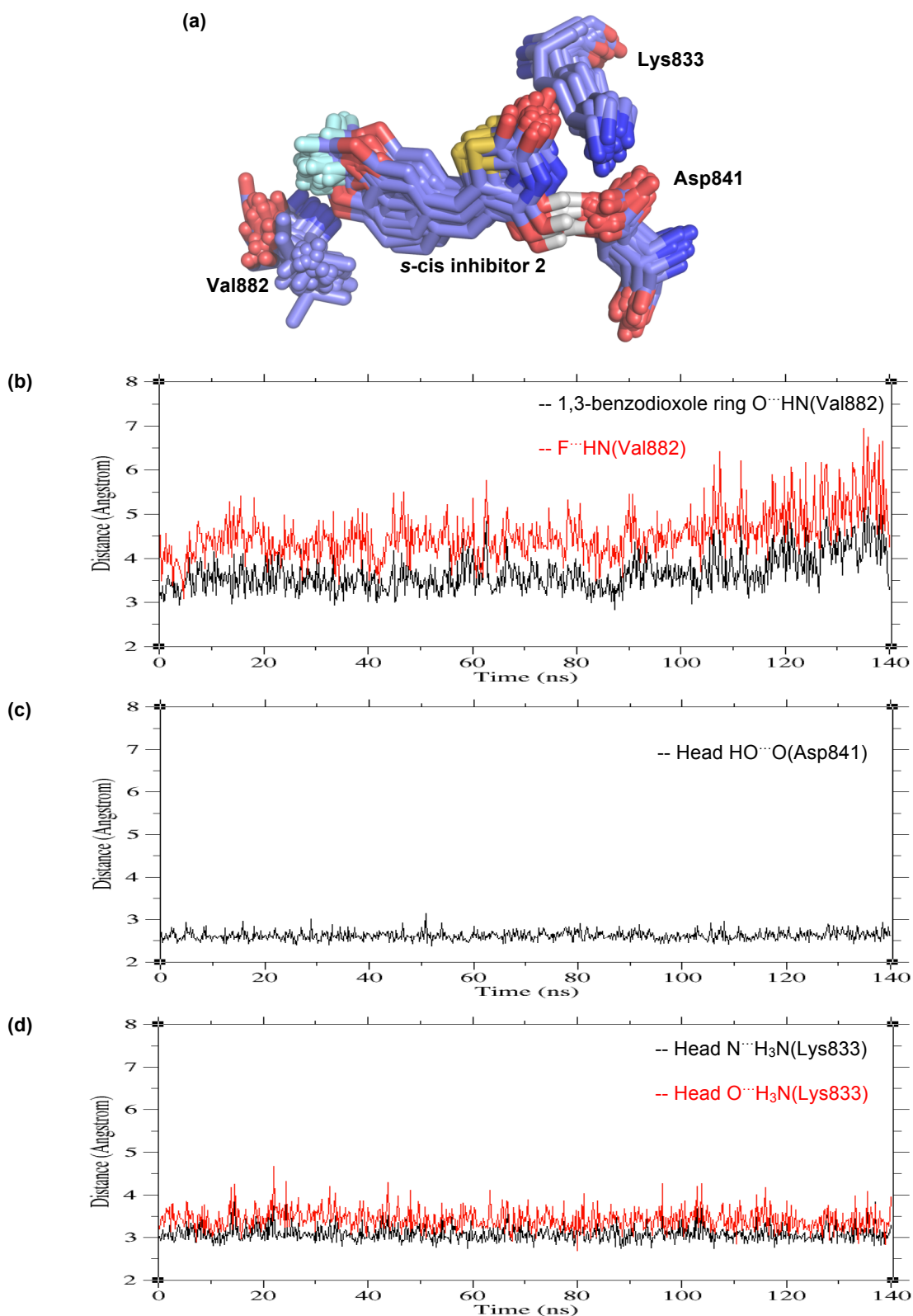
Figure 4.27: Protein-ligand interactions for simulation **II**. (a) Snapshots of the 35 equispaced MD conformations showing *s*-cis inhibitor **2** and residues Lys833, Asp841 and Val882. (b) Time series of the distances between the backbone amide nitrogen of Val882 and: the ligand 1,3-benzodioxole ring oxygen (black line), and the fluorine atom (red line). (c) Hydrogen bonding distances between Asp841 Oδ2 and the ligand protonated oxygen. (d) Time dependence of the interactions between Lys833 and: the ligand protonated oxygen (red line), and the ligand head nitrogen (black line).
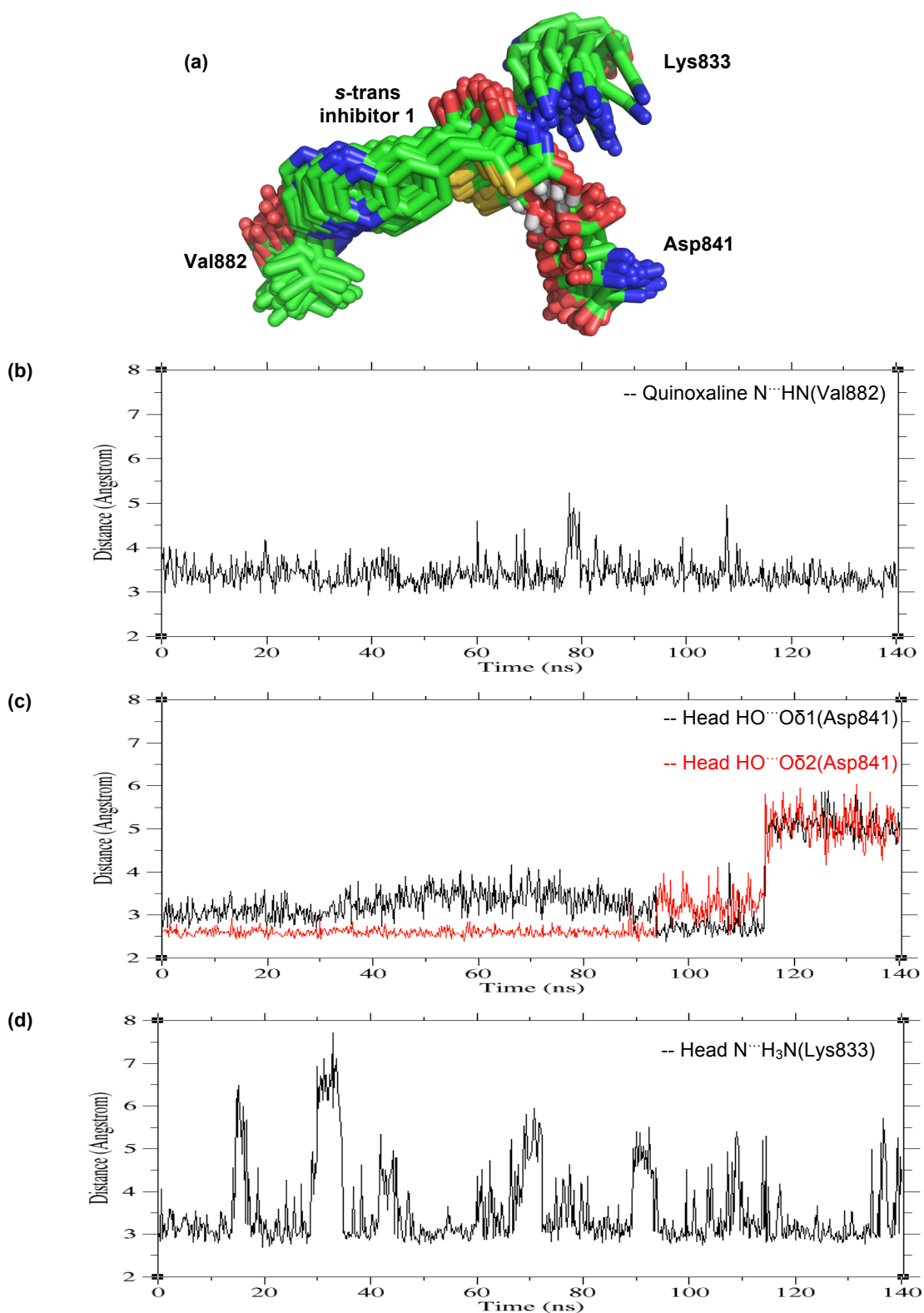
Figure 4.28: Protein-ligand interactions for simulation **III**. (a) Snapshots of the 35 equispaced MD conformations showing *s*-trans inhibitor **1** and PI3Kγ residues Lys833, Asp841 and Val882. (b) Time series of the distance between the backbone amide nitrogen of Val882 and the ligand quinoxaline nitrogen. (c) Hydrogen bonding distances between the ligand protonated oxygen and: Asp841 Oδ1 (black line), and Asp841 Oδ2 (red line). (d) Time dependence of the interaction between Lys833 and the ligand head nitrogen (black line).
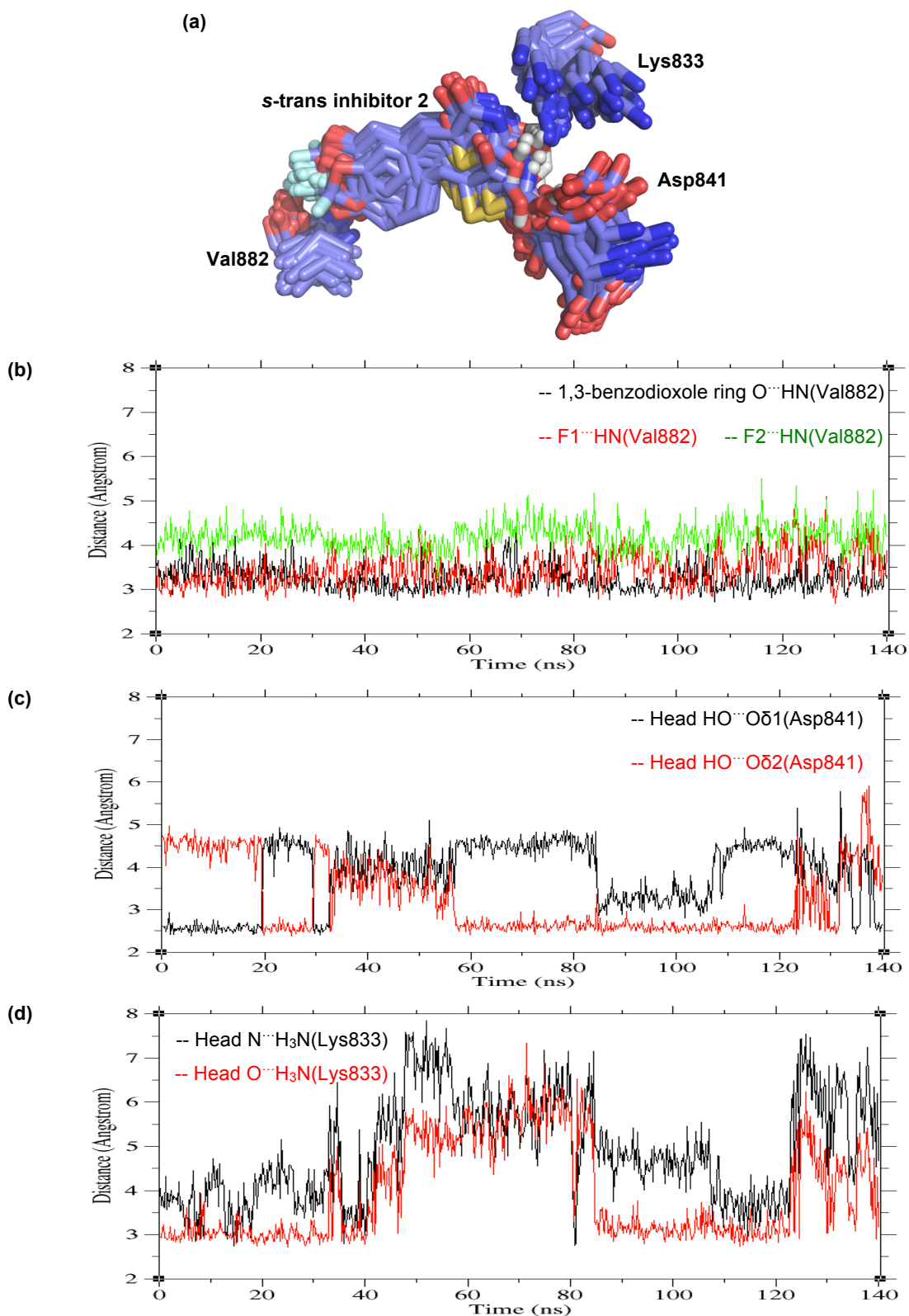
Figure 4.29: Protein-ligand interactions for simulation **IV**. (a) Snapshots of the 35 equispaced MD conformations showing *s*-trans inhibitor **2** and key residues. (b) Time series of the distances between the backbone amide nitrogen of Val882 and: ligand1,3-benzodioxole ring oxygen (black line); fluorine atom F1 (red line); and fluorine atom F2 (green line). (c) Hydrogen bonding distances between ligand protonated oxygen and: Asp841 Oδ1 (black line), and Asp841 Oδ2 (red line). (d) Interaction between Lys833 and: ligand protonated oxygen (black line), ligand head nitrogen (red line).

The hydrogen bond analysis for the PI3Kγ/*s*-cis inhibitor **2** system (i.e. simulation **II**) reveals the formation of five hydrogen bonds between the protein and the ligand (Table 4.8). Two of these bonds are formed in the hinge region of the active site, while the remaining three H-bonds involve interactions with the thiazolidinedione head of the inhibitor (Figure 4.27). As can be seen from the data in Table 4.8, the hydrogen-bonding profile of the PI3Kγ/*s*-cis inhibitor **2** system is very similar to the one presented by the PI3Kγ/*s*-cis inhibitor **1** system. Notably, the interactions to Asp841 and Lys833 present almost equivalent hydrogen bonding distances and occupancy values. However, the key difference is that *s*-cis inhibitor **2** forms two hydrogen bonds with Val882. A long-lived hydrogen bond is formed between the ligand 1,3-benzodioxole ring oxygen and the backbone amide nitrogen of Val882 (3.50 ± 0.25 Å, 81.7% occupancy). The second shorter-lived hydrogen bond is formed between one of the ligand fluorine atoms and the backbone amide nitrogen of Val882 (3.79 ± 0.17 Å, 12.7% occupancy). The formation of this F⋯H-N hydrogen bond is only speculative, and will be discussed in depth in chapter 5.

Another interesting observation for the systems containing the *s*-cis inhibitors is the formation of well-conserved hydrogen bonds between residues Lys833 and Asp841. This might contribute to the stabilisation of the binding conformations. For simulation **I**, the Oδ1 and Oδ2 of Asp841 form H-bonds with the positively charged side chain of Lys833, with occupancies of 98.5 and 43.3% (2.83 ± 0.14 Å and 2.97 ± 0.20 Å, respectively). Meanwhile, for simulation **II**, the same hydrogen bonds are conserved with very similar occupancies and average distances (2.81 ± 0.12 Å, 99% occupancy and 2.95 ± 0.19, 46.3% occupancy).

Turning now to the *s*-trans inhibitors, the results of the hydrogen bond analysis for simulation **III** (Table 4.9) suggest the formation of one stable hydrogen bond between *s*-trans inhibitor **1** and Val882, with occupancy of over 95%. In the phosphate-binding region of the active site, the protonated oxygen of the ligand thiazolidinedione moiety hydrogen bonds with Oδ1 and Oδ2 of Asp841 with occupancies of 36.0 and 71.1%, respectively. Also, the positively charged side chain of Lys833 forms one hydrogen bond to the thiazolidinedione nitrogen (3.07 ± 0.16 Å, 97.6% occupancy). It is apparent from this data that one crucial difference between the PI3Kγ/*s*-cis inhibitor **1** system and the PI3Kγ/*s*-trans inhibitor **1** system is that the later forms one less hydrogen bond to Lys833.

The 35 superimposed MD structures and the plots presented in Figure 4.28 display the dynamic changes and conformational mobility of the key active site residues for

simulation **III**. Of particular interest is the mobility of the flexible side chain of Lys833, which can be seen fluctuating away from the ligand (Figure 4.28d). This fluctuation might be caused by the formation of a hydrogen bond between the positively charged Lys833 and Oδ1/Oδ2 of Asp836 (Table 4.11, Figure 4.30a). In addition, the side chain of Asp841 also sustains interactions with residues Tyr867, Phe965 and Gly966 (Table 4.11). As a result of these extra interactions, the local binding pocket appears to have opened somewhat, which might be resulting in a relatively lower affinity with the ligand.
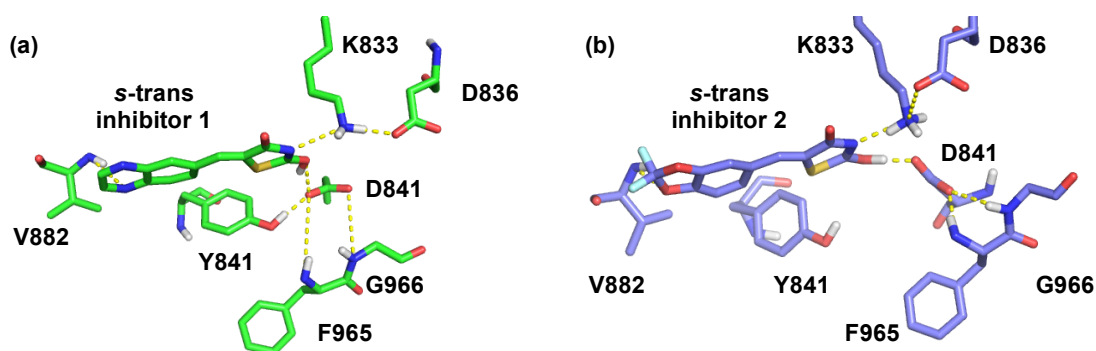
Figure 4.30: Representative snapshot from simulation **III** (a) and simulation **IV** (b) showing the hydrogen bond interactions for residues Asp841 and Lys833. The yellow dashed lines indicate hydrogen bond interactions.

Table 4.11: Hydrogen bond occupancies (%) and average distances (Å) of key interactions involving residues Asp841 and Lys833 for simulation **III**.

Donor	Acceptor	Simulation III	
		%occ	D_{HB} (Å)
Lys833	Asp836 Oδ1	52.6	2.81 (0.22)
Lys833	Asp836 Oδ2	33.7	2.89 (0.31)
Tyr867 OH	Asp841 Oδ1	33.4	2.74 (0.19)
Tyr867 OH	Asp841 Oδ2	90.0	2.88 (0.36)
Phe965 NH	Asp841 Oδ2	37.6	3.00 (0.34)
Gly966 NH	Asp841 Oδ2	33.4	2.89 (0.18)

D_{HB} average distances when in a hydrogen bond (Å); and percentage of occupancy (%occ). The analysis of hydrogen bonds is taken as donor-acceptor heavy-atom distance ≤ 4.0 Å and with an angle cutoff of 120°. The standard deviations (Å) are in parentheses.

It is also interesting to note the plot in Figure 4.28c, which apparently indicates the loss of the bond between Asp841 and the ligand protonated oxygen during the last 25 ns of the simulation trajectory. This data must be interpreted with caution because, as

it will be discussed in Section 4.2.4.3, this arises from the arrival of a water molecule, which causes the direct bond to switch to a water mediated interaction.

For the last complex, namely the PI3Kγ/*s*-trans inhibitor **2** system (simulation **IV**), the hydrogen bond analysis reveals the formation of three different hydrogen bonds to Val882 in the hinge region of the active site (Table 4.10). These are: a long-lived hydrogen bond to the ligand 1,3-benzodioxole ring oxygen; and two shorter-lived hydrogen bonds to each of the ligand fluorine atoms. The formation of the F··H-N hydrogen bonds is only speculative, and will be discussed in depth in chapter 5. Additionally, the thiazolidinedione head of the inhibitor formed hydrogen bonds to Lys833 and Asp841. However, as can be seen in Figure 4.29c and 4.29d, the distances between the corresponding hydrogen bonding atoms fluctuate significantly. This appears to be a result of the multiple interactions maintained by the side chains of Lys833 and Asp841, which are summarised in Table 4.12 and schematically depicted in Figure 4.30b. The conformational mobility of the ligand in the active site can be seen in Figure 4.29a. Consequently, the instability of the region might be giving rise to a poorly defined binding mode of ligand.

Table 4.12: Hydrogen bond occupancies (%) and average distances (Å) of key interactions involving residues Asp841 and Lys833 for simulation **IV**.

Donor	Acceptor	Simulation IV	
		%occ	D_{HB} (Å)
Lys833	Asp836 Oδ1	28.0	2.90 (0.29)
Lys833	Asp836 Oδ2	42.6	2.84 (0.22)
Lys833	Asp837 O	67.4	2.84 (0.18)
Tyr867 OH	Asp841 Oδ1	22.6	2.91 (0.36)
Tyr867 OH	Asp841 Oδ2	37.4	3.00 (0.42)
Phe965 NH	Asp841 Oδ1	24.8	3.24 (0.44)
Phe965 NH	Asp841 Oδ2	46.7	3.09 (0.34)
Gly966 NH	Asp841 Oδ1	11.7	3.63 (0.34)
Gly966 NH	Asp841 Oδ2	16.6	3.26 (0.35)

D_{HB} average distances when in a hydrogen bond (Å); and percentage of occupancy (%occ). The analysis of hydrogen bonds is taken as donor-acceptor heavy-atom distance ≤ 4.0 Å and with an angle cutoff of 120°. The standard deviations (Å) are in parentheses.

4.2.4 Water-mediated interactions

To identify protein-bound water molecules and hydrogen-bonded networks in the active site, water density maps were created. The results will be discussed in this section.

4.2.4.1 Simulation I

Water density maps for the PI3Kγ systems were used to identify distinctive regions of persisting water in the protein active site. For the PI3Kγ/*s*-cis inhibitor **1** system (i.e. simulation **I**), an area of persisting water is maintained in the vicinity of the ligand thiazolidinedione moiety (purple region in Figure 4.31a). As detailed in the water network shown in Figure 4.31b, this area of water density is attributed to a single resident water molecule which hydrogen bonds to the backbone amide hydrogen of Phe965 and to the backbone amide hydrogen of Gly966. Additionally, the simulation predicts that this water molecule hydrogen bonds to the carboxylate oxygen of Asp841, to the amide oxygen of Ile963 and to the protonated oxygen of the ligand thiazolidinedione moiety.
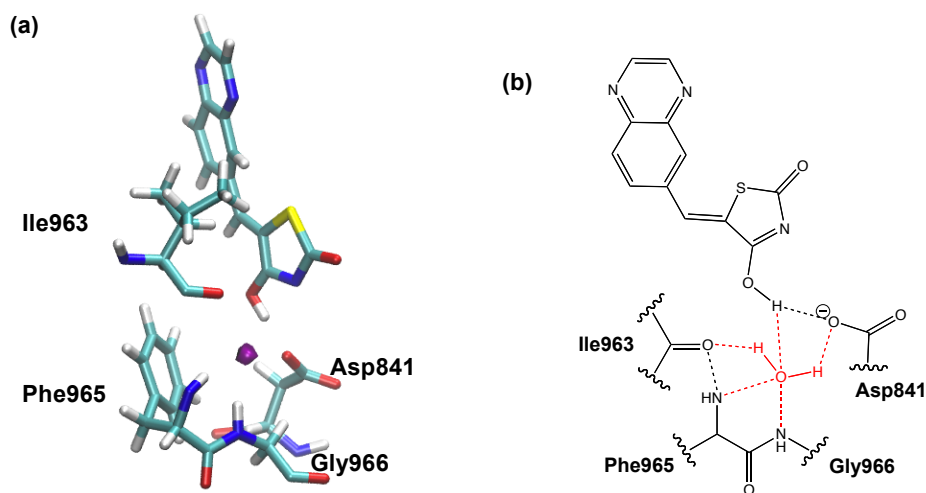
Figure 4.31: Persisting water site in the binding site of the PI3Kγ/*s*-cis inhibitor **1** system. (a) Water density map. The area of localised water presence is shown in purple. Note that the hinge of the active site is pointing upwards. (b) Predicted water-mediated hydrogen bonding network (in red dashed lines). The hydrogen bond between ligand-OH and Asp841 (100% occupancy), as well as the interaction between Ile963 and Phe965 (28.8% occupancy), are shown in black dashed lines.

By monitoring these hydrogen bonds along the MD trajectory, it is observed that the interactions of the water molecule to residues Phe965 and Gly966 are conserved with high occupancies (87.9% and 84.7%, respectively, Table 4.13). The occupancies for

its interaction with Asp841 and Ile963 along the trajectory are 88.2% and 65.9%, respectively. Meanwhile only 36.2% occupancy is found for the water hydrogen bond to the protonated oxygen of the ligand thiazolidinedione moiety (Table 4.13). The average hydrogen bond distances, together with their respective standard deviations are listed in Table 4.13.

Table 4.13: Hydrogen bond occupancies (%) and average distances (Å) of key interactions involving water molecule WAT for simulation **I**.

Donor	Acceptor	Simulation I	
		%occ	D_{HB} (Å)
Phe965 NH	WAT O	87.9	3.03 (0.19)
Gly966 NH	WAT O	84.7	3.05 (0.18)
WAT O-H1	Asp841 Oδ2	44.7	2.71 (0.15)
WAT O-H2	Asp841 Oδ2	43.5	2.71 (0.14)
WAT O-H1	Ile963 O	33.5	2.96 (0.25)
WAT O-H2	Ile963 O	32.4	2.92 (0.23)
Ligand OH	WAT O	36.2	3.69 (0.19)

D_{HB} average distances when in a hydrogen bond (Å); and percentage of occupancy (%occ). The analysis of hydrogen bonds is taken as donor-acceptor heavy-atom distance ≤ 4.0 Å and with an angle cutoff of 120°. The standard deviations (Å) are in parentheses.

Furthermore, Figure 4.32 offers a more detailed picture of the time-dependent distances in the water-mediated hydrogen bonding network. It can be seen that the water network is only disrupted towards the end of the MD simulation. This persisting water molecule is attributable to a non-crystallographic water, which is believed to drift into position during the equilibration stage.

Inspection of the crystal structure suggests water-mediating hydrogen bonding interactions between the carboxylate oxygen of Asp964 and the ligand thiazolidinedione oxygen (refer to Figure 4.16). However, as discussed in Section 4.2.2, Asp964 swings away from the ligand (refer to Figure 4.20) presumably causing this water to become disordered.
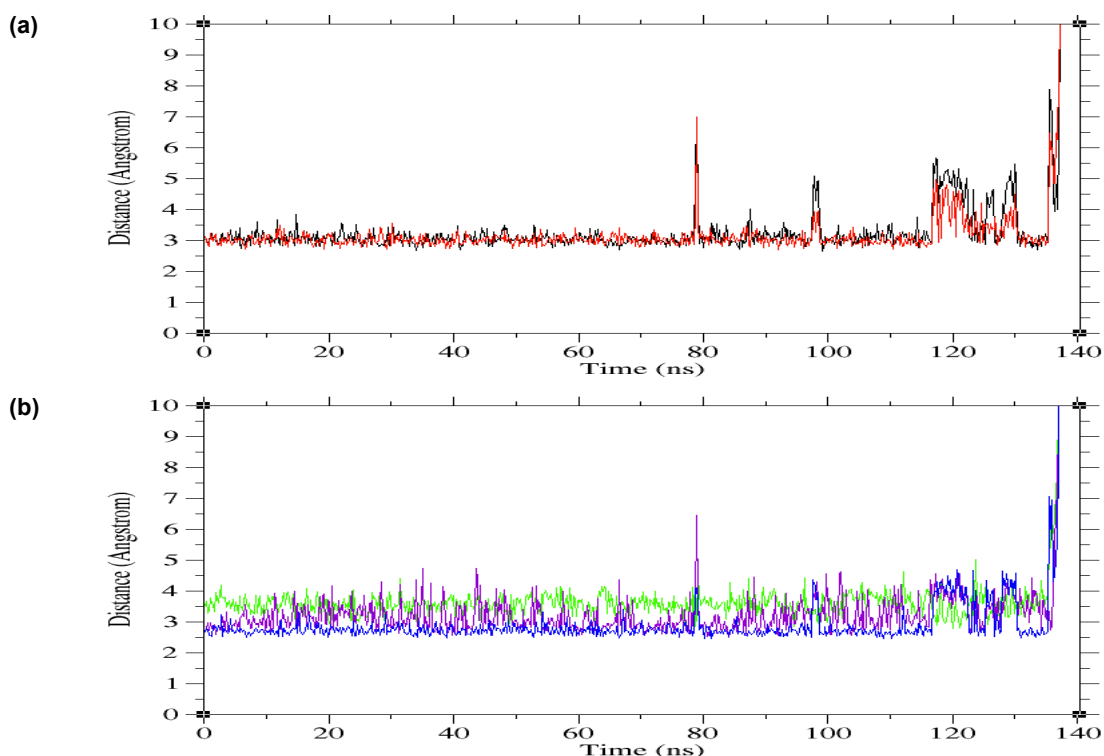
Figure 4.32: Time dependence of the distances in the water-mediated hydrogen bonding network for simulation **I**. Distance between the oxygen atom of the water molecule WAT and: backbone amide nitrogen of Phe965 (a, red line); backbone amide nitrogen of Gly966 (a, black line); Asp841 Oδ2 (b, blue line); backbone amide oxygen of Ile963 (b, purple line); and the ligand protonated oxygen (b, green line).

4.2.4.2 Simulation II

The water density maps for the PI3Kγ/*s*-cis inhibitor **2** system (i.e. simulation **II**) also reveal one area of persisting water density in the proximity of the ligand (Figure 4.33). The water site exhibits an analogous location to the one previously described for the PI3Kγ/*s*-cis inhibitor **1** system. Correspondingly, this area of water density is attributed to a single resident water molecule which hydrogen bonds to the backbone amide hydrogens of residues Phe965 and Gly966. Moreover, this localised water molecule hydrogen bonds to the carboxylate oxygen of Asp841, the backbone amide oxygen of Ile963, and to the protonated thiazolidinedione oxygen of *s*-cis inhibitor **2** as depicted in Figure 4.33b.
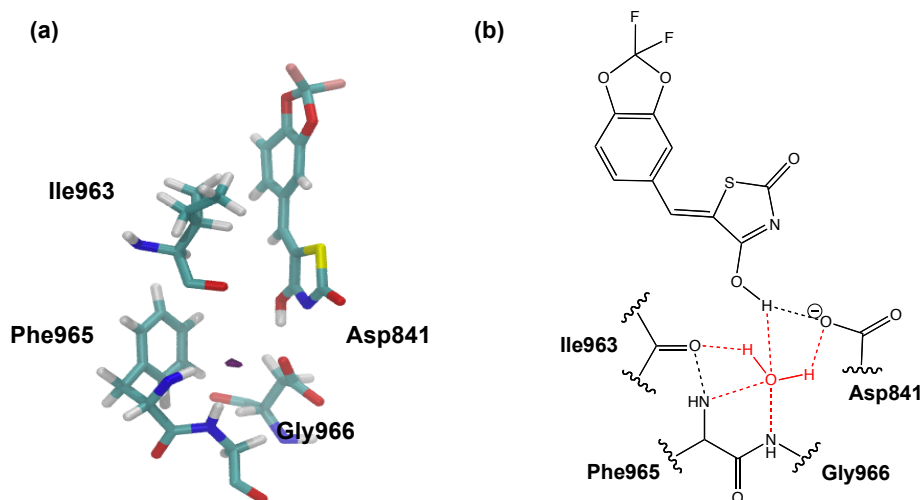
Figure 4.33: Persisting water site in the binding site of the PI3Kγ/*s*-cis inhibitor **2** system. (a) Water density map. The area of localised water presence is shown in purple. (b) Predicted water-mediated hydrogen bonding network (in red dashed lines). The hydrogen bond between the ligand and Asp841 (100% occupancy), as well as the interaction between Ile963 and Phe965 (45.7% occupancy) are shown in black dashed lines.

The results obtained from the hydrogen bond analysis are presented in Table 4.14. It is apparent from this table that the water interactions along the simulation trajectory are conserved with occupancies ranging from 18.2% to 35.4%. Notably, these occupancies are lower than the ones shown by the analogous water site in the PI3Kγ/*s*-cis inhibitor **1** system, which range from 32.4% to 87.9% (Table 4.13).

Table 4.14: Hydrogen bond occupancies (%) and average distances (Å) of key interactions involving water molecule WAT for simulation **II**.

Donor	Acceptor	Simulation II	
		%occ	D_{HB} (Å)
Phe965 NH	WAT O	34.7	3.04 (0.18)
Gly966 NH	WAT O	33.0	3.06 (0.18)
WAT O-H1	Asp841 Oδ2	21.3	2.72 (0.15)
WAT O-H2	Asp841 Oδ2	14.1	2.72 (0.17)
WAT O-H1	Ile963 O	19.0	2.86 (0.22)
WAT O-H2	Ile963 O	11.1	2.93 (0.25)
Ligand OH	WAT O	18.3	3.74 (0.17)

D_{HB} average distances when in a hydrogen bond (Å); and percentage of occupancy (%occ). The analysis of hydrogen bonds is taken as donor-acceptor heavy-atom distance ≤ 4.0 Å and with an angle cutoff of 120°. The standard deviations (Å) are in parentheses.

126

The identified area of water density is attributable to a non-crystallographic water, which drifts into position after ~90 ns (Figure 4.34).

Inspection of the crystal structure suggested water-mediating hydrogen bonding interactions between the carboxylate oxygen of Asp964 and the ligand thiazolidinedione oxygen (refer to Figure 4.17). However, as discussed in Section 4.2.2, Asp964 swings away from the ligand (refer to Figure 4.21) presumably causing this water to become disordered.
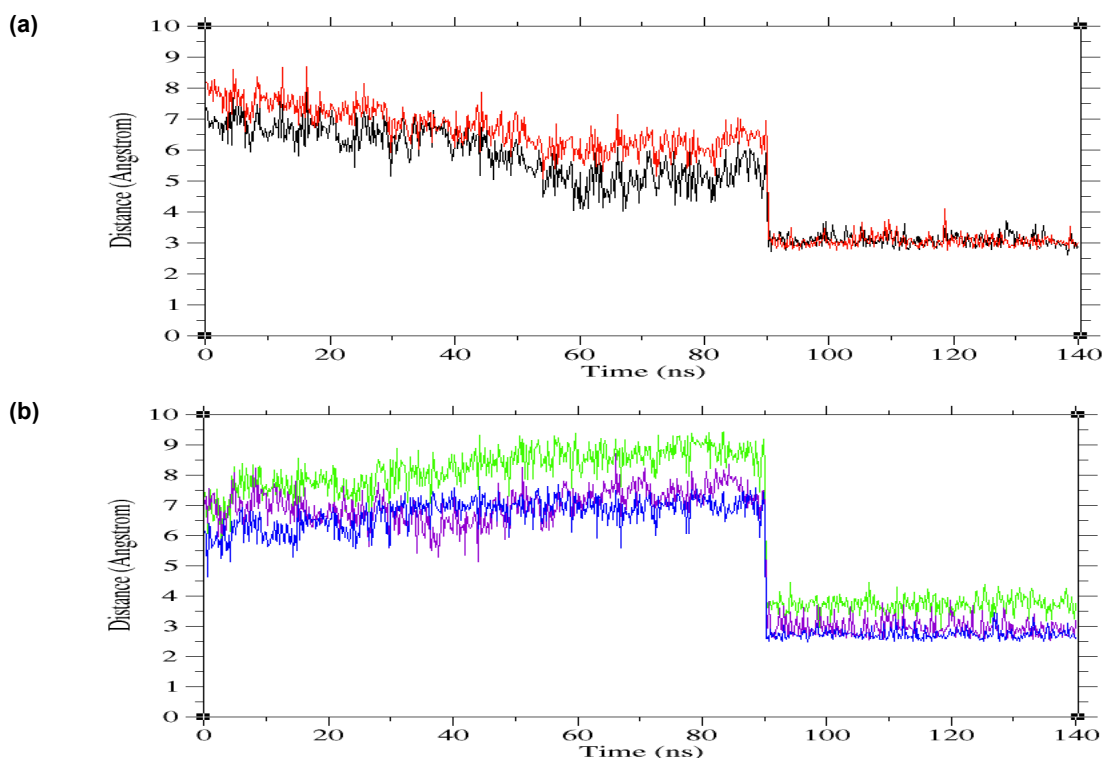
Figure 4.34: Time dependence of the distances in the water-mediated hydrogen bonding network for simulation **II**. Distance between the oxygen atom of the water molecule WAT and: backbone amide nitrogen of Phe965 (a, red line); backbone amide nitrogen of Gly966 (a, black line); Asp841 Oδ2 (b, blue line); backbone amide oxygen of Ile963 (b, purple line); and the ligand protonated oxygen (b, green line).

4.2.4.3 Simulation III

Turning now to the PI3Kγ/*s*-trans inhibitor **1** system (i.e. simulation **III**), the resulting water density map reveals an area of persisting water in the active site, adjacent to the protonated ligand oxygen (Figure 4.35a). This area of water density is attributed to a single persisting water molecule that forms a water network with the ligand protonated oxygen and residues Asp841 and Ile963 (Figure 4.35b).
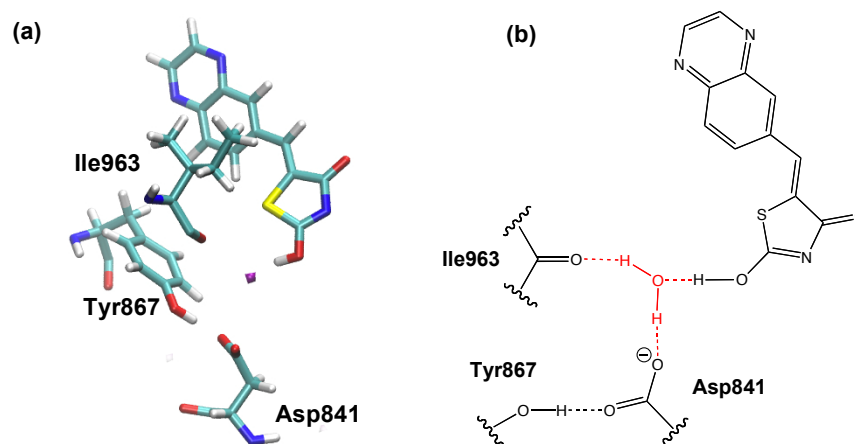
Figure 4.35: Persisting water site in the binding site of the PI3Kγ/*s*-trans inhibitor **1** system. (a) Water density map. The area of localised water presence is shown in purple. (b) Predicted water-mediated hydrogen bonding network (in red dashed lines). The hydrogen bond between Asp841 and Tyr867 is also shown (black dashed line).

The localised water molecule hydrogen bonds to the protonated ligand oxygen and to Ile963 with very similar occupancies (approximately 18%). Furthermore, the occupancy for the interactions with Oδ1 and Oδ2 of Asp841 are 18.3 and 13.5%, respectively (Table 4.15).

Table 4.15: Hydrogen bond occupancies (%) and average distances (Å) of key interactions involving water molecule WAT for simulation **III**.

Donor	Acceptor	Simulation III	
		%occ	D_{HB} (Å)
WAT O-H1	Asp841 Oδ1	10.3	2.88 (0.24)
WAT O-H2	Asp841 Oδ1	8.0	2.90 (0.22)
WAT O-H1	Asp841 Oδ2	7.1	3.39 (0.30)
WAT O-H2	Asp841 Oδ2	6.4	3.38 (0.27)
WAT O-H1	Ile963 O	7.9	2.82 (0.16)
WAT O-H2	Ile963 O	10.3	2.80 (0.12)
Ligand OH	WAT O	18.3	2.64 (0.08)

D_{HB} average distances when in a hydrogen bond (Å); and percentage of occupancy (%occ). The analysis of hydrogen bonds is taken as donor-acceptor heavy-atom distance ≤ 4.0 Å and with an angle cutoff of 120°. The standard deviations (Å) are in parentheses.

The identified area of water density corresponds to a non-crystallographic water, which, as can be seen in Figure 4.36, drifts into position towards the end of the simulation at 114 ns. An interesting observation is that the arrival of the water molecule disturbs a previously stable hydrogen bond between Asp841 and the

128

protonated ligand oxygen (Figure 4.37). In the hydrogen bonding rearrangement, Asp841 receives a proton from the water molecule and forms only a water-mediated interaction with the ligand.

(a)

(b)
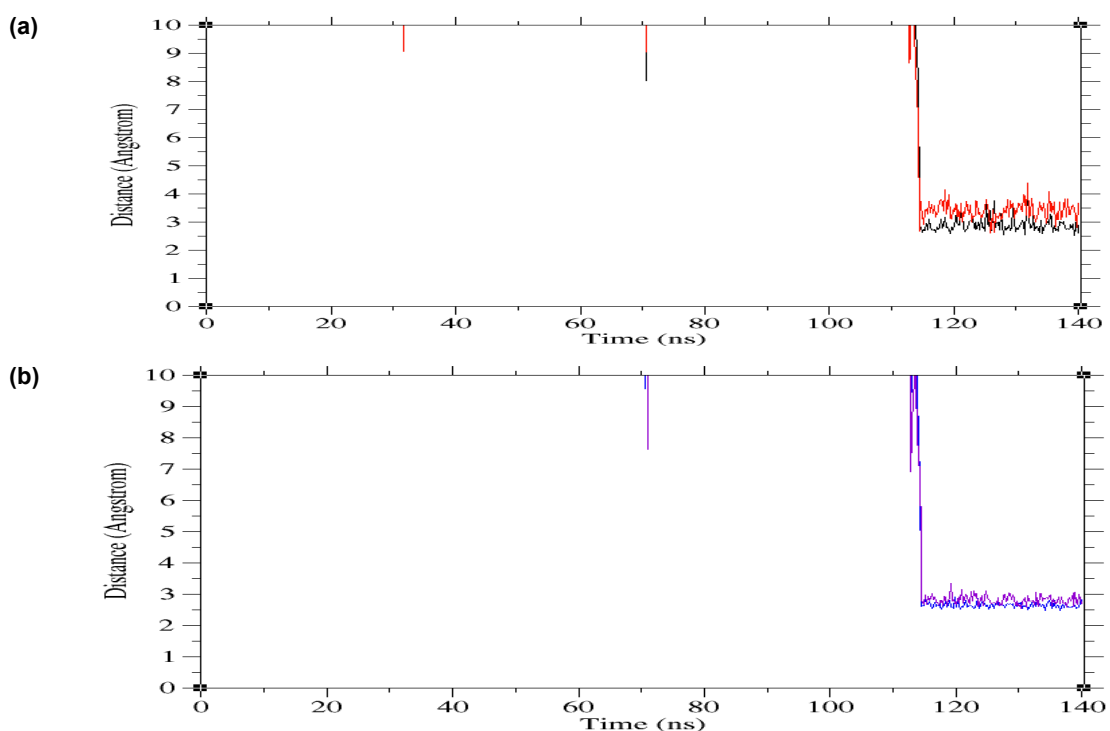
Figure 4.36: Time dependence of the distances in the water-mediated hydrogen bonding network for simulation **III**. Distance between the oxygen atom of the water molecule WAT and: Asp841 Oδ1 (a, black line); Asp841 Oδ2 (a, red line); backbone amide oxygen of Ile963 (b, purple line); and the protonated ligand oxygen (b, blue line).

(a)

(continued on the next page)

129

(b)
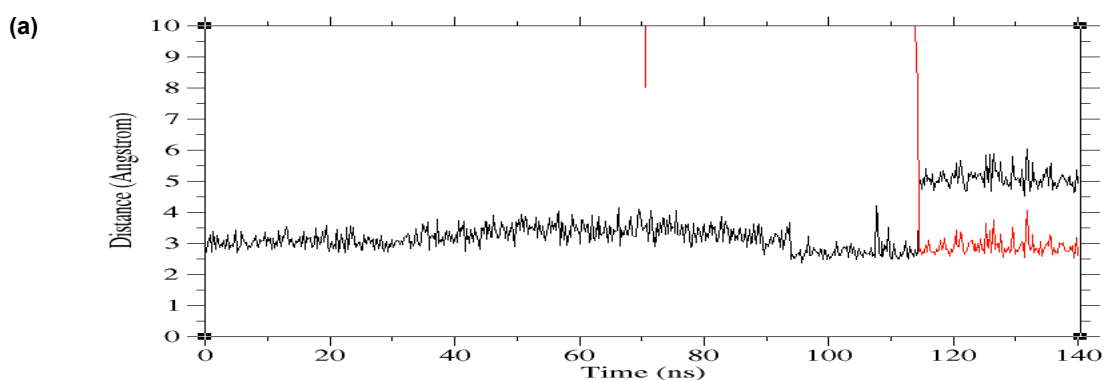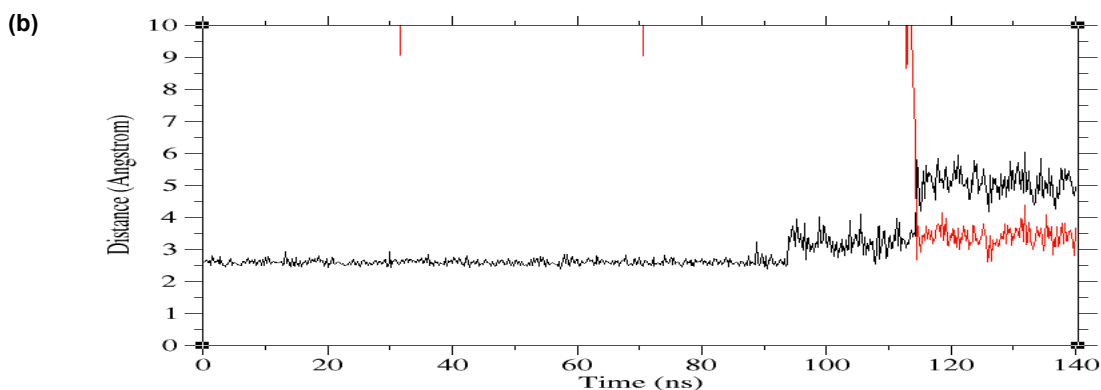
Figure 4.37: Time dependence of the distance between the ligand protonated oxygen and: Asp841 Oδ1 (a, black line); water molecule (a, red line); Asp841 Oδ2 (a, red line); and water molecule (b, red line).

4.2.4.4 Simulation IV

For the last complex, namely the PI3Kγ/*s*-trans inhibitor **2** system (i.e. simulation **IV**), the resulting water density map indicates one area of persisting water density adjacent to residues Asp841, Tyr867 and Ile963 (Figure 4.38a). The network of water-mediated hydrogen bonds formed in this region is illustrated in Figure 4.38b. Significantly, the resident water molecule does not form direct hydrogen bonds with the ligand.
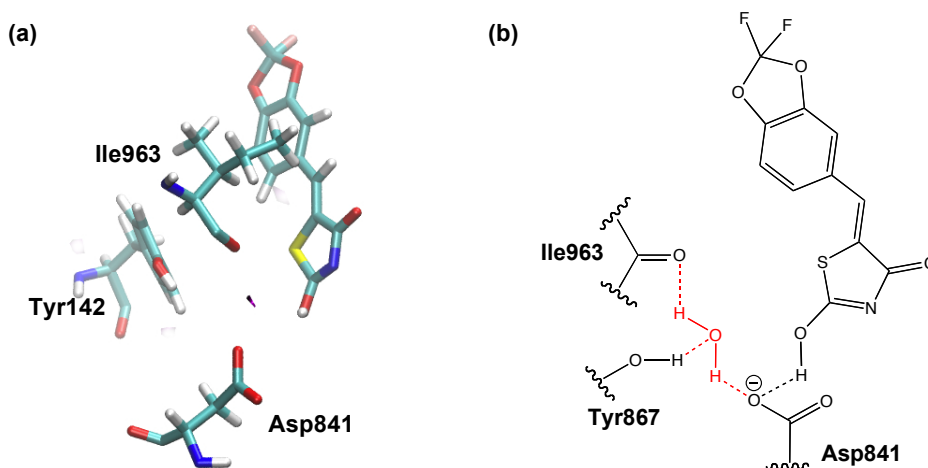
Figure 4.38: Persisting water site in the binding site of the PI3Kγ/*s*-trans inhibitor **2** system. (a) Water density map. The area of localised water presence is shown in purple. (b) Predicted water-mediated hydrogen bonding network (in red dashed lines). The hydrogen bond between Asp841 and Tyr867 is shown in black dashed lines.

The identified area of water density for the PI3Kγ/*s*-trans inhibitor **2** system lies in a similar location to the one presented by the water site described for the PI3Kγ/*s*-trans inhibitor **1** system. However, a comparison of Table 4.15 and Table 4.16 reveals that

130

the localised water molecule for the PI3Kγ/*s*-trans inhibitor **2** system has shorter occupancies. Another crucial difference is that in the PI3Kγ/*s*-trans inhibitor **1** system, the key hydrogen bond between the protonated ligand oxygen and Asp841 is broken by the arrival of the water molecule (see Figure 4.37); whereas in the PI3Kγ/*s*-trans inhibitor **2** system, the localised water molecule does not bind directly to the ligand, and the hydrogen bond between the protonated ligand oxygen and Asp841 is conserved.

The identified area of water density is attributable to a non-crystallographic water, which drifts into position towards the end of the simulation trajectory (Figure 4.39).

Table 4.16: Hydrogen bond occupancies (%) and average distances (Å) of key interactions involving water molecule WAT for simulation **VI**.

Donor	Acceptor	Simulation IV	
		%occ	D_{HB} (Å)
Tyr867 OH	WAT O	13.0	2.84 (0.18)
WAT O-H1	Ile963 O	6.1	2.79 (0.15)
WAT O-H2	Ile963 O	6.3	2.75 (0.14)
WAT O-H1	Asp841 Oδ2	3.3	2.83 (0.26)
WAT O-H2	Asp841 Oδ2	2.3	2.89 (0.25)

D_{HB} average distances when in a hydrogen bond (Å); and percentage of occupancy (%occ). The analysis of hydrogen bonds is taken as donor-acceptor heavy-atom distance ≤ 4.0 Å and with an angle cutoff of 120°. The standard deviations (Å) are in parentheses.
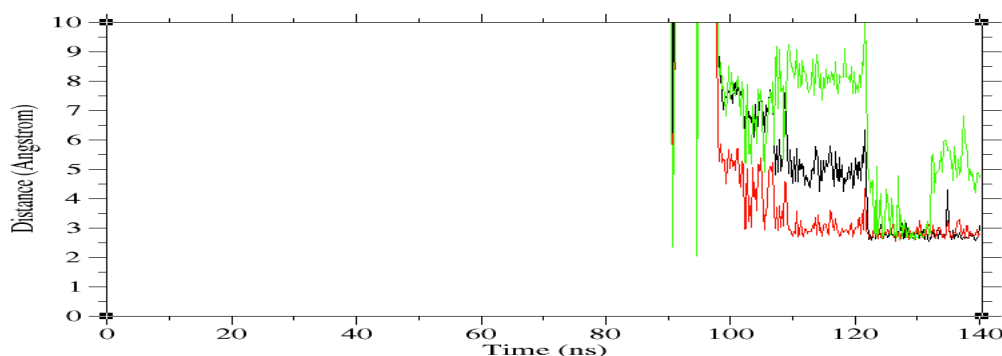

Figure 4.39: Time dependence of the distances in the water-mediated hydrogen bonding network for simulation **IV**. Distance between the oxygen atom of the water molecule WAT and: side-chain oxygen atom of Try867 (red line); backbone amide oxygen of Ile963 (black line); and Asp841 Oδ2 (green line).

From the analysis presented in this section, we may conclude that, for the systems involving the *s*-cis inhibitors (i.e. simulations **I** and **II**), the solvent in the active site plays a role in assisting ligand binding. The identified area of persisting water density lies in analogous positions for both simulations, and participates in very similar

hydrogen bonding networks, which include a stable interaction with the protonated oxygen in the thiazolidinedione head of the ligand. With regards to the systems involving the *s*-trans inhibitors (i.e. simulations **III** and **IV**), it is also possible to obtain information on the structuring of water in the active site. However, the identified water sites have lower occupancies.

4.2.5 Conclusions

The work described in this chapter aimed to provide a dynamical insight into the subtle basis of molecular recognition by PI3Kγ. Four stable fully solvated molecular dynamics simulations of PI3Kγ complexed with the four different inhibitors were performed. Based on the RMS deviations, RMSF profiles, temperatures and qualitative observation of the molecular dynamics trajectories, it was concluded that all four simulations are stable and suitable for investigation into the basis of molecular recognition by PI3Kγ.

The findings discussed in this chapter indicate that the structures of the four different protein-ligand systems sampled by MD simulations are quite similar at a global level (Section 4.2.1). However, larger differences are observed at the level of the ATP-binding pocket (Section 4.2.2). It was also found that the conformations sampled by MD are predominantly consistent with those reported by the X-ray crystal structures.

The evidence from this study suggests that the ligand and the residues involved in the key hydrogen bond interactions (Lys833, Asp841 and Val882) display lower conformational mobility during simulations **I** and **II** (i.e. systems with the *s*-cis inhibitors) than during simulations **III** and **IV** (i.e. systems involving the *s*-trans inhibitors).

One of the most interesting observations from the MD simulations is the movement of the DFG aspartate Asp964 (located at the beginning of the activation loop). The available crystallographic evidence (e.g. lack of resolved density and high B-factors) supporting this movement has been discussed. Furthermore, the static X-ray crystal structures are unlikely to accurately represent the dynamic nature of proteins. This is particularly true for regions characterised by an elevated degree of flexibility such as the activation loop of PI3Kγ. Thus, the differences in the activation loop dynamics observed during the MD simulations could be attributed to PI3Kγ exploring different parts of the normally accessible phase space over the timescale of the simulations.

The results reported in this chapter suggested that for the systems involving the *s*-cis inhibitors (i.e. simulations **I** and **II**), Asp964 moves away from the ligand. However, the rest of the DFG loop, namely Phe965 and Gly966, remains pointing towards the ligand. By comparison, for the systems comprising the *s*-trans inhibitors (i.e. simulations **III** and **IV**), the Asp964 displacement has larger fluctuations of up to 5 Å (refer to Figure 4.22c and 4.23c).

Throughout the four different MD simulations, the hydrogen bond interactions to Lys833, Asp841 and Val882 are maintained very well. Moreover, the cavity lined by the hydrophobic residues is well conserved during the trajectories (Section 4.2.3). The following two conclusions can be drawn from the evaluation and comparison of the hydrogen-bonding profiles for the four different trajectories:

Firstly, the hydrogen-bonding profile of the PI3Kγ/*s*-cis inhibitor **1** system (simulation **I**) is remarkably similar to the one presented by the PI3Kγ/*s*-cis inhibitor **2** system (simulation **II**). The key difference is that *s*-cis inhibitor **1** formed one hydrogen bond between its quinoxaline nitrogen and the backbone NH of Val882, while *s*-cis inhibitor **2** formed two hydrogen bonds with Val822 (one to the 1,3-benzodioxole ring oxygen and the other to a fluorine atom). Likewise, the hydrogen-bonding profile of the PI3Kγ/*s*-trans inhibitor **1** system (simulation **III**) is similar to the one presented by the PI3Kγ/*s*-trans inhibitor **2** system (simulation **IV**). Again, the key difference resides in the number of hydrogen bonds to Val882. Essentially, when comparing the same conformations of the inhibitors (namely simulation **I** *vs.* simulation **II**; and simulation **III** *vs.* simulation **IV**) the data from this study highlights the importance of the polar ligand-Val882 interactions.

Secondly, a comparison of the hydrogen-bonding profiles of the *s*-cis and *s*-trans conformations for each inhibitor (namely simulation **I** *vs.* simulation **III**; and simulation **II** *vs.* simulation **IV**) reveals that the *s*-cis isomers form more stable hydrogen bond interactions with PI3Kγ.

Because the MD simulations were performed in explicit solvent (TIP3P), it was possible to investigate the role played by solvent in the active site (Section 4.2.4). By using water density maps it was possible to identify one region of persisting water on each simulation model. The identified area of persisting water density in simulations **I** and **II** lies in analogous positions, and participates in very similar hydrogen bonding networks which include a stable interaction with the protonated oxygen in the thiazolidinedione head of the ligand. With regards to simulations **III** and **IV**, it was also

possible to obtain information on the structuring of water in the active site. However, the identified water site present lower occupancies.

Taken together, the findings reported in this chapter indicate that all four MD simulations are stable and suitable for use in the energetic calculations that will be discussed in chapter 5.

5. Energetic calculations

In this chapter, the main goal is to compare the binding affinity of PI3Kγ with a series of inhibitors (Table 5.1). For this purpose, the first objective is to calculate the binding free energy of the four different protein-ligand systems using the MM-PBSA method. Subsequently, to evaluate the contribution of each residue to the binding free energy, an energy decomposition analysis is performed. Following this, molecular mechanical and quantum chemical treatments are utilised to examine the crucial ligand-Val882 interaction.

Table 5.1: Composition of each MD simulation system.

Simulation	Protein-ligand model
I	PI3Kγ /*s*-cis inhibitor **1**
II	PI3Kγ /*s*-cis inhibitor **2**
III	PI3Kγ /*s*-trans inhibitor **1**
IV	PI3Kγ /*s*-trans inhibitor **2**

5.1 Methods

5.1.1 MM-PBSA analysis of binding free energies

The binding free energies for the PI3Kγ systems described in the previous chapter, and summarised in Table 5.1, were calculated using the MM-PBSA method[112], as implemented in the MMPBSA.py program[155] of AMBER 11[131]. A single trajectory approach was used; namely, the energies of the protein-ligand complexes, together with the energies of the free protein and the free ligand, were calculated from snapshots of the same simulation. For each protein-ligand system, a total of 35 equispaced snapshots were used. These were taken from the 140 ns MD trajectories with an interval of 4 ns. The analysis did not include an estimation of the solute entropy contributions to the binding free energy, as it was assumed these would be fairly similar across all complexes. This is a common approximation in calculations involving a set of different ligands binding to the same protein[156].

The molecular mechanics contribution to the binding free energy was calculated with the AMBER *ff99SB* force field[86] as used in the MD simulations. The electrostatic contribution to solvation was calculated by solving the Poisson-Boltzmann equation[157]

using the *pbsa* module of AMBER 11[131]. All calculations employed a solute internal dielectric constant of 1.0 and an external dielectric constant of 80.0. The modified Bondi radii[158, 159] were used to represent solute atoms. A probe radius of 1.4 Å was used to define both the dielectric boundary between the solute and solvent, as well as the solvent accessible surface of the solute. The PB equation was solved numerically using 1000 linear steps of finite difference. The non-polar contribution to the solvation free energy was calculated with a solvent-accessible surface area (SASA) dependent term with a coefficient of 0.0072 kcal·mol^{-1}Å^{-2}.

5.1.2 Per-residue free energy decomposition analysis

To obtain a detailed view of the protein-ligand binding and to identify the key residues responsible for the binding process, free-energy decomposition to each residue was performed. For this purpose, a per-residue decomposition based on MM-PBSA was used. The MM-PBSA analysis was performed using the same methodology as in Section 5.1.1. For each protein-ligand system, a total of 35 snapshots were used. These were taken from the 140 ns MD trajectories with an interval of 4 ns.

For decomposition, the contribution of each residue includes four terms: electrostatic contribution (ΔG_{elec}), van der Waals contribution (ΔG_{vdw}), polar solvation contribution (ΔG_{PB}), and non-polar solvation contribution (ΔG_{np}):

$$\Delta G_{residue} = \Delta G_{elec} + \Delta G_{vdw} + \Delta G_{PB} + \Delta G_{np} \qquad (5.1)$$

However, the PBSA surface area terms are not decomposable. Therefore, for each residue, the non-polar solvation contribution, which is calculated based on SASA, is reported as zero.

5.1.3 QM analysis of key hydrogen bond to Val882

Quantum mechanics (QM) calculations were used to examine the key ligand-Val882 interaction. The initial atomic coordinates were obtained from a total of 35 equispaced snapshots from each of the 140 ns MD trajectories; namely, the same ensembles of structures as the ones used for the MM-PBSA calculations. QM calculations were only carried out for the simulation systems involving the *s*-cis isomers of inhibitors **1** and **2** (i.e. simulation **I** and **II**). The QM interaction energy was evaluated using only the bicyclic moieties of the inhibitors (i.e. the quinoxaline moiety of *s*-cis inhibitor **1** and

the 1,3-benzodioxole ring of *s*-cis inhibitor **2**) and a protein fragment containing the peptide bond Ile881-Val882 capped with hydrogen atoms (Figure 5.1).
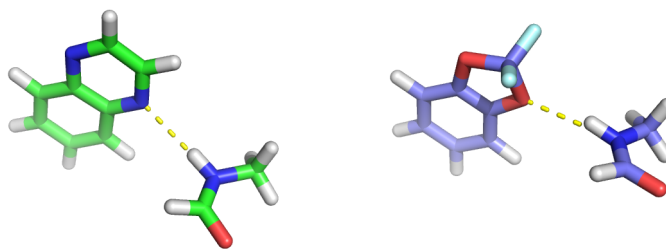
Figure 5.1: Bicyclic moieties of *s*-cis inhibitor **1** (left, in green) and *s*-cis inhibitor **2** (right, in purple) and corresponding protein fragments (containing peptide bond Ile881-Val882) treated quantum mechanically at the M06L/MG3S level in order to study the key Val882-ligand interaction.

In order to allow a direct comparison with MD geometries, the QM calculations were energy optimisations with all heavy atoms frozen. These were performed utilising the Gaussian09 package[114]. For all calculations, the M06L functional of Truhlar *et al.*[115] was used, along with the MG3S basis set[116]. The MG3S basis set is identical to the MG3 basis set[117,118] except that in the former, diffuse functions on hydrogen are absent; and the MG3 basis set is the same as 6-311++G(3*d2f*,2*df*,2*p*)[119,120] for H–Si systems, but improved[121] for P–Ar systems.

The interaction energy calculations were performed with counterpoise corrections[160, 161] to account for basis set superposition error (BSSE). In a dimer *AB*, the typical uncorrected interaction energy between monomers *A* and *B* would be calculated as:

$$\Delta E_{int} = E_{AB}^{AB}(AB) - E_A^A(A) - E_B^B(B) \qquad (5.2)$$

where the superscripts denote the basis set used, the subscripts denote the geometry, and the symbol in parentheses denotes the chemical system considered. Accordingly, $E_{AB}^{AB}(AB)$ represents the total energy of the dimer *AB* calculated with the full basis set *AB* of the dimer at that geometry. Notably, monomers *A* and *B* are each evaluated at their own geometries in their own basis sets. This results in a phenomenon called basis set superposition error. As monomer *A* approaches monomer *B*, the dimer *AB* can be artificially stabilised as monomer *A* utilises the extra basis functions from monomer *B* to describe its electron distribution, and vice versa. Thus without BSSE correction interaction energies will often be overestimated[162].

The counterpoise corrected interaction energy is:

$$\Delta E_{int}^{CP} = E_{AB}^{AB}(AB) - E_A^{AB}(A) - E_B^{AB}(B) \tag{5.3}$$

where $E_A^{AB}(A)$ and $E_B^{AB}(B)$ denote the total energies of monomers A and B, respectively, computed with the dimer basis set AB. The counterpoise correction is realised by introducing so-called ghost orbitals, namely, the ghost subsystem has basis set functions placed at the position of its atoms but no electrons or atomic charges.

5.2 Results and discussion

5.2.1　MM-PBSA analysis of binding free energies

To compare the binding affinity of the different ligands with PI3Kγ, the binding free energies (ΔG_{bind}) for the four different protein/ligand systems were calculated using the MM-PBSA method (Table 5.2).

As indicated in Table 5.2, PI3Kγ binds most strongly to inhibitor **1** in its *s*-cis isoform (simulation **I**), since this system displays the lowest calculated ΔG_{bind} (-26.9 ± 2.6 kcal·mol^{-1}) of the four ligands. Conversely, of the four systems, the weakest binding occurs between PI3Kγ and *s*-trans inhibitor **1** (simulation **III**). This protein-ligand system displays the highest calculated ΔG_{bind} (-18.9 ± 5.9 kcal·mol^{-1}), which corresponds to decreased binding affinity.

It is interesting to note that for both inhibitors the *s*-cis isomers were predicted to be the most favourable conformations, since their calculated binding energies were lower than the calculated values for their *s*-trans counterparts. For inhibitor **1**, the difference in calculated binding free energy between the *s*-cis and *s*-trans isomers is -8.0 kcal·mol^{-1}. On the other hand, for inhibitor **2**, the difference in calculated binding free energy between the *s*-cis and *s*-trans conformations was -2.9 kcal·mol^{-1}.

Table 5.2: Binding free energy calculations for the four different PI3Kγ complexes.

Simulation	Ligand	[a] ΔG_{bind}		[b] ΔG_{elec}		[c] ΔG_{vdw}		[d] ΔG_{PB}		[e] ΔG_{np}	
I	*s*-cis **1**	-26.9	(2.6)	-45.5	(5.4)	-34.8	(3.4)	55.6	(4.5)	-2.2	(0.1)
II	*s*-cis **2**	-24.9	(2.4)	-42.9	(4.4)	-33.6	(3.1)	53.8	(3.5)	-2.2	(0.1)
III	*s*-trans **1**	-18.9	(5.9)	-52.2	(7.9)	-31.6	(2.8)	67.1	(5.2)	-2.2	(0.1)
IV	*s*-trans **2**	-22.0	(8.6)	-35.8	(17.0)	-35.6	(2.5)	51.6	(10.3)	-2.2	(0.1)

All energies are in kcal·mol^{-1} and were calculated using 35 equispaced snapshots from 140 ns MD simulations. Standard deviations are in parentheses. [a] Total MM-PBSA binding free energy, excluding entropic contributions. [b] Electrostatic contribution to the binding free energy. [c] van der Waals contribution to the binding free energy. [d] Poisson-Boltzmann electrostatic solvation contribution to the binding free energy. [e] Total non-polar contribution to the binding free energy.

The computed energetics also suggest that *s*-cis inhibitor **1** is a stronger binder than *s*-cis inhibitor **2**, with an average binding energy of -26.9 ± 2.6 kcal·mol⁻¹ (simulation **I**) and -24.9 ± 2.4 kcal·mol⁻¹ (simulation **II**), respectively. Furthermore, the difference in calculated binding affinity is -2.0 kcal·mol⁻¹, which is consistent with the trend observed experimentally (-1.9 kcal·mol⁻¹, Table 5.3). Given the assumptions in the approach, the quantitative agreement with the experimental value of -1.9 kcal·mol⁻¹ is somewhat fortuitous.

The trend of experimental activities for inhibitors **1** and **2** was derived from K_i values taken from the literature[30]. For direct comparison with the calculated affinities, ΔG_{exp} was estimated using the following equation:

$$\Delta G_{exp} \approx -RT \ ln(K_i) \tag{5.4}$$

where *R* is the ideal gas constant (1.985 x 10⁻³ kcal·K⁻¹mol⁻¹) and *T* is the temperature (in K). The estimated ΔG_{exp} values are presented in Table 5.3.

Table 5.3: Estimated experimental (ΔG_{exp}) binding affinities.

Inhibitor	K_i (µM)	ΔG_{exp} (kcal·mol⁻¹)
1	0.0078	-11.0
2	0.1800	-9.1

It is important to emphasise that the experimental K_i values reported by Camps *et al.*[30] most likely correspond to a mixture of the *s*-cis and *s*-trans isomers of each bound and unbound inhibitor. However, as it has been discussed in chapter 3, on the basis of the available crystallographic data, for inhibitors **1** and **2** the *s*-cis conformation is found to be the preferred one. Hence, we assume the estimated ΔG_{exp} values in Table 5.3 are the experimental binding free energies for the *s*-cis isoforms.

It should be pointed out that, whereas the relative magnitudes of the calculated binding free energies for the *s*-cis isomers (ΔG_{bind}, Table 5.2) and the binding free energies derived from experimental values (ΔG_{exp}, Table 5.3) agree well, their absolute values do not. Such difference is a known behaviour already observed in many other investigations, and it can be ascribed to the omission of the entropic contribution (which is a typical approximation for these studies). However, the aim of this work was not to predict absolute free energies of binding in quantitative agreement with experiment. Rather, the aim was to examine the basis on which PI3Kγ distinguishes the different inhibitors. In this context, comparison between calculated

and experimental data show that MD simulations coupled with MM-PBSA calculations were able to reproduce the trend in affinity of *s*-cis inhibitor **1** and *s*-cis inhibitor **2** for PI3Kγ.

Further analysis of the energy terms and the binding free energy values (Table 5.2) reveals that electrostatic interactions (ΔG_{elec}) are favourable to binding. This is in good agreement with the structural and hydrogen bond analysis presented in Section 4.2.3. Additionally, the intermolecular van der Waals interactions (ΔG_{vdw}) were also favourable to binding. This is in agreement with the fact that the inhibitors have significant hydrophobic-aromatic portions interacting favourably with hydrophobic residues of the protein, i.e. the quinoxaline ring of inhibitor **1** and the 1,3-benzodioxole ring of inhibitor **2**; as well as the thiazolidinedione moiety for both cases (refer to Figures 4.16 to 4.19, panel c).

The electrostatic part of solvation free energy (ΔG_{PB}) opposes the binding by unfavourable contributions. The complexes involving inhibitor **1** (i.e. simulations **I** and **III**) exhibit higher ΔG_{PB} than the systems containing inhibitor **2** (i.e. simulations **II** and **IV**). Thus, this indicates that desolvation presents greater opposition to binding for inhibitor **1** than for inhibitor **2**. Lastly, the non-polar solvation energies (ΔG_{np}), which are responsible for the burial of the solvent accessible surface area (SASA) upon binding, contribute slightly favourably. All four systems present the same ΔG_{np} value, -2.2 ± 0.1 kcal·mol^{-1}.

By comparing the ΔG_{bind} standard deviations obtained for the four different protein-ligand systems, we can make an interesting observation. The complexes involving the *s*-trans inhibitors (i.e. simulation **III** and **IV**) exhibit higher standard deviation values, which indicates higher variability of the structural ensemble. These results are consistent with the previously made observations that *s*-trans systems presented larger conformational mobility in the active site (refer to Figures 4.26 to 4.29, panel a).

A comparison of the binding free energy contributions for simulation **I** (PI3Kγ/*s*-cis inhibitor **1**) and simulation **II** (PI3Kγ/*s*-cis inhibitor **2**) reveals that the electrostatic energy (ΔG_{elec}) in the *s*-cis inhibitor **1** complex was 2.6 kcal·mol^{-1} more favourable than that of the *s*-cis inhibitor **2** complex. However, this favourable electrostatic interaction is counteracted by the solvation effect. The total electrostatic contribution (ΔG_{elec} + ΔG_{PB}) was 10.1 kcal·mol^{-1} for the *s*-cis inhibitor **1** complex and 10.9 kcal·mol^{-1} for the *s*-cis inhibitor **2** complex. Therefore, after offset the difference in the electrostatic interactions decreased to 0.8 kcal·mol^{-1}. On the other hand, the van der Waals

interaction energy in the *s*-cis inhibitor **1** complex is 1.2 kcal·mol^{-1} more favourable than that of the *s*-cis inhibitor **2** complex (Table 5.2). This calculated difference in van der Waals interaction together with the fact that both inhibitors present very similar binding modes suggests that the quinoxaline moiety of *s*-cis inhibitor **1** interacts more favourably with the hydrophobic residues of the PI3Kγ active site than the 1,3-benzodioxole ring of *s*-cis inhibitor **2**.

As Table 5.2 shows, there is a significant difference (8.0 kcal·mol^{-1}) between the calculated binding free energy for simulation **I** (PI3Kγ/*s*-cis inhibitor **1**) and that for simulation **III** (PI3Kγ/*s*-trans inhibitor **1**). Interestingly, simulation **III** has the smallest, and therefore most favourable, predicted electrostatic energy contribution (-52.2 ± 7.9 kcal·mol^{-1}). However, the favourable Coulombic interactions within the protein-inhibitor complex were counteracted by a very unfavourable electrostatics of desolvation (67.1 ± 5.2 kcal·mol^{-1}). The total electrostatic contribution (ΔG_{elec} + ΔG_{PB}) in the *s*-cis inhibitor **1** complex is 4.8 kcal·mol^{-1} more favourable than that of the *s*-trans inhibitor **1** complex. The van der Waals interaction energies between PI3Kγ and the *s*-cis inhibitor **1**/*s*-trans inhibitor **1** are -34.8 ± 3.4 and 31.6 ± 2.8 kcal·mol^{-1}, respectively. In other words, the *s*-trans inhibitor **1** model leads to a large increase (by 3.2 kcal·mol^{-1}) in the van der Waals energy. This could partly be a result of the larger fluctuations of the *s*-trans inhibitor **1** in the active site (refer to Figure 4.28), which may cause the hydrophobic-aromatic portions of the ligand to interact less favourably with the hydrophobic residues of the PI3Kγ. A more detailed possible explanation for this observed difference in van der Waals energy will be given in the following section.

When analysing the binding energy contributions for simulation **IV** (PI3Kγ/*s*-trans inhibitor **2**) it is interesting to note that the van der Waals contribution to binding is almost equal to the electrostatic contribution (-35.6 ± 2.5 and 35.8 ± 17.0 kcal·mol^{-1}, respectively). By comparing the binding free energy contributions for simulation **II** (PI3Kγ/*s*-cis inhibitor **2**) and simulation **IV** (PI3Kγ/*s*-trans inhibitor **2**) it is found that the electrostatic energy (ΔG_{elec}) in the *s*-cis inhibitor **2** complex is 7.1 kcal·mol^{-1} more favourable than that of the *s*-trans inhibitor **2** complex. This result is partly due to the fact that the key hydrogen bond interactions (Asp841, Lys833 and Val882) between the protein and the ligand are better conserved in the *s*-cis inhibitor **2** model (Section 4.2.3). The total electrostatic contribution (ΔG_{elec} + ΔG_{PB}) in the *s*-trans inhibitor **2** complex is 4.9 kcal·mol^{-1} more than that of the *s*-cis inhibitor **2** complex. On the other hand, the van der Waals interaction energy in the *s*-trans inhibitor **2** complex is 2.0 kcal·mol^{-1} less than that of the *s*-cis inhibitor **2** complex. A conceivable explanation for

this result is that the location of the thiazolidinedione moiety of *s*-trans inhibitor **2** allows the formation of more favourable packing of the hydrophobic residues of the PI3Kγ active site.

5.2.2 Per-residue free energy decomposition analysis

To gain further insight into the contributions of each residue to the protein-ligand interactions, the binding free energy in the PI3Kγ complexes was decomposed per residue using the MM-PBSA approach. Free energy decomposition can provide information about the local interactions of a system in addition to global free energies[163]. Therefore it is a valuable tool that can be used to identify the key residues responsible for the binding affinity of inhibitors. This knowledge can then be applied to structure-based drug design.

The results of the decomposition analysis for each of the four protein-ligand systems are graphically presented in Figure 5.2. Only the residues with energy contributions over 0.5 kcal·mol^{-1} or under -0.5 kcal·mol^{-1} are labelled. Additionally, this decomposition of energy terms and the binding free energy for the key residues with energy contributions over 0.5 kcal·mol^{-1} or under -0.5 kcal·mol^{-1} are presented in Tables 5.4 to 5.7. It should be noted that the PB non-polar solvation energies are not yet decomposable within the MMPBSA.py program[155] of AMBER 11[131]. Consequently, they are not reported in the tables below. Because of this, this data must be interpreted with caution.

To further aid the discussion into contributions of the most relevant residues to binding free energy, comparisons between the representative structures of the different simulation trajectories are presented below. The aligned snapshots are shown in Figures 5.3 to 5.5. Note that the representative structure corresponds to the snapshot closest to the average structure over all the MD configurations.

It is apparent from the graphs presented in Figure 5.2 that, in all four protein-ligand complexes, the major favourable energy contributions originate predominantly from residues Met804, Ile831, Lys833, Asp841, Ile879, Ile881, Val882, Met953, Phe961 and Ile963. Conversely, residue Glu880 contributes unfavourably to the binding in all four protein-ligand systems.
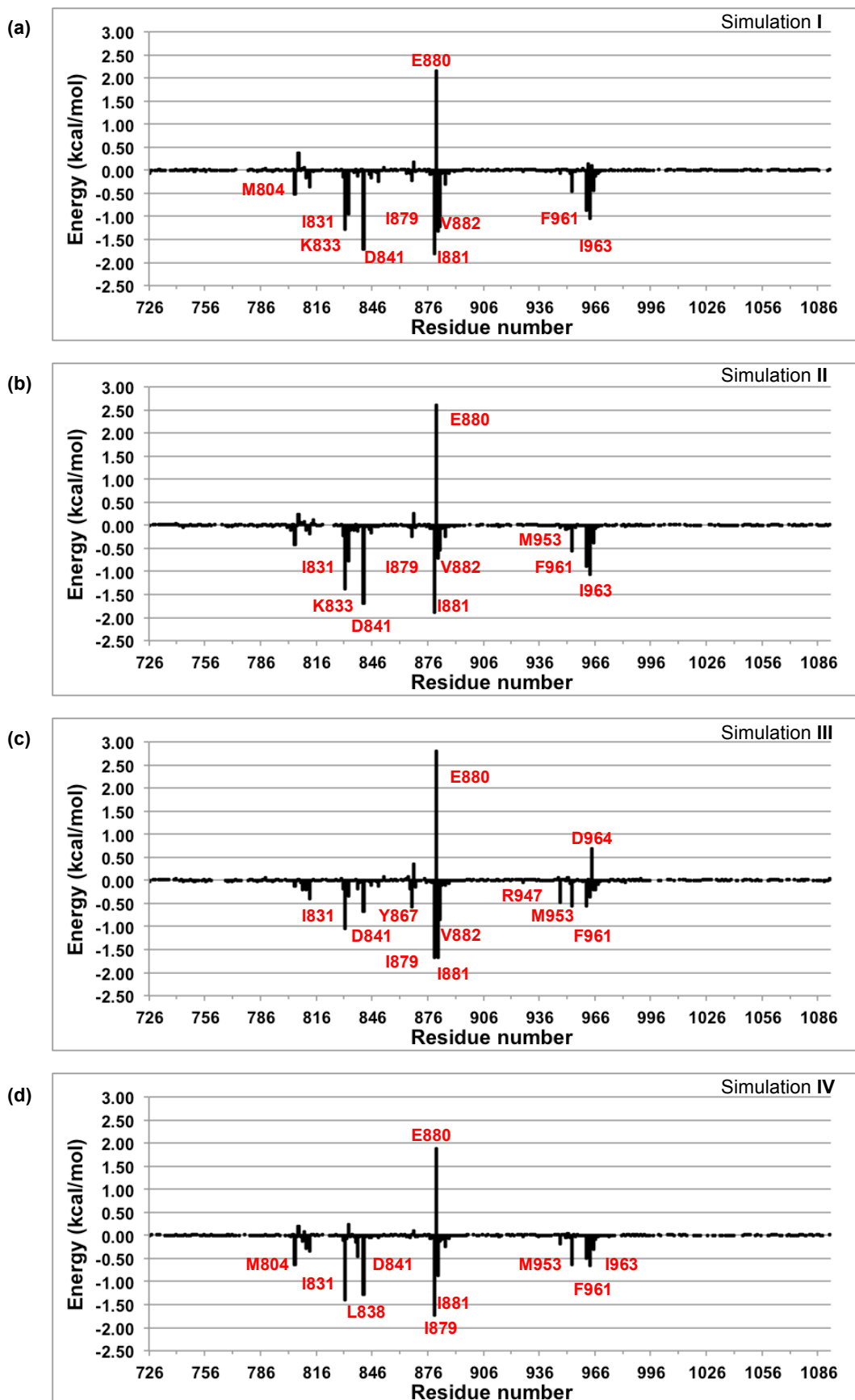
Figure 5.2: Per-residue energy decomposition for simulations **I** to **IV**. Residues with energy contributions over 0.5 kcal·mol^{-1} or under -0.5 kcal·mol^{-1} are labelled.

Table 5.4: Energy contributions of PI3Kγ residues to the binding of *s*-cis inhibitor **1** (data from Simulation **I**).

Residue		[a] ΔG_{bind}		[b] ΔG_{elec}		[c] ΔG_{vdw}		[d] ΔG_{PB}	
Ile	879	-1.8	(5.4)	0.3	(2.5)	-1.7	(2.0)	-0.4	(0.5)
Asp	**841**	**-1.7**	**(9.4)**	**-8.1**	**(6.3)**	**0.9**	**(1.8)**	**5.5**	**(5.1)**
Ile	881	-1.3	(8.5)	-0.4	(5.1)	-1.4	(1.5)	0.4	(1.2)
Ile	831	-1.3	(8.5)	-0.2	(5.0)	-1.3	(1.6)	0.2	(0.4)
Val	**882**	**-1.2**	**(5.4)**	**-1.6**	**(2.9)**	**-1.0**	**(1.2)**	**1.4**	**(1.3)**
Ile	963	-1.1	(4.7)	0.1	(1.8)	-2.7	(1.5)	1.5	(1.1)
Lys	**833**	**-1.0**	**(12.7)**	**-7.8**	**(8.8)**	**-0.8**	**(2.1)**	**7.7**	**(8.0)**
Phe	961	-0.9	(4.3)	0.1	(1.6)	-1.0	(1.7)	0.1	(0.4)
Met	804	-0.5	(5.1)	0.1	(3.0)	-0.8	(1.3)	0.2	(1.8)
Glu	880	2.2	(34.8)	-1.2	(23.6)	-0.5	(1.6)	3.8	(25.3)

Showing residues with energy contributions over 0.5 kcal·mol[-1] or under -0.5 kcal·mol[-1]. The contribution of residues Lys833, Asp841 and Val882, which form hydrogen bonds with the ligand, is shown in red. All energies are in kcal·mol[-1] and were calculated using 35 equispaced snapshots from 140 ns MD simulations. Standard deviations are in parentheses. [a] Total MM-PBSA binding free energy, excluding entropic contributions. [b] Electrostatic contribution to the binding free energy. [c] van der Waals contribution to the binding free energy. [d] Poisson-Boltzmann electrostatic solvation contribution to the binding free energy. Note that the non-polar contribution to the binding free energy is non-decomposable with Amber 11.

Table 5.5: Energy contributions of PI3Kγ residues to the binding of *s*-cis inhibitor **2** (data from Simulation **II**).

Residue		[a] ΔG_{bind}		[b] ΔG_{elec}		[c] ΔG_{vdw}		[d] ΔG_{PB}	
Ile	879	-1.9	(5.9)	0.4	(2.9)	-1.8	(1.8)	-0.5	(0.6)
Asp	**841**	**-1.7**	**(9.0)**	**-8.4**	**(5.4)**	**0.7**	**(1.6)**	**6.0**	**(4.9)**
Ile	831	-1.4	(5.7)	-0.2	(2.5)	-1.4	(1.8)	0.2	(0.3)
Ile	963	-1.1	(4.7)	0.1	(1.7)	-2.5	(1.6)	1.3	(1.0)
Phe	961	-0.9	(4.3)	0.1	(1.5)	-1.1	(1.4)	0.1	(0.6)
Lys	**833**	**-0.8**	**(12.0)**	**-6.7**	**(8.3)**	**-0.9**	**(1.4)**	**6.7**	**(7.1)**
Ile	881	-0.7	(5.3)	-0.4	(2.1)	-0.8	(1.4)	0.5	(1.3)
Met	953	-0.6	(4.2)	-0.3	(2.1)	-0.9	(1.3)	0.6	(0.7)
Val	**882**	**-0.5**	**(6.3)**	**0.1**	**(3.4)**	**-1.0**	**(1.2)**	**0.4**	**(1.4)**
Glu	880	2.6	(10.9)	-2.2	(7.7)	-0.6	(1.6)	5.4	(6.7)

Showing residues with energy contributions over 0.5 kcal·mol[-1] or under -0.5 kcal·mol[-1]. The contribution of residues Lys833, Asp841 and Val882, which form hydrogen bonds with the ligand, is shown in red. All energies are in kcal·mol[-1] and were calculated using 35 equispaced snapshots from 140 ns MD simulations. Standard deviations are in parentheses. [a] Total MM-PBSA binding free energy, excluding entropic contributions. [b] Electrostatic contribution to the binding free energy. [c] van der Waals contribution to the binding free energy. [d] Poisson-Boltzmann electrostatic solvation contribution to the binding free energy. Note that the non-polar contribution to the binding free energy is non-decomposable with Amber 11.

Table 5.6: Energy contributions of PI3Kγ residues to the binding of *s*-trans inhibitor **1** (data from Simulation **III**).

Residue		[a] ΔG_{bind}		[b] ΔG_{elec}		[c] ΔG_{vdw}		[d] ΔG_{PB}	
Ile	881	-1.7	(6.5)	-0.2	(3.5)	-1.9	(1.8)	0.4	(2.9)
Ile	879	-1.7	(7.3)	0.4	(4.7)	-1.4	(1.7)	-0.7	(0.9)
Ile	831	-1.1	(7.1)	-0.3	(4.8)	-1.1	(1.4)	0.3	(0.5)
Val	**882**	**-0.9**	**(5.4)**	**-1.7**	**(2.7)**	**-1.0**	**(1.2)**	**1.8**	**(1.5)**
Asp	**841**	**-0.7**	**(22.4)**	**-12.8**	**(15.9)**	**0.7**	**(1.4)**	**11.4**	**(10.9)**
Tyr	867	-0.6	(6.3)	0.7	(3.6)	-1.6	(1.2)	0.3	(2.5)
Met	953	-0.6	(5.1)	-0.3	(2.7)	-0.8	(1.3)	0.6	(0.9)
Phe	961	-0.6	(4.4)	0.0	(1.8)	-0.5	(1.3)	-0.1	(0.5)
Arg	947	-0.5	(12.8)	0.8	(9.1)	0.0	(1.9)	-1.3	(6.6)
Lys	**833**	**-0.3**	**(15.4)**	**-9.1**	**(11.2)**	**-0.6**	**(1.9)**	**9.3**	**(9.4)**
Asp	964	0.7	(22.5)	0.1	(16.8)	-0.8	(1.7)	1.4	(12.5)
Glu	880	2.8	(14.1)	-1.2	(10.0)	-0.6	(1.6)	4.6	(8.9)

Showing residues with energy contributions over 0.5 kcal·mol[-1] or under -0.5 kcal·mol[-1]. The contribution of residues Lys833, Asp841 and Val882, which form hydrogen bonds with the ligand, is shown in red. All energies are in kcal·mol[-1] and were calculated using 35 equispaced snapshots from 140 ns MD simulations. Standard deviations are in parentheses. [a] Total MM-PBSA binding free energy, excluding entropic contributions. [b] Electrostatic contribution to the binding free energy. [c] van der Waals contribution to the binding free energy. [d] Poisson-Boltzmann electrostatic solvation contribution to the binding free energy. Note that the non-polar contribution to the binding free energy is non-decomposable with Amber 11.

Table 5.7: Energy contributions of PI3Kγ residues to the binding of *s*-trans inhibitor **2** (data from Simulation **IV**).

Residue		[a] ΔG_{bind}		[b] ΔG_{elec}		[c] ΔG_{vdw}		[d] ΔG_{PB}	
Ile	879	-1.7	(7.5)	0.4	(4.0)	-1.6	(1.9)	-0.5	(0.6)
Ile	831	-1.4	(5.2)	0.0	(2.7)	-1.5	(1.5)	0.1	(0.5)
Asp	**841**	**-1.3**	**(23.5)**	**-10.0**	**(13.4)**	**0.5**	**(2.1)**	**8.1**	**(16.7)**
Ile	881	-0.9	(6.0)	-0.4	(2.6)	-0.9	(1.8)	0.5	(2.2)
Ile	963	-0.7	(6.0)	-0.6	(2.2)	-2.6	(1.4)	2.6	(1.3)
Met	953	-0.6	(5.1)	-0.3	(2.3)	-1.0	(1.5)	0.7	(0.9)
Met	804	-0.6	(5.7)	-0.2	(3.3)	-0.8	(1.6)	0.3	(1.8)
Phe	961	-0.5	(4.6)	0.1	(2.2)	-0.6	(1.4)	0.0	(0.5)
Leu	838	-0.5	(6.5)	-0.1	(3.6)	-0.5	(1.5)	0.2	(1.2)
Val	**882**	**-0.1**	**(6.4)**	**0.0**	**(3.3)**	**-1.4**	**(0.9)**	**1.2**	**(1.3)**
Lys	**833**	**0.2**	**(25.8)**	**-5.5**	**(20.2)**	**-0.7**	**(1.6)**	**6.4**	**(15.3)**
Glu	880	1.9	(13.3)	-2.0	(9.2)	-0.6	(1.5)	4.4	(8.6)

Showing residues with energy contributions over 0.5 kcal·mol[-1] or under -0.5 kcal·mol[-1]. The contribution of residues Lys833, Asp841 and Val882, which form hydrogen bonds with the ligand, is shown in red. All energies are in kcal·mol[-1] and were calculated using 35 equispaced snapshots from 140 ns MD simulations. Standard deviations are in parentheses. [a] Total MM-PBSA binding free energy, excluding entropic contributions. [b] Electrostatic contribution to the binding free energy. [c] van der Waals contribution to the binding free energy. [d] Poisson-Boltzmann electrostatic solvation contribution to the binding free energy. Note that the non-polar contribution to the binding free energy is non-decomposable with Amber 11.

Figure 5.3: Comparison between representative structures from the MD trajectories of the PI3Kγ/*s*-cis inhibitor **1** system (simulation **I**, shown in green) and the PI3Kγ/*s*-cis inhibitor **2** system (simulation **II**, shown in purple). (a) Cartoon representation of the enzymes, showing the ligands as sticks. (b) Active site of PI3Kγ showing the key hydrogen bonding residues. The yellow dashed lines indicate the key hydrogen bond interactions. For clarity, only the interactions of the PI3Kγ/*s*-cis inhibitor **2** system are indicated. (c) Active site of PI3Kγ showing the water mediating interactions. The water molecules are shown as spheres. (d) Active site of PI3Kγ showing the hydrophobic interactions. (e) Active site of PI3Kγ showing residues in the hinge region. Note that the view is rotated to highlight the different protein-ligand interactions.

As can be seen from the plots in Figures 5.2a and 5.2b, the interaction spectra for simulation **I** (PI3Kγ/*s*-cis inhibitor **1**) and simulation **II** (PI3Kγ/*s*-cis inhibitor **2**) are quite similar. However, a comparison of the energy contributions of the key PI3Kγ residues to the binding of *s*-cis inhibitor **1** (Table 5.4) and to the binding of *s*-cis inhibitor **2** (Table 5.5) reveals three residues whose contributions differ the most. These are Glu880 ($\Delta\Delta G_{bind}$ -0.4 kcal·mol^{-1}), Ile881 ($\Delta\Delta G_{bind}$ -0.6 kcal·mol^{-1}) and Val882 ($\Delta\Delta G_{bind}$ -0.7 kcal·mol^{-1}). Note that these three residues are located in the hinge region of the

active site and form interactions with the bicyclic moieties of the inhibitors (Figure 5.3e).

Firstly, residue Val882 forms stronger electrostatic interactions with *s*-cis inhibitor **1** (ΔG_{elec} -1.6 ± 2.9 kcal·mol^{-1}) than with *s*-cis inhibitor **2** (ΔG_{elec} 0.1 ± 3.4 kcal·mol^{-1}). This estimated difference in electrostatic energy ($\Delta\Delta G_{elec}$ -1.7 kcal·mol^{-1}) is consistent with the occupancies and average hydrogen bonding distances for the quinoxaline N···HN(Val882) hydrogen bond of the *s*-cis inhibitor **1** (3.21 ± 0.21 Å, 99.4% occupancy); and the 1,3-benzodioxole ring O···HN(Val882) hydrogen bond of the *s*-cis inhibitor **2** (3.50 ± 0.25 Å, 81.1% occupancy). In Section 2.2.3, it was speculated that a fluorine atom from *s*-cis inhibitor **2** might form a shorter-lived hydrogen bond with Val822 (3.79 ± 0.17 Å, 12.7% occupancy). This, however, seems unlikely for the following two reasons. First, it has been reported[164, 165] that due to its low polarisability and tightly contracted lone pairs, fluorine is unable to compete with stronger hydrogen bond acceptors such as oxygen or nitrogen. Furthermore, it has also been reported[166] that that F···H-O and F···H-N contacts with distances larger than 2.5 Å are very weak and come close in energy to van der Waals complexes. As such, when attributing a particular stabilising significance to interactions of this length or longer, the authors strongly recommend to act with caution.

Secondly, residue Ile881 forms stronger van der Waals interactions with *s*-cis inhibitor **1** than with *s*-cis inhibitor **2** ($\Delta\Delta G_{elec}$ -0.6 kcal·mol^{-1}). This calculated difference is the result of the quinoxaline moiety of *s*-cis inhibitor **1** interacting more favourably than the 1,3-benzodioxole ring of *s*-cis inhibitor **2** with the hydrophobic Ile881 (Figure 5.3e). An interesting observation is that Ile881 in PI3Kγ corresponds to a valine in PI3Kα, PI3Kβ and PI3Kδ. Even if Ile and Val are both hydrophobic residues, the smaller size Val may create a cavity in the PI3Kγ active site and be partly responsible for the weaker inhibition of the other isoforms by the γ-selective inhibitors **1** and **2** (refer to Table 1.2).

And thirdly, residue Glu880 contributes unfavourably to the binding in both protein-ligand systems. However, the total electrostatic contribution (ΔG_{elec} + ΔG_{PB}) for Glu880 in the *s*-cis inhibitor **1** complex was 0.6 kcal·mol^{-1} less unfavourable than that of the *s*-trans inhibitor **1** complex. A possible explanation for this is that the ligand is stopping Glu880 from being accessible to solvent.

A comparison of the energy contributions of the key PI3Kγ residues to the binding of *s*-cis inhibitor **1** (Table 5.4) and to the binding of *s*-trans inhibitor **1** (Table 5.6) shows significant differences in the contributions of the hydrogen bonding residues Lys833

($\Delta\Delta G_{bind}$ -0.7 kcal·mol^{-1}), Asp841 ($\Delta\Delta G_{bind}$ -1.0 kcal·mol^{-1}) and Val882 ($\Delta\Delta G_{bind}$ -0.3 kcal·mol^{-1}). This result was to be expected and is consistent with the hydrogen bond analysis presented in Section 4.2.3. The orientation of the thiazolidinedione moiety of *s*-trans inhibitor **1** contributed to the formation of stronger ($\Delta\Delta G_{bind}$ -0.4 kcal·mol^{-1}) van der Waals interactions with Tyr867 (Figure 5.4c). However, as discussed in Section 4.2.3, during MD simulations of the PI3Kγ/*s*-trans inhibitor **1** complex the local binding pocket presented large conformational mobility, which might be a reason for the relatively lower affinity of the *s*-trans inhibitor **1**. Overall, *s*-trans inhibitor **1** is a weaker binder than *s*-cis inhibitor **1**, the difference in calculated binding free energy between the *s*-cis and *s*-trans isomers is -8.0 kcal·mol^{-1}.
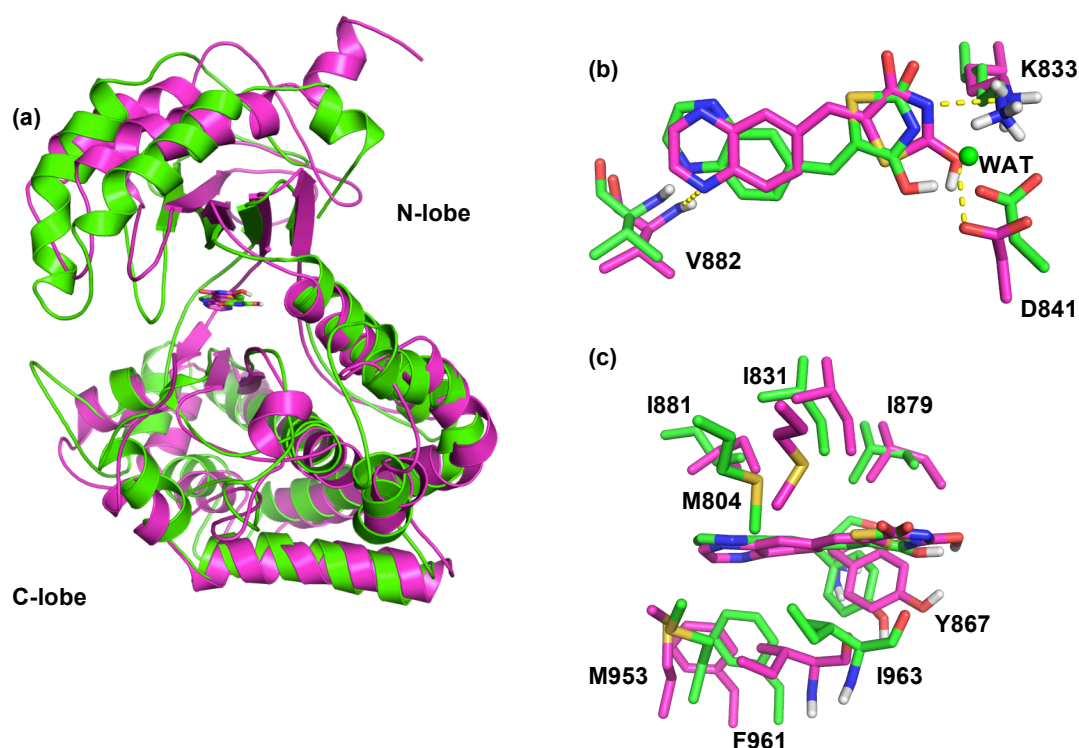
Figure 5.4: Comparison between representative structures from the MD trajectories of the PI3Kγ/*s*-cis inhibitor **1** system (simulation **I**, shown in green) and the PI3Kγ/*s*-trans inhibitor **1** system (simulation **III**, shown in magenta). (a) Cartoon representation of the enzymes, showing the ligands as sticks. (b) Active site of PI3Kγ showing the key hydrogen bonding residues. The yellow dashed lines indicate the key hydrogen bond interactions. For clarity, only the interactions of the PI3Kγ/*s*-trans inhibitor **1** system are indicated. (c) Active site of PI3Kγ showing the hydrophobic interactions. Note that the view is rotated to highlight the different protein-ligand interactions.

Turning now to the comparison of the complexes involving inhibitor **2**, the per-residue free energy decomposition analysis reveals a significant difference in the contribution of residues Lys833, Asp841 and Val822 to the binding of *s*-cis inhibitor **2** and *s*-trans inhibitor **2** ($\Delta\Delta G_{bind}$ -1.0 kcal·mol^{-1}, $\Delta\Delta G_{bind}$ -0.4 kcal·mol^{-1}, $\Delta\Delta G_{bind}$ -0.4 kcal·mol^{-1},

respectively). As is apparent from the results presented in Tables 5.6 and 5.8, *s*-cis inhibitor **2** is a stronger binder than *s*-trans inhibitor **2**. This is in agreement with the hydrogen bond analysis presented in Section 4.2.3. The orientation of the bicyclic ring of *s*-trans inhibitor **2** is predicted to result in less unfavourable interaction with Glu880 (ΔG_{bind} is 2.6 kcal·mol^{-1} for the *s*-cis inhibitor **2** complex and 1.9 kcal·mol^{-1} for the *s*-trans inhibitor **2** complex). Likewise, the slight displacement of the thiazolidinedione moiety of *s*-trans inhibitor **2** (see Figure 5.5c) is results in the formation of stronger van der Waals interactions with the side chain of Leu838 ($\Delta\Delta G_{bind}$ -0.4 kcal·mol^{-1}). At the same time, however, this same displacement of the *s*-trans inhibitor **2** causes the decrease in van der Waals interactions with the side chain of Phe961 ($\Delta\Delta G_{bind}$ 0.4 kcal·mol^{-1}; Figure 5.5c).

Figure 5.5: Comparison between representative structures from the MD trajectories of the PI3Kγ/*s*-cis inhibitor **2** system (simulation **II**, shown in purple) and the PI3Kγ/*s*-trans inhibitor **2** system (simulation **VI**, shown in cyan). (a) Cartoon representation of the enzymes, showing the ligands as sticks. (b) Active site of PI3Kγ showing the key hydrogen bonding residues. The yellow dashed lines indicate the key hydrogen bond interactions. For clarity, only the interactions of the PI3Kγ/*s*-trans inhibitor **2** system are indicated. (c) Active site of PI3Kγ showing the hydrophobic interactions. Note that the view is rotated to highlight the different protein-ligand interactions.

5.2.3 QM analysis of key hydrogen bond to Val882

The key ligand-Val882 interaction (Figure 5.6) was evaluated *via* quantum mechanics (QM) at the M06L/MG3S level. Energy optimisations with all heavy atoms frozen were performed to calculate the interaction energy between the bicyclic moieties of the inhibitors (i.e. the quinoxaline moiety of *s*-cis inhibitor **1** and the 1,3-benzodioxole ring of *s*-cis inhibitor **2**) and a protein fragment containing the peptide bond Ile881-Val882 (Figure 5.6). The 35 equispaced snapshots from the MD trajectories discussed in chapter 4, were used to create time-ordered structures from which the initial atomic coordinates were taken. Note that for the QM calculations only the data from simulations **I** and **II** were used.

The M06L functional was chosen because of its excellent performance, when compared to other DFT methods such as B3LPY, BLYP and BP86, for calculating non-covalent interactions[115, 122]. These calculations used the MG3S basis set[115].
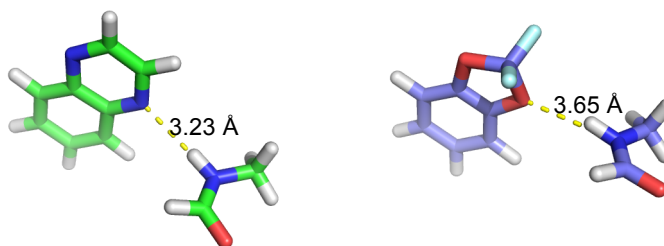
Figure 5.6: Bicyclic moieties of *s*-cis inhibitor **1** (left, in green) and *s*-cis inhibitor **2** (right, in purple) and corresponding protein fragments (containing peptide bond Ile881-Val882). The MD average distance for the *s*-cis inhibitor **1** quinoxaline N⋯HN(Val882) hydrogen bond is 3.23 Å, whereas the average distance for the *s*-cis inhibitor **2** 1,3-benzodioxole ring O⋯HN(Val882) hydrogen bond is 3.65 Å.

As indicated in Table 5.8, the QM average estimate of the interaction energy, $\Delta E_{QM[m06L]}$, predicted for the *s*-cis inhibitor **1** complex is more favourable than the one predicted for the *s*-cis inhibitor **2** complex (-4.1 ± 1.2 kcal·mol^{-1} and -1.2 ± 0.5 kcal·mol^{-1}, respectively, Table 5.8). Furthermore, the difference in the QM average estimate of the interaction energy, $\Delta\Delta E_{QM[m06L]}$, is -2.9 kcal·mol^{-1}, which is in agreement with the trend observed experimentally (-1.9 kcal·mol^{-1}, $\Delta\Delta G_{bind[exp]}$, Table 5.8).

Additionally, the key ligand-Val882 interaction (Figure 5.6) was evaluated *via* molecular mechanics, using the *ff99sb* force field. The MM average estimate of the interaction energy, $\Delta E_{MM[ff99SB]}$, predicted for the *s*-cis inhibitor **1** complex is more favourable than the one predicted for the *s*-cis inhibitor **2** complex (-3.8 ± 0.8

kcal·mol^{-1} and -1.9 ± 0.8 kcal·mol^{-1}, respectively, Table 5.8). These results indicate a reduced preference $\Delta\Delta E_{MM[ff99SB]}$ for *s*-cis inhibitor **1** of 1.9 kcal·mol^{-1} (Table 5.8).

Interestingly, a comparison of molecular mechanical and quantum chemical treatments of the ligand-valine interaction suggests that the *ff99sb* force field underestimates the relative strength of the quinoxaline N···HN(Val882) hydrogen bond of the *s*-cis inhibitor **1** by 0.3 kcal·mol^{-1}; and overestimates the strength of the 1,3-benzodioxole ring O···HN(Val882) hydrogen bond of the *s*-cis inhibitor **2** by 0.7 kcal·mol^{-1}.

Therefore, we correct the gas-phase protein-ligand interaction to account for inaccuracies of the force field in modelling the ligand-Val882 hydrogen bond. The corrected total binding free energy in the absence of entropic contribution, $\Delta G'_{bind\,[MM-PBSA]}$, estimated for the PI3Kγ/*s*-cis inhibitor **1** system is -27.2 kcal·mol^{-1}, and for the PI3Kγ/*s*-cis inhibitor **2** system is -27.2 kcal·mol^{-1} (Table 5.9). This correction leads to a preference in total binding energy of 2.9 kcal·mol^{-1} for *s*-cis inhibitor **1** ($\Delta\Delta G'_{bind\,[MM-PBSA]}$, Table 5.9).

Table 5.8: Energetic analysis for the PI3Kγ complexes.

Sim	Ligand	[a]$\Delta E_{QM[m06L]}$		[b]$\Delta E_{MM[ff99SB]}$		[c]$\Delta E_{elec[ff99SB]}$		[d]$\Delta E_{vdw[ff99SB]}$		[e]$\Delta G_{bind[exp]}$
I	*s*-cis **1**	-4.1	(1.2)	-3.8	(0.8)	-2.6	(0.8)	-1.2	(0.4)	-11.0
II	*s*-cis **2**	-1.2	(0.5)	-1.9	(0.8)	-0.8	(0.4)	-1.1	(0.4)	-9.1

All energies are in kcal·mol^{-1} and were calculated using 35 equispaced snapshots from 140 ns MD simulations (see text for details). Standard deviations are in parentheses. [a] Estimate of the energetic difference as estimated by QM at the M06L/MG3S level. [b] Estimate of the energetic difference as estimated by MM using the *ff99SB* force field. [c] Electrostatic component of the estimate of the energetic difference as estimated by MM using the *ff99SB* force field. [d] van der Waals components of the estimate of the energetic difference as estimated by MM using the *ff99SB* force field. [e] Estimated ΔG calculated from experimental data as explained by equation 5.4.

Table 5.9: QM and MM energetic analysis and binding free energy calculations calculated using MM-PBSA for the PI3Kγ complexes.

Sim	Ligand	[a]$\Delta E_{QM[m06L]}$		[b]$\Delta E_{MM[ff99SB]}$		[c]$\Delta G_{MM[MMPBSA]}$		[d]$\Delta G_{solv[MMPBSA]}$		[e]$\Delta G_{bind[MMPBSA]}$		[f]$\Delta G'_{bind[MMPBSA]}$	[g]$\Delta G_{bind[exp]}$
I	*s*-cis **1**	-4.1	(1.2)	-3.8	(0.8)	-80.3	(6.4)	-53.4	(4.5)	-26.9	(2.9)	-27.2	-11.0
II	*s*-cis **2**	-1.2	(0.5)	-1.9	(0.8)	-76.5	(5.4)	-51.6	(3.5)	-24.9	(2.4)	-24.2	-9.1

All energies are in kcal·mol^{-1} and were calculated using 35 equispaced snapshots from 140 ns MD simulations (see text for details). Standard deviations are in parentheses. Energetic analysis: [a] Estimate of the energetic difference as estimated by QM at the M06L/MG3S level. [b] Estimate of the energetic difference as estimated by MM using the *ff99SB* force field. Binding free energies: [c] Gas phase energy contribution to the binding free energy, $\Delta G_{MM\,[MM-PBSA]} = \Delta G_{ele\,[MM-PBSA]} + \Delta G_{vdw\,[MM-PBSA]}$. [d] Solvation energy contribution to the binding free energy, including polar and non-polar contributions. [e] Total MM-PBSA binding free energy, excluding entropic contributions. [f] Total binding free energy in the absence of entropic contribution, with corrected values as per the QM. [g] Estimated ΔG calculated from experimental data as explained by equation 5.4.

5.3 Conclusions

The work described in this chapter seeks to provide an insight into the origin of the binding affinity of PI3Kγ to inhibitors. Using the MM-PBSA approach, the binding free energy was predicted for four protein-ligand complexes.

The following conclusions can be drawn from the results of the binding free energy calculations (Section 5.2.1). First, for both inhibitors, the *s*-cis isomers are predicted to be the most favourable conformations, since their calculated binding energies were lower than the calculated values for their *s*-trans counterparts. Second, *s*-cis inhibitor **1** is a stronger binder than *s*-cis inhibitor **2**. The difference in their predicted binding free energies ($\Delta\Delta G_{bind}$) was -2.0 kcal·mol^{-1}, which is consistent with the trend observed experimentally.

To estimate the contribution of each residue to the binding free energy, an energy decomposition analysis was performed using the MM-PBSA method (Section 5.2.2). In all four protein-ligand complexes the major favourable energy contributions originate predominantly from residues Met804, Ile831, Lys833, Asp841, Ile879, Ile881, Val882, Met953, Phe961 and Ile963. Conversely, residue Glu880 contributes unfavourably to the binding in all four protein-ligand systems.

From the results of the per-residue decomposition analysis it was possible to identify the key residues whose contributions differ the most when comparing the different protein-ligand systems. When comparing the PI3Kγ/*s*-cis inhibitor **1** system *versus* the PI3Kγ/*s*-cis inhibitor **2** system, it appears from the energetic analysis that residues Glu880, Ile881 and Val882 are responsible for the higher affinity of PI3Kγ for *s*-cis inhibitor **1** over *s*-cis inhibitor **2**.

The crucial ligand-Val882 interaction of the PI3Kγ/*s*-cis inhibitor **1** complex and the PI3Kγ/*s*-cis inhibitor **2** complex was examined utilising molecular mechanical and quantum chemical treatments (Section 5.2.3). The results indicate that the *ff99sb* force field slightly underestimates the relative strength of the quinoxaline N···HN(Val882) hydrogen bond of the *s*-cis inhibitor **1**; and overestimates the strength of the 1,3-benzodioxole ring O···HN(Val882) hydrogen bond of the *s*-cis inhibitor **2**. Correction of the gas-phase protein-ligand interaction to account for inaccuracies of the force field in modelling the ligand-Val882 hydrogen bond leads to a larger preference of 2.9 kcal·mol^{-1} for *s*-cis inhibitor **1** ($\Delta\Delta G'_{bind\ [MM-PBSA]}$, Table 5.9).

Finally, a number of important limitations need to be considered. First, during the calculation of binding free energy using the MM-PBSA approach, the entropic contribution was excluded. Hence, the calculated binding free energy was lower than that derived from the experiment. However, this is a common approximation in calculations involving a set of different ligands binding to the same protein[156, 167].

Second, the fact that a single trajectory approach was used to estimate the total binding free energies (i.e. a single simulation of the solvated protein-ligand complex instead of three separate simulations of the protein, ligand and protein-ligand complex in explicit water) is a potential source of error. In the single trajectory approach, only ligand-protein distances contribute to the calculated change in potential energy (because all other interatomic distances stay exactly the same). In other words, this approach assumes the same conformational distribution for the free and bound ligand[108]. It has been suggested[167] that the separate trajectory approach is more accurate than the single trajectory approach. However, it is also more computationally expensive.

An alternative approach for the per-residue energy decomposition could have been to use the MM-GBSA (Molecular Mechanics Generalized Born Surface Area) method[163, 168]. In this way, the nonpolar solvation contributions could have been decomposed. However, MM-GBSA is an approximation that is more efficient but may be slightly less accurate than MM-PBSA[167, 169].

6. References

1. Cantley, L. C., The phosphoinositide 3-kinase pathway. *Science* **2002,** *296*, 1655-1657.

2. Engelman, J. A.; Luo, J.; Cantley, L. C., The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nature Reviews Genetics* **2006,** *7*, 606-619.

3. Drees, B. E.; Mills, G. B.; Rommel, C.; Prestwich, G. D., Therapeutic potential of phosphoinositide 3-kinase inhibitors. *Expert Opinion on Therapeutic Patents* **2004,** *14*, 703-732.

4. Rueckle, T.; Schwarz, M. K.; Rommel, C., PI3K gamma inhibition: towards an 'aspirin of the 21st century'? *Nature Reviews Drug Discovery* **2006,** *5*, 903-918.

5. Knight, Z. A.; Shokat, K. M., Chemically targeting the PI3K family. *Biochemical Society Transactions* **2007,** *35*, 245-249.

6. Dhand, R.; Hiles, I.; Panayotou, G.; Roche, S.; Fry, M. J.; Gout, I.; Totty, N. F.; Truong, O.; Vicendo, P.; Yonezawa, K.; Kasuga, M.; Courtneidge, S. A.; Waterfield, M. D., PI 3-kinase is a dual specificity enzyme: autoregulation by an intrinsic protein-serine kinase activity. *Embo Journal* **1994,** *13*, 522-533.

7. Hunter, T., Protein kinases and phosphatases - the ying and yang of protein phosphorylation and signaling *Cell* **1995,** *80*, 225-236.

8. Ciesla, J.; Fraczyk, T.; Rode, W., Phosphorylation of basic amino acid residues in proteins: important but easily missed. *Acta Biochimica Polonica* **2011,** *58*, 137-147.

9. Steeg, P. S.; Palmieri, D.; Ouatas, T.; Salerno, M., Histidine kinases and histidine phosphorylated proteins in mammalian cell biology, signal transduction and cancer. *Cancer Letters* **2003,** *190*, 1-12.

10. Cohen, P., The origins of protein phosphorylation. *Nature Cell Biology* **2002,** *4*, E127-E130.

11. Endicott, J. A.; Noble, M. E. M.; Johnson, L. N., The Structural Basis for Control of Eukaryotic Protein Kinases. In *Annual Review of Biochemistry*, Kornberg, R. D., Ed. 2012; Vol. 81, pp 587-613.

12. Eyster, K. M., The membrane and lipids as integral participants in signal transduction: lipid signal transduction for the non-lipid biochemist. *Advances in Physiology Education* **2007,** *31*, 5-16.

13. Wymann, M. P.; Schneiter, R., Lipid signalling in disease. *Nature Reviews Molecular Cell Biology* **2008,** *9*, 162-176.

14. Toker, A.; Cantley, L. C., Signalling through the lipid products of phosphoinositide-3-OH kinase. *Nature* **1997,** *387*, 673-676.

15. Sundstrom, T. J.; Anderson, A. C.; Wright, D. L., Inhibitors of phosphoinositide-3-kinase: a structure-based approach to understanding potency and selectivity. *Organic & Biomolecular Chemistry* **2009,** *7*, 840-850.

16. Hirsch, E.; Braccini, L.; Ciraolo, E.; Morello, F.; Perino, A., Twice upon a time: PI3K's secret double life exposed. *Trends in Biochemical Sciences* **2009,** *34*, 244-248.

17. Hirsch, E.; Ciraolo, E.; Ghigo, A.; Costa, C., Taming the PI3K team to hold inflammation and cancer at bay. *Pharmacology & Therapeutics* **2008,** *118*, 192-205.

18. Yap, T. A.; Garrett, M. D.; Walton, M. I.; Raynaud, F.; de Bono, J. S.; Workman, P., Targeting the PI3K-AKT-mTOR pathway: progress, pitfalls, and promises. *Current Opinion in Pharmacology* **2008,** *8*, 393-412.

19. Carracedo, A.; Pandolfi, P. P., The PTEN-PI3K pathway: of feedbacks and cross-talks. *Oncogene* **2008,** *27*, 5527-5541.

20. Vanhaesebroeck, B.; Waterfield, M. D., Signaling by distinct classes of phosphoinositide 3-kinases. *Experimental Cell Research* **1999,** *253*, 239-254.

21. Ward, S. G.; Finan, P., Isoform-specific phosphoinositide 3-kinase inhibitors as therapeutic agents. *Current Opinion in Pharmacology* **2003,** *3*, 426-434.

22. Hirsch, E.; Costa, C.; Ciraolo, E., Phosphoinositide 3-kinases as a common platform for multi-hormone signaling. *Journal of Endocrinology* **2007,** *194*, 243-256.

23. Williams, R.; Berndt, A.; Miller, S.; Hon, W.-C.; Zhang, X., Form and flexibility in phosphoinositide 3-kinases. *Biochemical Society Transactions* **2009,** *37*, 615-626.

24. Vanhaesebroeck, B.; Guillermet-Guibert, J.; Graupera, M.; Bilanges, B., The emerging mechanisms of isoform-specific PI3K signalling. *Nature Reviews Molecular Cell Biology* **2010,** *11*, 329-341.

25. Pacold, M. E.; Suire, S.; Perisic, O.; Lara-Gonzalez, S.; Davis, C. T.; Walker, E. H.; Hawkins, P. T.; Stephens, L.; Eccleston, J. F.; Williams, R. L., Crystal structure and functional analysis of Ras binding to its effector phosphoinositide 3-kinase gamma. *Cell* **2000,** *103*, 931-943.

26. Ohashi, P. S.; Woodgett, J. R., Modulating autoimmunity: pick your Pl3 kinase. *Nature Medicine* **2005,** *11*, 924-925.

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

28. Liu, Z. N.; Roberts, T. M., Human tumor mutants in the p110 alpha subunit of PI3K. *Cell Cycle* **2006,** *5*, 675-677.

29. Jackson, S. P.; Schoenwaelder, S. M.; Goncalves, I.; Nesbitt, W. S.; Yap, C. L.; Wright, C. E.; Kenche, V.; Anderson, K. E.; Dopheide, S. M.; Yuan, Y. P.;

Sturgeon, S. A.; Prabaharan, H.; Thompson, P. E.; Smith, G. D.; Shepherd, P. R.; Daniele, N.; Kulkarni, S.; Abbott, B.; Saylik, D.; Jones, C.; Lu, L.; Giuliano, S.; Hughan, S. C.; Angus, J. A.; Robertson, A. D.; Salem, H. H., PI 3-kinase p110 beta: a new target for antithrombotic therapy. *Nature Medicine* **2005,** *11*, 507-514.

30. Camps, M.; Ruckle, T.; Ji, H.; Ardissone, V.; Rintelen, F.; Shaw, J.; Ferrandi, C.; Chabert, C.; Gillieron, C.; Francon, B.; Martin, T.; Gretener, D.; Perrin, D.; Leroy, D.; Vitte, P. A.; Hirsch, E.; Wymann, M. P.; Cirillo, R.; Schwarz, M. K.; Rommel, C., Blockade of PI3K gamma suppresses joint inflammation and damage in mouse models of rheumatoid arthritis. *Nature Medicine* **2005,** *11*, 936-943.

31. Barber, D. F.; Bartolome, A.; Hernandez, C.; Flores, J. M.; Redondo, C.; Fernandez-Arias, C.; Camps, M.; Ruckle, T.; Schwarz, M. K.; Rodriguez, S.; Martinez-A, C.; Balomenos, D.; Rommel, C.; Carrera, A. C., PI3K gamma inhibition blocks glomerulonephritis and extends lifespan in a mouse model of systemic lupus. *Nature Medicine* **2005,** *11*, 933-935.

32. Kuehnel, I.; Pundt, N.; Baier, A.; Bondeva, T.; Meinecke, I.; Gay, S.; Rueckle, T.; Camps, M.; Schwarz, M. K.; Rommel, C.; Wetzker, R.; Pap, T., Phosphatidylinositol 3-kinase gamma regulates matrix metalloproteinases production in rheumatoid arthritis synovial fibroblasts. *Annals of the Rheumatic Diseases* **2007,** *66*, A63-A63.

33. Rommel, C.; Camps, M.; Ji, H., PI3K delta and PI3K gamma: partners in crime in inflammation in rheumatoid arthritis and beyond? *Nature Reviews Immunology* **2007,** *7*, 191-201.

34. Wei, X.; Han, J.; Chen, Z.-z.; Qi, B.-w.; Wang, G.-c.; Ma, Y.-h.; Zheng, H.; Luo, Y.-f.; Wei, Y.-q.; Chen, L.-j., A phosphoinositide 3-kinase-gamma inhibitor, AS605240 prevents bleomycin-induced pulmonary fibrosis in rats. *Biochemical and Biophysical Research Communications* **2010,** *397*, 311-317.

35. Walker, E. H.; Perisic, O.; Ried, C.; Stephens, L.; Williams, R. L., Structural insights into phosphoinositide 3-kinase catalysis and signalling. *Nature* **1999,** *402*, 313-320.

36. Walker, E. H.; Pacold, M. E.; Perisic, O.; Stephens, L.; Hawkins, P. T.; Wymann, M. P.; Williams, R. L., Structural determinants of phosphoinositide 3-kinase inhibition by wortmannin, LY294002, quercetin, myricetin, and staurosporine. *Molecular Cell* **2000,** *6*, 909-919.

37. Huang, C.-H.; Mandelker, D.; Schmidt-Kittler, O.; Samuels, Y.; Velculescu, V. E.; Kinzler, K. W.; Vogelstein, B.; Gabelli, S. B.; Amzel, L. M., The structure of a human p110 alpha/p85 alpha complex elucidates the effects of oncogenic PI3K alpha mutations. *Science* **2007,** *318*, 1744-1748.

38. Berndt, A.; Miller, S.; Williams, O.; Le, D. D.; Houseman, B. T.; Pacold, J. I.; Gorrec, F.; Hon, W.-C.; Liu, Y.; Rommel, C.; Gaillard, P.; Rueckle, T.; Schwarz, M. K.; Shokat, K. M.; Shaw, J. P.; Williams, R. L., The p110 delta structure: mechanisms for selectivity and potency of new PI(3)K inhibitors. *Nature Chemical Biology* **2010,** *6*, 117-124.

39. Zhang, X.; Vadas, O.; Perisic, O.; Anderson, K. E.; Clark, J.; Hawkins, P. T.; Stephens, L. R.; Williams, R. L., Structure of Lipid Kinase p110 beta/p85 beta

Elucidates an Unusual SH2-Domain-Mediated Inhibitory Mechanism. *Molecular Cell* **2011,** *41*, 567-578.

40. Krugmann, S.; Hawkins, P. T.; Pryer, N.; Braselmann, S., Characterizing the interactions between the two subunits of the p101/p110 gamma phosphoinositide 3-kinase and their role in the activation of this enzyme by G(beta gamma) subunits. *Journal of Biological Chemistry* **1999,** *274*, 17152-17158.

41. Rodriguezviciana, P.; Warne, P. H.; Dhand, R.; Vanhaesebroeck, B.; Gout, I.; Fry, M. J.; Waterfield, M. D.; Downward, J., Phosphatidylinositol-3-OH kinase as a direct target of Ras. *Nature* **1994,** *370*, 527-532.

42. Nalefski, E. A.; Falke, J. J., The C2 domain calcium-binding motif: Structural and functional diversity. *Protein Science* **1996,** *5*, 2375-2390.

43. Djordjevic, S.; Driscoll, P. C., Structural insight into substrate specificity and regulatory mechanisms of phosphoinositide 3-kinases. *Trends in Biochemical Sciences* **2002,** *27*, 426-432.

44. Saraste, M.; Sibbald, P. R.; Wittinghofer, A., The P-loop: a common motif in ATP- and GTP-binding proteins. *Trends in Biochemical Sciences* **1990,** *15*, 430-434.

45. Han, M.; Zhang, J. Z. H., Class I Phospho-inositide-3-kinases (PI3Ks) Isoform-Specific Inhibition Study by the Combination of Docking and Molecular Dynamics Simulation. *Journal of Chemical Information and Modeling* **2010,** *50*, 136-145.

46. Scapin, G., Structural biology in drug design: selective protein kinase inhibitors. *Drug Discovery Today* **2002,** *7*, 601-611.

47. Bondeva, T.; Pirola, L.; Bulgarelli-Leva, G.; Rubio, I.; Wetzker, R.; Wymann, M. P., Bifurcation of lipid and protein kinase signals of PI3K gamma to the protein kinases PKB and MAPK. *Science* **1998,** *282*, 293-296.

48. Pirola, L.; Zvelebil, M. J.; Bulgarelli-Leva, G.; Van Obberghen, E.; Waterfield, M. D.; Wymann, M. P., Activation loop sequences confer substrate specificity to phosphoinositide 3-kinase alpha (PI3K alpha) - Functions of lipid kinase-deficient PI3K alpha in signaling. *Journal of Biological Chemistry* **2001,** *276*, 21544-21554.

49. Kannan, N.; Neuwald, A. F., Did protein kinase regulatory mechanisms evolve through elaboration of a simple structural component? *Journal of Molecular Biology* **2005,** *351*, 956-972.

50. Shan, Y.; Seeliger, M. A.; Eastwood, M. P.; Frank, F.; Xu, H.; Jensen, M. O.; Dror, R. O.; Kuriyan, J.; Shaw, D. E., A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proceedings of the National Academy of Sciences of the United States of America* **2009,** *106*, 139-144.

51. Amzel, L. M.; Huang, C.-H.; Mandelker, D.; Lengauer, C.; Gabelli, S. B.; Vogelstein, B., Structural comparisons of class I phosphoinositide 3-kinases. *Nature Reviews Cancer* **2008,** *8*, 665-669.

52. Alaimo, P. J.; Knight, Z. A.; Shokat, K. M., Targeting the gatekeeper residue in phosphoinositide 3-kinases. *Bioorganic & Medicinal Chemistry* **2005,** *13*, 2825-2836.

53. Dar, A. C.; Shokat, K. M., The Evolution of Protein Kinase Inhibitors from Antagonists to Agonists of Cellular Signaling. In *Annual Review of Biochemistry*, Kornberg, R. D.; Raetz, C. R. H.; Rothman, J. E.; Thorner, J. W., Eds. 2011; Vol. 80, 769-795.

54. Liu, Y.; Gray, N. S., Rational design of inhibitors that bind to inactive kinase conformations. *Nature Chemical Biology* **2006,** *2*, 358-364.

55. Levinson, N. M.; Kuchment, O.; Shen, K.; Young, M. A.; Koldobskiy, M.; Karplus, M.; Cole, P. A.; Kuriyan, J., A Src-like inactive conformation in the Abl tyrosine kinase domain. *Plos Biology* **2006,** *4*, 753-767.

56. Maira, S.-M.; Stauffer, F.; Schnell, C.; Garcia-Echeverria, C., PI3K inhibitors for cancer treatment: where do we stand? *Biochemical Society Transactions* **2009,** *37*, 265-272.

57. Wu, P.; Hu, Y., Small molecules targeting phosphoinositide 3-kinases. *MedChemComm* **2012,** *3*, 1337-1355.

58. Arcaro, A.; Wymann, M. P., Wortmannin is a potent phosphatidylinositol 3-kinase inhibitor: the role of phosphatidylinositol 3,4,5-trisphosphate in neutrophil responses. *Biochemical Journal* **1993,** *296*, 297-301.

59. Vlahos, C. J.; Matter, W. F.; Hui, K. Y.; Brown, R. F., A specific inhibitor of phosphatidylinositol 3-kinase, 2-(4-morpholinyl)-8-phenyl-4H-1-benzopyran-4-one (LY294002). *Journal of Biological Chemistry* **1994,** *269*, 5241-5248.

60. Bragg, W. L.; Thomson, J. J., The diffraction of short electromagnetic waves by a crystal. *Proceedings of the Cambridge Philosophical Society* **1914,** *17*, 43-57.

61. Rhodes, G., *Crystallography Made Crystal Clear*. Third ed.; U.S.A., 2006;

62. Leslie, A. G. W., The integration of macromolecular diffraction data. *Acta Crystallographica Section D* **2006,** *62*, 48-57.

63. Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C., A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **1958,** *181*, 662-666.

64. Perutz, M., Isomorphous replacement and phase determination in non-centrosymmetric space groups. *Acta Crystallographica* **1956,** *9*, 867-873.

65. delaFortelle, E.; Bricogne, G., Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Macromolecular Crystallography, Pt A* **1997,** *276*, 472-494.

66. Rossmann, M. G.; Blow, D. M., The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica* **1962,** *15*, 24-31.

67. Rossmann, M. G., The Molecular Replacement Method *Acta Crystallographica Section A* **1990,** *46*, 73-82.

68. Wang, B. C., Resolution of phase ambiguity in macromolecular crystallography. *Methods in Enzymology* **1985,** *115*, 90-112.

69. Hendrickson, W. A.; Smith, J. L.; Sheriff, S., Direct phase determination based on anomalous scattering *Methods in Enzymology* **1985,** *115*, 41-55.

70. Drenth, J., *Principles of Protein X-Ray Crystallography*. Third ed.; 2007.

71. Taylor, G. L., Introduction to phasing. *Acta Crystallographica Section D* **2010,** *66*, 325-38.

72. Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M., Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *Febs Journal* **2008,** *275*, 1-21.

73. Bhat, T., Calculation of an OMIT map. *Journal of Applied Crystallography* **1988,** *21*, 279-281.

74. Davis, A. M.; Teague, S. J.; Kleywegt, G. J., Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angewandte Chemie-International Edition* **2003,** *42*, 2718-2736.

75. Brünger, A. T., Free R value: Cross-validation in crystallography. *Methods in Enzymology* **1997,** *277*, 366-396.

76. Brünger, A. T., Free R-value - A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **1992,** *355*, 472-475.

77. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V., Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* **1963,** *7*, 95-99.

78. Schrödinger, E., Quantisierung als eigenwertproblem. *Annals of Physics* **1926,** *79*.

79. Atkins, P. W., *Physical Chemistry*. 6th ed.; Oxford University Press: 2001.

80. McGervey, J. D., *Quantum mechanics*. Third ed.; Academic Press: London UK, 1995.

81. Born, M.; Oppenheimer, R., Zur Quantentheorie der Molekeln. *Annals of Physics* **1927,** *84*, 457-484.

82. Leach, A. R., *Molecular Modelling-Principles and Applications*. First ed.; Longman: 1996.

83. Lewars, E. G., *Computational Chemistry-Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. 2nd ed.; Springer: 2011.

84. Jensen, F., *Introduction to Computational Chemistry*. 2nd ed.; John Wiley & Sons Ltd: 2007.

85. Cramer, C. J., *Essentials of Computational Chemistry*. Second ed.; John Wiley & Sons Ltd: 2004.

86. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins-Structure Function and Bioinformatics* **2006,** *65*, 712-725.

87. Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004,** *25*, 1157-1174.

88. Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F., A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry* **2004,** *25*, 1656-1676.

89. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **1998,** *102*, 3586-3616.

90. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A., Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *Journal of Computational Chemistry* **1995,** *16*, 1357-1377.

91. Jones, J. E., On the determination of molecular fields - II From the equation of state of a gas. *Proceedings of the Royal Society of London Series a-Containing Papers of a Mathematical and Physical Character* **1924,** *106*, 463-477.

92. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems. *Journal of Chemical Physics* **1993,** *98*, 10089-10092.

93. Li, J., Basic Molecular Dynamics. In *Handbook of Materials Modeling*, Yip, S., Ed. Springer: The Netherlands, 2005; 565-588.

94. Hockney, R. W., The potential calculation and some applications. *Methods in Computational Physics* **1970,** *9*, 136-211.

95. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* **1983,** *79*, 926-935.

96. Fogolari, F.; Brigo, A.; Molinari, H., The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *Journal of Molecular Recognition* **2002,** *15*, 377-392.

97. Nicholls, A.; Honig, B., A rapid Finite-Difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *Journal of Computational Chemistry* **1991,** *12*, 435-445.

98. Davis, M. E.; McCammon, J. A., Solving the Finite-Difference linearized Poisson-Boltzmann equation – a comparison of relaxation and conjugate gradient methods. *Journal of Computational Chemistry* **1989,** *10*, 386-391.

99. Onufriev, A.; Case, D. A.; Bashford, D., Effective Born radii in the generalized Born approximation: The importance of being perfect. *Journal of Computational Chemistry* **2002,** *23*, 1297-1304.

100. Floris, F.; Tomasi, J., Evaluation of the dispersion contribution to the solvation energy. A simple computational model in the continuum approximation. *Journal of Computational Chemistry* **1989,** *10*, 616-627.

101. Floris, F. M.; Tomasi, J.; Ahuir, J. L. P., Dispersion and repulsion contributions to the solvation energy: Refinements to a simple computational model in the continuum approximation *Journal of Computational Chemistry* **1991,** *12*, 784-791.

102. Toukmaji, A. Y.; Board, J. A., Ewald summation techniques in perspective: A survey. *Computer Physics Communications* **1996,** *95*, 73-92.

103. Andersen, H. C., Molecular dynamics simulations at constant pressure and/or temperature. *Journal of Chemical Physics* **1980,** *72*, 2384-2393.

104. Adelman, S. A.; Doll, J. D., Generalized Langevin equation approach for atom-solid-surface scattering - General formulation for classical scattering off harmonic solids *Journal of Chemical Physics* **1976,** *64*, 2375-2388.

105. Hunenberger, P., Thermostat algorithms for molecular dynamics simulations. In *Advanced Computer Simulation Approaches for Soft Matter Sciences I*, Holm, C.; Kremer, K., Eds. 2005; Vol. 173, 105-147.

106. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **1984,** *81*, 3684-3690.

107. Kollman, P., Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews* **1993,** *93*, 2395-2417.

108. Gilson, M. K.; Zhou, H.-X., Calculation of protein-ligand binding affinities. In *Annual Review of Biophysics and Biomolecular Structure*, 2007; Vol. 36, pp 21-42.

109. Zwanzig, R. W., High‐Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *Journal of Chemical Physics* **1954,** *22*, 1420-1426.

110. Straatsma, T. P.; McCammon, J. A., Multiconfiguration thermodynamic integration. *Journal of Chemical Physics* **1991,** *95*, 1175-1188.

111. Åqvist, J.; Medina, C.; Samuelsson, J. E., A new method for predicting binding affinity in computer-aided drug design. *Protein Engineering* **1994,** *7*, 385-391.

112. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., Calculating structures and free energies of complex

molecules: Combining molecular mechanics and continuum models. *Accounts of Chemical Research* **2000,** *33*, 889-897.

113. Case, D. A., Normal-mode analysis of protein dynamics *Current Opinion in Structural Biology* **1994,** *4*, 285-290.

114. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; M. Bearpark; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; J. Tomasi; M. Cossi; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.;. Foresman, J. B; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision B.01*, Gaussian, Inc.: Wallingford CT, 2010.

115. Zhao, Y.; Truhlar, D. G., A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *Journal of Chemical Physics* **2006,** *125*.

116. Lynch, B. J.; Zhao, Y.; Truhlar, D. G., Effectiveness of diffuse basis functions for calculating relative energies by density functional theory. *Journal of Physical Chemistry A* **2003,** *107*, 1384-1388.

117. Fast, P. L.; Sanchez, M. L.; Truhlar, D. G., Multi-coefficient Gaussian-3 method for calculating potential energy surfaces. *Chemical Physics Letters* **1999,** *306*, 407-410.

118. Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A., Gaussian-3 theory using reduced Møller-Plesset order. *Journal of Chemical Physics* **1999,** *110*, 4703-4709.

119. Frisch, M. J.; Pople, J. A.; Binkley, J. S., Self-consistent molecular-orbital methods 25. Supplementary functions for Gaussian basis sets. *Journal of Chemical Physics* **1984,** *80*, 3265-3269.

120. Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V., Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li-F *Journal of Computational Chemistry* **1983,** *4*, 294-301.

121. Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A., Gaussian-3 (G3) theory for molecules containing first and second-row atoms. *Journal of Chemical Physics* **1998,** *109*, 7764-7776.

122. Remya, K.; Suresh, C. H., Which Density Functional Is Close to CCSD Accuracy to Describe Geometry and Interaction Energy of Small Noncovalent Dimers? A Benchmark Study Using Gaussian09. *Journal of Computational Chemistry* **2013,** *34*, 1341-1353.

123. Carey, F.; Sundberg, R., *Advanced Organic Chemistry. Part A: Structure and Mechanisms*. 4th ed.; 2002.

124. Smith, M. B.; March, J., *March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*. 6th ed.; Wiley: 2007.

125. Kleywegt, G. J.; Harris, M. R.; Zou, J. Y.; Taylor, T. C.; Wahlby, A.; Jones, T. A., The Uppsala Electron-Density Server. *Acta Crystallographica Section D* **2004,** *60*, 2240-2249.

126. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* **2004,** *60*, 2126-2132.

127. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallographica Section D* **2010,** *66*, 486-501.

128. Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L.-W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C., PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallographica Section D* **2002,** *58*, 1948-1954.

129. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D* **2012,** *68*, 352-367.

130. Painter, J.; Merritt, E. A., Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallographica Section D* **2006,** *62*, 439-450.

131. Case, D. A.; Darden, T. A.; III, T. E. C.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitber, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 11*, University of California, San Francisco, 2010.

132. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics & Modelling* **2006,** *25*, 247-260.

133. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; J. R. Cheeseman; J. A. Montgomery, J.; 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;

Johnson B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision D.02*, Gaussian, Inc.: Wallingford CT, 2004.

134. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *Journal of Physical Chemistry* **1993,** *97*, 10269-10280.

135. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics & Modelling* **1996,** *14*, 33-38.

136. Schaftenaar, G.; Noordik, J. H., Molden: a pre- and post-processing program for molecular and electronic structures. *Journal of Computer-Aided Molecular Design* **2000,** *14*, 123-134.

137. *Grace*, Version 5.1.22; http://plasma-gate.weizmann.ac.il/Grace/: 1996-2008.

138. *The PyMOL Molecular Graphics System*, Version 1.3; Schrödinger, LLC: 2009-2010.

139. D'Abramo, M.; Rabal, O.; Oyarzabal, J.; Gervasio, F. L., Conformational Selection versus Induced Fit in Kinases: The Case of PI3K-gamma. *Angewandte Chemie-International Edition* **2012,** *51*, 642-646.

140. *Sybyl*, Version 7.3; Tripos Inc. : St. Louis, MO, 2004.

141. Fletcher, D. A.; McMeeking, R. F.; Parkin, D., The United Kingdom Chemical Database Service. *Journal of Chemical Information and Computer Sciences* **1996,** *36*, 746-749.

142. Kanolt, C. W., Ionization of water at 0 degrees, 18 degrees and 25 degrees derived from conductivity measurements of the hydrolysis of the ammonium salt of diketotetrarydrothiazole. *Journal of the American Chemical Society* **1907,** *29*, 1402-1416.

143. Pinson, J.-A.; Schmidt-Kittler, O.; Zhu, J.; Jennings, I. G.; Kinzler, K. W.; Vogelstein, B.; Chalmers, D. K.; Thompson, P. E., Thiazolidinedione-Based PI3K alpha Inhibitors: An Analysis of Biochemical and Virtual Screening Methods. *ChemMedChem* **2011,** *6*, 514-522.

144. Pomel, V.; Klicic, J.; Covini, D.; Church, D. D.; Shaw, J. P.; Roulin, K.; Burgat-Charvillon, F.; Valognes, D.; Camps, M.; Chabert, C.; Gillieron, C.; Francon, B.; Perrin, D.; Leroy, D.; Gretener, D.; Nichols, A.; Vitte, P. A.; Carboni, S.; Rommel, C.; Schwarz, M. K.; Rueckle, T., Furan-2-ylmethylene thiazolidinediones as novel, potent, and selective inhibitors of phosphoinositide 3-kinase gamma. *Journal of Medicinal Chemistry* **2006,** *49*, 3857-3871.

145. Tahmassebi, D., Theoretical studies of the tautomerism in oxazolidine and thiazolidine heterocyclic systems. *Journal of Molecular Structure: THEOCHEM* **2003,** *638*, 11-20.

146. Sundriyal, S.; Khanna, S.; Saha, R.; Bharatam, P. V., Metformin and glitazones: does similarity in biomolecular mechanism originate from tautomerism in these drugs? *Journal of Physical Organic Chemistry* **2008,** *21*, 30-33.

147. Enchev, V.; Chorbadjiev, S.; Jordanov, B., Comparative Study of the Structure of Rhodanine, Isorhodanine, Thiazolidine-2,4-dione, and Thiorhodanine. *Chemistry of Heterocyclic Compounds* **2002,** *38*, 1110-1120.

148. Valls, N.; Segarra, V. M.; Alcalde, E.; Marin, A.; Elguero, J., Synthesis, spectroscopy and tautomeric study of thiazoles substituted in position-2 and position-4 by hydroxy, mercapto and amino-groups. *Journal für Praktische Chemie* **1985,** *327*, 251-260.

149. McNaught;, A. D.; Wilkinson, A., *IUPAC. Compendium of Chemical Terminology*. 2nd ed.; Blackwell Scientific Publications: Oxford, 1997.

150. Martin, Y. C., Let's not forget tautomers. *Journal of Computer-Aided Molecular Design* **2009,** *23*, 693–704.

151. Minkin, V. I.; Garnovskii, A. D.; Elguero, J.; Katritzky, A. R.; Denisko, O. V., The tautomerism of heterocycles: Five-membered rings with two or more heteroatoms. *Advances in Heterocyclic Chemistry* **2000,** *76*, 157-323.

152. Elguero, J.; Katritzky, A. R.; Denisko, O. V., Prototropic tautomerism of heterocycles: Heteroaromatic tautomerism - General overview and methodology. In *Advances in Heterocyclic Chemistry*, Katritzky, A. R., Ed. 2000; Vol. 76, 1-84.

153. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977,** *23*, 327-341.

154. Loncharich, R. J.; Brooks, B. R.; Pastor, R. W., Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methyl-amide. *Biopolymers* **1992,** *32*, 523-535.

155. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **2012,** *8*, 3314-3321.

156. Muzzioli, E.; Del Rio, A.; Rastelli, G., Assessing protein kinase selectivity with molecular dynamics and mm-pbsa binding free energy calculations. *Chemical Biology & Drug Design* **2011,** *78*, 252-259.

157. Gilson, M. K.; Sharp, K. A.; Honig, B. H., Calculating the Electrostatic Potential of Molecules in Solution: Method and Error Assessment. *Journal of Computational Chemistry* **1988,** *9*, 327-335.

158. Tsui, V.; Case, D. A., Molecular dynamics simulations of nucleic acids with a generalized born solvation model. *Journal of the American Chemical Society* **2000,** *122*, 2489-2498.

159. Tsui, V.; Case, D. A., Theory and applications of the Generalized Born solvation model in macromolecular simulations. *Biopolymers* **2000,** *56*, 275-291.

160. Boys, S. F.; Bernardi, F., Calculation of small molecular interactions by differences of separate total energies. Some procedures with reduced errors. *Molecular Physics* **1970,** *19*, 553-566.

161. Schwenke, D. W.; Truhlar, D. G., Systematic study of basis set superposition errors in the calculated interaction energy of two HF molecules. *Journal of Chemical Physics* **1985,** *82*, 2418-2426.

162. Hobza, P.; Müller-Dethlefs, K., *Non-covalent Interactions: Theory and Experiment*. The Royal Society of Chemistry: UK, 2009.

163. Gohlke, H.; Kiel, C.; Case, D. A., Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *Journal of Molecular Biology* **2003,** *330*, 891-913.

164. Dunitz, J. D.; Taylor, R., Organic fluorine hardly ever accepts hydrogen bonds. *Chemistry-A European Journal* **1997,** *3*, 89-98.

165. Dunitz, J. D., Organic fluorine: Odd man out. *ChemBioChem* **2004,** *5*, 614-621.

166. Howard, J. A. K.; Hoy, V. J.; Ohagan, D.; Smith, G. T., How good is fluorine as a hydrogen bond acceptor? *Tetrahedron* **1996,** *52*, 12613-12622.

167. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* **2011,** *51*, 69-82.

168. Gohlke, H.; Case, D. A., Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *Journal of Computational Chemistry* **2004,** *25*, 238-250.

169. Hou, T.; McLaughlin, W. A.; Wang, W., Evaluating the potency of HIV-1 protease drugs to combat resistance. *Proteins-Structure Function and Bioinformatics* **2008,** *71*, 1163-1174.

Part II: Synthesis of novel fluorescent Ca^{2+} indicators

1. Introduction

This chapter begins with a brief introduction to the importance of intracellular calcium, followed by a summary of the fundamental theoretical principles required to understand the phenomenon of fluorescence. Subsequently, the concept of fluorescent indicators will be introduced, and a brief review of current fluorescent Ca^{2+} indicators will then be provided. This is followed by a description of fluorescence resonance energy transfer (FRET) and the development of FRET systems using green fluorescent protein (GFP). Finally, the concept of fluorescent indicators will be linked to GFP and the importance of this approach will be discussed, leading to the aims of the research project.

1.1 Importance of calcium in biology

The calcium ion (Ca^{2+}) plays an important role in the regulation of cell physiology, controlling aspects of virtually all cellular functions[1, 2]. Increases in cytosolic calcium trigger muscle contraction and the secretion of hormones, switch genes on and off, control neuronal signalling, provoke key events in the cell cycle, induce changes in the egg after fertilisation, help cells to defend themselves by removing damaging molecules from the cell surface, and control cell death (apoptosis)[3-8].

To understand the crucial roles that calcium plays, it is necessary to measure its concentration inside living cells. For this purpose, sensitive methods of quantifying changes in intracellular calcium levels in ways that do not disrupt cell function are required. Many modern means of determining calcium levels depend on fluorescence[9, 10]. The principles of fluorescence spectroscopy will now be briefly introduced.

1.2 The phenomena of fluorescence

Some molecules are capable of being excited, via absorption of light energy, to a higher energy state, also called an excited state. Fluorescence[11-13] is the emission of

light that occurs from an excited singlet state, that is to say, when the electron in the excited orbital has the opposite spin orientation to the ground-state electron.

1.2.1 Electronic transitions and the Jablonski diagram

A substance that is capable of undergoing electronic transitions that ultimately result in fluorescence is called a fluorophore[14]. The Jablonski energy diagram[15, 16] presented in Figure 1.1 provides an excellent illustration of the entire range of allowed electronic transitions that occur between the absorption and emission of light by a fluorophore.
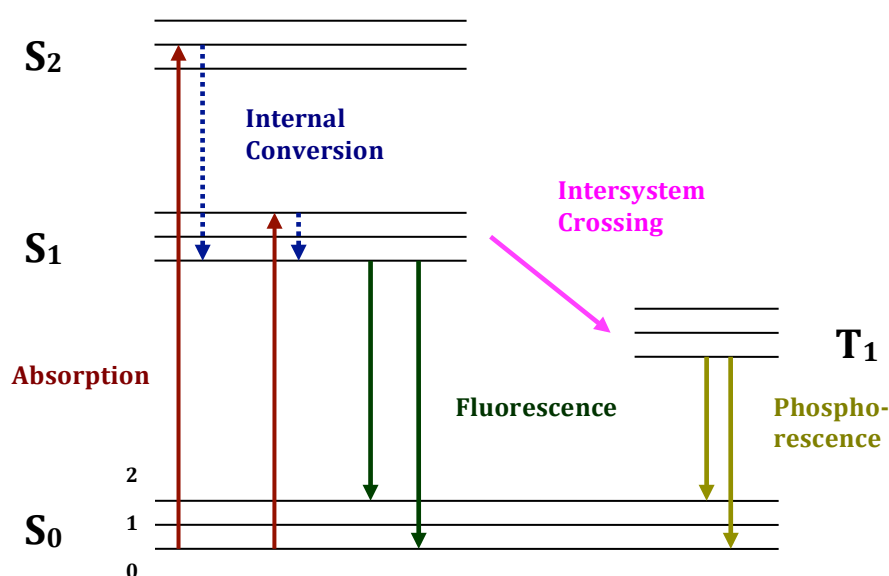
Figure 1.1: Jablonski energy diagram. Edited from ref[11].

In its ground state (S_0), the fluorophore molecule is in a stable configuration, with all the electrons in as low energy orbitals as possible. When a photon from an external source (such as electromagnetic radiation in the near IR, visible or UV region) strikes a fluorophore molecule, the molecule can absorb the light energy. The increase in energy is equal to the energy of the photon as expressed by the following relation:

$$E = h\nu = \frac{hc}{\lambda}$$

(1.1)

where h is the Planck's constant, ν and λ are the frequency and the wavelength of the radiation, respectively, and c is the speed of light. Following photon absorption, several processes usually occur.

The fluorophore is commonly excited to the higher electronic levels of the first (S_1) or the second (S_2) vibrational levels. In general, molecules in condensed phases rapidly relax from any higher electronic levels to the lowest vibrational energy level of the first excited state, S_1. This vibrational relaxation process is called internal conversion. It involves the loss of energy in the absence of light emission and generally occurs within 10^{-12} seconds or less[11].

Subsequently, the fluorophore rearranges from the first excited state (S_1) back to the ground state (S_0) by giving up its excess energy as fluorescence radiation. It should be noted that during the absorption and excitation processes, the electron did not change its spin direction; it merely passed from ground singlet state to excited singlet state. This means that the radiative deactivation pathway takes place between two states of identical spin multiplicity. Therefore, fluorescence is a spin-allowed transition that occurs rapidly. The lifetime of a fluorophore, namely the average time between its excitation and return to the ground state, is usually around 10 ns[17].

However, other relaxation pathways can occur. Molecules in the first excited state (S_1) may undergo intersystem crossing, which involves a transition to the lowest excited triplet state (T_1). This event involves spin-flip. The emission of a photon from T_1 is called phosphorescence. As transition from T_1 back to the ground state (S_0) is spin-forbidden, the rate constants for triplet emission are several orders of magnitude smaller than those for fluorescence.

Alternatively, the molecule may follow another relaxation pathway. The excited state energy could be dissipated non-radiatively as heat or it could be transferred to another molecule via fluorescence resonance energy transfer (FRET). The latter will be discussed in section 1.4.

1.2.2   Absorption and emission spectra

Each fluorescent molecular species absorbs a range of wavelengths of light energy, and also emits a range of wavelengths. Within these ranges the fluorophore has one or more optimal wavelengths for excitation (excitation maximum) and a related optimum wavelength of emission (emission maximum). The excitation maximum usually (but not always) coincides with the wavelength of maximum absorption[12].

The range of excitation and emission wavelengths can be represented in the form of a fluorescence spectrum, such as the one depicted in Figure 1.2. These spectral

characteristics of the fluorophore depend upon its chemical structure and the solvent in which it is dissolved.

There are three rules generally followed by compounds exhibiting fluorescence[11, 18]. Firstly, the emission spectra are generally independent of the excitation wavelength (Kasha's rule). This is a consequence of the rapid relaxation processes occurring in the fluorescent molecule.

Secondly, the light energy emitted is always of a longer wavelength than the light energy absorbed. This phenomenon is called the Stokes shift (Figure 1.2). The causes of the Stokes shift depend on the fluorophore and its environment. However, one common cause is the energy lost during the transient excited lifetime (i.e. the rapid decay to the lowest vibrational level S_1).

And thirdly, the emission spectrum of a molecule is typically a mirror image of its absorption spectrum (the mirror image rule). This occurs because emission usually takes place at higher vibrational ground states (S_0), and because electronic excitation does not greatly alter the nuclear geometry.
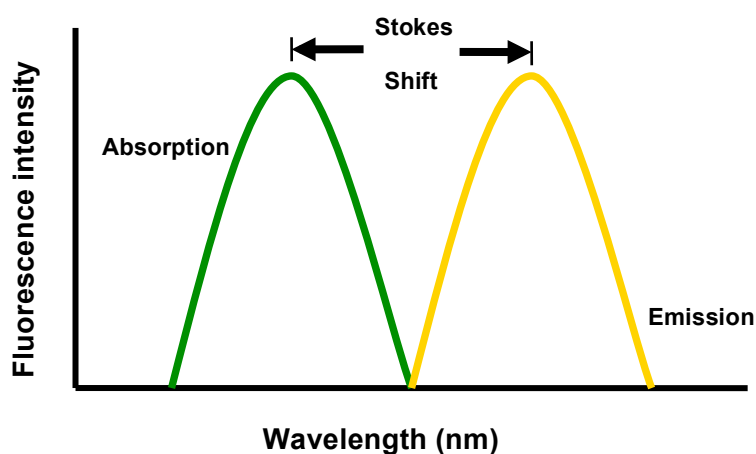
Figure 1.2: Representation of a fluorescence spectrum: the emitted photons have less energy and are shifted to longer wavelengths. This means that the colour of the light that is emitted is different from the colour of the light that has been absorbed.

To obtain an emission spectrum the different wavelengths of fluorescent light emitted by a sample (the fluorophore) are measured by holding the excitation light at a constant wavelength. Conversely, to obtain an excitation spectrum the emission light is held at a constant wavelength, and the excitation light is scanned through many different wavelengths.

The next section is devoted to a particular type of fluorescent molecule whose spectral characteristics are currently used for measurement of intracellular Ca^{2+} levels.

1.3 Fluorescent Ca^{2+} indicators

Fluorescent indicators[19, 20] are molecules able to transform chemical information (for example, the concentration of calcium ion inside a living cell) into an analytically useful signal, i.e. fluorescent emission.

A fluorescent indicator, or probe, for Ca^{+2} recognition, is formed from two different components: a recognition moiety (termed ionophore) and a signalling moiety, the fluorophore[21, 22]. The recognition moiety is responsible for the selectivity and efficiency of ion binding. Its design depends on ligand topology, the properties of the cation of interest (such as ionic radius, charge, hardness or coordination number) and the nature of the solvent system. On the other hand, the signalling moiety or fluorophore acts as a signal transducer. Namely, it converts the information (recognition event) into an optical signal (fluorescence emission) expressed as a change in its photophysical properties. The photophysical changes of a fluorescent probe upon ion binding can involve various photoinduced processes such as electron transfer (PET) or charge transfer (PCT)[23, 24].

Fluorescent probes enable the investigation of changes in Ca^{2+} concentrations because their physicochemical characteristics in both the free and bound forms are sufficiently different. Depending on the spectral changes that occur upon ion binding, fluorescent probes can be categorised as either ratiometric or single wavelength indicators[11, 19].

Single wavelength indicators (also known as non-ratiometric dyes) exhibit significant Ca^{2+} dependent changes in fluorescence intensity without shifting their excitation or emission wavelengths (Figure 1.3). Consequently, monitoring relative changes in fluorescence intensity allows the determination of Ca^{+2} concentration, $[Ca^{+2}]$. The mathematical relationship is expressed by the following equation[9]:

$$[Ca^{2+}] = K_d \ x \ \frac{F - F_{min}}{F_{max} - F}$$

(1.2)

where *F* represents the value of the fluorescence intensity, *Fmax* and *Fmin* are calibration constants and correspond to the fluorescence intensity at saturating and zero Ca^{2+} concentrations, respectively, and K_d, is the dissociation constant.
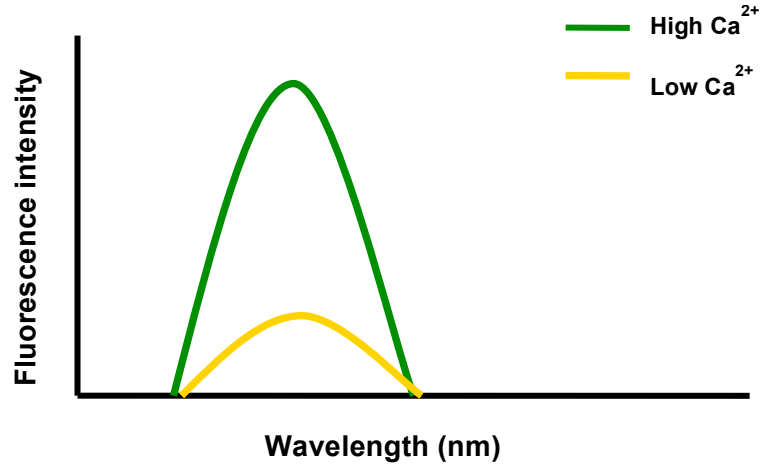
Figure 1.3: Schematic representation of the emission spectra of single wavelength indicators.

Conversely, ratiometric indicators shift the peak wavelength of either their excitation or emission spectrum according to the free Ca^{+2} concentration (Figure 1.4). Therefore, Ca^{+2} concentration is calculated as the ratio between two fluorescence intensity values that are taken at two different wavelengths. The mathematical relationship between the measured ratio (*R*) and the calcium ion concentration is expressed as follows:

$$[Ca^{2+}] = K_d \; x \; \frac{R - R_{min}}{R_{max} - R} \; x \; \frac{S_{f2}}{S_{b2}}$$

(1.3)

where K_d is the dissociation constant, *R* is the measured ratio of intensity between two wavelengths ($\lambda 1$ and $\lambda 2$), the values of R_{max} and R_{min} are the ratio values measured under conditions of saturating calcium levels and in the absence of calcium, respectively, and S_{f2}/S_{b2} is the ratio of calcium-free to calcium-bound fluorescence intensities at wavelength two ($\lambda 2$). The derivation of equation 1.3 thus depends on four parameters that have to be independently calibrated.

The use of ratriometric indicators allows for correction of uneven dye loading, dye leakage, photobleaching and changes in cell volume, and therefore permits a very accurate quantification of the Ca^{2+} concentration, $[Ca^{+2}]$.
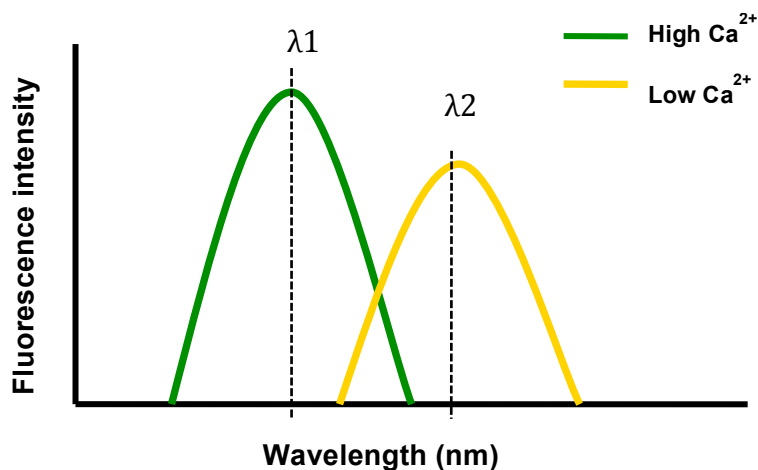
Figure 1.4: Schematic representation of the emission spectra of ratiometric indicators.

An important consideration when choosing an indicator is its Ca^{2+}-binding affinity, which is reflected in the dissociation constant[25]. The dissociation constant of the indicator (K_d) is the ionic concentration at which half of the probe has bound a Ca^{2+} ion, and has Molar units. Matching the dissociation constant to the concentration range of interest is therefore a crucial requirement. The dissociation constant is sensitive to pH, temperature, viscosity, ionic strength, competing ions, and cellular interactions. Thus, to quantify $[Ca^{+2}]$ the dissociation constant of the fluorescent probe should ideally be determined both *in situ* and in the specific cell used.

Chemical indicators that are now available allow for intracellular Ca^{2+} detection over a very large range (<50 nM to >50 μM). Higher affinity indicators can be used to quantify Ca^{2+} levels in the cytosol while lower affinity indicators can be optimised for measuring Ca^{2+} in subcellular compartments with higher concentrations[10, 19].

1.3.1 High-affinity Ca^{2+} indicators

In terms of the recognition moiety, a subset of compounds that have proven to be particularly well suited for cytosolic Ca^{2+} measurements has evolved from EGTA (**1**), the best-known calcium ion chelator. Its octa-coordinate tetracarboxylated binding site has selectivity for Ca^{2+} over Mg^{2+} of five orders of magnitude[26]. It has been reported[27] that this discrimination presumably stems from a binding cavity that has the right size for Ca^{2+} but not for Mg^{2+}, which is a much smaller ion. Unfortunately, the Ca^{2+}-buffering level of EGTA is strongly dependent on pH, making it unsuitable for quantification of intracellular Ca^{2+} levels.

Tsien[28] developed a derivative of EGTA in which the constituent methylene groups were replaced by benzene rings. This replacement has little effect on the overall geometry of the molecule; hence the binding site possesses much the same desirable size and shape as EGTA. Due to the incorporation of the aromatic rings, the pK_a of the amine nitrogens is lowered from ~8 or 9 down to ~5 or 6, making the molecule much less pH sensitive at physiological pH. This molecule was named BAPTA (**2**) as an acronym for bis(*o*-aminophenoxy)ethane-*N,N,N',N'*-tetraacetic acid.
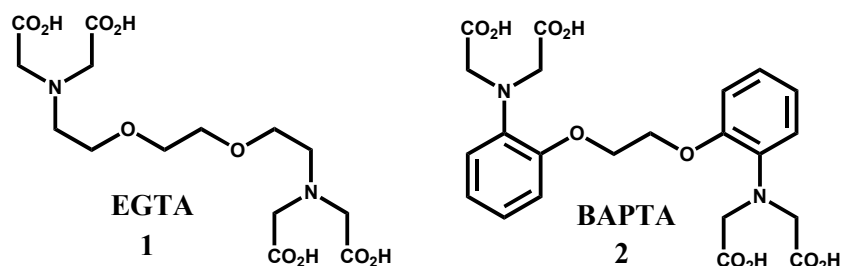
Figure 1.5: Structures of the calcium chelators EGTA (**1**) and BAPTA (**2**).

BAPTA absorbs light in the far ultraviolet (UV) spectrum[9]. Thus it cannot be used as an intracellular Ca^{2+} indicator. However, several fluorescent probes based on the BAPTA moiety have been developed. Among these, Fura-2[29] (**3**), Indo-1[29] (**4**) and Fluo-3[30] (**5**) are the most widely known. Fura-2 and Indo-1 are ratiometric probes, while Fluo-3 belongs to the category of single wavelength indicators.
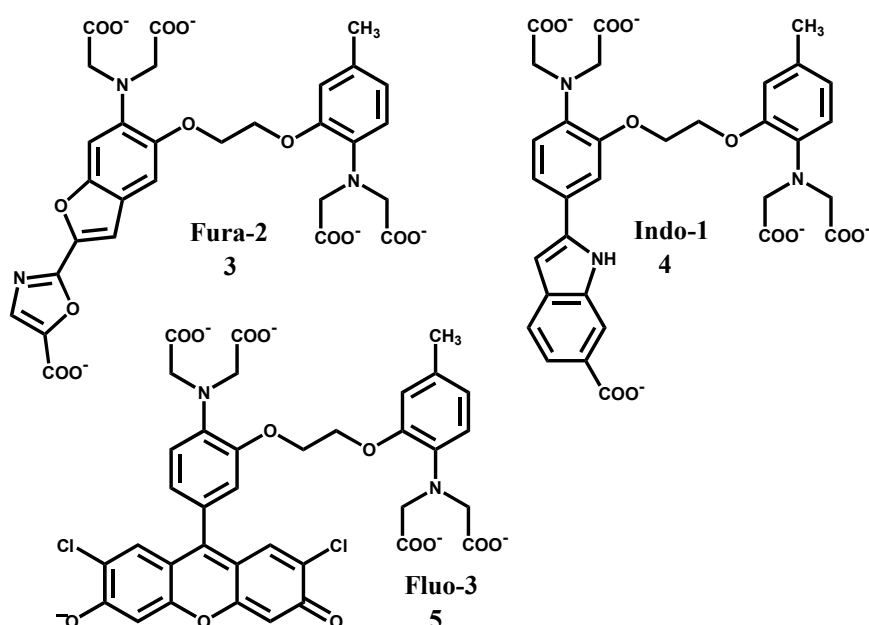
Figure 1.6: Structures of the calcium indicators Fura-2 (**3**), Indo-1 (**4**) and Fluo-3 (**5**).

Upon Ca^{2+} binding, the conformational change caused to BAPTA's carboxyl groups is transmitted to the fluorophore (or signalling moiety) and results in changes to the

excitation and/or emission properties of the fluorescent probe. The type of spectral changes which occur vary depending on the indicator. For example, Fura-2 undergoes a shift in the absorption peak, whereas Indo-1 suffers a shift in both absorbance and emission spectra. On the other hand, Fluo-3 being a single wavelength indicator, presents an increase or decrease in the efficiency of fluorescence. These three indicators are classified as high affinity Ca^{2+} indicators because they have dissociation constant values in the nanomolar scale ($K_d < 1$ µM).

Fluorescent indicators containing a BAPTA binding site can be loaded into a cell by masking the four carboxylates groups (**6**) as acetoxymethyl esters (AM, **7**)[31]. Once inside, non-specific esterases hydrolyse them to their free acid form, in which the indicator is able to bind calcium ions with great selectivity.

6: X=O⁻
7: X=OCH₂OCOCH₃
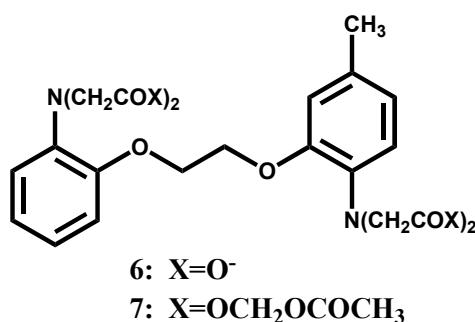
Figure 1.7: Structures of the calcium chelator (**6**) and its acetoxymethyl tetraester (**7**). Hydrolysis of the ester yields the chelator tetra-anion, four molecules of acetate, and formaldehyde and eight protons.

1.3.2 Low affinity coumarin-type fluorescent probes

Low affinity fluorescent probes ($K_d > 1$ µM) are useful for detecting calcium levels in the micromolar range that would saturate the response of high-affinity indicators such as the ones described in the previous section. Thus, this type of probe is used to measure Ca^{2+} in subcellular compartments where relatively high levels of intracellular Ca^{2+} are expected[19, 32]. As the ion dissociation rates of low-affinity indicators are faster, they are more suitable for tracking rapid Ca^{2+} flux kinetics than their high affinity counterparts[10].

In this section, two low affinity Ca^{2+} indicators crucial to this research project will be described. These compounds also have the Ca^{2+} chelating structure of BAPTA as recognition moiety. However, the signalling molecule is now coumarin-based.

Coumarin (**8**) consists of a benzene ring fused to an α-pyrone. This chromophore is stable upon irradiation and has long excitation and emission wavelengths[33]. Additionally, coumarins exhibit Stokes shifts large enough to avoid significant overlap of the excitation and emission spectra[34]. These characteristics offered the potential for the application of the coumarin chromophore as ion detectors in biological systems.

8
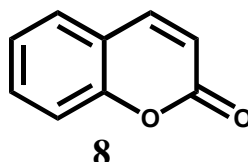
Figure 1.8: Chemical structure of coumarin.

The first coumarin-type polycarboxylate fluorescent ion probe, BTC (**9**), was synthesised by Katerinopoulous and co-workers[35]. It retained the BAPTA framework as ionophore, and incorporated a 3-benzothiazolyl coumarin moiety as fluorophore.

9 BTC Z=O
10 BTIC Z=NH
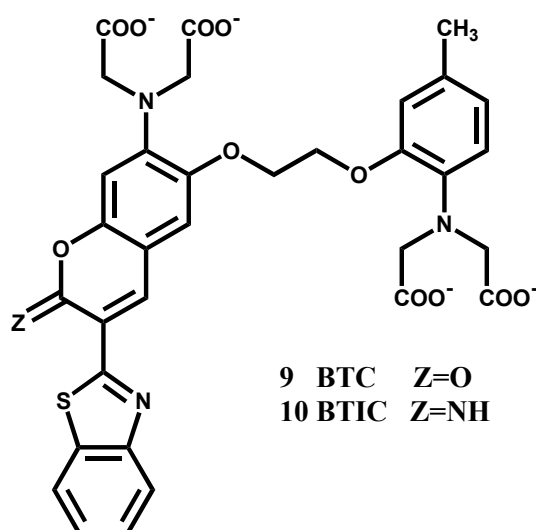
Figure 1.9: Structures of the low affinity calcium indicator BTC (**9**) and its analogue BTIC (**10**).

BTC offers ratiometric measurements using visible light excitation. The excitation maximum shifts from 462 to 401 nm upon binding Ca^{2+}. Its emission wavelength is 525 nm. It has been reported[10] that the higher excitation wavelength of BTC compared with that of Fura-2 or Indo-1 should lower cellular cytotoxicity and autofluorescence.

BTC's dissociation constant is 7.0 µM, which is much greater than that of Fura-2, Indo-1 or Fluo-3 (Table 1.1). Its lower calcium affinity enables accurate measurements of higher concentrations of calcium ion, as well as analysis of prompt

changes in Ca^{2+} concentration[36-38]. However, it should be noted that BTC might be too weak to accurately detect low Ca^{2+} concentrations.

Table 1.1: Dissociation constants for different Ca^{2+} indicators[27, 39].

Indicator	K_d (µM)
Fura-2	0.145
Indo-1	0.230
Fluo-3	0.390
BTC	7.000
BTIC	5.400

Katerinopoulous and co-workers[39] also synthesised an iminocoumarin analogue of BTC, called BTIC (**10**). They envisaged that the substitution of the carbonyl oxygen with an imino nitrogen had the advantage of enhancing the compound's polarity. This in turn could increase the solubility of the probe in the cytosol.

The excitation maximum of the free form of BTIC is at 471 nm, whereas its bound form has a slightly lower wavelength maximum at 414 nm (Table 1.2). It should be noted that this compound exhibits a 953% increase in fluorescence intensity upon calcium binding.

BTIC's dissociation constant, 5.4 µM, is slightly smaller than that for BTC (Table 1.1). The difference has been attributed to the lower electronegativity of the immino-coumarin nitrogen compared to the carbonyl oxygen of BTC[39].

Table 1.2: Properties of some fluorescent Ca^{2+} indicators[39].

Indicator	$\lambda_{excitation}$ (nm)		$\lambda_{emission}$ (nm)	
	Free indicator	Ca^{2+} complex	Free indicator	Ca^{2+} complex
BTC	462	401	525	525
BTIC	471	414	535	535

BTC and BTIC are two examples of the series of available coumarin-type fluorescent Ca^{2+} indicators. These two low-affinity ratiometric indicators with visible excitation wavelengths are of particular importance for the present project.

1.4 Fluorescence resonance energy transfer

A fluorophore in the excited state can, as discussed in section 1.2.2, follow different relaxation pathways. One process that can occur is fluorescence resonance energy transfer (FRET)[40-44]. This phenomenon takes place when two suitable fluorophores are in an appropriate relative orientation such that the excited fluorophore (called the donor) can transfer energy to the second, longer-wavelength fluorophore (acceptor) in a non-radiative manner.

It is important to emphasise that the transfer of energy in FRET does not involve the emission of a photon by the donor with subsequent absorbance by the acceptor. There is no intermediate photon in the FRET mechanism. This is a process mediated by long-range dipole-dipole coupling between the donor and the acceptor molecules.

Resonance energy transfer is possible when the emission spectrum of the donor molecule overlaps with the absorption spectrum of the acceptor molecule. This is termed the overlap integral.
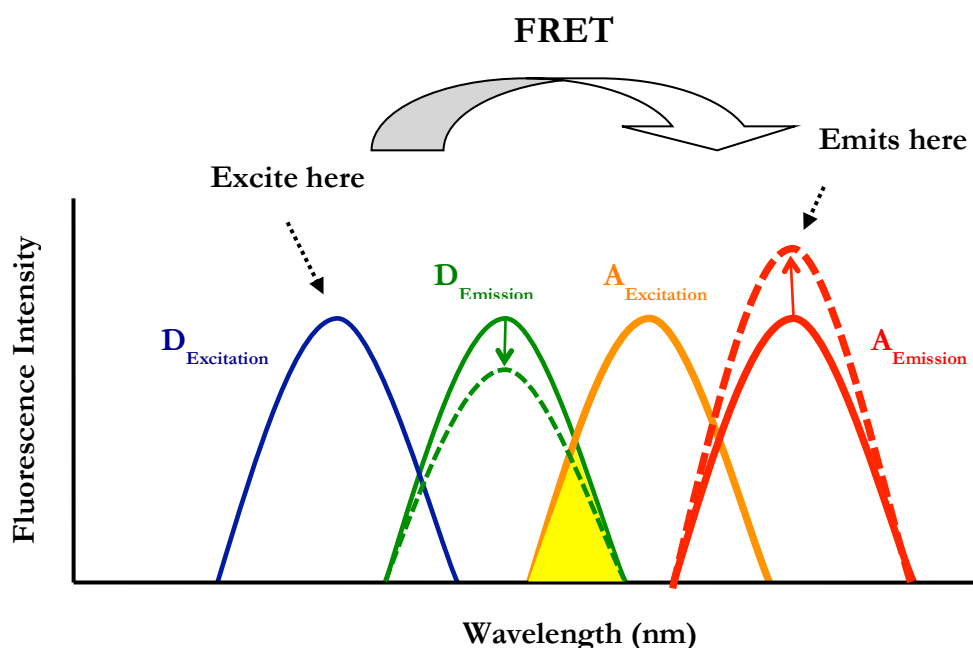
Figure 1.10: Schematic representation of the FRET process between the donor (D) and acceptor (A) molecules. The result is an increase in the emission intensity of the acceptor with accompanying loss of the emission intensity of the donor (dashed lines). The overlap integral between the donor emission and the acceptor absorption spectra is shown in yellow.

Presented in Figure 1.10 are the changes to the emission intensities of the spectral profiles from a donor and an acceptor molecule undergoing FRET. Firstly, light energy is absorbed at the excitation frequency for the donor molecule. Subsequently, the

donor transfers non-radiatively some of its energy to the acceptor, which then re-emits the light at its own emission wavelength. The net result is that the donor emits less energy than it normally would (since some of the energy it would radiate as light gets transferred to the acceptor instead), while the acceptor emits more light energy at its excitation frequency (because it is getting extra energy input from the donor fluorophore). Therefore, FRET is manifested with emission at the longer wavelengths (lower energies) characteristic of the acceptor molecule.

Three features determine the efficiency of FRET[43, 45, 46]. These are: the distance separating the donor and acceptor, which must be in the range 10-100 Å. Secondly, the degree of spectral overlap. And thirdly, the relative orientations of the donor emission transition dipole and the acceptor absorption transition dipole.

The efficiency of energy transfer for a single donor-acceptor pair at a fixed distance, according to Förster's theory[11, 40], is given by equation 1.4:

$$E = \frac{R_0{}^6}{R_0{}^6 + r^6}$$

(1.4)

where r is the distance between the centers of the donor and the acceptor molecules and R_0 (Förster distance) is the distance at which the energy transfer efficiency (E) is 50%. The Förster distance depends on the overlap integral of the donor emission spectrum with the acceptor absorption spectrum and their mutual orientation as expressed by equation 1.5:

$$R_0 = (J K^2 Q_0 n^{-4})^{1/6} \; x \; 9.7 \; x \; 10^{-3}$$

(1.5)

where J is the spectral overlap integral; K^2 the orientation factor for a dipole-dipole interaction; Q_0 the quantum yield of fluorescence of the energy donor in the absence of the acceptor; and n the refractive index of the medium between the donor and the acceptor.

As FRET is a nondestructive spectroscopic method, it can be applied to intact cells while they are performing physiological functions. Any biochemical signal that changes the distance between the two fluorophores, or the relative orientation of their transition dipoles will modulate the efficiency of FRET, thus changing the ratio between donor and acceptor emission intensities. Monitoring these changes allows molecular interactions to be measured with reasonable resolution (limited by the 10-100 Å distance requirement).

As a result of the above, measurements of the efficiency of FRET have been extensively used in biological applications. Some examples follow. FRET has been utilised as a "spectroscopic ruler" to assess distances and orientations between specific labelling sites within a single macromolecule or between two separate molecules[47-49]. Additionally, it has been used to investigate the structure of proteins or protein complexes[50-52], conformational molecular changes[53], interactions between molecules[54, 55], and as a powerful indicator of biochemical events[56-58].

In this study, the focus is on the use of FRET between GFP (green fluorescent protein) and a second fluorophore to study molecular interactions.

1.5 Green fluorescent protein

GFP is an extraordinary protein that exhibits green fluorescence when excited by ultraviolet or blue light. It was first isolated in 1962 from the bioluminescent jellyfish *Aequorea victoria* by Shimomura *et al.*[59] The wild type GFP has a major excitation peak at a wavelength of 395 nm and a minor one at 475 nm. Its emission peak is at a wavelength 508 nm[60].

The X-ray crystal structure of GFP[61, 62] shows a unique barrel-like structure, consisting of 11 strands of β-sheet with an α-helix inside. It is within this barrel that the chromophore is located, completely protected from bulk solvent. Most remarkably, the chromophore is formed in an autocatalytic cyclisation from a tri-peptide motif (Ser65-Tyr66-Gly67) in the primary structure of GFP. This spontaneous intramolecular process only requires oxygen and does not depend on the presence of enzymes or other auxiliary factors. As a result, in any biological or artificial system (except under strict anaerobic conditions), expression of the gene coding for GFP leads to the appearance of a fluorescent signal. This outstanding feature, together with GFP's stability and non-toxicity, enable its use for *in vivo* imaging. It is possible to express the gene for GFP within cells[63], and use it as a reporter gene. Additionally, it is possible to obtain proteins that are synthesised with attached GFP[64, 65]. This allows the visualisation of dynamic cellular events, as well as enabling protein localisation. More recently, it has been even possible to express GFP in entire organisms[66, 67].

1.5.1 GFP in FRET systems

Mutations have been introduced to the amino acid sequence of the wild type GFP (wtGFP) to produce various GFP-like molecules with higher photostability and different excitation and emission spectra throughout the visible region[68-70]. Some examples include the blue (BFP), cyan (CFP) and yellow (YFP) fluorescent proteins. These GFP mutants have been used as donor/acceptor pairs for FRET to develop a number of intracellular sensors, enabling monitoring of the local environment (for example calcium concentration[71, 72] or pH[73]).

When both the GFP donor and the acceptor are fused to the same host molecule, intramolecular FRET can be measured[74, 75]. Heim and Tsien[68] carried out an early application of this technique by fusing BFP and GFP in the same molecule, separated by a 25-residue linker containing a trypsin cleavage site. The intact fusion protein showed FRET. This was demonstrated because ultraviolet excitation produced green emission from the acceptor molecule. However, after proteolytic cleavage of the spacer, which caused the two protein domains to diffuse apart and thus disrupted FRET, the green emission from the acceptor GFP almost completely disappeared, whereas the blue emission from the donor BFP was enhanced. This construct provided a ratio-based detection method that was used to show the change in FRET following cleavage.

Alternative approaches have employed chemical fluorophores as intramolecular FRET partners[76]. Previously it has been mentioned that the distance between the donor and acceptor molecule determines the efficiency of FRET. When using two GFP mutants, the closest approach for the FRET pair is limited by their own protein structure. Attempts to overcome this limitation are exemplified by the FlAsH[77, 78] and ReAsH[79] methods.

1.5.2 Chemical modification of GFP

Another alternative for improving intramolecular FRET systems was recently described by Suzuki *et al.*[80] Their approach consists in the introduction of a range of fluorescent dyes (for example, BODIPY, Alexa Fluor, rhodamine and eosin) onto a GFP mutant at a carefully chosen site on the protein surface. The attached chemical label then serves as a FRET partner to the natural GFP chromophore. Furthermore, the introduction of a cleavage site that is responsive to chemical[81] or enzymatic

action[82, 83] provides an efficient ratio-based detection method that relies on the change in FRET following cleavage.

The starting point for the construction of these probes was a GFP mutant, GFPuv5, which has seven point mutations relative to wtGFP and fluoresces 240 times more strongly than the latter[84]. It was then necessary to prepare a mutated GFP with a unique cysteine residue that could undergo chemical derivatisation, thus serving as attachment point for the chemical FRET partner. For this purpose and to ensure that the site for chemical derivatisation was unique, the native Cys48 and Cys70 residues of GFPuv5 were replaced by non-thiolic amino acids (either Ser or Ala). Next, to obtain a suitable distance for an intramolecular FRET partner, a new Cys residue was introduced either by mutation of Ile229 (localised at the C-terminal region) or Glu6 (localised at the N-terminal region).

Subsequent chemical derivatisation of the cysteine sites was straightforward; the thiol group of the engineered cysteine can be attached to a thiol-reactive fluorescent dye. The separation between the centroids of attached chemical and GFP chromophore in mutant Ile229Cys and Glu6Cys was measured as approximately 23 and 21 Å, respectively. Despite these very similar distances, computational methods[85] as well as experiments[82] have shown increased energy transfer of the Ile229Cys mutant (called uv5GFPC48I229) compared with the Glu6Cys mutant.

An example of the effectiveness of these protein constructs as intramolecular FRET systems, and as bioprobes, will now be given. The work by Suzuki *et al.*[83] describes a GFP mutant (uv5casS00tag), which possess a unique cysteine site as well as a substrate sequence susceptible to caspase-3 localised in the C-terminal region. The mutated GFP was chemically modified by a maleimide derivative of Alexa Fluor 532. In the resulting system, the native GFP chromophore acts as FRET donor and the appended chemical dye as FRET acceptor.

Following excitation at 488 nm, a strong intramolecular FRET signal at 550 nm is exhibited. At this stage, the emission of the native GFP fluorophore is quenched, and FRET is manifested by the emission at the longer wavelengths characteristic of the acceptor molecule. However, after caspase-3 cleavage, the components providing FRET are physically separated and FRET is destroyed causing the emission wavelength to switch back to that of the native GFP fluorophore (511 nm).

The effectiveness of this system was demonstrated *in vitro* (in cell lysates) and *in vivo* in living HeLa cells undergoing apoptosis. In both cases, *in vivo* and *in vitro*, the new

intramolecular-FRET-based bioprobes have proven their utility as sensitive indicators of caspase-3 activity. This approach is not only versatile (many other cleavage sites amenable to enzymatic or chemical cleavage can be introduced), but also time-saving. Running the GFP signal as a result of a fused gene construct and gene transfection requires the cell line to reach confluence, which usually takes a number of days. On the other hand, a chemically modified GFP mutant can now be introduced into a cell in a fully constructed and folded form using cell penetration enhancers, in only a few hours[83].

Besides providing efficient ratio-based detection *in vitro* and in living cells, these systems have another remarkable advantage: the formation of two physically separated molecules. This opens the possibly of detection by molecular size of fragment changes using fluorescence correlation spectroscopy (FCS), a method which has not yet been applied using these constructs.

The methods described provide the basis for the design of novel bioprobes using intramolecular FRET between GFP and carefully sited chemical labels. During this research project this approach will be extended to try to develop a fluorophore that could allow calcium quantification inside living cells.

1.6   Aims of the Project

The aim of this research project is to synthesise a new fluorescence Ca^{2+} indicator that can be selectively attached to the surface cysteine residue of green fluorescent protein mutant uv5GFPC48I229.

If a fluorescent Ca^{2+} indicator could be chemically conjugated with a screening protein such as green fluorescent protein, this would create a method that allows Ca^{2+} visualisation in living cells as they perform functions such as apoptosis, while at the same time enabling quantification of calcium levels. Further, the same chemistry could allow the attachment of the Ca^{2+} indicator to surfaces (e.g. 96-well plate formats) and other media (e.g. nucleic acids). It should also be possible to simultaneously incorporate enzyme-cleavable sites within the construct to target multiplexing applications[86].

The new fluorescent probes proposed are a modification of BTC (**9**) and its iminocoumarin analogue BTIC (**10**). This decision was taken based on the probes dissociation constants and spectral profile. The proposed new indicators retain the

tetracarboxylated ionophore and incorporate a structurally modified fluorophore (Figure 1.11 and 1.12). The modified fluorophore must contain a SH-reactive site to site-specifically attach it to the surface cysteine residue of the mutant GFP.

9 BTC   **X=H**
11 Derivative   **X=CH₂-CH₂-NH-CO-CH₂-Br**
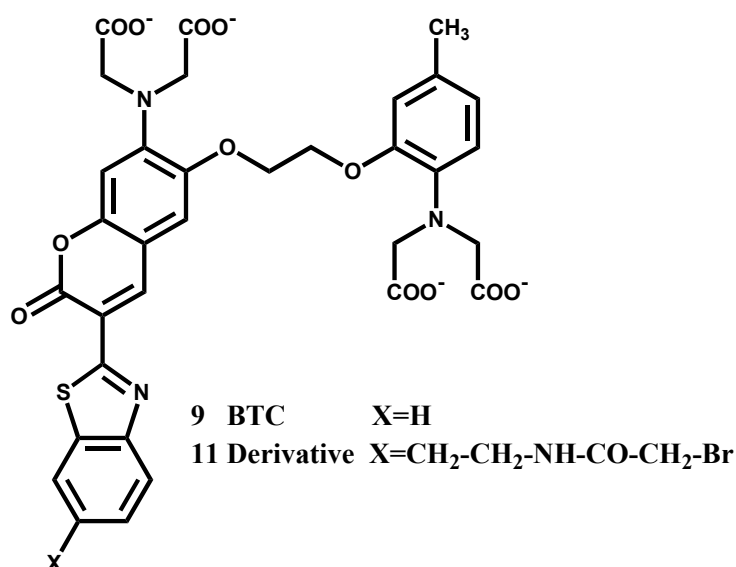
Figure 1.11: Structure of BTC (**9**) and the proposed thiol-reactive new indicator (**11**).

Two different approaches will be taken to derivatise BTC and BTIC. It was envisaged that the incorporation of a α-haloketo moiety to the benzothiazolyl group of BTC would result in a thiol-reactive probe (**11**, Figure 1.11). Similarly, the incorporation of a halogen moiety into the benzothiazolyl group of BTIC was envisaged to result in a derivatisable probe (**12-14**, Figure 1.12).

10 BTIC   **X=H**
12 Derivative   **X=Cl**
13 Derivative   **X=Br**
14 Derivative   **X=I**
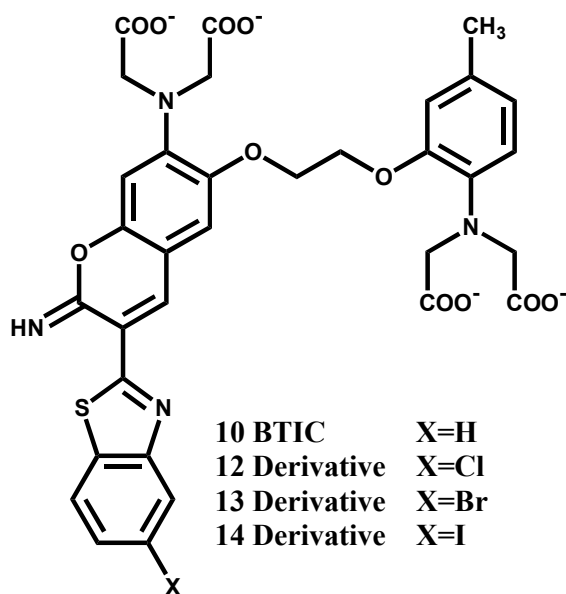
Figure 1.12: Structure of BTIC (**10**) and its halogenated derivatives (**12-14**).

The proposed methods of attachment to GFP are depicted in Figure 1.13. The fluorescent probe derivative bearing an α-haloketo moiety could be attached to the surface cysteine of GFP in a S_N2 type reaction. On the other hand, the use of click chemistry[87, 88] would enable the attachment of the halogenated fluorescent probe to the single SH site of GFP. Additionally, further transformation of the halide moiety (X) into an olefin (for example by Heck reaction[89, 90]) would enable the use of thiol-ene Michael addition reaction[91, 92].
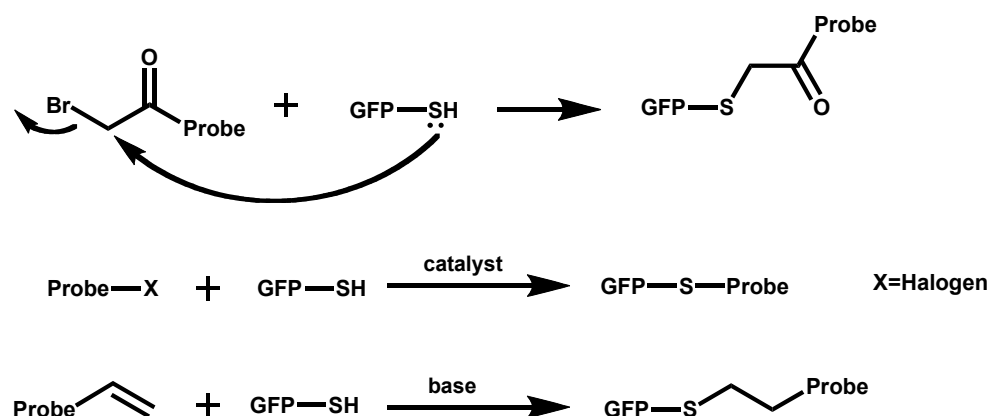
Figure 1.13: Attachment of derivatised probe to the surface cysteine of GFP.

It is important to note that this work takes place as the first step towards a longer four-stage research project. These stages involve:

(1) Chemical synthesis of the new indicator.

(2) Evaluation of the spectral properties of the newly synthesised fluorophore as well as binding affinities for calcium and other ions such as K^+, Na^+ or Mg^{2+}.

(3) Chemical conjugation of the new fluorophore and GFP mutant uv5GFPC48I229.

(4) Introduction of the chemically modified fluorophore-GFP into the cell.

This work specifically deals with the first objective, and exclusively aimed to progress towards the synthesis of the novel fluorescent calcium indicator.

1.6.1 Proposed chemical synthetic routes

The proposed synthetic route to potential fluorescent probe **11** involves the condensation of esterified benzothiazole **16** with salicylaldehyde **15** by a Knoevenagel-type reaction to first give **17** (Scheme 1).
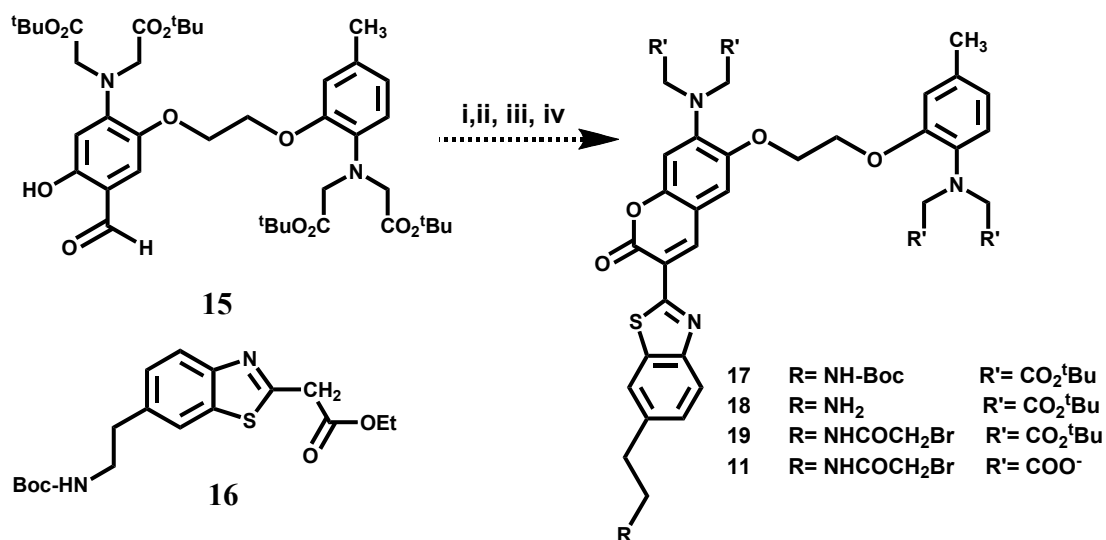
i) **Piperidine, EtOH, r.t. 12h; ii) Et$_3$SiH; iii) bromoacetyl bromide; iv) TFA.**

Scheme 1: Proposed synthetic route to potential fluorescent probe **11**.

Selective N-Boc-deprotection of **17** using triethylsilane and TFA would be expected to yield amine **18**, which would then be treated with bromoacetyl bromide to form α-haloamide **19**. Finally, acid cleavage of the o-tertiary-butyl groups will yield the tetraacid **11**. It may not be possible to de-protect the NH-Boc and Bu^{t}O$_2$C sites selectively. If so, an interesting option would be to try to use Ca^{2+} to selectively protect the released CO$_2$H group while the NH$_2$ is functionalised.

The synthetic route to halogenated iminocoumarin derivative **21** is also expected to proceed *via* a Knoevenagel condensation between the versatile precursor **15** and halogenated benzothiazole **20**. Deprotection of the four ^{t}butyl esters would furnish the desired potential fluorescent probe **12**.
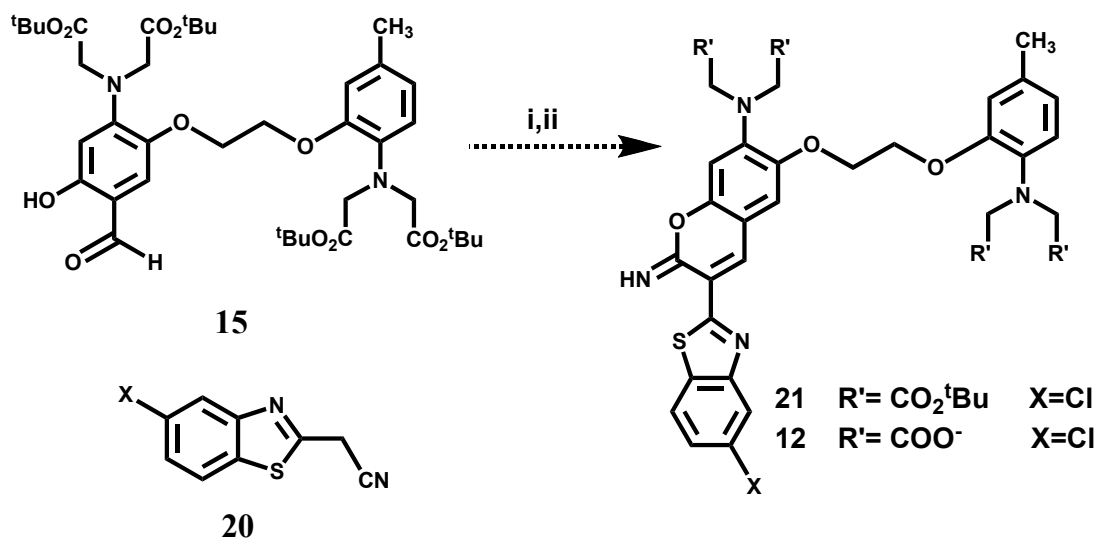
i) **Piperidine, EtOH, r.t. 12h; ii) TFA.**

Scheme 2: Proposed synthetic route to potential fluorescent probe **12**.

The synthesis of salicylaldehyde **15** is a seven-step procedure (Scheme 3), which had been successfully completed in our laboratory by Dr. Laura L. Etchells. Commercially available 2,5-dibenzyloxy-nitrobenzene (**22**) was transformed to **23** by selective debenzylation in the presence of trifluoroacetic acid. 5-Methyl-2-nitrophenol (**24**) was treated with 1,2-dibromoethane affording **25**. Subsequent coupling of **23** and **25** gave product **26**. Hydrogenation of the nitro groups using a palladium catalyst generated the amino functionalities in **27**, which underwent *N*-esterification in the presence of *t*-butyl bromoacetate and proton sponge (**28**). Finally, a Vilsmeier formylation produced **29** and hydrogenation using 5% acetic acid and palladium as catalyst afforded salicylaldehyde **15**. This thesis contains description of attempts to provide a derivative suitable for derivatisation of the reagent **15** from Dr. Etchells' work.
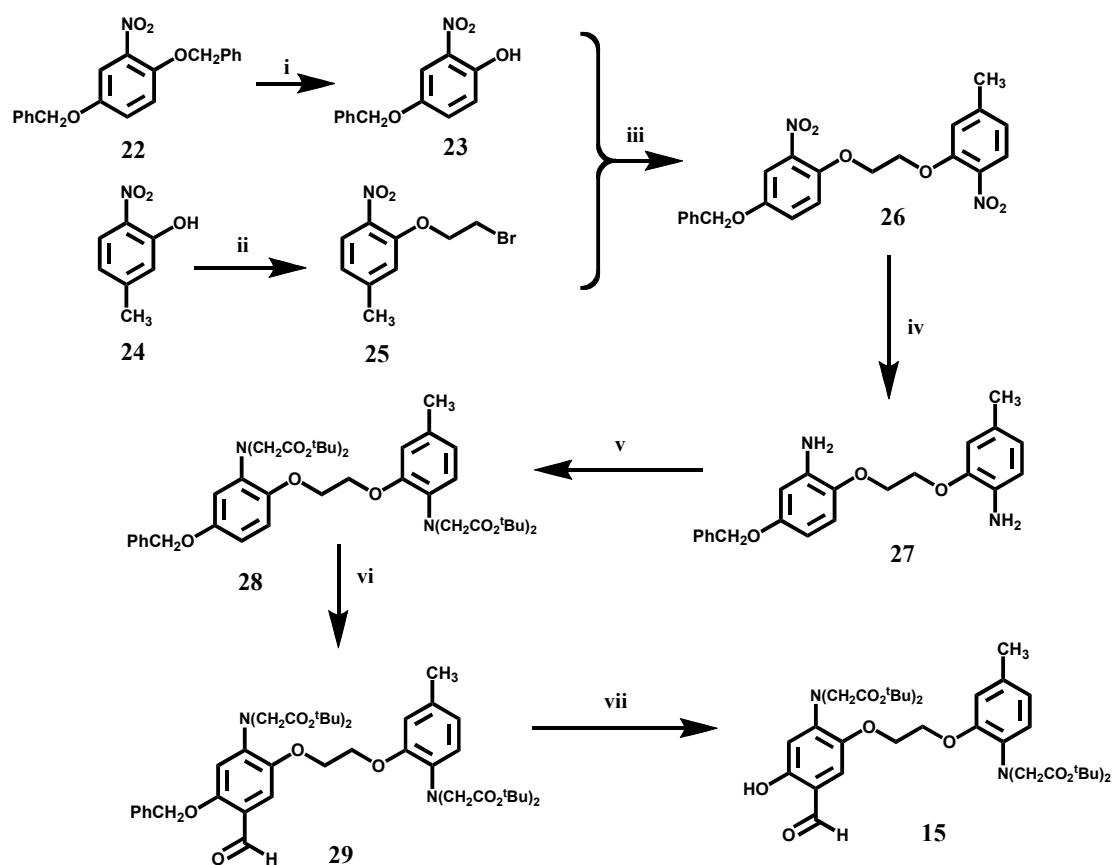
i) TFA, CHCl$_3$, r.t., 56% ii) 1,2-dibromoethane, K$_2$CO$_3$, DMF, reflux, 45% iii) K$_2$CO$_3$, DMF, 91% iv) H$_2$ 10% Pt/C, THF/MeOH, 87% v) ^{t}butyl bromoacetate, proton sponge, NaI, ACN, reflux, 62% vi) POCl$_3$, pyridine, DMF, 90% vii) H$_2$, 10% Pd/C, EtOAc, 5% acetic acid, 95%

Scheme 3: Synthetic route to salicylaldehyde **15**.

Progress towards the synthesis of potential fluorescent probes **11** and **12** is the primary aim of this project and will be reviewed in the following chapters.

2. Results and discussion

This chapter provides a discussion of the progress toward the derivatisation of BTC and BTIC. A detailed description of each synthetic step is followed by the description and analysis of mechanistic pathways as well as the corresponding spectroscopic data that confirm the structures of the individual products.

2.1 Progress towards the synthesis of BTC derivatives

As a first step, the proposed synthetic route to potential fluorescent probe **11** involves the condensation of esterified benzothiazole **16** with salicylaldehyde **15** by a Knoevenagel-type reaction (Scheme 4).

|		
17	R= NH-Boc	R'= CO$_2^t$Bu
18	R= NH$_2$	R'= CO$_2^t$Bu
19	R= NHCOCH$_2$Br	R'= CO$_2^t$Bu
11	R= NHCOCH$_2$Br	R'= COO$^-$
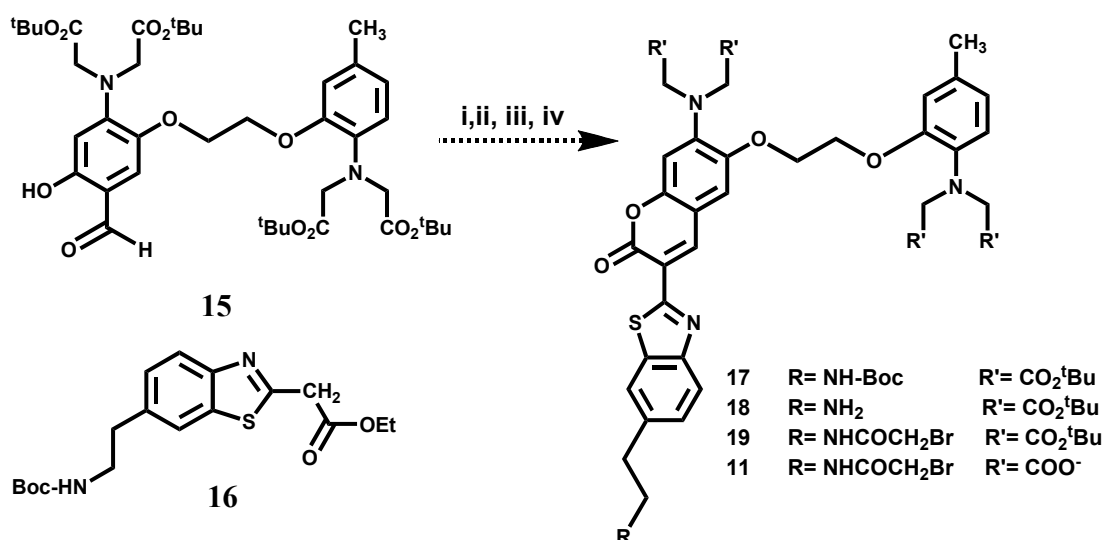
i) Piperidine, EtOH, r.t. 12h; ii) Et$_3$SiH; iii) bromoacetyl bromide; iv) TFA.

Scheme 4: Proposed synthetic route to potential fluorescent probe **11**.

A five-step synthesis of **16** was designed commencing with commercially available 4-aminophenethylamine **30** (Scheme 5). The key transformations in this synthesis are described in the following sections.
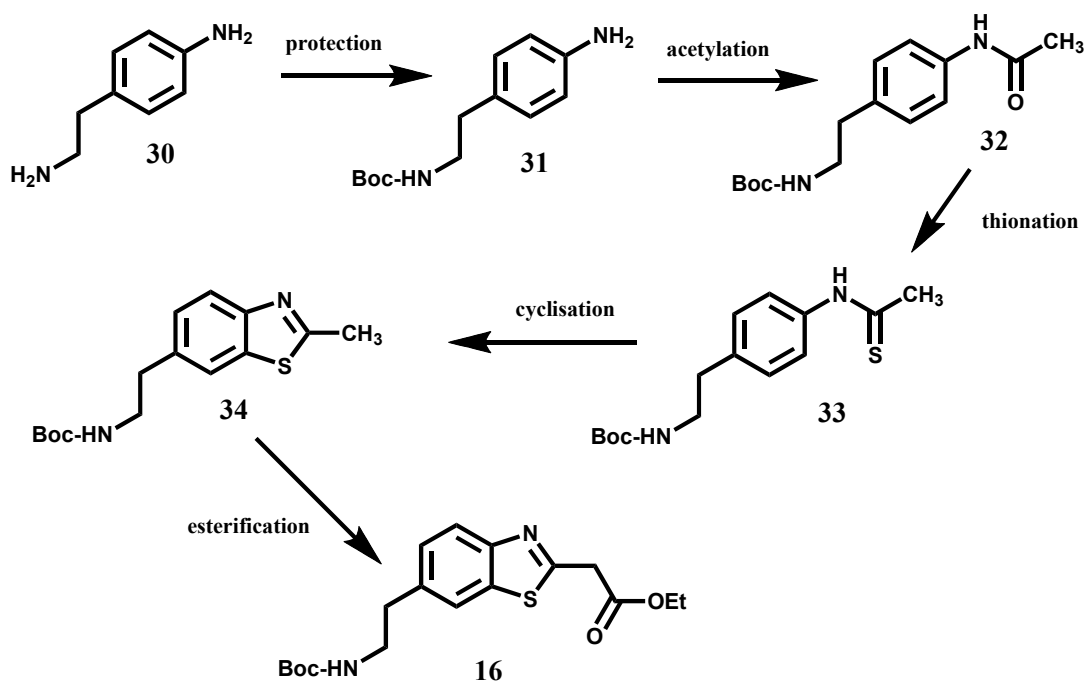
Scheme 5: Proposed synthetic route to benzothiazole **16.**

2.1.1 Selective Boc-protection of 4-aminophenethylamine (30)

The proposed starting material 4-aminophenethylamine (**30**) possesses two amine functional groups. As such, to ensure selective reactivity in the subsequent synthetic steps, the introduction of a protective group onto one of the amine functionalities was necessary.
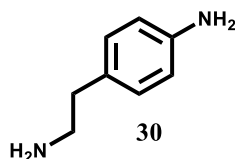
Scheme 6: 4-aminophenethylamine.

It was desirable to protect the alkyl amine group. To this end, it was required to have a protecting group that could overcome the great nucleophilicity of nitrogen, as well as the weaker leaving group ability of an ammonium species. An additional requisite was that deprotection (to release the original amine) could occur under mild conditions. The N-*tert*-butoxycarbonyl (N-*t*-Boc) protecting group fulfils these characteristics and was therefore chosen.

The N-*t*-Boc carbonate ester reacts with amines to produce *tert*-butyl carbamates, which are stable in the presence of a wide range of nucleophiles and under alkaline

conditions[93]. Additionally, the N-*t*-Boc functionality is very labile and can easily liberate the parent amine under acidic conditions[94, 95].

Formation of N-*t*-Boc-protected amines is conducted by reaction with di-*tert*-butyldicarbonate (Boc$_2$O) (**35**). This amino-de-acyloxy-substitution is generally assumed[96] to proceed through a two-step mechanism: a tetrahedral addition intermediate is formed by nucleophilic attack of the amine, and then the leaving group is expelled.

As illustrated in Scheme 7, addition begins with donation of the amine nitrogen lone pair of electrons to the partially positively charged carbon atom belonging to Boc$_2$O (**35**). This forms a new σ bond to the nucleophile, giving rise to intermediate **36** with a fully negatively charged oxygen. Subsequently, the alkoxy group reverts back to a carbonyl group while at the same time expelling the leaving group, a carboxylate ion. The leaving group is now a free anion (**38a**, **38b**), which is unstable as a carbonate half ester, and it decomposes to carbon dioxide and a *tert*-butoxide anion. The reaction finishes with a proton transfer between the positively charged nitrogen of **37** and the oxygen from the *tert*-butoxide anion, generating the Boc-protected amine (**39**) and *tert*-butanol.
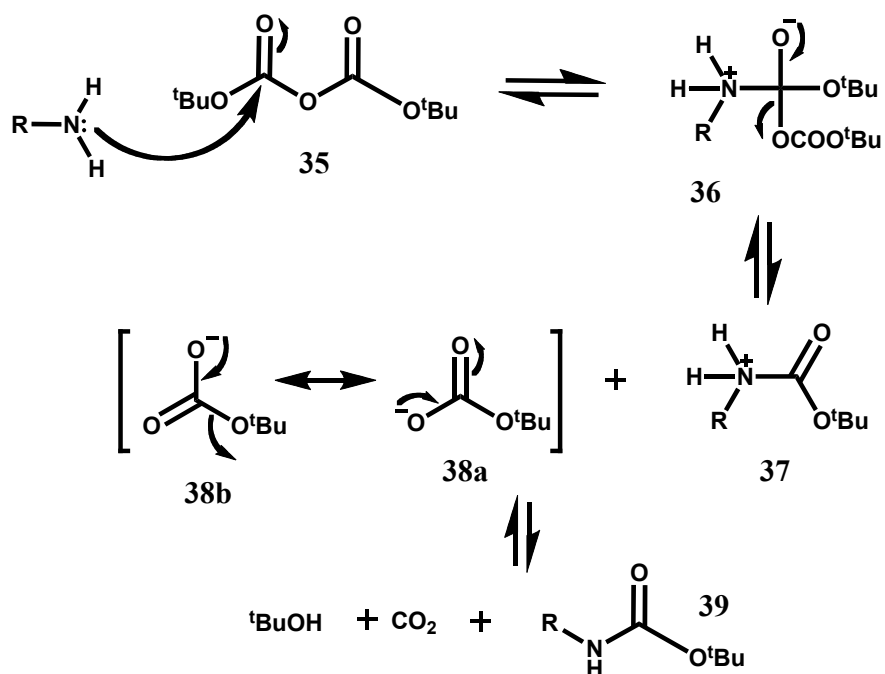
Scheme 7: Mechanism of the N-*t*-Boc protection of amines.

Proton transfers between oxygen and nitrogen acids and bases are usually extremely fast, hence the rate-determining step for these reactions is assigned to the attack of the amine on the *tert*-butyl dicarbonate[97, 98].

Commercially available 4-aminophenethylamine (**30**) was treated with di-*tert*-butyldicarbonate (Boc$_2$O) in dichloromethane to afford the corresponding protected amine **31** in 95% yield (Scheme 8). The product did not require further purification.
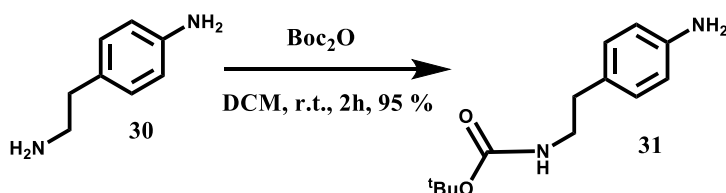
Scheme 8: N-*t*-Boc protection of 4-aminophenethylamine.

The selectivity of this reaction is due to the higher pK$_a$ of an alkylamine (ca.10-11) compared to the lower pK$_a$ of an arylamine (ca. 3-5)[99, 100]. With its higher pK$_a$, the alkylamine is considered to be the more basic. As the basicity of an organic compound governs its ability to donate an electron pair, the more basic alkylamine will be the attacking moiety during the first step of the mechanism.

Synthesis of protected amine **31** was confirmed by spectroscopic analysis, which included ^{1}H NMR and ^{13}C NMR, as well as electrospray mass spectroscopy. The ^{1}H NMR spectrum of **31** (Figure 2.1) contained a series of well defined signals corresponding to those reported in the literature[101]. The most significant signal is the singlet at δ 1.43 ppm, which corresponds to the nine hydrogen atoms of the N-*t*-Boc moiety (Figure 2.1). The presence of amine protons was confirmed by hydrogen-deuterium exchange using D$_2$O.

The ^{13}C NMR spectrum further confirmed the formation of **31**. In this case, the singlet at δ 155.90 ppm corresponds to the carbonyl group in the N-*t*-Boc functionality (Figure 2.2).

The melting point of mono-Boc-protected amine **31** was found to be 83-85 $^{\circ}$C, which corresponds to the melting point reported (83-85 $^{\circ}$C) by Dannhardt and co-workers[102]. Additionally, in accordance to the expected value, the electrospray mass spectroscopy identified [M+Na]$^+$ = 259.

After this Boc-protection step, the protected amine functionality became stable towards hydrogenation, bases and most nucleophiles.
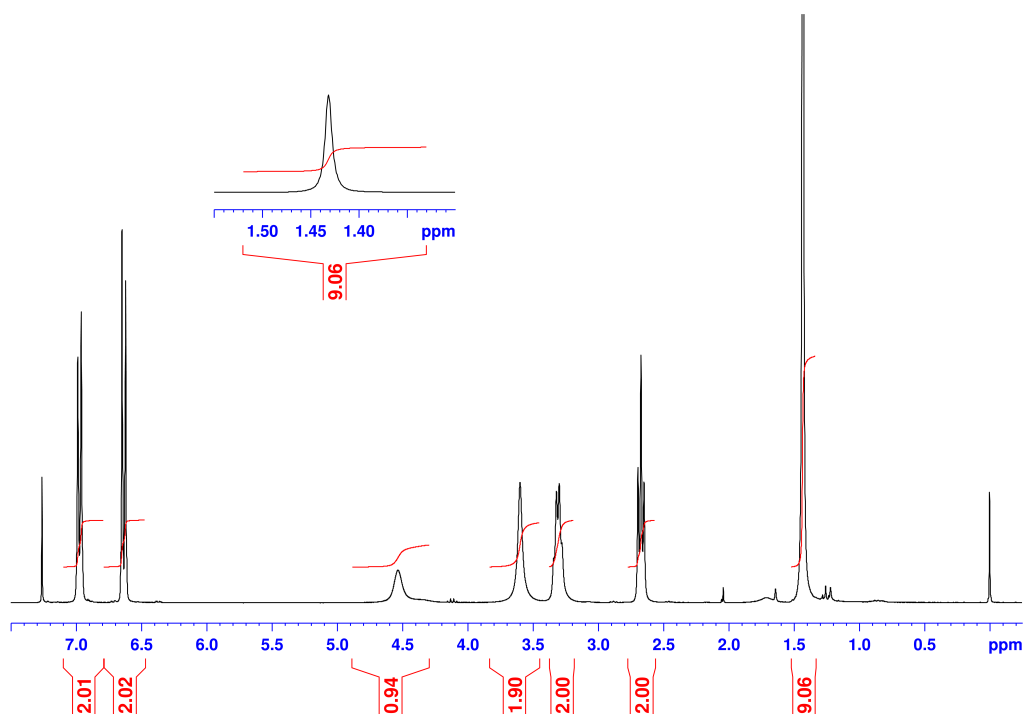
Figure 2.1: ^{1}H NMR spectrum of *tert*-butyl 4-aminophenethylcarbamate (**31**) in CDCl$_3$.
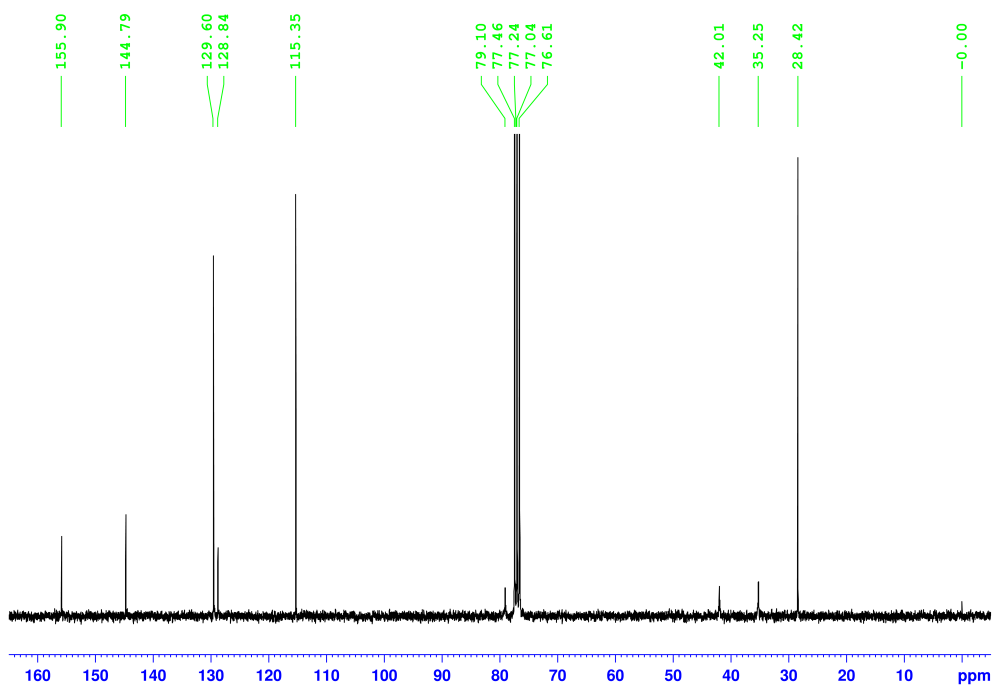
Figure 2.2: ^{13}C NMR spectrum of *tert*-butyl 4-aminophenethylcarbamate (**31**) in CDCl$_3$.

2.1.2   Acetylation of *tert*-butyl 4-aminophenethylcarbamate (31)

The next step of the synthetic route was acetylation of the unprotected amine with a carboxylic acid chloride in the presence of triethylamine acting as an organic base. This nucleophilic acyl substitution is shown in Scheme 9.
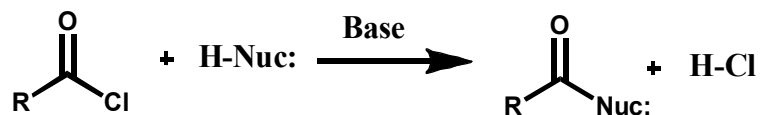
Scheme 9: Nucleophilic acyl substitution under basic conditions.

The conversion of an acid chloride into an amide can be achieved using base assistance to speed up the reaction and to remove acid by-products. Bulky, non-nucleophilic bases, such as triethylamine (Et_3N) or *N,N*-diisopropylethylamine (DIEA), are typically used.

The mechanism is presented in Scheme 10. Firstly, the amine nucleophile attacks the electrophilic C=O carbon atom belonging to the acid chloride. Rehybridisation of the carbonyl carbon from sp^2 to sp^3 then occurs, and a tetrahedral alkoxide ion intermediate is produced. An electron pair from the oxygen displaces the leaving group Cl^-, generating a new carbonyl compound (an amine) as product. The main role of Et_3N is to remove HCl and drive the reaction to completion. Et_3N can also speed up the reaction by displacing Cl^- from the acid chloride to generate a positively charged species $[R(C=O)NEt_3]^+$ which is significantly more reactive than the acid chloride with $R-NH_2$.
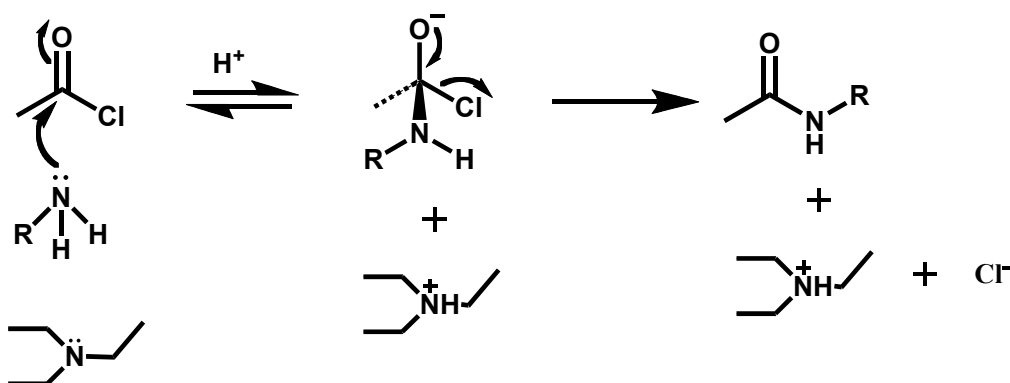
Scheme 10: Mechanism of the acetylation of an amine with a carboxylic acid chloride in the presence of triethylamine.

Following the standard procedure proposed by Lee and co-workers[103], compound **31** was treated with acetyl chloride in the presence of triethylamine (Scheme 11). Product **32** was obtained in 93% yield.
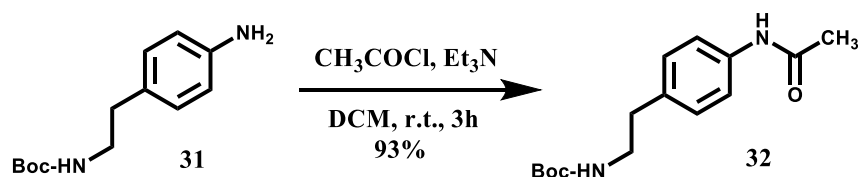
Scheme 11: Acetylation of **31**.
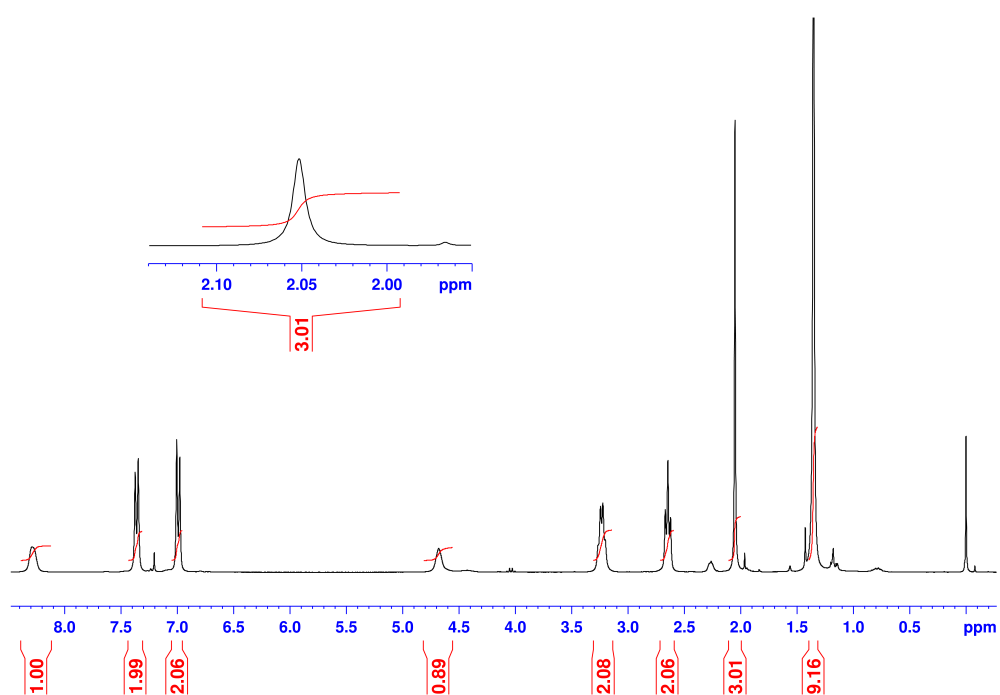
Figure 2.3: ^{1}H NMR spectrum of *tert*-butyl 4-acetamidophenethylcarbamate **(32)** in CDCl$_3$.

The ^{1}H NMR spectrum of **32** (Figure 2.3) shows a singlet of methyl protons (δ 2.05 ppm) confirming the presence of an acetyl group in the molecule. As expected, it also shows the following: two sets of doublets corresponding to the aromatic protons (δ 7.00 ppm, 7.37 ppm); two broad singlets corresponding to the proton of an N-H bond (δ 4.68 and 8.28 ppm); a quartet for a methylene group (δ 3.24 ppm) and a triplet of the other methylene protons (δ 2.64 ppm); and a singlet of nine methyl protons for the *tert*-butyl moiety (δ 1.35 ppm). The presence of two NH groups was confirmed by deuterium exchange using D$_2$O.
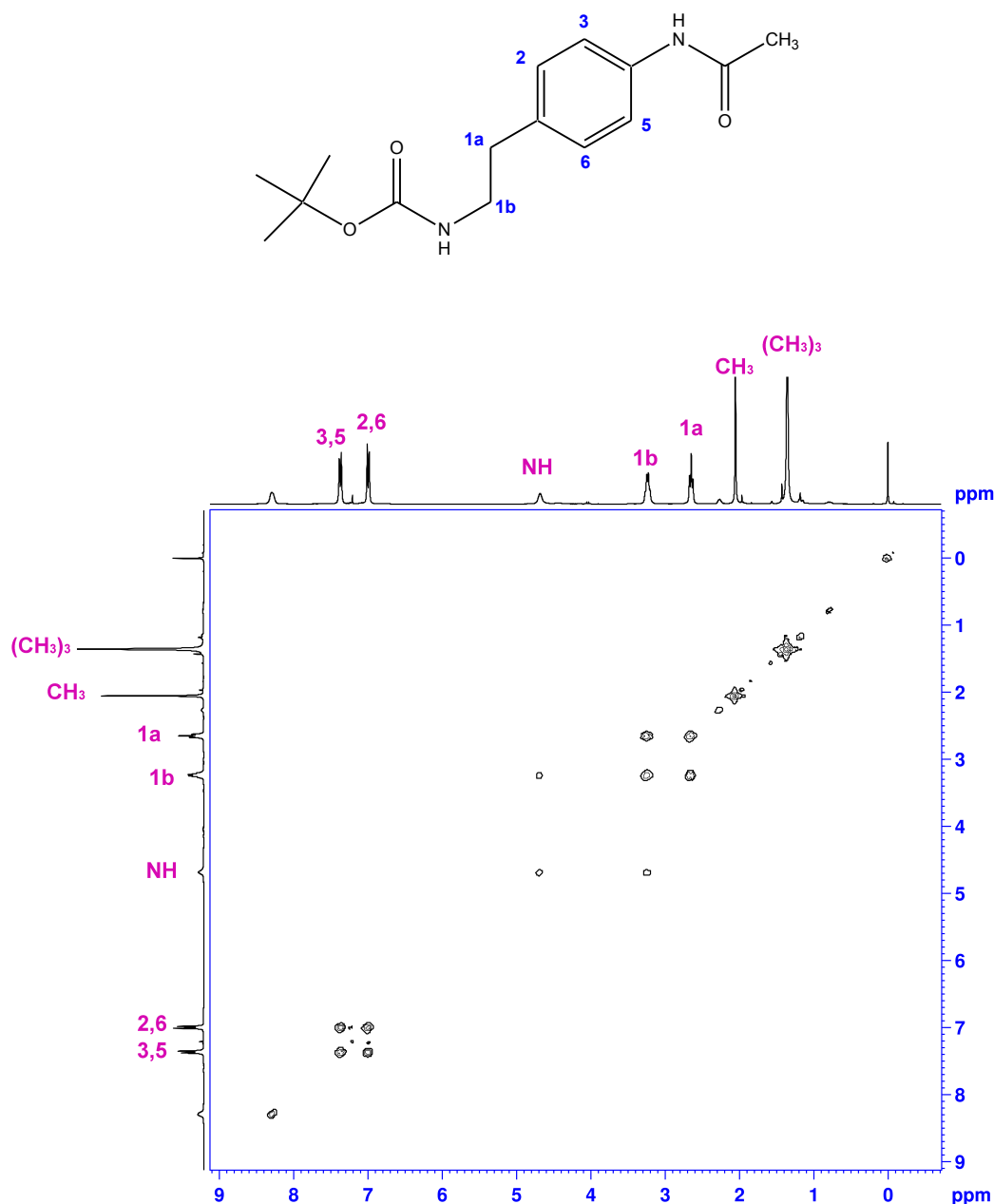
Figure 2.4: ^1H- ^1H COSY NMR spectrum of *tert*-butyl 4-acetamidophenethylcarbamate (**32**) in $CDCl_3$.

The assignment of the ^1H NMR spectrum signals was aided by the use of a 2D ^1H-^1H COSY experiment, shown in Figure 2.4. The COSY (correlation spectroscopy) experiment is an efficient way of establishing the connectivity of a molecule by determining which protons are spin-spin coupled. The experiment's result is a two-dimensional contour map, each dimension representing proton chemical shifts and the contours indicating signal intensity. The diagonal peaks correspond to the standard 1D peaks in a ^1H NMR experiment (i.e. the same as in Figure 2.3), while a cross peak indicates coupling between a pair of protons. As depicted in Figure 2.4, cross peaks exist between protons **1a** and **1b**, and between **1b** and the **NH**. There are

also well defined cross peaks between the aromatic protons. On the other hand, there are no cross peaks for the protons for the *tert*-butyl moiety (δ 1.35 ppm), nor for the methyl protons (δ 2.05 ppm).

The ^{13}C spectrum offers further characterisation of the molecule as it relates directly to the carbon skeleton. The successful synthesis of compound **32** is also confirmed by the presence of a peak at δ 24.56 ppm in the ^{13}C spectrum, which corresponds to the carbon atom of a methyl group (Figure 2.5).
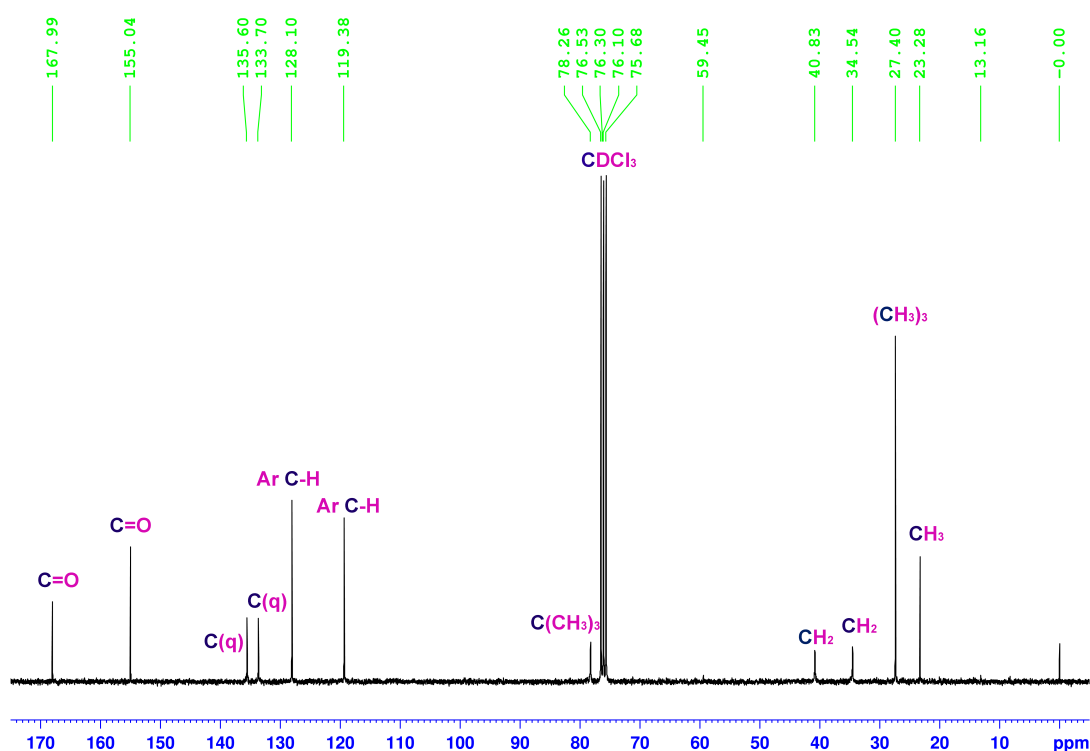
Figure 2.5: ^{13}C NMR spectrum of *tert*-butyl 4-acetamidophenethylcarbamate (**32**) in CDCl$_3$.

In addition, a DEPT-135 spectrum (Figure 2.6) corroborates that the peak at δ 24.56 ppm is indeed a CH$_3$ group. The DEPT (distortionless enhancement by polarisation transfer) experiment is a very useful method to establish the multiplicity of the carbon atom, that is, to determine the presence of primary, secondary and tertiary carbons. The technique relies on polarisation transfer, namely the transfer of proton magnetisation onto the directly bound carbon nucleus. In a DEPT-135 spectrum, signals from CH$_2$ groups are negative, while CH and CH$_3$ groups generate positive signals. Since quaternary carbon atoms have no attached protons, they are not seen in DEPT spectra.
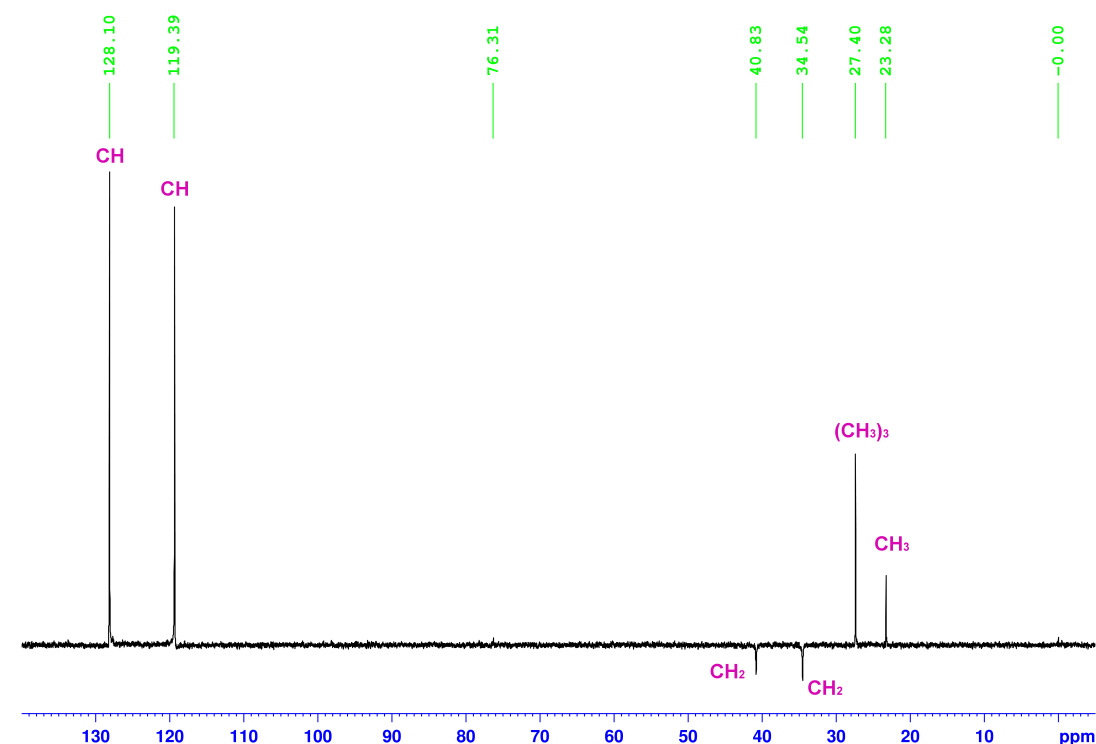
Figure 2.6: DEPT-135 spectrum of *tert*-butyl 4-acetamidophenethylcarbamate (**32**) in CDCl$_3$.

Lastly, a ^{1}H-^{13}C HMQC experiment (Figure 2.7) was carried out for compound **32**. The HMQC (heteronuclear multiple quantum correlation) experiment is a heteronuclear correlation technique that allows the identification of H-C connectivity within a molecule. This technique relies on magnetisation transfer from the proton to its directly attached carbon atom, and back onto the proton (for higher sensitivity). The results are displayed in a similar manner to those from the COSY experiment explained above. However, in this case one dimension of the 2D map corresponds to ^{13}C chemical shifts while the other dimension denotes the ^{1}H chemical shifts. Consequently, the cross peaks define which carbon atom a particular proton (or group of protons) is attached to. In this way, the correlation between heteronuclei is established.

As this technique requires two data sets (a proton chemical shift and a carbon chemical shift) to be sensitive, quaternary carbons, as well as protons bound to non-carbon atoms (i.e. NH, OH) do not produce signals in the HMQC spectrum. In Figure 2.7 it is possible to see the connectivity between the previously identified singlet of methyl protons at δ 2.05 ppm and the peak at δ 24.56 ppm found in the ^{13}C spectra.
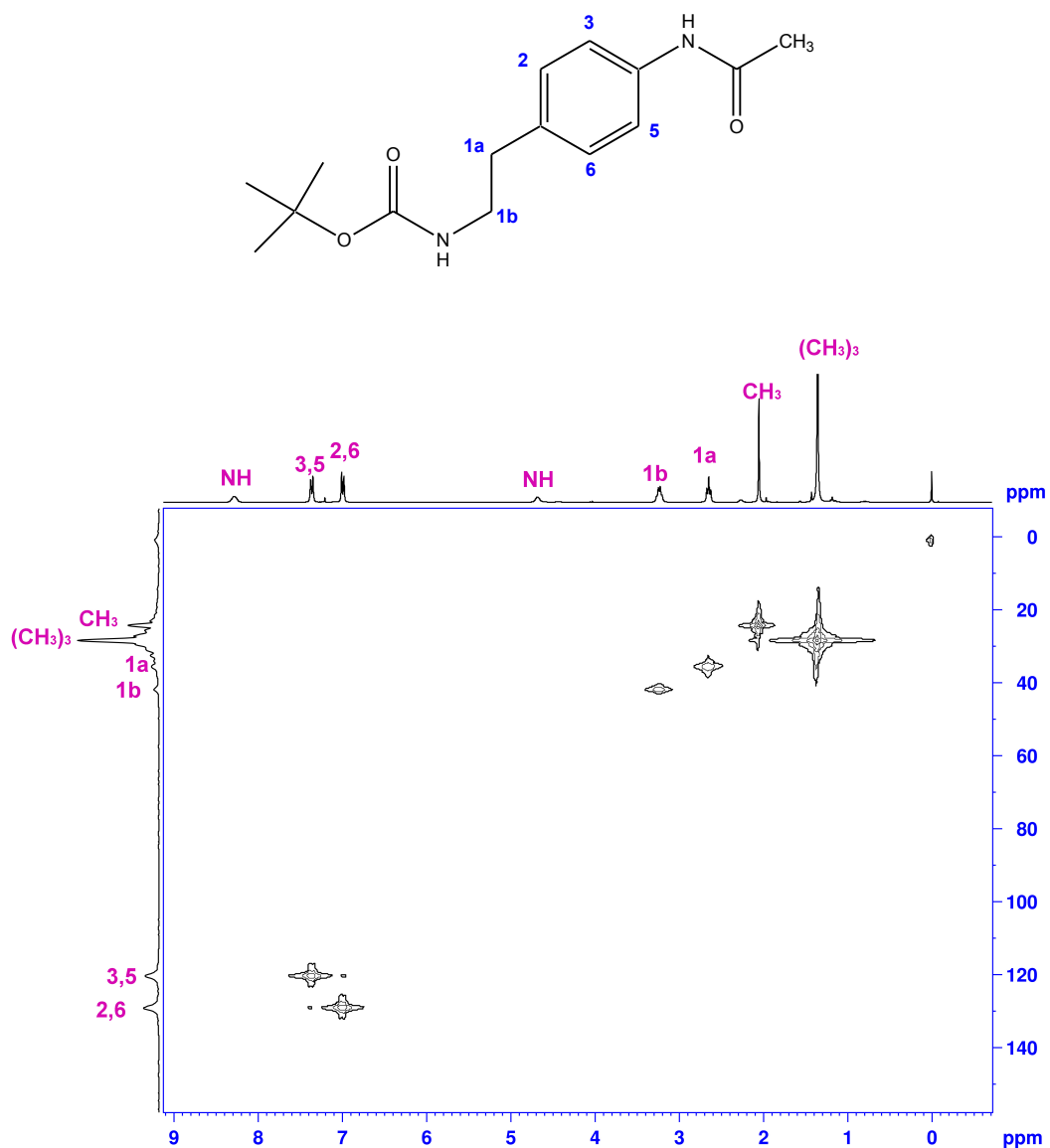
Figure 2.7: HMQC spectrum of *tert*-butyl 4-acetamidophenethylcarbamate (**32**) in CDCl$_3$.

In addition to the NMR experiments, the electrospray mass spectroscopy also identified the expected peaks at 337 [M+ACN+NH$_4$]$^+$, and 277 [M-H$^+$]$^-$. Lastly, the formation of compound **32** was further corroborated by its IR spectrum, which contained the absorption band of a primary amide (3308 cm^{-1}), a band for a C=O bond in an ester grouping (1668 cm^{-1}), a second band for a C=O bond (1603 cm^{-1}) corresponding to the acetyl group, and an intense absorption band of an ester group (1250 cm^{-1}).

2.1.3 Thionation using Lawesson's reagent

The third step of the synthetic route was the transformation of the carbonyl functional group of **32** into its corresponding thiocarbonyl. The two most widely used thionation reagents are phosphorus pentasulfide (P_4S_{10}), **40**, and Lawesson's reagent (2,4-bis(*p*-methoxy-phenyl)-1,3-dithiaphosphetane 2,4-disulfide), **41**. It has been suggested[104-106] that Lawesson's reagent has advantages over P_4S_{10} in terms of requirements for excess P_4S_{10} and longer reaction times. Moreover, Lawesson's reagent has proven[107] to be a superior agent for the smooth, selective transformation of aromatic, aliphatic and unsaturated amides to their corresponding thio-analogues. Furthermore, the selectivity of Lawesson's reagent has enabled the extensive, successful use of the reagent for the selective thionation of amide moieties in molecules containing N-*t*-Boc-protected amino groups[105, 107]. For these reasons Lawesson's reagent was selected as the thionating agent.
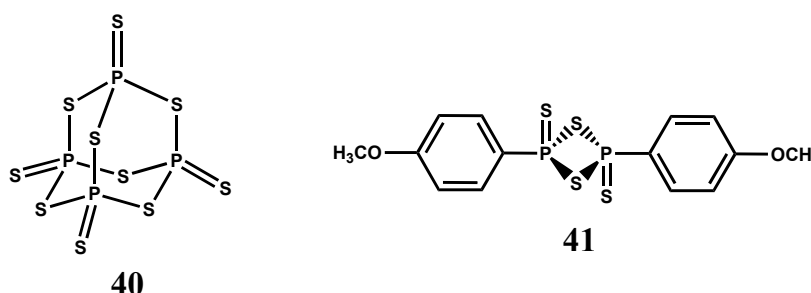
Scheme 12: Thionation reagents phosphorus pentasulfide, **40**, and Lawesson's reagent, **41**.

Lawesson's Reagent was first introduced into chemical synthesis by Lawesson and co-workers[108-110] in 1978 and since then it has been widely used to synthesise thioketones, thioamides, thioesters and dithiocarboxylic acids, thiopeptides, sulfur-containing heterocycles, thioxocyclanes, various phosphacyclanes and macrocyclic thioesters, which are often inaccessible by other methods. The attractiveness of Lawesson's reagent has been attributed to its ready availability, convenient handling, easy work up, high yields of sulfur-containing reaction products, and comparative ease of isolating them from reaction mixtures[105, 111].

The conversion of oxygen functionalities of aromatic, aliphatic, and unsaturated carboxamides by Lawesson's reagent into their corresponding thio analogues is a well-established process[107, 111]. The method is usually performed in refluxing benzene, toluene or xylene (Scheme 13).
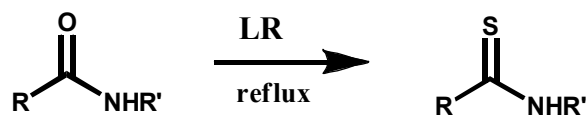
Scheme 13: General reaction of Lawesson's reagent (LR) with amides.

The mechanism for thionation using Lawesson's reagent[106, 107, 112-114] can be rationalised in terms of a nucleophilic attack of carbonyl oxygen upon the dissociation products of Lawesson's reagent. Initially, Lawesson's reagent (**41**) in solution is in equilibrium with a more reactive dithiophosphine ylide (**42, 43**):
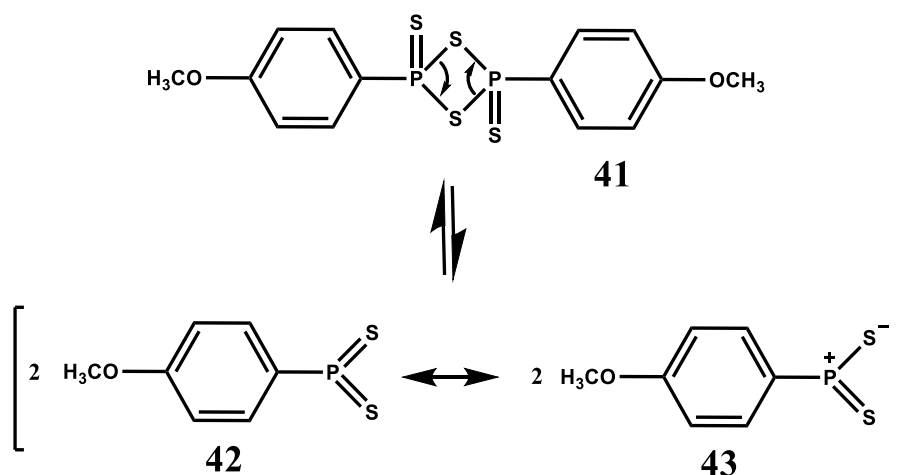
Scheme 14: Dissociation mechanism of Lawesson's reagent.
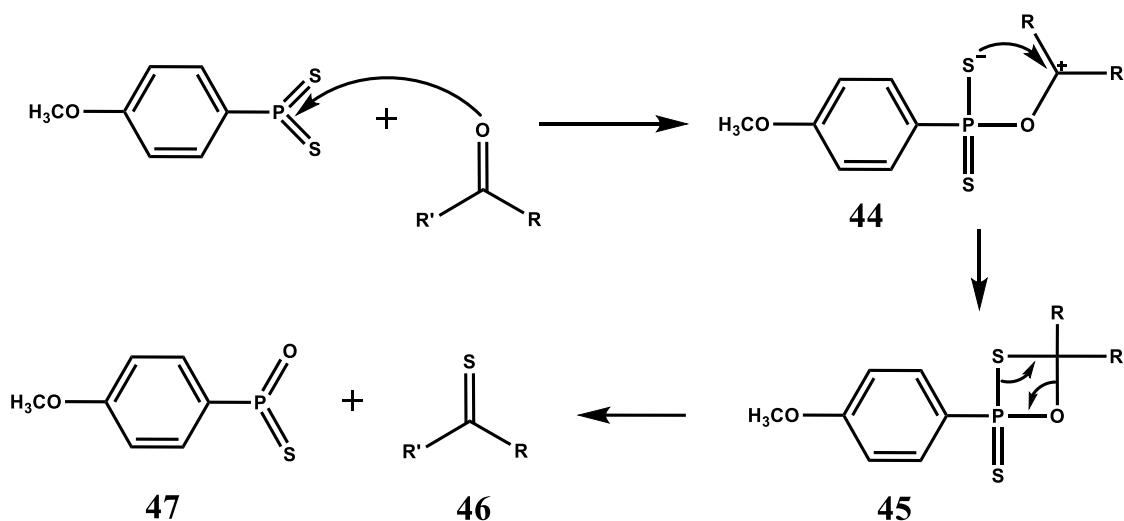
Scheme 15: Mechanism of the thionation reaction using Lawesson's reagent.

The reaction of the ylide with a carbonyl group gives rise to a thiaoxaphosphetane intermediate (a four-membered ring, **45,** Scheme 15), which decomposes to the corresponding thioketone (**46**) through a Wittig-analogous reaction.

This mechanism is similar to the proposed O-S exchange reactions of phosphorus dithioacids[105]. Thus the possibility of these thionation reactions taking place is based on the relative ease with which the phosphorus atom coordination number can increase in dithiophosphates from P(IV) to P(V) and in thionophosphines from P(III) to P(V).

The driving force of the thionation depicted in Scheme 15 is the formation of stronger P=O and P-O bonds instead of labile P=S and P-S bonds[107, 115]. This results in a thermodynamically more stable product **47**, favouring at the same time the formation of the thionated product **46** by exchanging sulfur for oxygen.

Lawesson and co-workers[110] as well as Ozturk and co-workers[106] were able to isolate the trimer *p*-methoxyphenylmetathiophosphonate **48,** a side-product of **47** which can be used as evidence to support the previously described mechanism. It should be noted that side-product **48** also contributes to the reaction energy. The reason for this is that six-membered cyclic phosphoryl (P=O) and thiophosphoryl (P=S) products are more stable than intermediate 4-membered heterocycles (**45**).
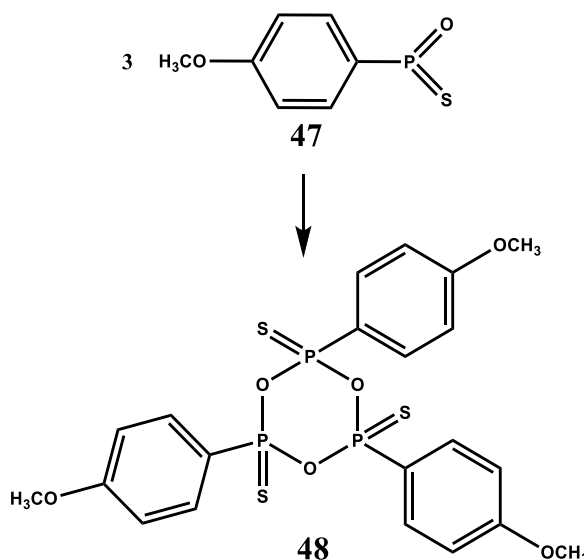
Scheme 16: Formation of side product **48**.

Thiobenzanilide **33** was prepared by sulfurization of **32** using Lawesson's reagent in tetrahydrofuran, following the method proposed by Csomos *et al.*[116] (Scheme 17).
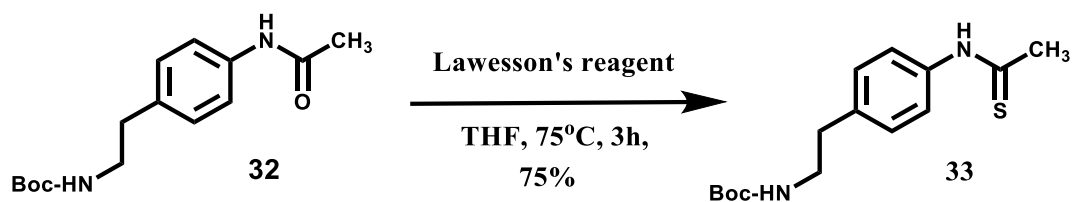
Scheme 17: Thionation of **32** using Lawesson's reagent.

Successful purification by flash column chromatography (3:1 petroleum ether 40-60: ethyl acetate) afforded product **33** (one spot by t.l.c.: R_f = 0.17, methanol: DCM, 3:97) in 75% yield. However, two sets of distinguishable signals were observed by ^{1}H NMR spectroscopy (Figure 2.8).
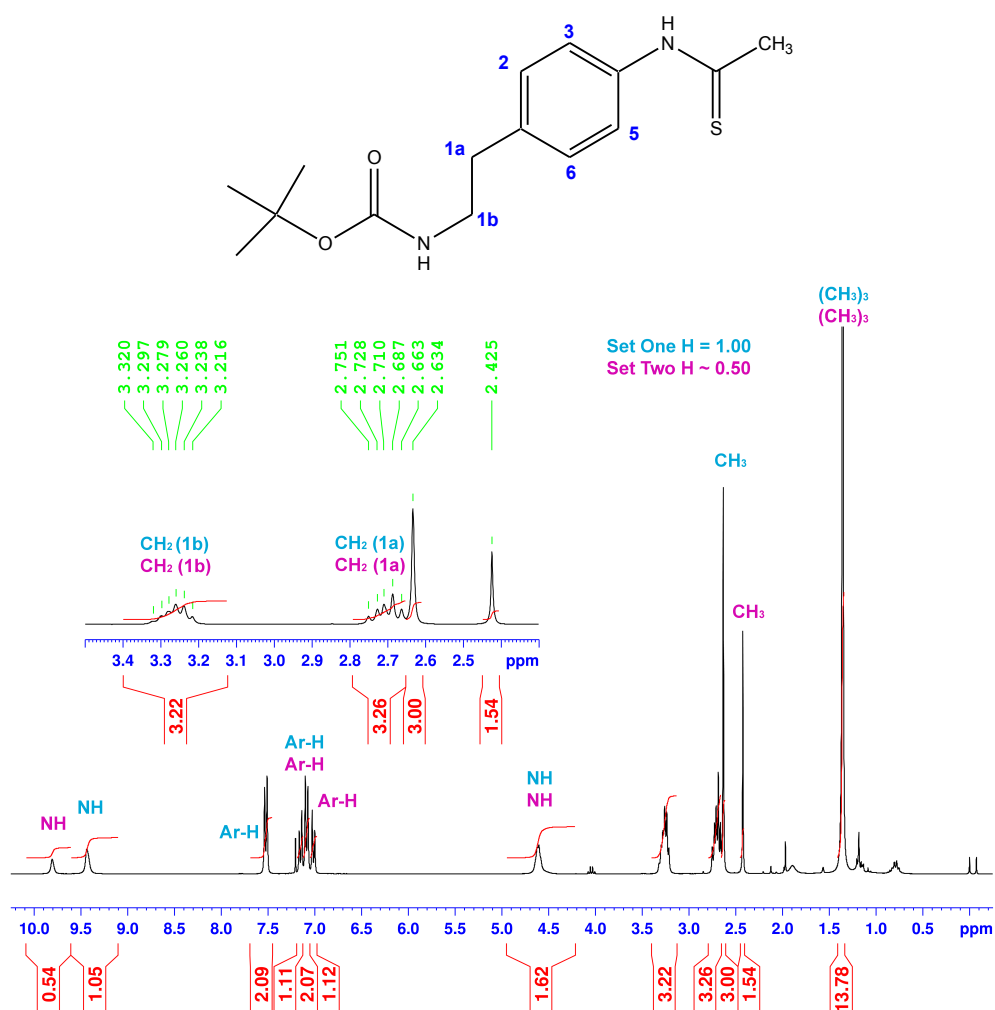
Figure 2.8: ^{1}H NMR spectrum of *tert*-butyl 4-thioacetamidophenethylcarbamate (**33**) in CDCl$_3$.

In the first set of signals, one proton integrates for one unit, whereas in the second set one proton corresponds approximately to 0.50. Besides the different integration values, both sets of signals are identical, that is to say, each contains all the expected

207

peaks for **33** (each are labelled in Figure 2.8). This strongly suggested the presence of two rotamers (discussed later in this section). Note that the presence of amine protons was confirmed by hydrogen-deuterium exchange using D_2O.

The assignment of ^{1}H NMR spectrum signals was aided by the ^{1}H-^{1}H COSY (Figure 2.9), [13]C (Figure 2.10), DEPT-135 (Figure 2.11), and HMQC (Figure 2.12) spectra. Discussion of each of the spectra follows.
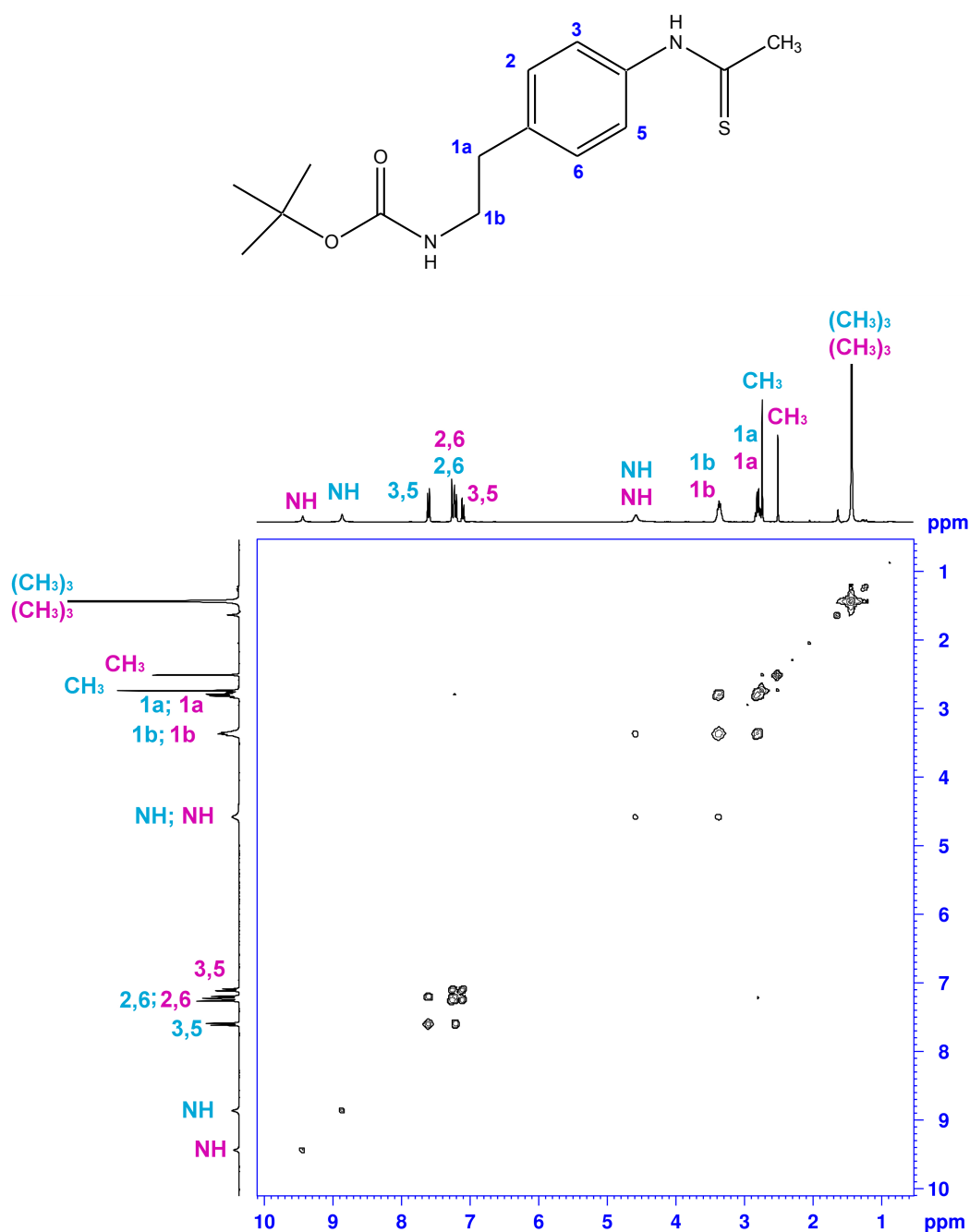
Figure 2.9: ^{1}H- ^{1}H COSY NMR spectrum of *tert*-butyl 4-thioacetamidophenethylcarbamate **(33)**

As previously mentioned, cross peaks exist in the ^{1}H-^{1}H COSY spectrum where there is spin-spin coupling between protons. In Figure 2.9 there are cross peaks between **1a** and **1b** hydrogens and between **1b** and NH. On the other hand there are no cross peaks between **1a** and NH. There are also well defined cross peaks between the doublet of aromatic protons corresponding to **3** and **5** from set one and the multiplet of protons corresponding to **2** and **6**. Additionally, there are cross peaks between the latter and the doublet of aromatic protons corresponding to **3** and **5** from set two. It is interesting to notice that, as expected, there are no cross peaks between the protons corresponding to **3** and **5** from set one and the protons corresponding to **3** and **5** from set two.
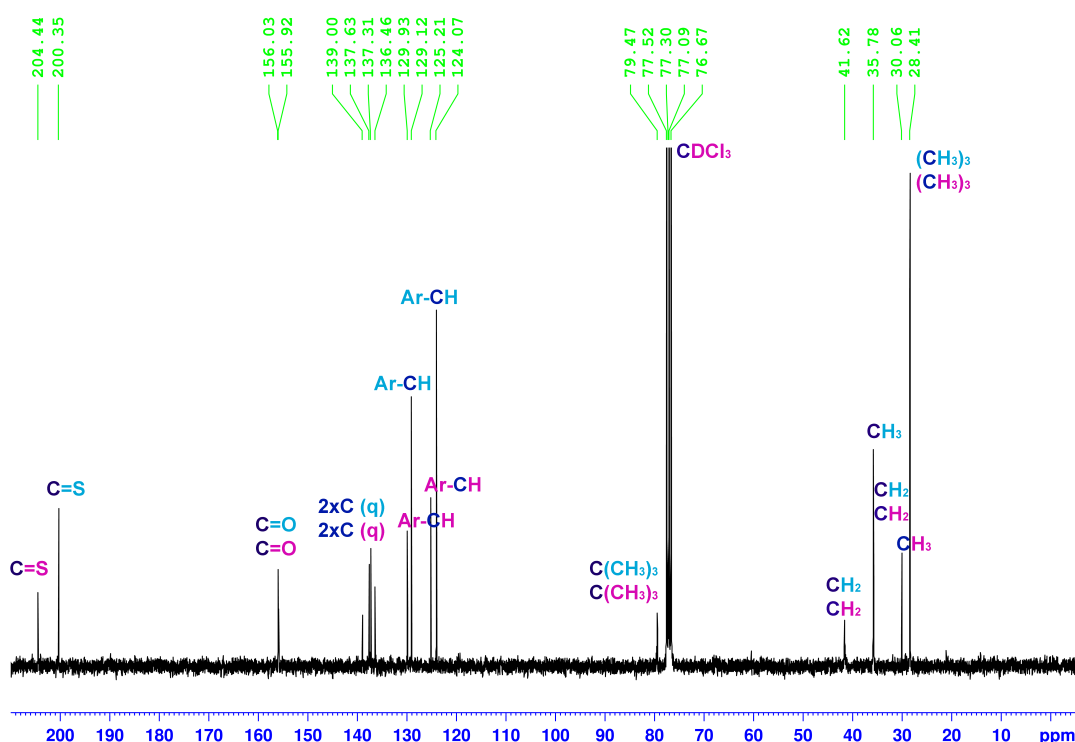
Figure 2.10: ^{13}C NMR spectrum of *tert*-butyl 4-thioacetamidophenethylcarbamate (**33**) in CDCl$_3$.

The ^{13}C NMR spectrum of compound **33** (Figure 2.10) also shows two distinguishable sets of signals corresponding to the desired product (each signal is labelled). The peaks at δ 200.3 and 204.4 ppm merit special attention. Both chemical shifts may correspond to a C=S, strongly supporting the presence of two rotamers. The DEPT-135 spectrum of the product, shown in Figure 2.11, confirms the presence of four –CH and four quaternary carbons in the aromatic region. Additionally, it also confirms the presence of two CH$_3$ groups, each corresponding to one of the rotamers.
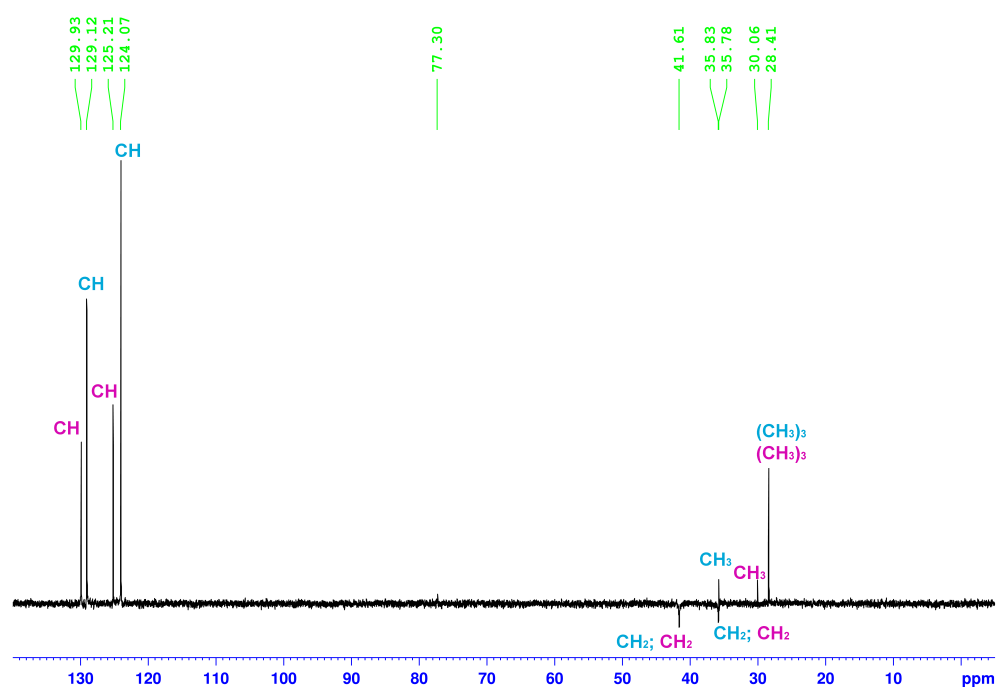
Figure 2.11: DEPT-135 spectrum of *tert*-butyl 4-thioacetamidophenethylcarbamate (**33**) in CDCl$_3$.

Lastly, using the ^{13}C-^{1}H HMQC spectrum (Figure 2.12) it was possible to establish the correlation between the protons found in the ^{1}H NMR spectrum and the carbon nuclei found in the ^{13}C and DEPT-135 spectra.

The IR spectrum of **33** contains the absorption bands of a primary amide (3246 cm^{-1}), a C=O bond in an ester grouping (1683 cm^{-1}), an intense absorption band of an ester group (1249 cm^{-1}) and most importantly, the absorption band of a thioamide C=S bond (1161 cm^{-1}). Finally, the electrospray mass spectroscopy identified the corresponding species (ES)$^+$ 353 [M+ACN+NH$_4$]$^+$ and (ES)$^-$ 293 [M-H$^+$]$^-$.
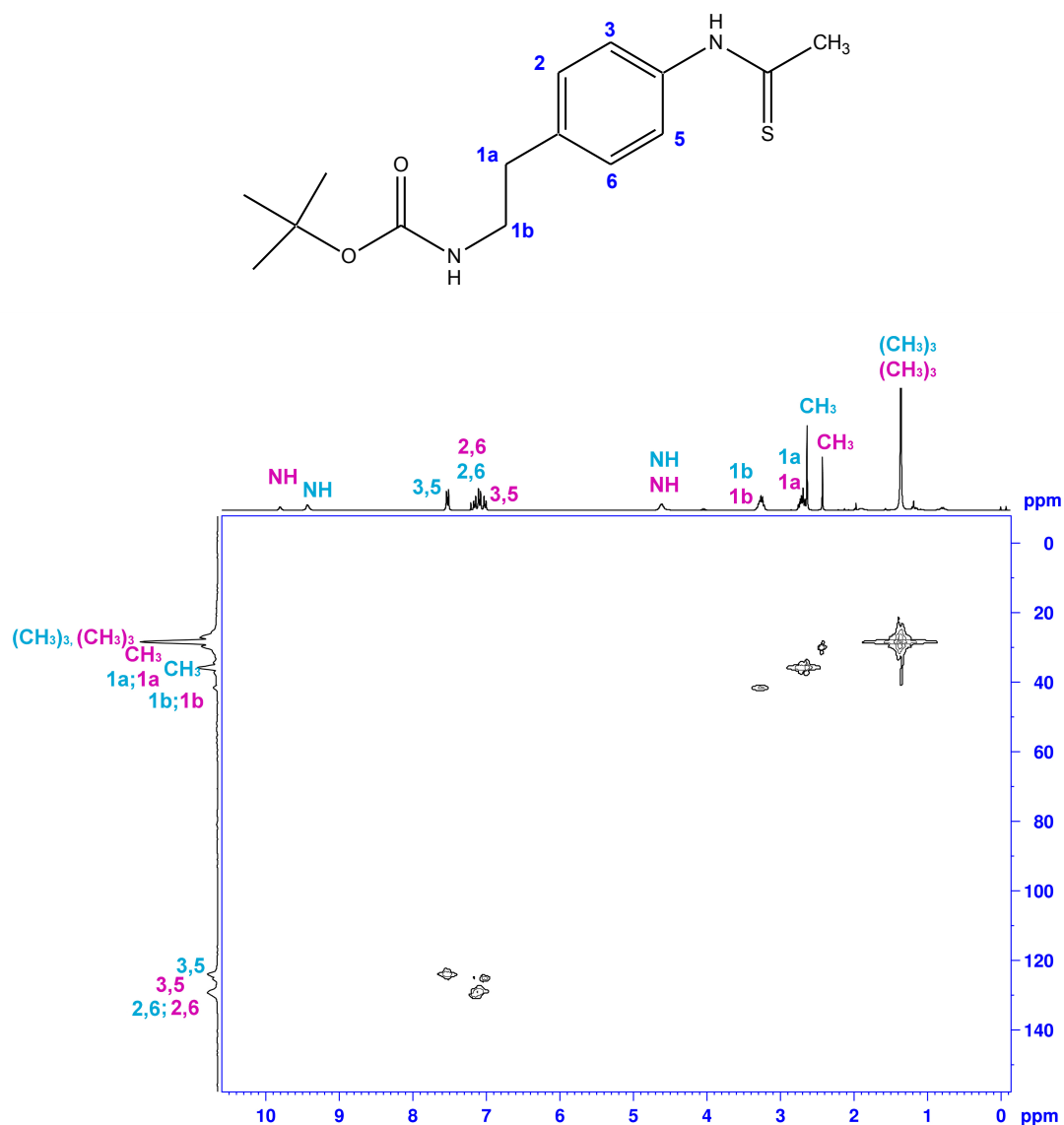
Figure 2.12: HMQC spectrum of *tert*-butyl 4-thioacetamidophenethylcarbamate (**33**) in CDCl$_3$.

Hence, it was possible to conclude, firstly, the success of the thionation reaction, and secondly, the presence of two rotamers (rotation about the N-C bond) of **33** (Scheme 18). Conformations of the rotamers are defined to be *anti* for rotamer A and *syn* for rotamer B.
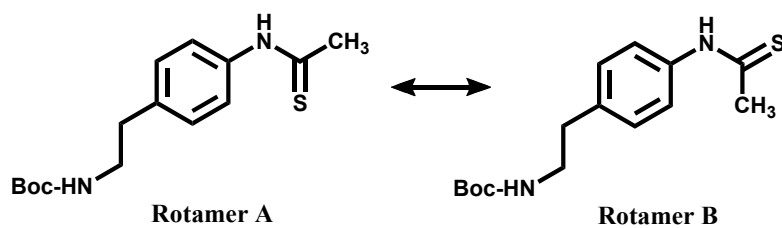
Scheme 18: Rotamers of compound **33**.

The naming is done according to the Klyne-Prelog system[117] for describing conformations about a single bond. *Anti* or *syn* indicates that the substituents with higher priority are on opposite sides or the same side, respectively. Priorities of the substituents are assigned by means of the sequence rules of Cahn, Ingold and Prelog[118].

Conformational isomerism of amides and thioamides as a result of hindered rotation about the C-N bond is a well-known phenomenon[119-122]. Under conditions such that rotation about the C-N bond is slow on the NMR time scale, well-separated NMR signals are observed for each rotamer (for an excellent review on this subject refer to the work of Stewart and Siddall[122]).

In solution, the rotational barriers about the C-N bond in amides are on the order of 16-20 kcal·mol^{-1} and are increased on going to polar solvents. By comparison, the barriers to rotation for thioamides have been found to be uniformly larger (5-7 kcal·mol^{-1} higher) than those for the corresponding amides[123-125]. There are wide variations in the relative abundance of the two isomers of unsymmetrically substituted amides and thioamides[122, 126]. For example, thioformanilide exists exclusively in the *syn* form indicating that the S···phenyl interaction is larger than the O···phenyl interaction in formanilide (which exists only as 50% *syn*). Yet, thioacetanilide shows 40% *syn* isomer compared with 5% in acetanilide[126]. It has been documented[122-125] that the variation of the relative abundance of the *syn* and *anti* rotameric forms depends not only on the chemical nature of the substituents, but also on the solvent polarity and temperature.

For the case of compound **33**, based on the integration values provided in Figure 2.8, the predominant isomer was found in a ratio of ~ 2 to 1 (at room temperature and in CDCl$_3$). However, it was not possible to prove which set of signals corresponds to rotamer A or B. Further work could involve NOESY spectra and variable temperature NMR experiments.

Varma and Kumar[127] reported on a rapid and solvent-free synthesis of thioketones, thioflavones, and thioisoflavones by mixing the substrates with Lawesson's reagent and subsequently exposing them to microwave irradiation. They achieved high yields within 3–4 minutes. In an attempt to improve the yield and decrease the reaction time, microwave-assistance was tried following the Varma and Kumar solvent-free method[127]. Unfortunately, ^{1}H NMR spectral analysis of the crude product indicated that the reaction had been unsuccessful.
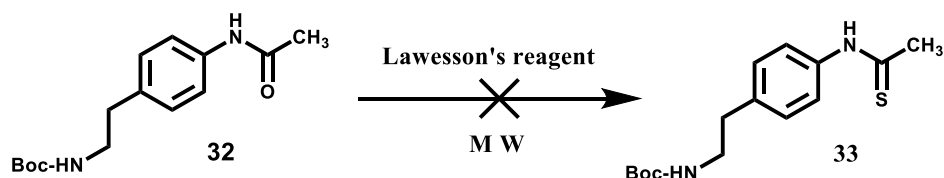
Scheme 19: Unsuccessful thionation of **32** using microwave irradiation.

2.1.4 Cyclisation of thioamide 33

The next synthetic stage involved the preparation of a substituted benzothiazole by cyclisation of **33**. Various methods for the synthesis of benzothiazoles are known[128-133]. Among these, the Jacobson synthesis has been reported[134-137] as an effective method for the formation of benzothiazoles by cyclisation of benzamides and therefore it was selected.

The Jacobson synthesis[138] of benzothiazoles involves oxidative cyclisation of thioanilides or *N*-arylthioanilides with an unsubstituted *ortho*-position, in the presence of potassium ferricyanide in basic medium (Scheme 20). It has been reported[139] that the best yields are obtained for arylamines bearing electron-donating groups which enhance the electron density *ortho* to the nitrogen atom.
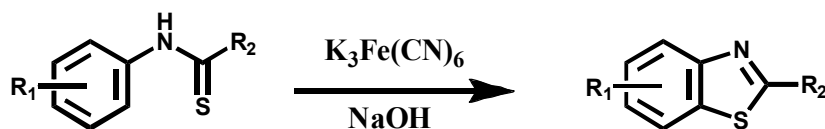
Scheme 20: The Jacobson synthesis.

Two likely mechanisms for the Jacobson synthesis can be found in the literature. That proposed by Metzger and Planck[140] (also reported by Jackson *et al.*[134]) is illustrated in Scheme 21. This mechanistic pathway involves an attack on the anionic complex of the hexacyanide ligand of the ferric iron (Fe^{3+}) ion, followed by a rearrangement to form a thioimidic cation (**49**). Then the benzene ring attacks **49**, with loss of a proton (aromatisation), yielding the benzothiazole (**50**).
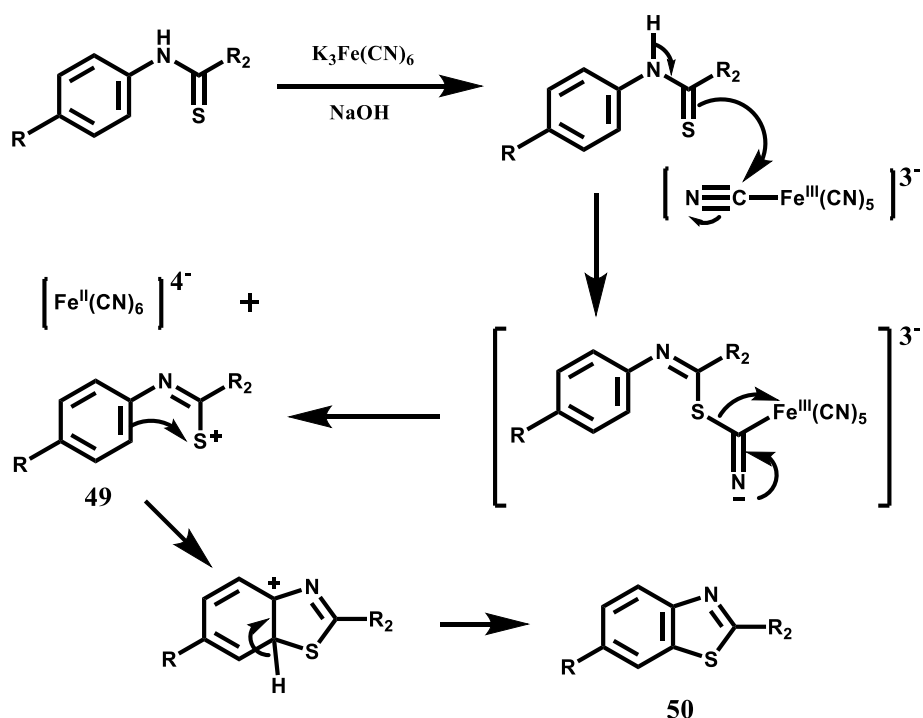
Scheme 21: Proposed mechanism for Jacobson synthesis: path A.

The second mechanism (Path B, Scheme 22), proposed by Stevens *et al.*[141], involves reaction of the thiobenzamide with a base to form a thiolate anion (**51**), which then undergoes one-electron oxidation to form the thiol radical **52**. This radical then attacks the unoccupied *ortho* position in the substrate. The elimination of a hydrogen radical from the reactive intermediate **53** forms the benzothiazole (**54**).
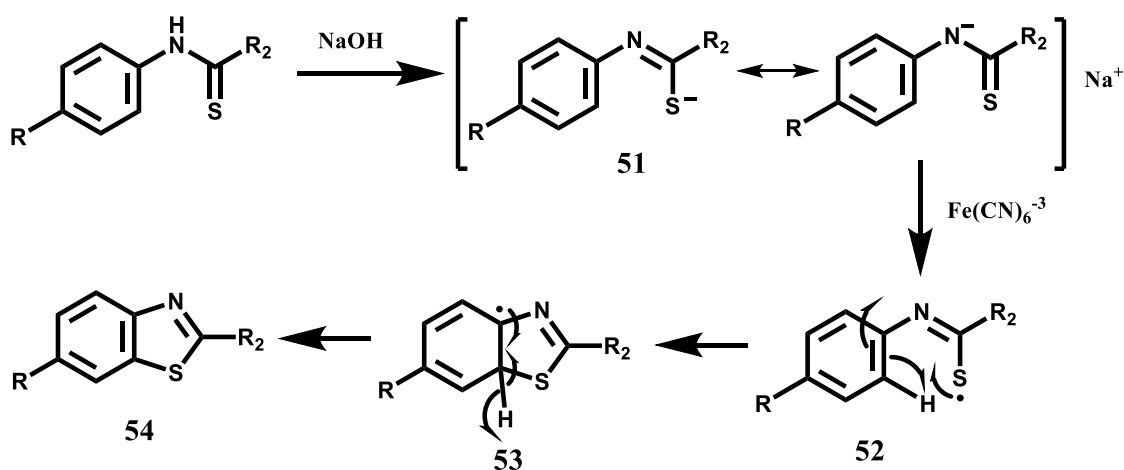
Scheme 22: Proposed mechanism for Jacobson synthesis: path B.

The cyclisation of thioamide **33** was carried out using a modification of the methodology proposed by Mortimer[137]. Compound **33** was dissolved in 1.5 M aq.

NaOH, treated with potassium ferricyanide [K$_3$Fe(CN)$_6$] and irradiated for 14 minutes in a Biotage microwave oven (120 °C) to produce benzothiazole **34** in 54% yield.
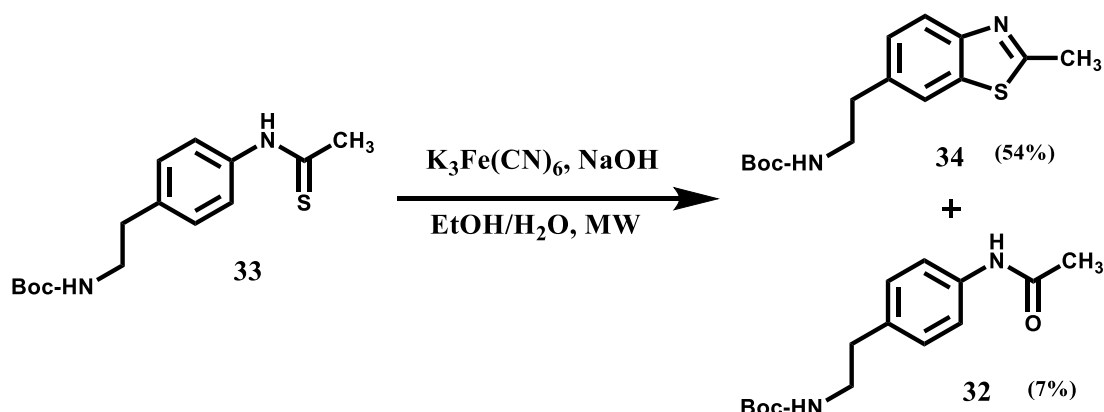
Scheme 23: Cyclisation of thioamide **33**.

Purification by flash column chromatography allowed isolation of **34** and **32.** Characterisation of **32** was possible through comparison of its spectroscopic data with the data of previously prepared *tert*-butyl 4-acetamidophenethylcarbamate (Section 2.1.2).

Formation of side-product **32** involves desulfurization by attack of hydroxide ion on the thiocarbonyl carbon atom, with subsequent loss of H$_2$S. This side reaction is likely to be a consequence of the nature of the ring substituents. It has been reported[134] that *p*-electron withdrawing groups clearly inhibit the thio-enolization (first step in Path A, Scheme 21), and thus ring closure is affected. In this case the presence of the *p*-alkyl group, a weakly activating group, may partially inhibit the thioenolization, generating **34** only in moderate yield and at the same time allowing desulfurization.

Synthesis of product **34** was confirmed by spectroscopic analysis. A series of well-defined signals was observed in the ^{1}H NMR spectrum of the product (Figure 2.13): a singlet of methyl protons (δ 2.17 ppm); a second singlet of methyl protons of the *tert*-butyl group (δ 1.43 ppm); a triplet and multiplet of methylene protons (δ 2.84 ppm, 3.41-3.30 ppm); a broad singlet corresponding to the N-H (δ 4.49 ppm); and three peaks (doublet δ 7.86 ppm, singlet δ 7.63 ppm, broad doublet with a spike for CHCl$_3$ δ 7.25-7.28 ppm) corresponding to three aromatic protons. These latter merit special attention. The singlet at 7.63 ppm corresponds to an isolated aromatic proton, which indicates that the symmetry of the aromatic ring has been lost, thus suggesting the adjacent 5-membered ring formation. The presence of an NH group was confirmed by deuterium exchange using D$_2$O.
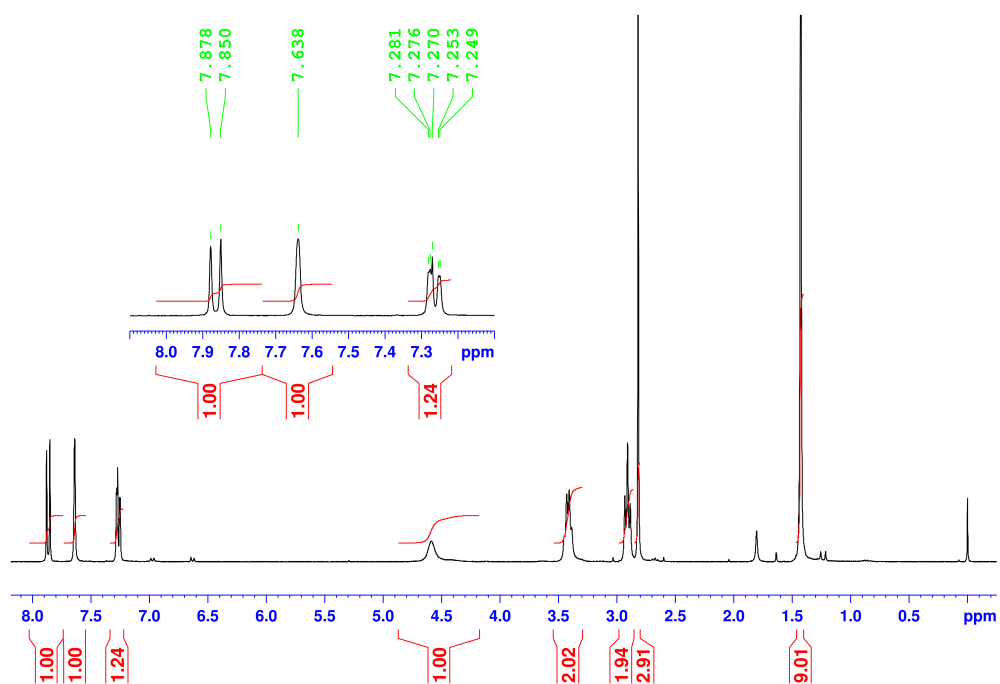
Figure 2.13: ^{1}H NMR spectrum of *tert*-butyl 2-(2-methylbenzo[*d*]thiazol-6-yl)ethylcarbamate (**34**) in CDCl$_3$.
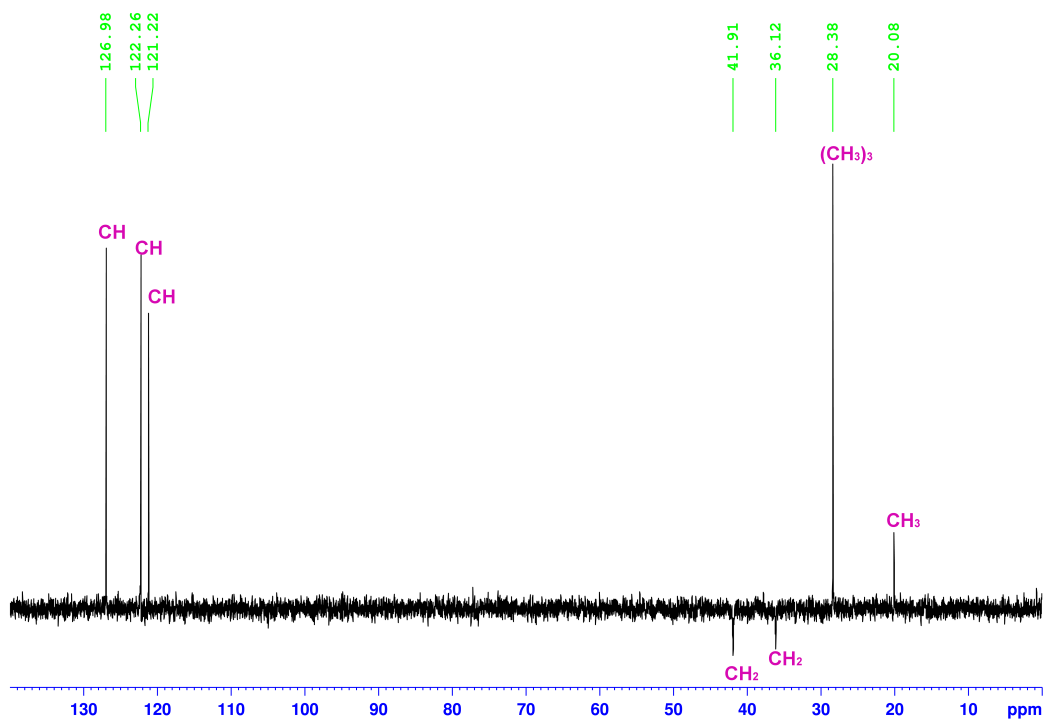
Figure 2.14: DEPT-135 spectrum of *tert*-butyl 2-(2-methylbenzo[*d*]thiazol-6-yl)ethylcarbamate (**34**) in CDCl$_3$.

The assignments for the ^{1}H NMR in Figure 2.13 were aided by use of ^{13}C and DEPT-135 spectra (Figure 2.14), ^{1}H-^{1}H COSY NMR (Figure 2.15), and [13]C-[1]H (HMQC) correlations (Figure 2.16).

The ^{1}H-^{1}H COSY spectrum of **34** is depicted in Figure 2.15. The most interesting region of this spectrum is that related to the connectivity of the aromatic protons, with well-defined cross peaks between the aromatic protons corresponding to either **4** or **5**.
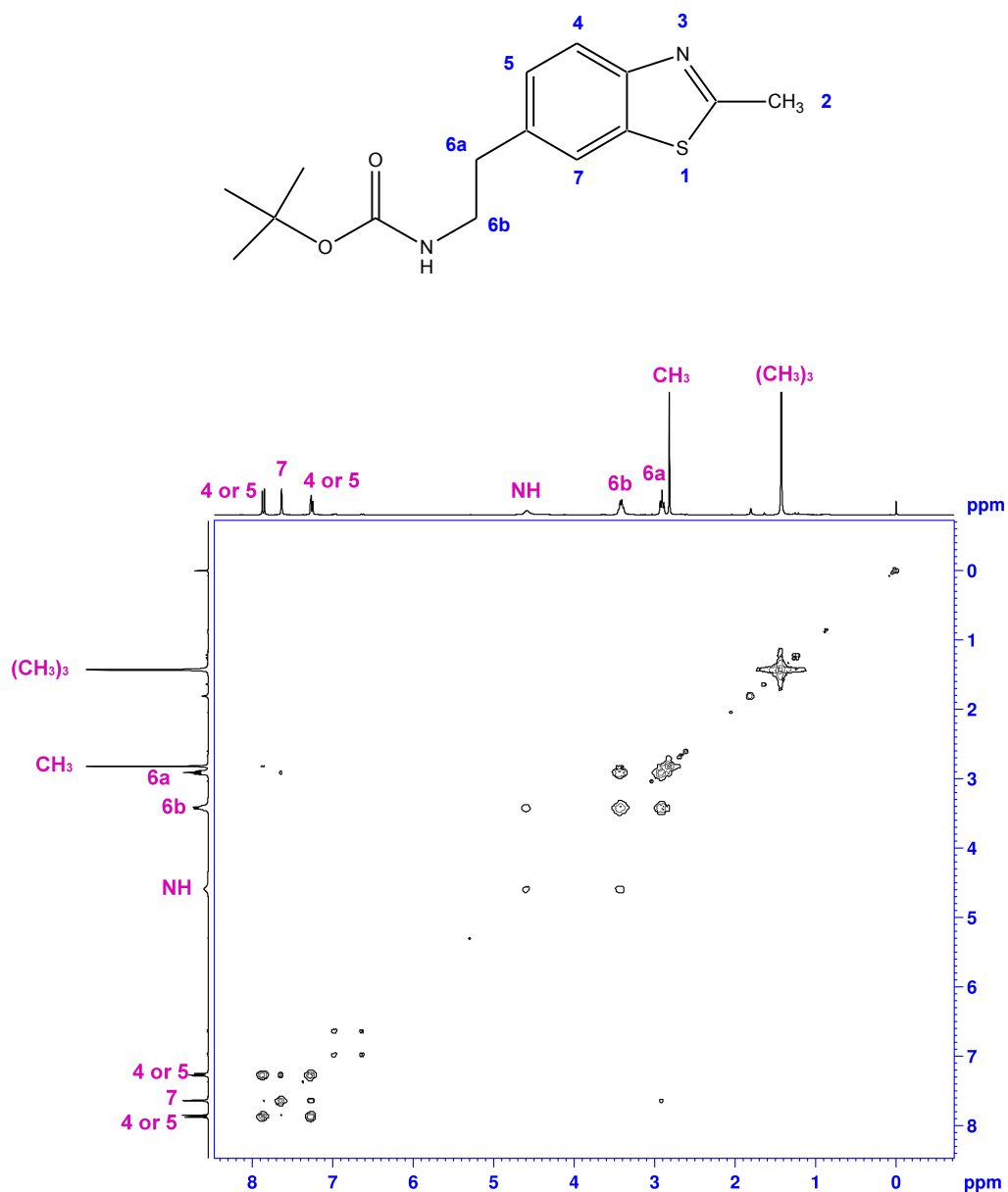
Figure 2.15: ^{1}H- ^{1}H COSY NMR spectrum of *tert*-butyl 2-(2-methylbenzo[*d*]thiazol-6-yl)ethylcarbamate (**34**) in CDCl₃.

The ^{13}C NMR spectrum of compound **34** shows six different peaks in the aromatic region (δ 152.13, 136.02, 135.86, 126.77, 122.17, 121.23 ppm), which confirms the presence of a non-symmetrical 6-membered ring. Furthermore, the DEPT-135

spectrum of **34** (Figure 2.14) corroborates the presence of three -CH and three quaternary carbons in the aromatic region.

By using ^{13}C-^1H (HMQC) correlations (Figure 2.16) it was possible establish the correlation between the protons found in the ^1H NMR and the carbon nuclei found in the aromatic region of the ^{13}C spectrum.
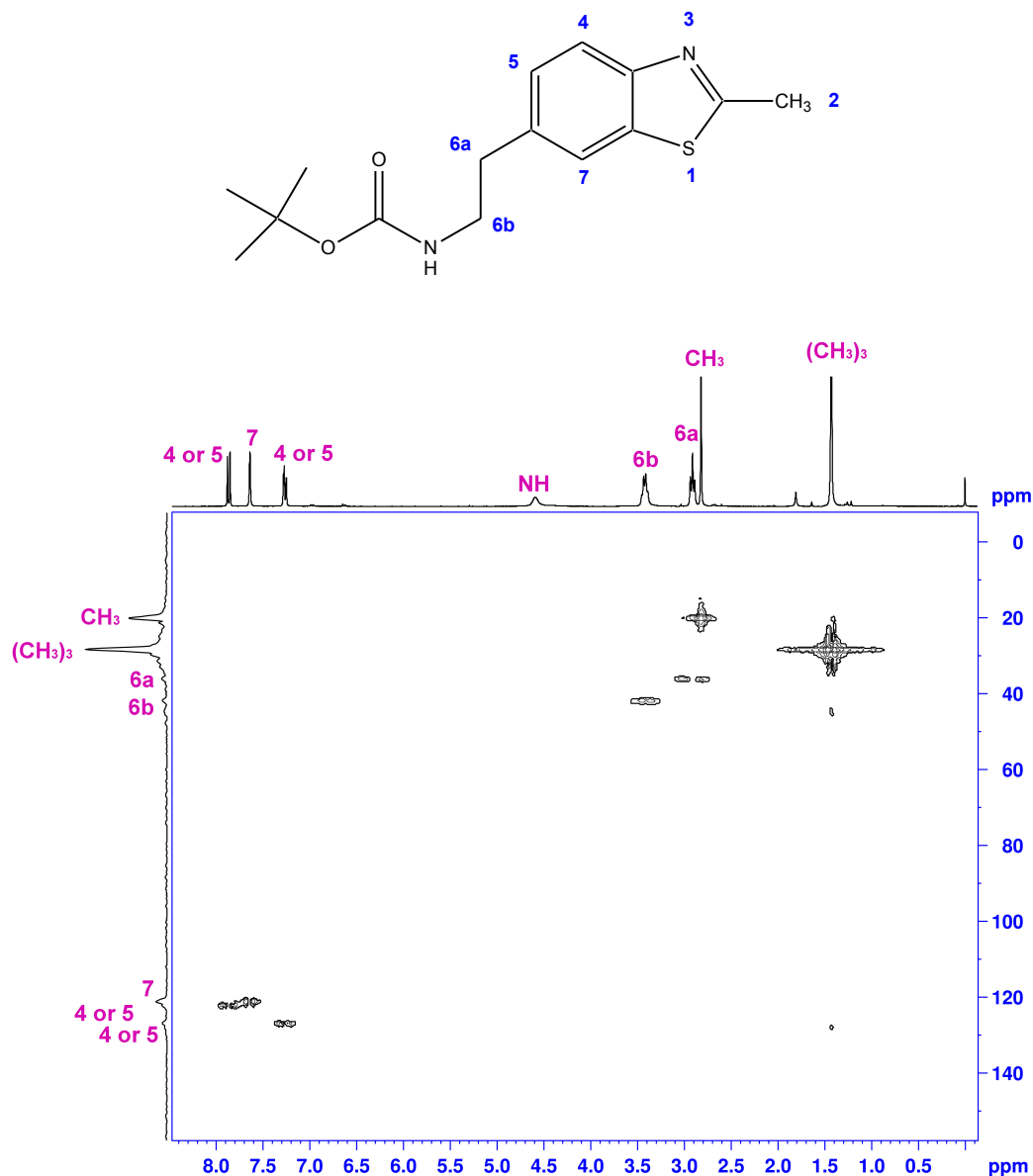
Figure 2.16: HMQC spectrum of *tert*-butyl 2-(2-methylbenzo[*d*]thiazol-6-yl)ethylcarbamate (**34**) in CDCl$_3$.

A close-up view of the aromatic region of the HMQC spectrum of **34** is shown in Figure 2.17. In this expansion, the singlet corresponding to **7** (δ 7.63 ppm) appears first in the ^{13}C spectrum (δ 121.22 ppm), followed by the doublet corresponding to either **4** or **5** (δ 7.86 correlated to δ 122.26 ppm), and lastly the multiplet

corresponding to either **4** or **5** (δ 7.25-7.28 correlated to δ 126.9 ppm). Establishing the connectivity in the aromatic region provided the final evidence on synthesis of product **34**.
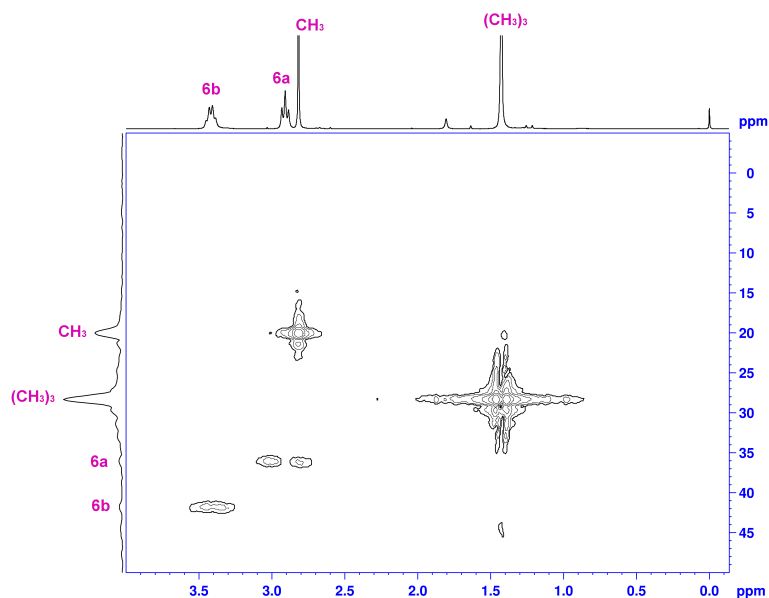
Figure 2.17: Expanded view of aliphatic region of the HMQC spectrum of *tert*-butyl 2-(2-methylbenzo[*d*]thiazol-6-yl)ethylcarbamate (**34**) in CDCl₃
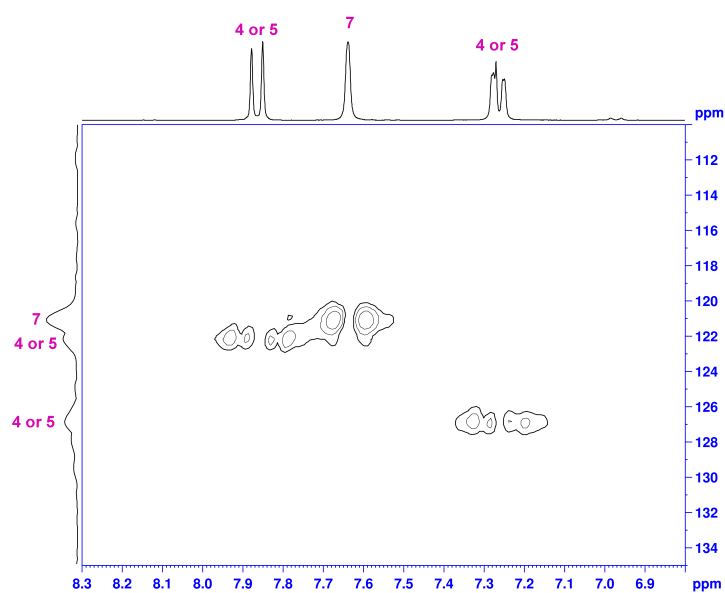
Figure 2.18: Expanded view of the aromatic region of the HMQC spectrum of *tert*-butyl 2-(2-methylbenzo[*d*]thiazol-6-yl)ethylcarbamate (**34**) in CDCl₃.

2.1.5 Esterification of benzothiazole 34

The last step in the synthetic route was the esterification of benzothiazole **34**. This step involved anion formation using lithium diisopropylamine and subsequent treatment with ethyl chloroformate.

Lithium diisopropylamide (LDA, **57**) in tetrahydrofuran is the most widely utilised amine anion base for deprotonation of weakly acidic organic compounds[142-144]. Its combination of low nucleophilicity and high kinetic basicity makes this hindered amide invaluable in proton abstraction applications.

LDA (**57**) can be easily prepared by reaction between *n*-butyllithium (**55**) and diisopropylamine (**56**) (Scheme 24). LDA is a very strong base because the conjugate acid of diisopropylamine has a pK_a ~ 35.

$$^nC_4H_9Li \;+\; ^iPr_2NH \xrightarrow[\text{0 °C}]{\text{THF}} \; ^iPr_2N^{\ominus} \; Li^{\oplus} \;+\; ^nC_4H_{10}$$

<div align="center">

55   **56**   **57**

</div>

<div align="center">
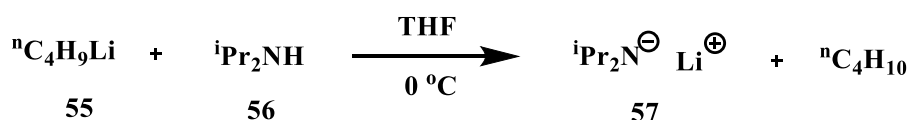
Scheme 24: Preparation of LDA.

</div>

Benzothiazole **34** was treated with a solution of LDA at 0 °C, and then with ethyl chloroformate at -78 °C yielding esterified product **16** (Scheme 25).

<div align="center">
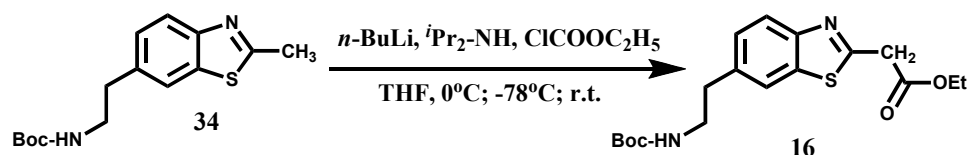
Scheme 25: Esterification of benzothiazole **34**.

</div>

In the first step of the mechanism, LDA deprotonates the α-carbon of **34.** The resulting carbanion (**58**) then attacks the carbonyl group of ethyl chloroformate affording **16** (Scheme 26).
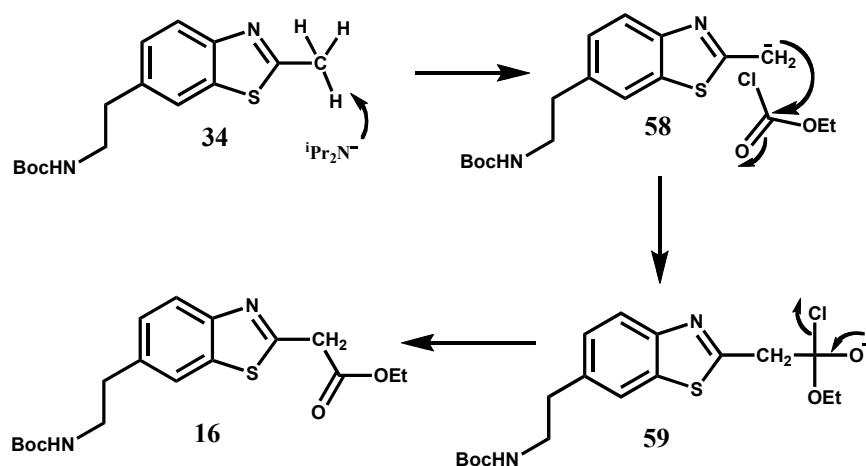
Scheme 26: Reaction mechanism for the esterification of benzothiazole **34**.

Separation of product **16** from its precursor **34** was problematic. Attempted flash column chromatography over silica gel with PET/EtOAc (3:1), DCM/Methanol 1%, PET/EtOAc (4:1) or with hexane/Et$_2$O (1:1) as eluents were not successful. However, purification by preparative t.l.c. with hexane/Et$_2$O (1:1) as eluent proved effective.

After successful purification by preparative t.l.c., identification of product **16** was possible by electrospray mass spectroscopy by identifying (ES)$^+$ 387 ([M+Na]$^+$ and (ES)$^-$ 363 ([M-H$^+$]$^-$. Due to the small amount of product obtained, very detailed spectroscopic analyses were not an option.

Figure 2.19 shows the ^{1}H NMR spectrum of compound **16** after unsuccessful purification by flash column chromatography (1:1 hexane:Et$_2$O, SiO$_2$). The spectrum shows two distinguishable sets of peaks, one of them corresponding to the starting material **34** and the second believed to be the spectrum of the desired product **16**. The most striking peak in this figure is the singlet at 4.09 ppm corresponding to CH$_2$ (**2a**) of **16**.
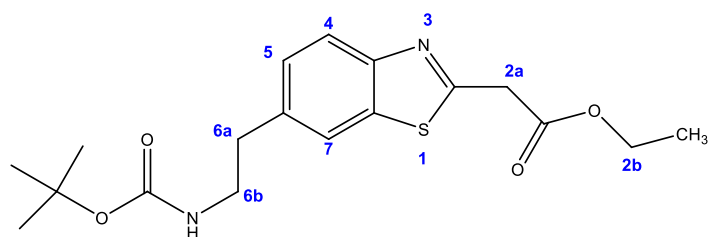
Desired Product (16)
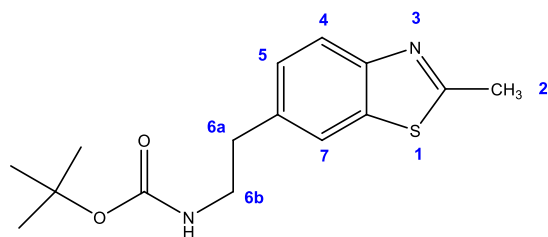
Starting material (34)
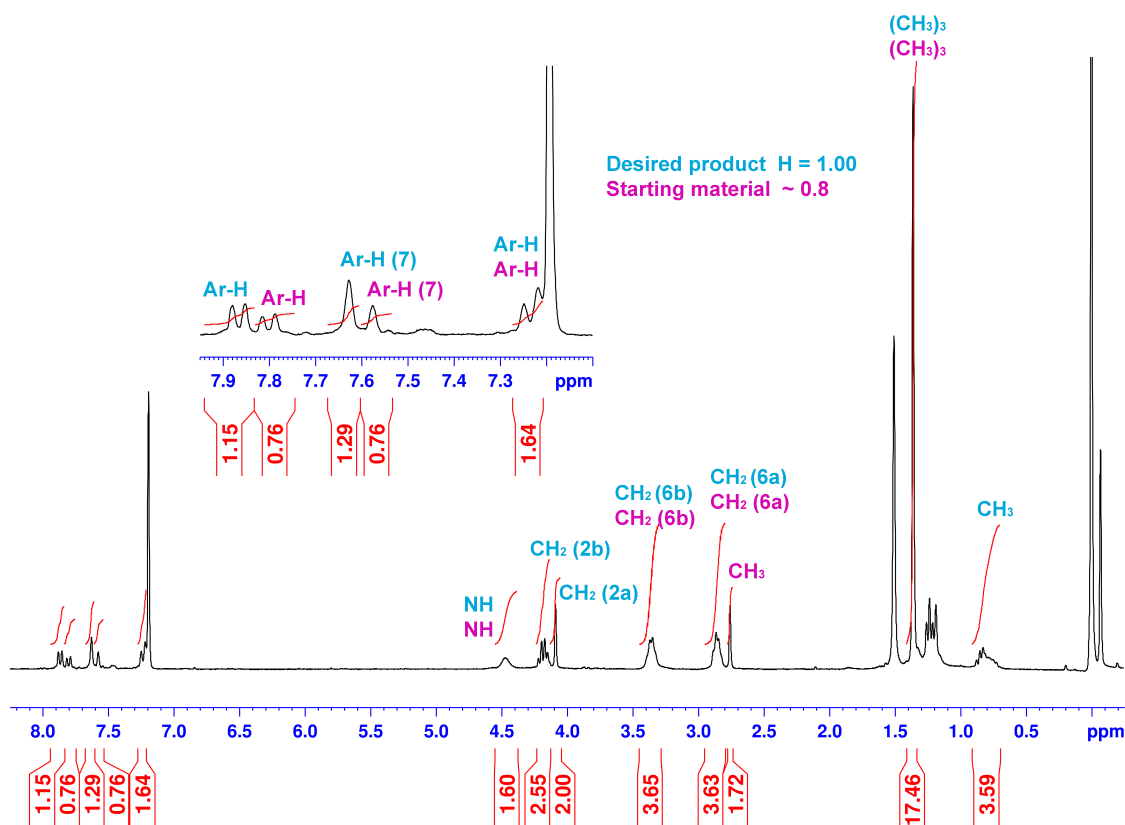
Figure 2.19: ^{1}H NMR spectrum of ethyl 2-(6-(2-(*tert*-butoxycarbonylamino)ethyl)benzo[*d*]thiazol-2-yl) (**16**) in CDCl$_3$.

Based on the hypothesis that the acidic silica may cause a problem with respect to stability of product **16**, an attempt was made to isolate the desired product using

different separating agents, e.g. alumina. However, this was unsuccessful. Further work would involve the assessment of HPLC purification.

The esterification was also tested with benzyl chloroformate, phenyl chloroformate and isobutyl chloroformate. Reactions were successful, but the purification protocol was not improved.

2.1.6 Conclusions of the benzothiazole 16 route

Ultimately, the difficulties in the last step impeded the successful completion of the synthetic route leading to benzothiazole **16**. This in turn hindered the total synthesis of BTC's derivative **11**, which required the condensation of the esterified benzothiazole **16** with the already available salicylaldehyde **15** by a Knoevenagel-type reaction, followed by selective Boc-deprotection and treatment with bromoacetyl bromide with subsequent acid cleavage of the tertiary butyl groups.

2.2 Progress towards the synthesis of BTIC derivatives

The objective in this section is to derivatise BTIC to produce a fluorophore that can be attached to a GFP mutant. It was envisaged that incorporation of a halogen moiety into the benzothiazolyl group would result in a derivatisable probe (**12**). To this end, salicylaldehyde **15** had to be condensed by a Knoevenagel-type reaction with benzothiazole **20** (Scheme 27). Each synthetic step is described in the following section.
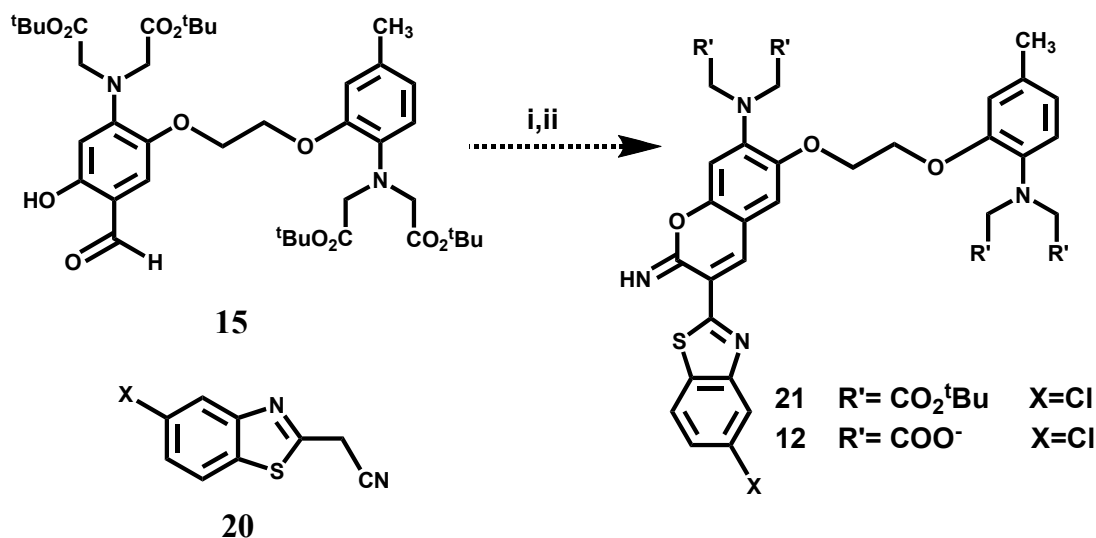
i) Piperidine, EtOH, r.t. 12h; ii) TFA.

Scheme 27: Proposed synthetic route to potential fluorescent probe **12**.

2.2.1 Preparation of benzothialoze 20 using 2-aminothiophenol

As mentioned in section 2.1.4, several methods of synthesis of 1,3-benzothiazoles have been described in the literature. One of them, the cyclisation of benzamides *via* the Jacobson synthesis has been successfully explored. In this section an alternative method involving the condensation of 2-aminothiophenol with a substituted nitrile will be assessed.

2-Aminothiophenol (**60**) contains a fragment with S and N atoms attached to the ortho-positions of the benzene ring. As this fragment is also present in the structure of 1,3-benzothiazole (**61**) and other sulfur-containing heterocyclic systems, several methods of benzothiazole synthesis use 2-aminothiophenol as starting material.
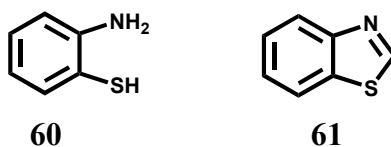
Scheme 28: 2-Aminothiophenol (**60**) and 1,3-benzothiazole (**61**).

To construct the benzothiazole ring a second reagent containing functional groups able to react with the thiol and amino groups of the 2-aminothiophenol is required. The second reagent should also be capable of furnishing the carbon atom that links the sulfur and nitrogen atoms. Benzothiazoles are synthesised by cyclocondensation of 2-

aminothiophenol with carboxylic acid derivatives[145, 146], aldehydes[147, 148], acyl chlorides[149] and nitriles[150, 151].

A convenient one-step cyclocondensation between 2-aminothiophenol (**60**) and malononitrile (**62**) in the presence of glacial acetic acid and ethanol has been reported by Saito and co-workers[152]. This method affords 1,3-benzothiazol-2-ylacetonitrile (**64**) in good yield (Scheme 29).
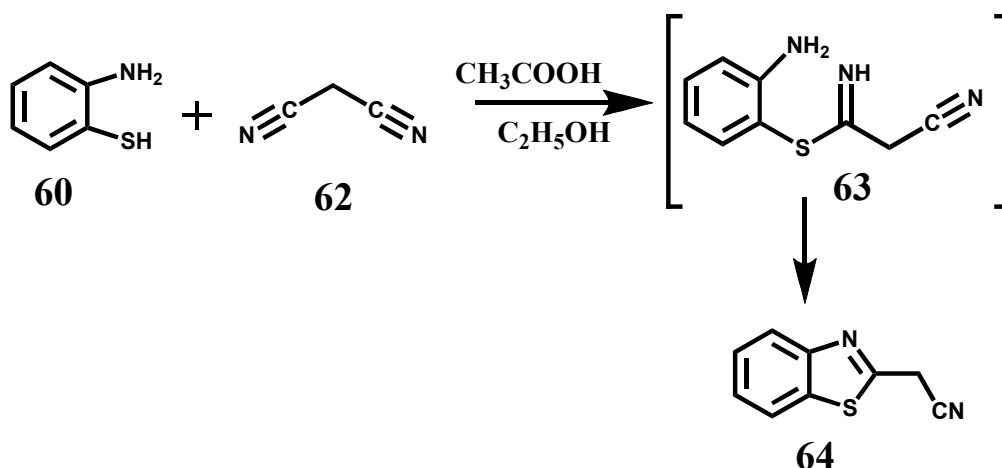
Scheme 29: Synthesis of 1,3-benzothiazol-2-ylacetonitrile.

The formation of **64** has been reported by Elgemeie *et al.*[153] to proceed *via* intermediacy of **63**. Based on this reported intermediate, the putative mechanism for this reaction can be rationalised as follows. Firstly, as illustrated in Scheme 30, the nitrile C≡N bond of **62** is activated by acid. In this way, it becomes more electrophilic. Two resonance structures (65a and 65b) can be drawn for the activated species.
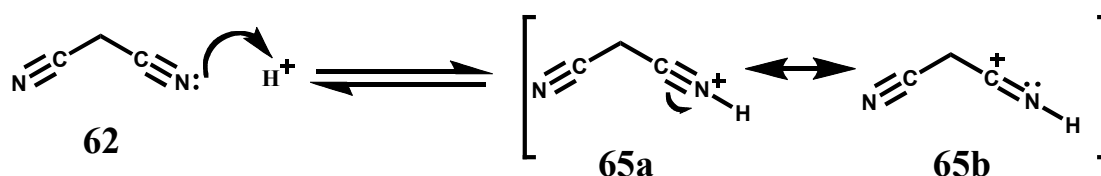
Scheme 30: Resonance structures of **62**.

The next stage involves a nucleophilic addition. It begins with the donation of one lone pair of electrons from the sulfur atom to the positively charged carbon of the activated nitrile. This forms a new σ bond to the nucleophile and gives rise to adduct **66**. Subsequently, deprotonation of the –SH group leads to the regeneration of the acid catalyst and the formation of intermediate **63**.
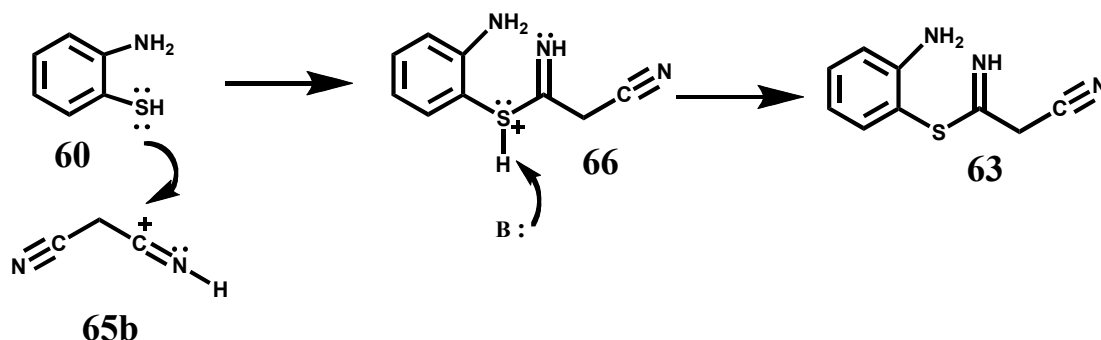
Scheme 31: Formation of intermediary **63**.

Lastly, because a second nucleophile is present (i.e. the intramolecular amine group) a second nucleophilic addition takes place[154]. Addition begins with donation of the amine nitrogen lone pair of electrons to the partially positive carbon atom belonging to the imine. This involves intramolecular ring closure and leads to the expulsion of the leaving group (NH_3).
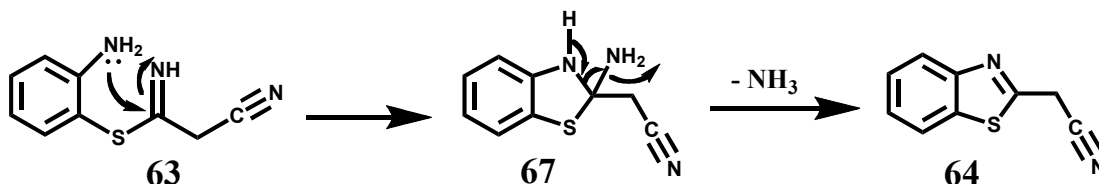
Scheme 32: Formation of product **64**.

According to Baldwin's rules[155], this ring formation reaction is classified as *5-exo-trig*. The numerical prefix describes the ring size, i.e. 5 atoms constitute the skeleton of the cycle. The prefix *exo* refers to the fact that the breaking bond is exocyclic to the formed ring. And finally, the suffix *trig* (trigonal, sp^2) indicates the geometry of the carbon atom undergoing the ring-closure reaction.

The synthesis of benzothiazole **20** was carried out following the procedure described by Saito and co-workers[152]. Commercially available 2-amino-4-chloro-benzenethiol (**68**) was treated with malononitrile (**62**) in the presence of acetic acid and ethanol to afford benzothiazole **20** in 61% yield (Scheme 33). The product did not require further purification.
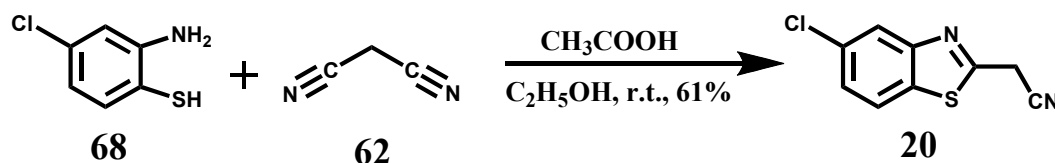
Scheme 33: Synthesis of benzothiazole **20**.

Synthesis of **20** was confirmed by spectroscopic analysis. A series of well-defined signals was observed in the ^{1}H NMR spectrum of the product (Figure 2.20). These are: a singlet of methylene protons (δ 4.63 ppm), and three peaks (doublet δ 7.56 ppm, singlet δ 8.05 ppm, doublet δ 8.13 ppm) corresponding to three aromatic protons. The two extra peaks in the spectrum correspond to residual solvent (acetone-d$_6$, quintet, δ 2.05 ppm) and residual water (δ 2.84 ppm).
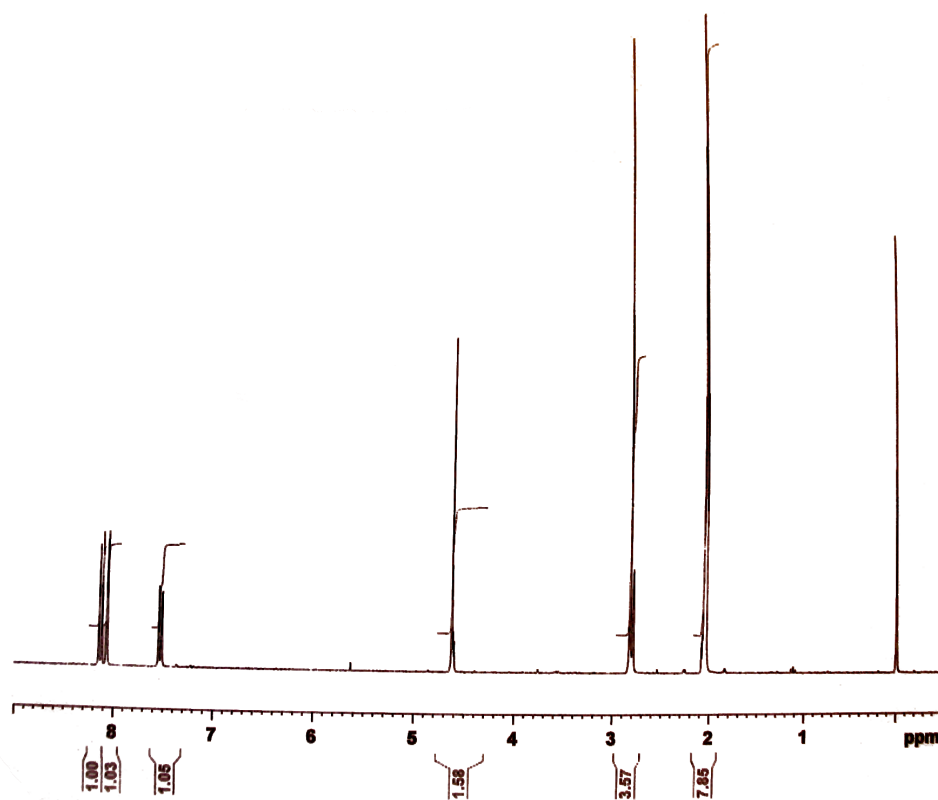
Figure 2.20: ^{1}H NMR spectrum of 2-(5-chlorobenzo[*d*]thiazol-2-yl)acetonitrile (**20**) in acetone-d$_6$.

The ^{1}H-^{1}H COSY spectrum of **20** is depicted in Figure 2.21. Well-defined cross peaks can be seen only in the aromatic region where spin-spin coupling between the protons was expected.
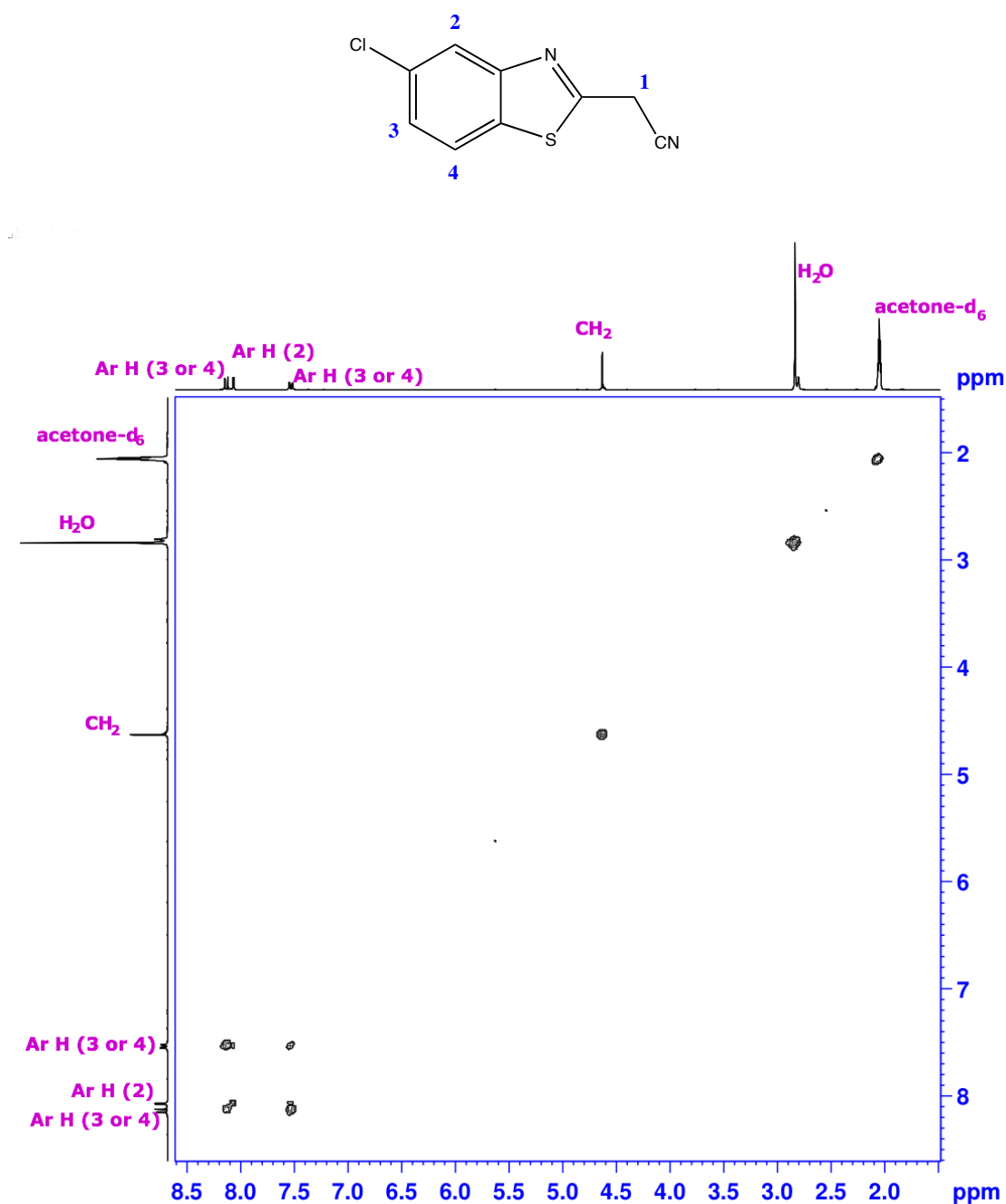
Figure 2.21: ^{1}H- ^{1}H COSY NMR spectrum of 2-(5-chlorobenzo[*d*]thiazol-2-yl)acetonitrile (**20**) in acetone-d₆.

The successful synthesis of product **20** is also confirmed by its ^{13}C NMR spectrum (Figure 2.22). As expected, nine different signals are shown in the spectrum. The peak at δ 23.33 corresponds to the carbon atom in a methylene group, while the peaks at δ 123.42, 124.28 and 126.90 belong to the aromatic -CH carbons. The remaining five signals (δ 116.35, 132.95, 135.17, 154.70, 163.57 ppm) correspond to quaternary carbons. These assignments were possible with the assistance of the DEPT-135 spectrum displayed in Figure 2.23.
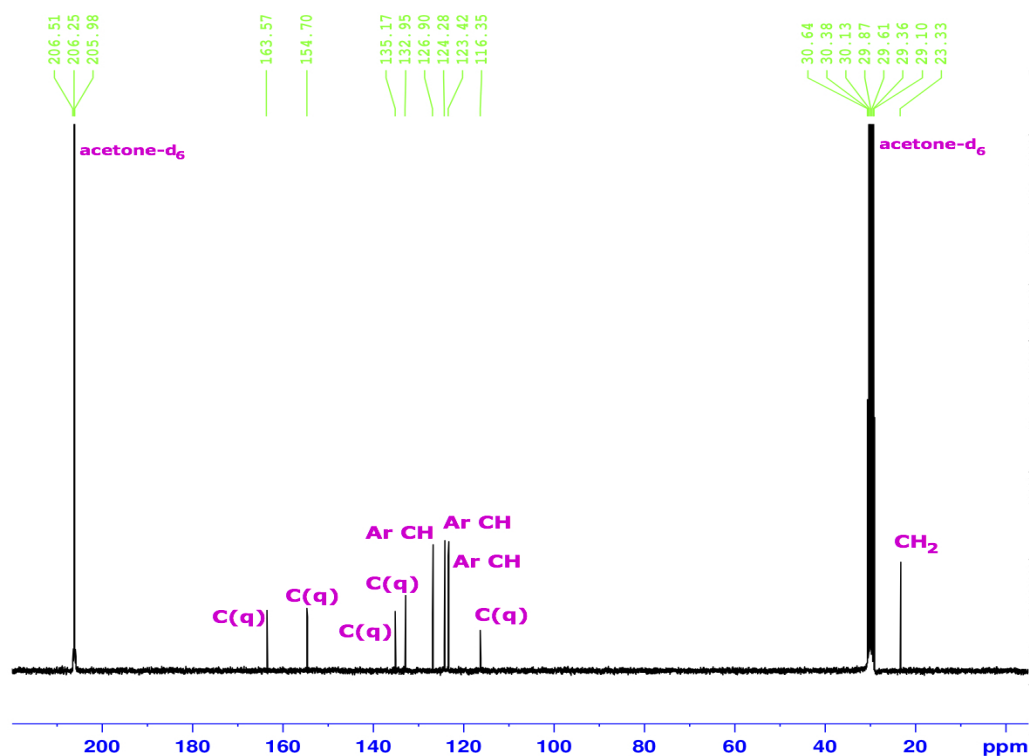
Figure 2.22: ^{13}C NMR spectrum of 2-(5-chlorobenzo[*d*]thiazol-2-yl)acetonitrile (**20**) in acetone-d$_6$.
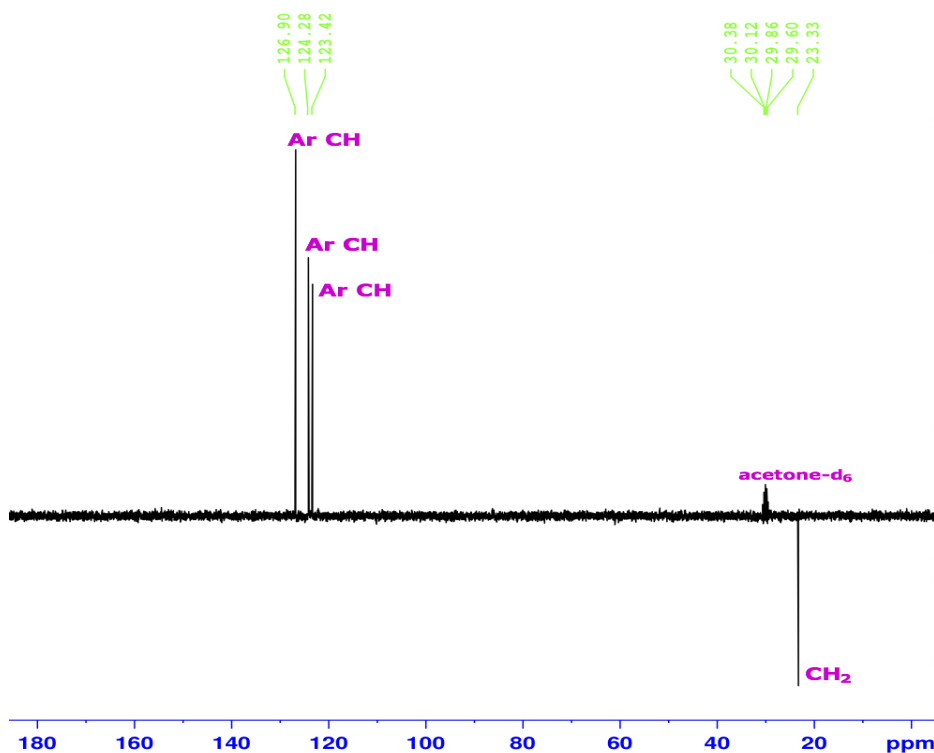
Figure 2.23: DEPT-135 spectrum of 2-(5-chlorobenzo[*d*]thiazol-2-yl)acetonitrile (**20**) in acetone-d$_6$.

Lastly, by using ^{13}C-^{1}H (HMQC) correlations (Figure 2.24) it was possible confirm the connectivity between the protons found in the ^{1}H NMR spectrum and the carbon

nuclei found in the ^{13}C spectrum. It is worth restating that the HMQC experiment requires two data sets (a proton chemical shift and a carbon chemical shift) to be sensitive. Thus quaternary carbons, as well as protons bound to non-carbon atoms (i.e. NH, OH) do not produce signals.
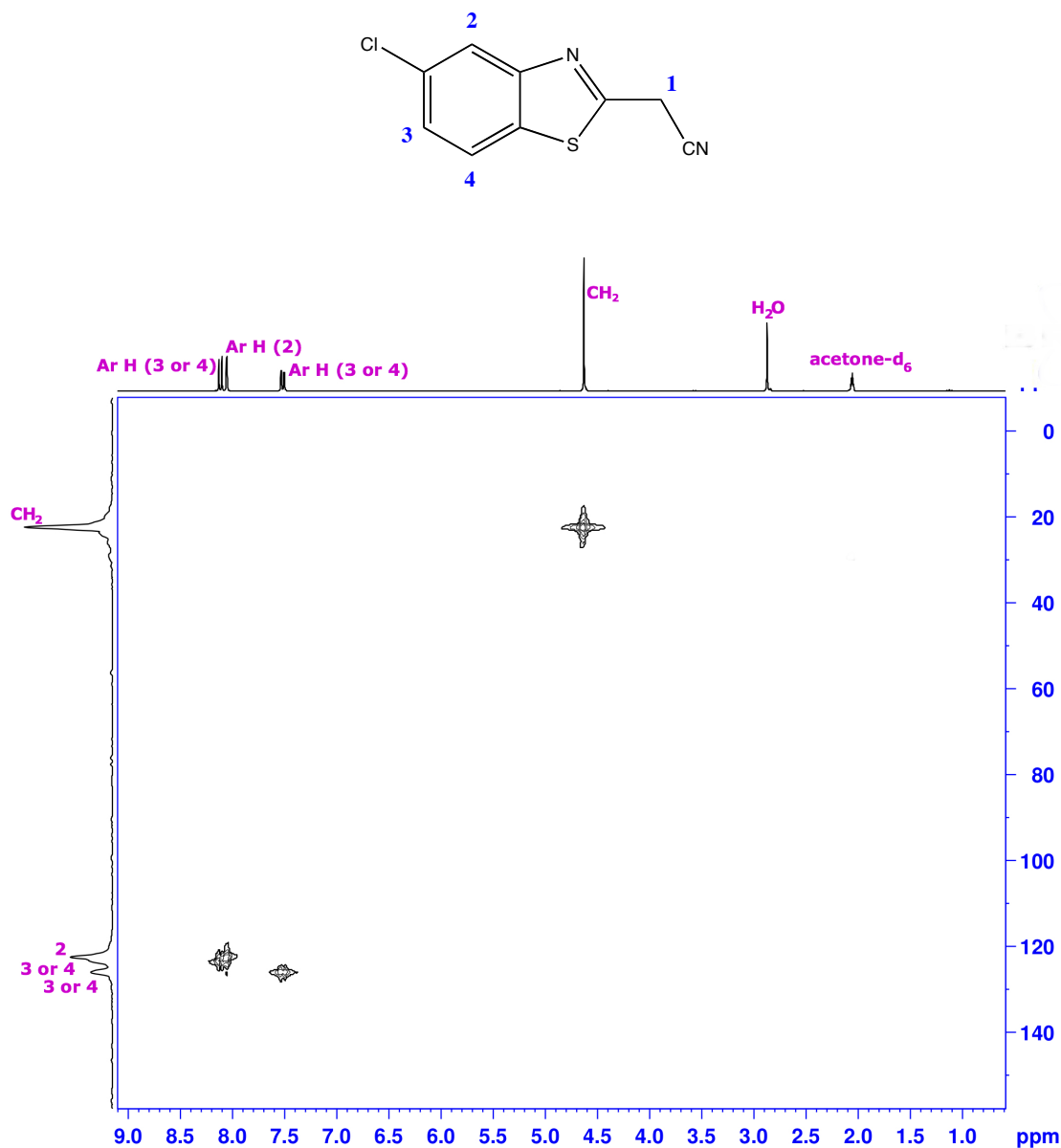
Figure 2.24: HMQC spectrum of 2-(5-chlorobenzo[*d*]thiazol-2-yl)acetonitrile (**20**) in acetone-d_6.

The formation of compound **20** was further corroborated by electrospray mass spectroscopy which identified the expected peaks at 231 [M+Na]$^+$, and 207 [M-H$^+$]$^-$.

2.2.2 Knoevenagel condensation

The next synthetic step involved a Knoevenagel condensation[156]. The condensation of aldehydes or ketones and active methylene compounds is called Knoevenagel reaction. When compounds contain a methylene group between two strong electron-withdrawing groups they are termed active methylene compounds. The methylene protons are called α-protons. Treatment with a suitable base (a base that has a conjugate acid with a pK_a greater than that of the α-proton) removes the α-proton and generates the corresponding enolate anion (Scheme 34). Extra resonance stabilisation of the enolate anion makes the α-protons more acidic.
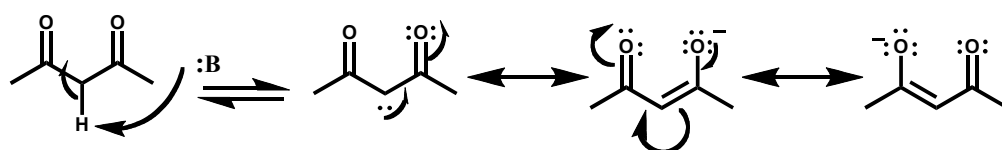
Scheme 34: Example of an active methylene compound.

As illustrated in Scheme 35, various coumarins have been prepared *via* Knoevenagel condensation of salicylaldehyde (**69**) with different active methylene compounds[157-160]. Usually, a secondary amine is used as the base, as it allows partial deprotonation of the active methylene compound, but not of a normal aldehyde, thus impeding self-condensation of the aldehyde.

X=CO$_2$Et, COMe, CN, p-C$_6$H$_4$NO$_2$
Y=CO$_2$Et
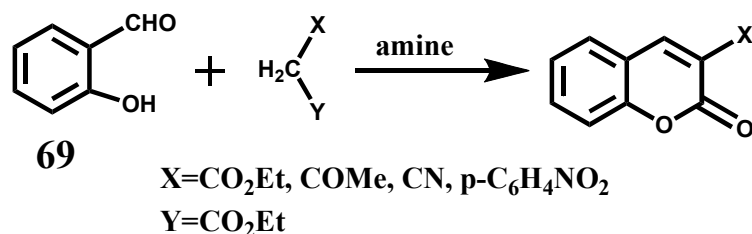
Scheme 35: Preparation of coumarins *via* Knoevenagel condensation.

This reaction has also been applied to the synthesis of iminocoumarins[161, 162], as shown in Scheme 36. Again, an amine is used to create a stabilised carbanion.

X= CN, p-C$_6$H$_4$NO$_2$
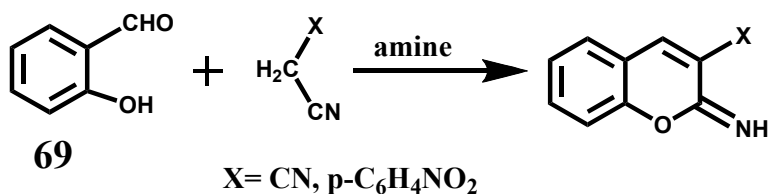
Scheme 36: Preparation of iminocoumarins *via* Knoevenagel condensation.

Following a modification of the methodology proposed by Liepouri and co-workers[39], product **21** was obtained in moderate yield by Knoevenagel-type reaction of benzothiazole **20** with salicylaldehyde **15** in anhydrous ethanol and in the presence of piperidine (Scheme 37).
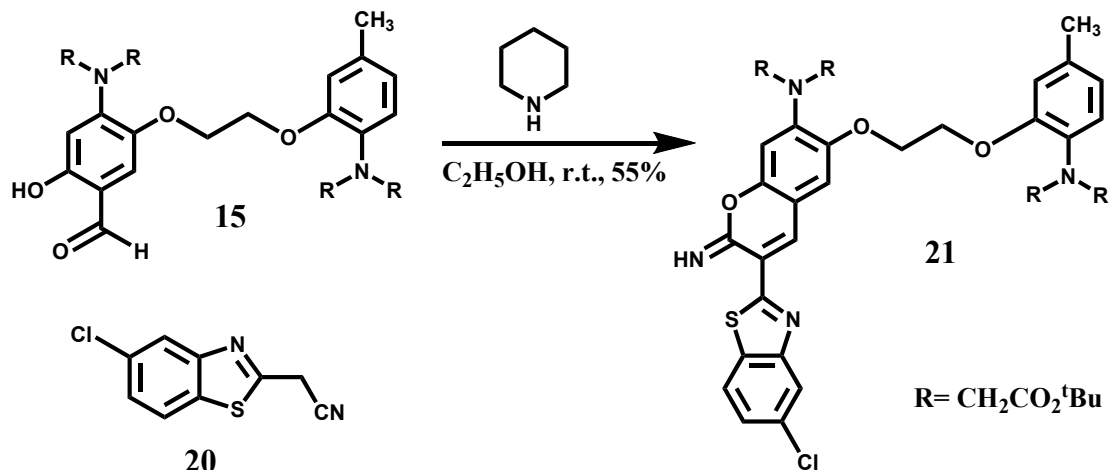
Scheme 37: Synthesis of **21** using a Knoevenagel-type reaction.

The first step of the mechanism for this reaction is the deprotonation of **20** by piperidine. Two resonance structures (**70a** and **70b**) can be drawn for the activated species.
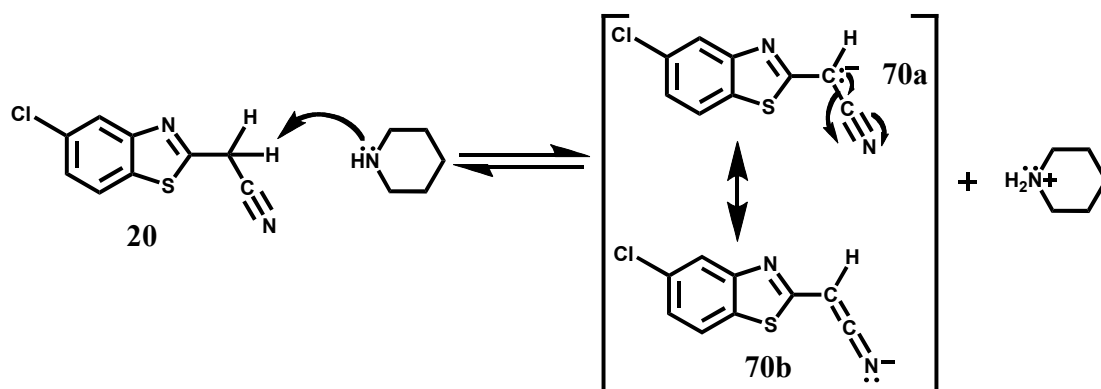
Scheme 38: Deprotonation of **20** by piperidine.

The second stage involves attack by the carbanion on the carbonyl group of salicylaldehyde **15**, forming new σ bond to the nucleophile. The resulting adduct (**71**) undergoes subsequent base-catalysed dehydration. Finally, intermediate **73** undergoes intramolecular ring closure to form the coumarin ring. The ring formation begins by donation of the hydroxyl lone pair of electrons to the partly positive nitrile carbon, and concludes by proton transfer affording product **21**.
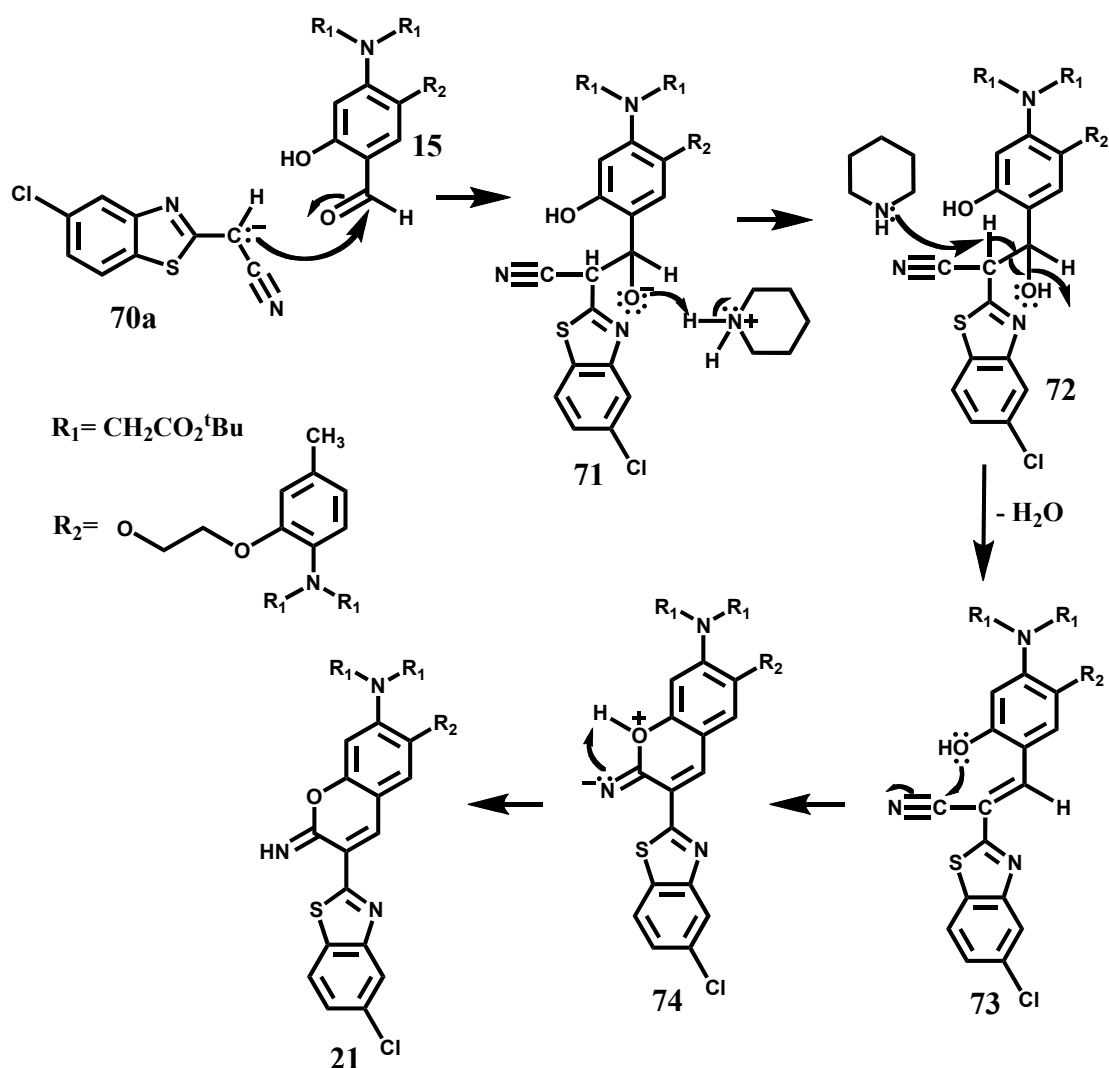
Scheme 39: Reaction mechanism for the synthesis of **21** using a Knoevenagel-type reaction.

Strong evidence to support the successful synthesis of iminocoumarin **21** was found in the ^{1}H NMR spectrum of the product. It displayed a series of well-defined signals that are labelled in Figure 2.25. These are: two singlets at δ 1.42 and 1.48 ppm each corresponding to the 18 hydrogens of the tetra ^{t}butyl esters; a singlet of three methyl protons (δ 2.25 ppm); three singlets of methylene protons, each integrating for four hydrogens (δ 4.05, 4.14 and 4.37 ppm); and eight peaks (doublet δ 6.48 ppm, doublet δ 6.72 ppm, doublet δ 6.82 ppm, singlet δ 7.10 ppm, doublet δ 7.32 ppm, singlet δ 7.55 ppm, doublet δ 7.83 ppm, singlet δ 7.98 ppm) corresponding to nine aromatic protons; lastly, a broad singlet corresponding to the proton of an N-H bond (δ 8.54 ppm).
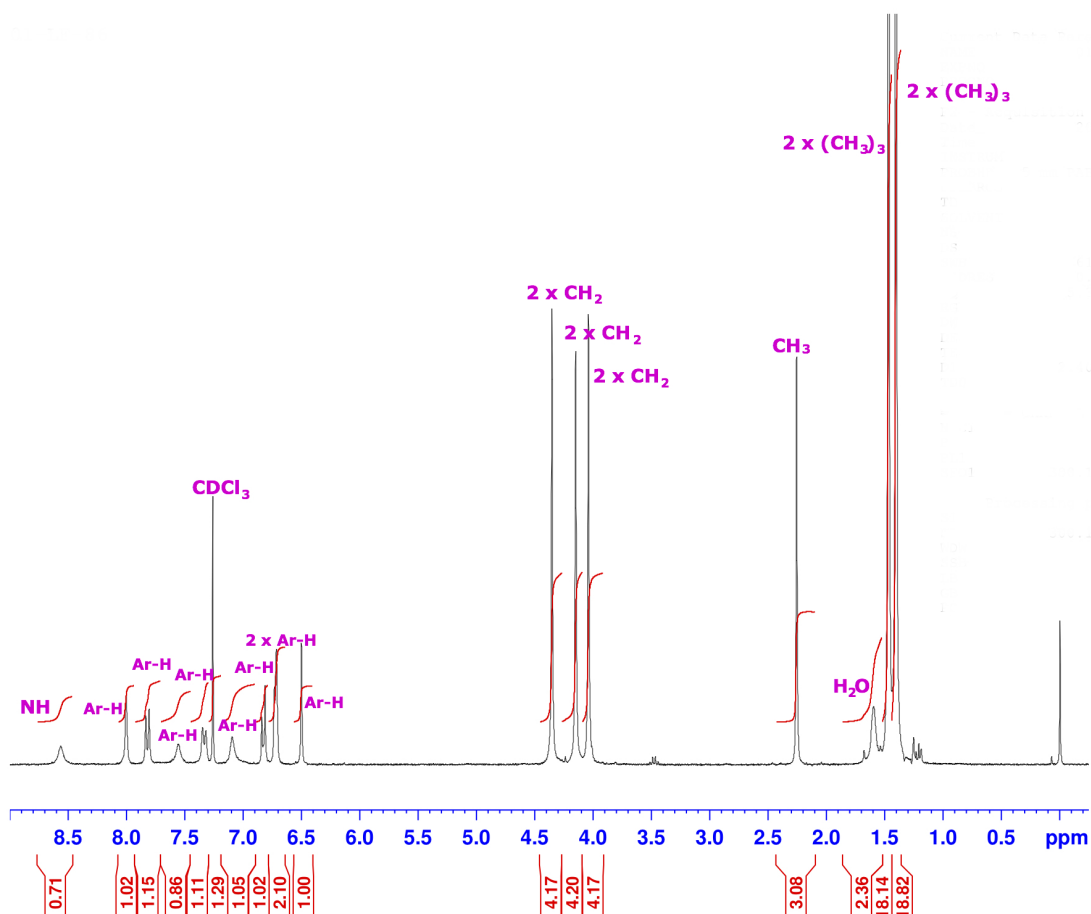
Figure 2.25: ^{1}H NMR spectrum of BTIC derivative **21** in CDCl$_3$

Product **21** was visibly fluorescent. Unfortunately, it appears to be photosensitive, and as such further spectroscopic analyses were not possible.

2.2.3 Conclusions of the progress towards derivatisable probe 12

The first attempt to synthesise **21** has been successfully completed. Deprotection of the tetra ^{t}butyl esters would furnish potential fluorescent probe **12**. The spectral properties and binding affinities for Ca^{2+} could then be determined.

However, a key issue that has been made obvious by the studies of this thesis is the unexpected (light-induced) instability of **21**. This in reality precludes its eventual use as a GFP-bound probe and a different fluorophore skeleton should be sought.

Overall, a complex and many-step synthesis was achieved, and indicated that this particular framework (compound **21**) is not suitable for a fluorophore to link to proteins (e.g. GFP mutant uv5GFPC48I229), surfaces (e.g. 96-well plate formats) or other

media (e.g. nuclei acids). This however, could not have been assessed in any other way than a trial synthesis. In this respect, this research removes what to all appearances was a likely candidate.

Further work could involve the synthesis of other BTIC derivatives using the versatile precursor **15** with other halogenated benzothiazoles (X= Br, I). Additionally, the ability of the iminocoumarin moiety to be extended via nitrogen substitution with a lipophilic chain could also be exploited.

3. Experimental

In this chapter a description of techniques and apparatus used for the purpose of the work discussed in this thesis is presented along with a description of experimental procedures and data for the synthesis of compounds.

3.1 Chemicals and reagents

Chemicals and reagents (highest quality available) were purchased from the suppliers indicated below:

Acros Organics, Loughborough, Leicestershire, UK: *n*-butyl lithium.

BHD Ltd., Poole, UK: acetyl chloride, potassium hydrogen carbonate, sodium hydroxide pellets.

Fisher Scientific, Loughborough, Leicestershire, UK: magnesium sulphate, sodium sulphate, ethyl acetate, petroleum ether 40-60°C, dichloromethane, hexane, diethyl ether, methanol, ethanol, chloroform.

Fluka Chemicals Ltd. Dorset, UK: isobutyl chloroformate, Lawesson's reagent.

Fluorochem Ltd., Glossop, Derbyshire, UK: Silica gel 40-63u, 60 Å.

Sigma-Aldrich Co. Ltd., Dorset, UK: 4-aminophenethylamine, benzyl chloroformate, diisopropylamine, di-*tert*butyl dicarbonate, ethyl chloroformate, phenyl chloroformate, potassium ferricyanide, triethylamine, 2-amino-4-chloro-benzenethiol, malononitrile, piperidine, tetrahydrofuran anhydrous, anhydrous dichloromethane, chloroform-D, acetone-d_6.

3.2 Instrumentation

3.2.1 Melting point determination

Melting points, recorded using a Köfler Melting Point apparatus microscope (Reichert, Austria), are uncorrected.

3.2.2 Infrared spectra

Infrared spectra were recorded as thin films (solution in $CHCl_3$) on a JASCO 4100 FTIR spectrometer. Absorption maxima (υ_{max}) are recorded in wavenumbers (cm^{-1}) and the following abbreviations are used: w, weak; m, medium; s, strong; br, broad.

3.2.3 Nuclear magnetic resonance

NMR spectra were recorded using a Bruker Avance-300 spectrometer operating at 300 MHz for ^{1}H-NMR and 75 MHz for ^{13}C-NMR spectroscopies using XWIN NMR system software. Proton assignments were supported by ^{1}H-^{1}H COSY spectra. The presence of amine protons was confirmed by hydrogen-deuterium exchange using D_2O and is indicated with an asterisk (*). Carbon assignments are supported by DEPT-135 editing and where necessary ^{13}C-^{1}H (HMQC) correlations. Chemical shifts (δ) are reported in parts per million (ppm) to the nearest 0.01 ppm and are referenced to tetramethylsilane or the appropriate residual solvent peak. Coupling constants (J) are reported in Hertz to the nearest 0.1 Hz. Data are reported as follows: chemical shift, integration, multiplicity (br, broad; s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet; or as a combination of these e.g. dd, double doublet, dt, double triplet, etc. In all cases '~' signifies approximate or appearing as, where signal should be dd or other multiplicity, but appears on the spectrum as t or other multiplicity.

3.2.4 Mass spectrometry

Mass spectrometry was performed at the department of Chemistry, University of Manchester. A Micromass Trio 2000 (Mosslynx software) instrument was used for electron impact (EI) and chemical ionisation (CI) conditions and a Micromass Platform for positive and negative ion electrospray. Only molecular ions, fractions from molecular ions and other major peaks are reported as mass/charge (m/z) ratios.

3.2.5 UV-visible spectrophotometry

UV and visible spectra were recorded in 1 cm path-length quartz cuvettes using a Peltier-thermostatted cuvette holder in a Cary 4000 UV-visible spectrophotometer.

3.2.6 Flash column chromatography

Flash column chromatography was carried out using a slight modification of the procedure reported by Still and co-workers[163] (the column was loaded with the prepared slurry). The silica gel was Kieselgel 60, particle size 0.035-0.070 mm. Solvents used as eluents are indicated in the experimental descriptions. Where mixtures of solvents are given, such as 1:3, this refers to v/v.

3.2.7 Thin layer chromatography (t.l.c.)

Reaction progress was monitored by t.l.c. on silica gel with fluorescent indicator (POLYGRAM SILG/ UV254, Macherey-Nagel). Visualisation was by the quenching of UV fluorescence (λ_{max}=254 nm) or by staining with vanillin, followed by heating. Retention factors (R_f) are quoted to 0.1.

3.2.8 General

Experimental details are written in accordance with Journal of the Chemical Society, Perkin Transactions 1 Instructions for Authors (2003).

All reactions were carried out with exclusion of water in an inert nitrogen atmosphere using a vacuum nitrogen manifold unless otherwise stated. All glassware, syringes and needles were pre-dried in an oven (110 °C) and cooled in a nitrogen atmosphere prior to use. Solvents were evaporated on a Büchi RE111 Rotavapor equipped with a Büchi 461 water bath. Distilled water was obtained from a Millipore Elix 5 system.

3.3 Chemical synthesis

tert-Butyl 4-aminophenethylcarbamate (31)

$C_{13}H_{20}N_2O_2$, MW: 236.31
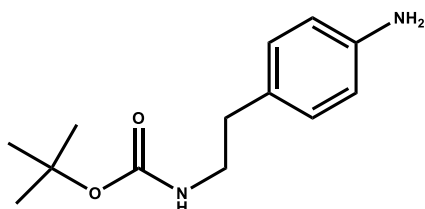
A mixture of 4-aminophenethylamine (**30**) (200 mg, 1.46 mmol) and di-*tert*butyl dicarbonate (353 mg, 1.61 mmol) in anhydrous DCM (2 mL) was stirred at r.t. with monitoring by t.l.c. On completion of reaction, the mixture was concentrated *in vacuo*. The residue was taken up in ethyl acetate (10 mL), washed with water (2 x 10 mL) and then brine (2 x 10 mL). The organic layer was dried (MgSO$_4$), filtered and concentrated *in vacuo* to afford the title compound as a yellow solid (332 mg, 95%). m.p. 81-83 ºC[101]; R_f 0.25 (methanol: DCM, 3:97); ν_{max}(film)/cm^{-1} 3375 w (N-H), 3197 w (N-H, primary amine), 3197 w (N-H, primary amine), 2979 w (C$_{Ar}$-H), 2870 w (C-H), 1683 s (C=O), 1513 s (C=C), 1365 m (C$_{Ar}$-N), 1244 s (C-O), 1163 s, 907 m; δ_H (300 MHz, CDCl$_3$) 6.98 (2H, d, *J* 7.4, C(2,6)H), 6.64 (2H, d, *J* 7.4, C(3,5)H), 4.53 (1H, br s*, NH), 3.59 (2H, br s*, NH$_2$), 3.36-3.25 (2H, m, C(1b)H$_2$), 2.68 (2H, t, *J* 6.8, C(1a)H$_2$), 1.43 (9H, s, C(CH$_3$)$_3$); δ_C (75 MHz,CDCl$_3$) 155.90 (C(O)OC(CH$_3$)$_3$), 144.79 (quaternary C), 129.60 (C(2)H and C(6)H), 128.84 (quaternary C), 115.35 (C(3)H and C(5)H), 79.10 (OC(CH$_3$)$_3$), 42.01 (C(1b)H$_2$), 35.25 (C(1a)H$_2$), 28.42 ((CH$_3$)$_3$); *m/z* (ES)$^+$ 259 ([M+Na]$^+$, 100%)

tert-Butyl 4-acetamidophenethylcarbamate (32)

C$_{15}$H$_{22}$N$_2$O$_3$, MW: 278.35
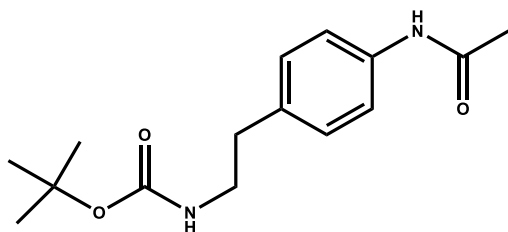
To a solution of *tert*-butyl 4-aminophenethylcarbamate **(31)** (322mg, 1.36 mmol) and triethylamine (278 μL, 1.66 mmol) in anhydrous DCM (5 mL) at 0 °C was added acetyl chloride (131 μL, 1.51 mmol) dropwise with stirring. The reaction mixture was stirred at r.t. for 3 h. The reaction mixture was concentrated *in vacuo* and the residue taken up in ethyl acetate (15 mL), washed with water (2 x 15 mL) and with brine (2 x 15 mL). The organic layer was dried (MgSO$_4$), filtered and concentrated *in vacuo*. Purification by flash column chromatography (3:1 petroleum ether 40-60: ethyl acetate, SiO$_2$) afforded the title compound as a white powder (360 mg, 93%). m.p. 130-131 ºC; R_f 0.15 (methanol: DCM, 3:97); ν_{max}(film)/cm^{-1} 3308 w (N-H), 2976 w (C$_{Ar}$-H), 2869 w (C-H), 1668 s (C=O), 1603 (C=O), 1530 s (C=C), 1409 m, 1365 m (C$_{Ar}$-N), 1250 s (C-O), 1163 s, 1051 m (C-N), 909 m; δ_H (300 MHz, CDCl$_3$) 7.37 (2H, d, *J* 8.2, C(3,5)H),

7.00(2H, d, *J* 8.2, C(2,6)H̲), 4.68 (1H, br s*, NH), 3.24 (2H, dt, *J* 6.4, C(1b)H̲₂), 2.64 (2H, t, *J* 7.0, C(1a)H̲₂), 2.05 (3H, s, COCH̲₃) 1.35 (9H, s, OC(CH̲₃)₃); δ_C (75 MHz,CDCl₃) 168.28 (C̲OCH₃), 155.87 (C̲(O)OC(CH₃)₃), 136.21 (quaternary C̲), 135.02 (quaternary C̲), 129.31 (C̲(2)H and C̲(6)H), 120.21 (C̲(3)H and C̲(5)H), 79.28 (OC̲(CH₃)₃), 41.74 (C̲(1b)H₂), 35.56 (C̲(1a)H₂), 28.40 ((C̲H₃)₃), 24.56 (COC̲H₃); *m/z* (ES)⁺ 337 ([M+ACN+NH₄]⁺, 100%), (ES)⁻ 277 ([M-H⁺]⁻, 100%).

tert-Butyl 4-thioacetamidophenethylcarbamate (33)

C₁₅H₂₂N₂O₂S, MW: 294.41
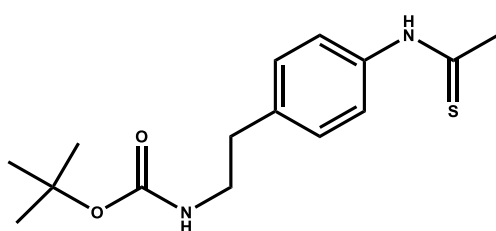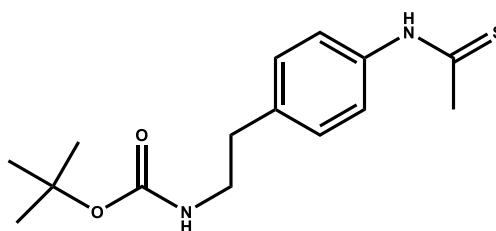
Rotamer A   **Rotamer B**

To a solution of *tert*-butyl 4-acetamidophenethylcarbamate **(32)** (306 mg, 1.10 mmol) in anhydrous THF (11 mL), Lawesson's reagent (267 mg, 0.66 mmol) was added in one portion. The reaction mixture was heated at reflux for 3.5 h then concentrated *in vacuo*. The residue was taken up in ethyl acetate (15 mL) and extracted with saturated aqueous sodium hydrogen carbonate (2 x 15 mL) and water (2 x 15 ml). The organic layer was dried (MgSO₄), filtered and concentrated *in vacuo*. Purification by flash column chromatography (3:1 petroleum ether 40-60: ethyl acetate, SiO₂) afforded the title compound as a yellow oil (243 mg, 75 %). R_f 0.17 (methanol: DCM, 3:97); ν_{max} (film)/cm⁻¹ 3246 w (N-H), 3029 w (C$_{Ar}$-H), 2973 w (C-H), 1683 s (C=O), 1509 s (C=C), 1409 m, 1364 m (C$_{Ar}$-N), 1249 s (C-O), 1161 s (C=S), 1051 w (C-N), 909 m; *m/z* (ES)⁺ 353 ([M+ACN+NH₄]⁺, 100%), (ES)⁻ 293 ([M-H⁺]⁻, 100%). Two rotamers (rotation about the N-C bond) were identified by NMR spectroscopy. At room temperature, the predominant isomer was found in a ratio of ~ 2 to 1. See section 2.1.3 for details. Predominant isomer: δ_H (300 MHz, CDCl₃) 8.95 (1H, br s*, NH̲), 7.61 (2H, d, *J* 7.5, C(3,5)H̲), 7.17-7.28 (2H, m, C(2,6)H̲), 4.59 (1H, br s*, NH), 3.26-3.44 (2H, m, C(1b)H̲₂), 2.75-2.86 (2H, m, C(1a)H̲₂), 2.74 (3H, s, CSCH̲₃), 1.43 (9H, s, OC(CH̲₃)₃); δ_C (75 MHz, CDCl₃) 200.35 (C̲SCH₃), 155.92 (C̲(O)OC(CH₃)₃), 137.31 (quaternary C̲), 136.46 (quaternary C̲), 129.12 (Ar-H), 124.07 (Ar-H), 79.47

(O$\underline{C}$(CH$_3$)$_3$), 41.62 ($\underline{C}$(1b)H$_2$), 35.83 ($\underline{C}$(1a)H$_2$), 35.78 (CO$\underline{C}$H$_3$), 28.40 (($\underline{C}$H$_3$)$_3$). Minor isomer: δ_H (300 MHz, CDCl$_3$) 9.49 (1H, br s*, N$\underline{H}$), 7.17-7.28 (2H, m, C(2,6)$\underline{H}$), 7.10 (2H, d, J 7.5, C(3,5)$\underline{H}$), 4.59 (1H, br s*, NH), 3.26-3.44 (2H, m, C(1b)$\underline{H}_2$), 2.75-2.86 (2H, m, C(1a)$\underline{H}_2$), 2.51 (3H, s, CS$\underline{CH}_3$), 1.43 (9H, s, OC(C$\underline{H}_3$)$_3$); δ_C (75 MHz, CDCl$_3$) 204.44 ($\underline{C}$SCH$_3$), 156.03 ($\underline{C}$(O)OC(CH$_3$)$_3$), 139.00 (quaternary $\underline{C}$), 137.63 (quaternary $\underline{C}$), 129.93 (Ar-H), 125.21 (Ar-H), 79.47 (O$\underline{C}$(CH$_3$)$_3$), 41.62 ($\underline{C}$(1b)H$_2$), 35.83 ($\underline{C}$(1a)H$_2$), 30.06 (CO$\underline{C}$H$_3$), 28.40 (($\underline{C}$H$_3$)$_3$).

tert-Butyl 2-(2-methylbenzo[d]thiazol-6-yl)ethylcarbamate (34)

C$_{15}$H$_{20}$N$_2$O$_2$S, MW: 292.40
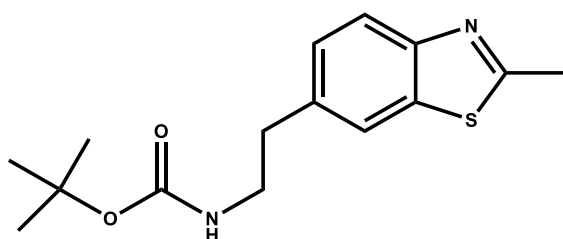
A solution of potassium ferricyanide (677 mg, 2.05 mmol) in water (3.5 mL) was stirred into a mixture of *tert*-butyl 4-thioacetamidophenethylcarbamate (**33**) (151 mg, 0.51 mmol) and sodium hydroxide (160 mg, 4.01 mmol) in water (6.5 mL) and ethanol (1mL). The reaction was carried out using a Biotage 2.0 microwave system at 120 °C for 14 min to produce a green, viscous precipitate. The precipitate was collected by filtration and washed with cold water (10 mL), then dissolved in ethyl acetate and insoluble material removed by filtration. The ethereal layer was dried (Na$_2$SO$_4$), filtered and concentrated *in vacuo*. Purification by flash column chromatography (3:1 petroleum ether 40-60: ethyl acetate, SiO$_2$) afforded the title compound as a yellow solid (84 mg, 56%). m.p. 76-77 °C; R_f 0.25 (methanol: DCM, 3:97); ν_{max} (film)/cm^{-1} 3334 w (N-H), 2974 w (C$_{Ar}$-H), 1693 s (C=O), 1603 (C=N), 1520 s (C=C), 1409 m, 1364 m (C$_{Ar}$-N), 1269 m, 1248 s (C-O), 1166 s, 1054 w (C-N), 963 m, 818 m (C-S); δ_H (300 MHz, CDCl$_3$) 7.86 (1H, d, J 8.2, C(4/5)$\underline{H}$), 7.53 (1H, s, C(7)$\underline{H}$), 7.28 (1H, d, J 6.5, C(4/5)$\underline{H}$), 4.49 (1H, br s*, NH), 3.41-3.30 (2H, m, C(6b)$\underline{H}_2$), 2.84 (2H, t, J 6.5, C(6a)$\underline{H}_2$), 2.17 (3H, s, C$\underline{H}_3$) 1.43 (9H, s, OC(C$\underline{H}_3$)$_3$); δ_C (75 MHz,CDCl$_3$) 166.49 (S$\underline{C}$CH$_3$), 155.84 ($\underline{C}$(O)OC(CH$_3$)$_3$), 152.13 (C=$\underline{C}$N), 136.02 (quaternary $\underline{C}$), 135.86 (C=$\underline{C}$S), 126.77 ($\underline{C}$(4)H/$\underline{C}$(5)H), 122.17 ($\underline{C}$(4)H/$\underline{C}$(5)H), 121.23 ($\underline{C}$(7)H), 79.32 (O$\underline{C}$(CH$_3$)$_3$), 41.91 ($\underline{C}$(6b)H$_2$), 36.12 ($\underline{C}$(6a)H$_2$), 28.38 (($\underline{C}$H$_3$)$_3$), 20.08 (C$\underline{C}$H$_3$); m/z (ES)$^+$ 315 ([M+Na]$^+$, 100%), (ES)$^-$ 291 ([M-H$^+$]$^-$, 100%).

Ethyl 2-(6-(2-(*tert*-butoxycarbonylamino)ethyl)benzo[d]thiazol-2-yl)acetate (35)

$C_{18}H_{24}N_2O_4S$, MW: 364.46
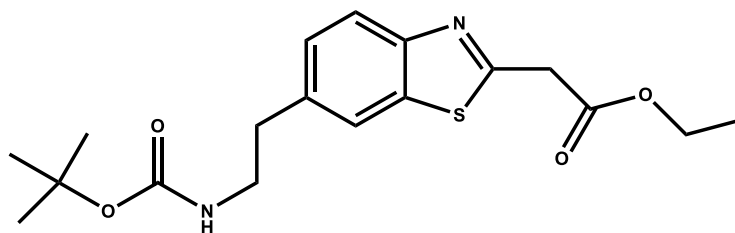
n-BuLi (246 μL, 0.31 mmol) was added to a solution of diisopropylamine (43 μL, 0.31 mmol) in THF (0.75 mL) at 0 °C. The reaction mixture was maintained at 0 °C for 20 min, cooled to -78 °C and *tert*-butyl 2-(2-methylbenzo[d]thiazol-6-yl)ethylcarbamate (**34**) (30 mg, 0.11 mmol) in THF (0.75 mL) added dropwise. The mixture was stirred at -78 °C for 30 min, after which ethyl chloroformate (13 μL, 0.13 mmol) was added dropwise and the mixture stirred at r.t. for 2.5 h. Saturated aq. ammonium chloride (10 mL) was added and the mixture extracted with diethyl ether (2 x 10 mL). The organic layer was dried (Na_2SO_4), filtered and concentrated *in vacuo* to afford the title compound (crude 34 mg, 85 %). R_f 0.14 (Et$_2$O: hexane, 1:1); ν_{max} (film)/cm^{-1} 3329 w (N-H), 2960 w (C_{Ar}-H), 2870 w (C-H), 1737 s (C=O), 1713 s (C=O), 1602 w (C=N), 1518 s (C=C), 1455 m, 1367 m (C_{Ar}-N), 1247 s (C-O), 1166 s, 1152 s, 1025 w (C-N), 910 m, 819 m (C-S); δ_H (300 MHz, CDCl$_3$) 7.87 (1H, d, *J* 8.2, C(4/5)<u>H</u>), 7.63 (1H, s, C(7)<u>H</u>), 7.23 (1H, d, *J* 9.0, C(4/5)<u>H</u>), 4.48 (1H, br s, NH), 4.18 (2H, q, *J* 7.1, C<u>H</u>$_2$CH$_3$), 4.09 (2H, s, C(2a)<u>H</u>), 3.42-3.30 (2H, m, C(6b)<u>H</u>$_2$), 2.90-2.80 (2H, m, C(6a)<u>H</u>$_2$), 1.36 (9H, s, OC(C<u>H</u>$_3$)$_3$), 0.89-0.71 (3H, m, CH$_2$C<u>H</u>$_3$); m/z (ES)$^+$ 387 ([M+Na]$^+$, 100%), (ES)$^-$ 363 ([M-H$^+$]$^-$, 100%).

Isobutyl 2-(6-(2-(*tert*-butoxycarbonylamino)ethyl)benzo[d]thiazol-2-yl)acetate

$C_{20}H_{28}N_2O_4S$, MW: 392.18
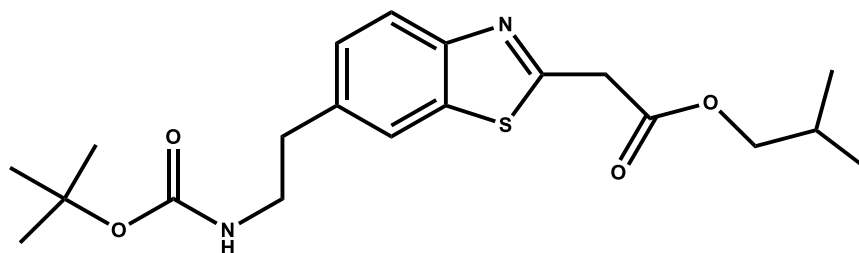
n-BuLi (246 μL, 0.31 mmol) was added to a solution of diisopropylamine (43 μL, 0.31 mmol) in THF (0.75 mL) at 0 °C. The reaction mixture was maintained at 0 °C for 20 min, cooled to -78 °C and *tert*-butyl 2-(2-methylbenzo[d]thiazol-6-yl)ethylcarbamate (**72**) (30 mg, 0.11 mmol) in THF (0.75 mL) added dropwise. The mixture was stirred at -78 °C for 30 min, after which isobutyl chloroformate (13 μL, 0.13 mmol) was added dropwise and the reaction mixture stirred at r.t. for 2.5 h. Saturated aq. ammonium chloride (10 mL) was added and the mixture extracted with diethyl ether (2 x 10 mL). The organic layer was dried (Na_2SO_4), filtered and concentrated *in vacuo* to afford the title compound as a yellow oil (crude 40.9 mg, 95%). R_f 0.26 (methanol: DCM, 3:97); δ_H (300 MHz, $CDCl_3$) 7.94 (1H, d, *J* 8.2, C(4/5)$\underline{H}$), 7.69 (1H, s, C(7)$\underline{H}$), 7.30 (1H, d, *J* 8.4, C(4/5)$\underline{H}$), 4.69 (1H, br s, NH), 4.17 (1H, s, C(2a)$\underline{H}$), 3.97 (2H, d, *J* 6.6, -C$\underline{H}_2$CHCH$_3$), 3.54-3.37 (2H, m, C(6b)$\underline{H}_2$), 3.04-2.87 (2H, m, C(6a)$\underline{H}_2$), 2.05-1.89 (1H, m, -CH$_2$C$\underline{H}$CH$_3$), 1.43 (6H, s, -CH$_2$CHC$\underline{H}_3$) 0.92 (9H, s, OC(C$\underline{H}_3$)$_3$).

2-(5-Chlorobenzo[*d*]thiazol-2-yl)acetonitrile (20)

$C_9H_5ClN_2S$, MW: 208.67
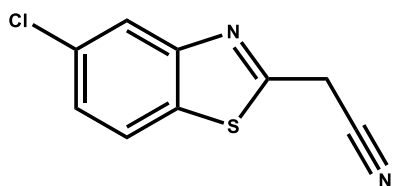
An equimolar mixture of 2-amino-4-chloro-benzenethiol (250 mg, 1.56 mmol), malononitrile (103.45 mg, 1.56 mmol), and acetic acid (84.6 μL, 1.56 mmol) in ethanol (1.5 mL) was magnetically stirred at r.t. until no further precipitate was produced, and then allowed to stand overnight. The resultant pale yellow precipitate was harvested by filtration and recrystallized from ethanol to afford the title compound (199.2 mg, 61%). R_f 0.62 (methanol: DCM, 3:97); m.p. 150-151 °C; ν_{max} (film)/cm^{-1} 3053 m (C_{Ar}-H), 2941 w (C_{Ar}-H), 2359 m (C≡N), 1622 m (C=N), 1500 s (C=C), 1435 s, 1393 m (C_{Ar}-N), 1294 m, 1164 s, 1075 w (C-N), 902 m, 812 m (C-S), 650 w (C-Cl); δ_H (300 MHz, $(CD_3)_2CO$) 8.13 (1H, d, C(3/4)$\underline{H}$), 8.05 (1H, s, C(2)$\underline{H}$), 7.56 (1H, d, C(3/4)$\underline{H}$), 4.63 (2H, s, C(1)$\underline{H}_2$); δ_C (75 MHz, $(CD_3)_2CO$) 163.57 (N=$\underline{C}$-CH$_2$), 154.70 (C=$\underline{C}$N), 135.17 (C=$\underline{C}$S), 132.95 ($\underline{C}$-Cl), 126.90 ($\underline{C}$(3)H/$\underline{C}$(4)H), 124.28 ($\underline{C}$(3)H/$\underline{C}$(4)H), 123.42 ($\underline{C}$(2)H), 116.35 ($\underline{C}$≡N), 23.33 ($\underline{C}$(1)H$_2$); m/z (ES)$^+$ 231 ([M+Na]$^+$, 100%), (ES)$^-$ 207 ([M-H$^+$]$^-$, 100%).

Halogenated BTIC derivative (21)

tert-butyl 2,2'-(2-(2-(7-(bis(2-*tert*-butoxy-2-oxoethyl)amino)-3-(5-chlorobenzo[*d*]thiazol-2-yl)-2-imino-2*H*-chromen-6-yloxy)ethoxy)-4-methylphenylazanediyl) diacetate

$C_{49}H_{61}ClN_4O_{11}S$, MW: 949.55
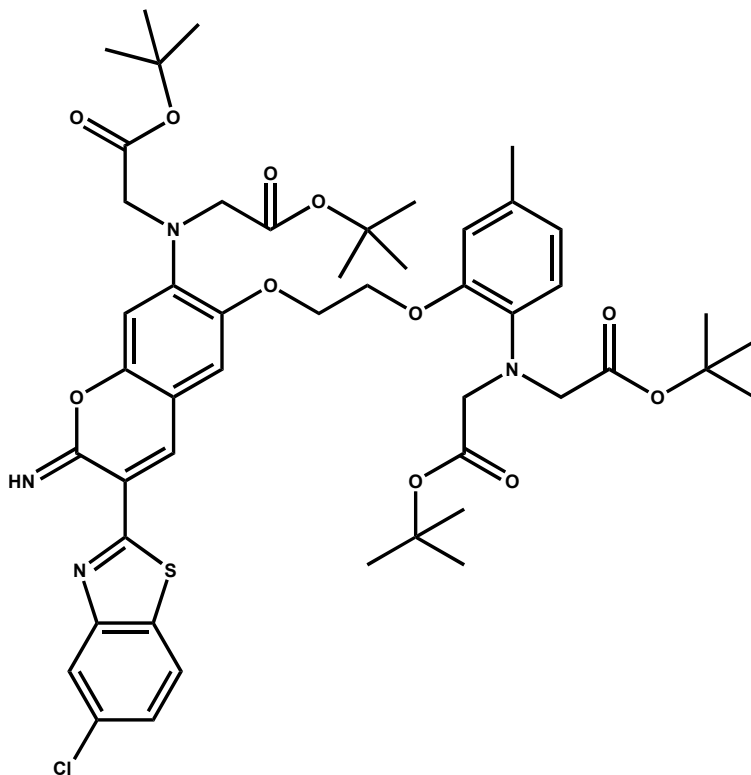
To a solution of 2-(5-chlorobenzo[*d*]thiazol-2-yl)acetonitrile (**20**) (7.14 mg, 0.034 mmol) in 0.5 mL dry ethanol was added piperidine (25.6 μL, 0.26 mmol) and the solution was stirred at r.t. for 10 min. The reaction mixture was then added to a solution of salicylaldehyde **15** (20 mg, 0.026 mmol) in dry ethanol (2 mL). This system was magnetically stirred overnight at r.t. The yellow precipitate was filtered, washed with dry ether and dried *in vacuo* to afford the title compound as a yellow solid (13.5 mg, 55%); R_f 0.21 (methanol: DCM, 3:97); m.p. 165-166 °C; δ_H (300 MHz, CDCl$_3$) 8.54 (1H, br s, N<u>H</u>), 7.98 (1H, s, Ar-H), 7.83 (1H, d, Ar-H), 7.55 (1H, s, Ar-H), 7.32 (1H, d, Ar-H), 7.10 (1H, s, Ar-H), 6.82 (1H, d, Ar-H), 6.72 (2H, d, Ar-H), 6.48 (1H, s, Ar-H), 4.37 (4H, s, 2 x CH$_2$), 4.14 (4H, s, (CH$_2$)$_2$), 4.05 (4H, s, (CH$_2$)$_2$), 2.25 (3H, s, CH$_3$), 1.48 (18H, s, ((CH$_3$)$_3$)$_2$), 1.42 (18H, s, ((CH$_3$)$_3$)$_2$).

4. References

1. Thomas, A. P.; Delaville, F., The use of fluorescent indicators for measurements of cytosolic-free calcium concentration in cell populations and single cells. In *Cellular calcium: a practical approach*, Cobbold, P. H.; McCormack, J. G., Eds. Oxford University Press: GB, 1991; 1-54.

2. Berridge, M. J.; Lipp, P.; Bootman, M. D., The versatility and universality of calcium signalling. *Nature Reviews Molecular Cell Biology* **2000,** *1*, 11-21.

3. Berridge, M. J.; Bootman, M. D.; Roderick, H. L., Calcium signalling: dynamics, homeostasis and remodelling. *Nature Reviews Molecular Cell Biology* **2003,** *4*, 517-529.

4. Lee, N. Y.; Sun, M. H.; Ko, W. S., Effect of hyperglycemia on the changes of intracellular $[Ca^{2+}]_i$ in heart myoblast. *Chinese Journal of Physiology* **2010,** *53*, 294-298.

5. MacLennan, D. H., Ca^{2+} signalling and muscle disease. *European Journal of Biochemistry* **2000,** *267*, 5291-5297.

6. Missiaen, L.; Robberecht, W.; van den Bosch, L.; Callewaert, G.; Parys, J. B.; Wuytack, F.; Raeymaekers, L.; Nilius, B.; Eggermont, J.; De Smedt, H., Abnormal intracellular Ca^{2+} homeostasis and disease. *Cell Calcium* **2000,** *28*, 1-21.

7. Nicotera, P.; Orrenius, S., The role of calcium in apoptosis. *Cell Calcium* **1998,** *23*, 173-180.

8. Zhivotovsky, B.; Orrenius, S., Calcium and cell death mechanisms: a perspective from the cell death community. *Cell Calcium* **2011,** *50*, 211-221.

9. Rudolf, R.; Mongillo, M.; Rizzuto, R.; Pozzan, T., Looking forward to seeing calcium. *Nature Reviews Molecular Cell Biology* **2003,** *4*, 579-586.

10. Takahashi, A.; Camacho, P.; Lechleiter, J. D.; Herman, B., Measurement of intracellular calcium. *Physiological Reviews* **1999,** *79*, 1089-125.

11. Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*. Third ed.; Springer: 2006.

12. Gell, C.; Brockwell, D.; Smith, A., *Handbook of single molecule fluorescence spectroscopy*. Oxford University Press: 2006.

13. Valeur, B., *Molecular Fluorescence: Principles and aplications* First ed.; Wiley-VCH: 2002.

14. Joo, C.; Balci, H.; Ishitsuka, Y.; Buranachai, C.; Ha, T., Advances in single-molecule fluorescence methods for molecular biology. *Annual Review of Biochemistry* **2008,** *77*, 51-76.

15. Jablonski, A., Efficiency of Anti-Stokes Fluorescence in Dyes. *Nature* **1933,** *131*, 839-840.

16.   Atkins, P. W., *Physical Chemistry*. 6th ed.; Oxford University Press: 2001.

17.   Fairclough, R. H.; Cantor, C. R., The use of singlet-singlet energy transfer to study macromolecular assemblies. *Methods in Enzymology* **1978,** *48*, 347-379.

18.   Royer, C. A., Approaches to teaching fluorescence spectroscopy. *Biophysical Journal* **1995,** *68*, 1191-1195.

19.   Paredes, R. M.; Etzler, J. C.; Watts, L. T.; Zheng, W.; Lechleiter, J. D., Chemical calcium indicators. *Methods* **2008,** *46*, 143-151.

20.   Kasten, F. H., Introduction to fluorescent probes: properties, history and applications. In *Fluorescent and Luminescent Probes for Biological Activity*, Second ed.; Mason, W. T., Ed. Academic Press: 1999; 12-33.

21.   Valeur, B.; Leray, I., Design principles of fluorescent molecular sensors for cation recognition. *Coordination Chemistry Reviews* **2000,** *205*, 3-40.

22.   Prodi, L.; Montalti, M.; Zaccheroni, N.; Dolci, L. S., Probes and sensors for cations. In *Advanced Concepts in Fluorescence Sensing*, Geddes;, C. D., Ed. Springer 2005; Vol. 9, 1-57.

23.   Bryan, A. J.; Prasanna de Silva, A.; De Silva, S. A.; Rupasinghe, R. A. D. D.; Sandanayake, K. R. A. S., Photo-induced electron transfer as a general design logic for fluorescent molecular sensors for cations. *Biosensors* **1989,** *4*, 169-179.

24.   Valeur, B., Principles of fluorescent probe design for ion recognition. In *Probe design and chemical sensing*, Lakowicz, J. R., Ed. Plenum: 1994; Vol. 4, 21-48.

25.   Haugland, R.; Johnson, I. D., Intracellular Ion Indicators. In *Fluorescent and Luminescent Probes for Biological Activity*, Second ed.; Mason, W. T., Ed. Academic Press: 1999; 12-33.

26.   Tsien, R. Y., Monitoring Cell Calcium. In *Calcium as a Cellular Regulator*, Carafoli, E.; Klee, C., Eds. Oxford University Press: New York, 1999; 28-54.

27.   Tsien, R. Y., Fluorescent probes of cell signaling. *Annual Review of Neuroscience* **1989,** *12*, 227-253.

28.   Tsien, R. Y., New calcium indicators and buffers with high selectivity against magnesium and protons: design, synthesis, and properties of prototype structures. *Biochemistry* **1980,** *19*, 2396-2404.

29.   Grynkiewicz, G.; Poenie, M.; Tsien, R. Y., A new generation of Ca^{2+} indicators with greatly improved fluorescence properties. *Journal of Biological Chemistry* **1985,** *260*, 3440-3450.

30.   Kao, J. P. Y.; Harootunian, A. T.; Tsien, R. Y., Photochemically generated cytosolic calcium pulses and their detection by Fluo-3. *Journal of Biological Chemistry* **1989,** *264*, 8179-8184.

31.   Tsien, R. Y., A non-disruptive technique for loading calcium buffers and indicators into cells. *Nature* **1981,** *290*, 527-528.

32. Hyrc, K. L.; Bownik, J. M.; Goldberg, M. P., Ionic selectivity of low-affinity ratiometric calcium indicators: mag-Fura-2, Fura-2FF and BTC. *Cell Calcium* **2000,** *27*, 75-86.

33. Keating, G.; O'Kennedy, R., The chemistry and occurrence of coumarins. In *Coumarins: Biology, Applications and Mode of Action*, First ed.; Thornes, R. D.; O'Kennedy, R., Eds. John Wiley & Sons: 1997; 23-66.

34. Katerinopoulos, H. E., The coumarin moiety as chromophore of fluorescent ion indicators in biological systems. *Current Pharmaceutical Design* **2004,** *10*, 3835-3852.

35. Iatridou, H.; Foukaraki, E.; Kuhn, M. A.; Marcus, E. M.; Haugland, R. P.; Katerinopoulos, H. E., The development of a new family of intracellular calcium probes. *Cell Calcium* **1994,** *15*, 190-198.

36. Ito, K.; Miyashita, Y.; Kasai, H., Micromolar and submicromolar Ca^{2+} spikes regulating distinct cellular functions in pancreatic acinar cells. *Embo Journal* **1997,** *16*, 242-251.

37. Zhao, M.; Hollingworth, S.; Baylor, S. M., Properties of tri- and tetracarboxylate Ca^{2+} indicators in frog skeletal muscle fibers. *Biophysical Journal* **1996,** *70*, 896-916.

38. Papazoglou, T. G.; Iatridou, H.; Foukaraki, E.; Katerinopoulos, H. E., Detection of cardiovascular calcified deposits via tetracarboxylate ion dye (BTC) probing and laser-induced fluorescence spectroscopy. *Journal of Photochemistry and Photobiology B: Biology* **1995,** *27*, 81-84.

39. Liepouri, F.; Foukaraki, E.; Deligeorgiev, T. G.; Katerinopoulos, H. E., Iminocoumarin-based low affinity fluorescent Ca^{2+} indicators excited with visible light. *Cell Calcium* **2001,** *30*, 331-335.

40. Stryer, L., Fluorescence energy transfer as a spectroscopic ruler. *Annual Review of Biochemistry* **1978,** *47*, 819-846.

41. Clegg, R. M., Fluorescence resonance energy transfer and nucleic acids. *Methods in Enzymology* **1992,** *211*, 353-388.

42. dos Remedios, C. G.; Moens, P. D. J., Fluorescence resonance energy transfer spectroscopy is a reliable "ruler" for measuring structural changes in proteins: dispelling the problem of the unknown orientation factor. *Journal of Structural Biology* **1995,** *115*, 175-185.

43. Scholes, G. D., Long-range resonance energy transfer in molecular systems. *Annual Review of Physical Chemistry* **2003,** *54*, 57-87.

44. Selvin, P. R., Fluorescence resonance energy transfer. *Methods in Enzymology* **1995,** *246*, 300-334.

45. Bottiroli, G.; Croce, A. C.; Ramponi, R., Fluorescence resonance energy transfer imaging as a tool for in situ evaluation of cell morphofunctional characteristics. *Journal of Photochemistry and Photobiology B: Biology* **1992,** *12*, 413-416.

46. Chirio-Lebrun, M. C.; Prats, M., Fluorescence resonance energy transfer (FRET): theory and experiments. *Biochemical Education* **1998,** *26*, 320-323.

47. Lorenz, M.; Diekmann, S., Distance determination in protein-DNA complexes using fluorescence resonance energy transfer. *Methods in Molecular Biology* **2006,** *335*, 243-255.

48. Haustein, E.; Jahnz, M.; Schwille, P., Triple FRET: A tool for studying long-range molecular interactions. *ChemPhysChem* **2003,** *4*, 745-748.

49. Tsien, R. Y.; Bacskai, B. J.; Adams, S. R., FRET for studying intracellular signalling. *Trends in Cell Biology* **1993,** *3*, 242-245.

50. Haas, E.; Wilchek, M.; Katchalski-Katzir, E.; Steinberg, I. Z., Distribution of end-to-end distances of oligopeptides in solution as estimated by energy transfer. *Proceedings of the National Academy of Sciences of the United States of America* **1975,** *72*, 1807-1811.

51. Chapman, E. R.; Alexander, K.; Vorherr, T.; Carafoli, E.; Storm, D. R., Fluorescence energy transfer analysis of calmodulin-peptide complexes. *Biochemistry* **1992,** *31*, 12819-12825.

52. Parsons, M.; Vojnovic, B.; Ameer-Beg, S., Imaging protein-protein interactions in cell motility using fluorescence resonance energy transfer (FRET). *Biochemical Society Transactions* **2004,** *32*, 431-433.

53. Heyduk, T., Measuring protein conformational changes by FRET/LRET. *Current Opinion in Biotechnology* **2002,** *13*, 292-296.

54. Raicu, V.; Jansma, D. B.; Miller, R. J.; Friesen, J. D., Protein interaction quantified in vivo by spectrally resolved fluorescence resonance energy transfer. *Biochemical Journal* **2005,** *385*, 265-277.

55. Chen, Y.; Mauldin, J. P.; Day, R. N.; Periasamy, A., Characterization of spectral FRET imaging microscopy for monitoring nuclear protein interactions. *Journal of Microscopy* **2007,** *228*, 139-152.

56. Bunt, G.; Wouters, F. S., Visualization of molecular activities inside living cells with fluorescent labels. *International Review of Cytology* **2004,** *237*, 205-277.

57. Klostermeier, D.; Millar, D. P., Time-resolved fluorescence resonance energy transfer: A versatile tool for the analysis of nucleic acids. *Biopolymers* **2002,** *67*, 159-179.

58. Herbst, K. J.; Ni, Q.; Zhang, J., Dynamic visualization of signal transduction in living cells: from second messengers to kinases. *IUBMB Life* **2009,** *61*, 902-908.

59. Shimomura, O.; Johnson, F. H.; Saiga, Y., Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology* **1962,** *59*, 223-239.

60. Tsien, R. Y., The green fluorescent protein. *Annual Review of Biochemistry* **1998,** *67*, 509-544.

61. Ormö, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J., Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **1996,** *273*, 1392-1395.

62. Yang, F.; Moss, L. G.; Phillips, G. N., Jr., The molecular structure of green fluorescent protein. *Nature Biotechnology* **1996,** *14*, 1246-1251.

63. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C., Green fluorescent protein as a marker for gene expression. *Science* **1994,** *263*, 802-805.

64. Zimmer, M., Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chemical Reviews* **2002,** *102*, 759-781.

65. Miyawaki, A.; Sawano, A.; Kogure, T., Lighting up cells: Labelling proteins with fluorophores. *Nature Cell Biology* **2003**, S1-S7.

66. Devgan, V.; Thomas, M.; Ullas, K. S.; Rao, M. R.; Seshagiri, P. B., Embryo culture-based generation of enhanced green fluorescent protein-transgenic mice. *Biochemical and Biophysical Research Communications* **2003,** *303*, 994-1001.

67. Kwon, M. S.; Koo, B. C.; Choi, B. R.; Lee, H. T.; Kim, Y. H.; Ryu, W. S.; Shim, H.; Kim, J. H.; Kim, N. H.; Kim, T., Development of transgenic chickens expressing enhanced green fluorescent protein. *Biochemical and Biophysical Research Communications* **2004,** *320*, 442-448.

68. Heim, R.; Tsien, R. Y., Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current Biology* **1996,** *6*, 178-182.

69. Matz, M. V.; Lukyanov, K. A.; Lukyanov, S. A., Family of the green fluorescent protein: journey to the end of the rainbow. *Bioessays* **2002,** *24*, 953-959.

70. Shaner, N. C.; Patterson, G. H.; Davidson, M. W., Advances in fluorescent protein technology. *Journal of Cell Science* **2007,** *120*, 4247-4260.

71. Romoser, V. A.; Hinkle, P. M.; Persechini, A., Detection in living cells of Ca^{2+}-dependent changes in the fluorescence emission of an indicator composed of two green fluorescent protein variants linked by a calmodulin-binding sequence. A new class of fluorescent indicators. *Journal of Biological Chemistry* **1997,** *272*, 13270-13274.

72. Miyawaki, A.; Llopis, J.; Helm, R.; McCaffery, J. M.; Adams, J. A.; Ikura, M.; Tsien, R. Y., Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature* **1997,** *388*, 882-887.

73. Kneen, M.; Farinas, J.; Li, Y.; Verkman, A. S., Green fluorescent protein as a noninvasive intracellular pH indicator. *Biophysical Journal* **1998,** *74*, 1591-1599.

74. Truong, K.; Ikura, M., The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes in vivo. *Current Opinion in Structural Biology* **2001,** *11*, 573-578.

75. Miyawaki, A., Development of probes for cellular functions using fluorescent proteins and fluorescence resonance energy transfer. *Annual Review of Biochemistry* **2011,** *80*, 357-373.

76. Pollok, B. A.; Heim, R., Using GFP in FRET-based applications. *Trends in Cell Biology* **1999,** *9*, 57-60.

77. Griffin, B. A.; Adams, S. R.; Tsein, R. Y., Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **1998,** *281*, 269-272.

78. Griffin, B. A.; Adams, S. R.; Jones, J.; Tsien, R. Y., Fluorescent labeling of recombinant proteins in living cells with FlAsH. *Macromolecular Crystallography, Pt A* **2000,** *327*, 565-578.

79. Adams, S. R.; Tsien, R. Y., Preparation of the membrane-permeant biarsenicals FlAsH-EDT2 and ReAsH-EDT2 for fluorescent labeling of tetracysteine-tagged proteins. *Nature Protocols* **2008,** *3*, 1527-1534.

80. Suzuki, M.; Ito, Y.; Savage, H. E.; Husimi, Y.; Douglas, K. T., Intramolecular fluorescent resonance energy transfer (FRET) by BODIPY chemical modification of cysteine-engineered mutants of green fluorescent protein. *Chemistry Letters* **2003,** *32*, 306-307.

81. Zhang, W.; Suzuki, M.; Ito, Y.; Douglas, K. T., A chemically modified Green-Fluorescent Protein that responds to cleavage of an engineered disulphide bond by fluorescence resonance energy transfer (FRET)-based changes. *Chemistry Letters* **2005,** *34*, 766-767.

82. Suzuki, M.; Ito, Y.; Savage, H. E.; Husimi, Y.; Douglas, K. T., Protease-sensitive signalling by chemically engineered intramolecular fluorescent resonance energy transfer mutants of green fluorescent protein. *Biochimica et Biophysica Acta* **2004,** *1679*, 222-229.

83. Suzuki, M.; Ito, Y.; Sakata, I.; Sakai, T.; Husimi, Y.; Douglas, K. T., Caspase-3 sensitive signaling in vivo in apoptotic HeLa cells by chemically engineered intramolecular fluorescence resonance energy transfer mutants of green fluorescent protein. *Biochemical and Biophysical Research Communications* **2005,** *330*, 454-460.

84. Mitchell, F. L.; Marks, G. E.; Bichenkova, E. V.; Douglas, K. T.; Bryce, R. A., Molecular probes: insights into design and analysis from computational and physical chemistry. *Biochemical Society Transactions* **2008,** *36*, 46-50.

85. Mitchell, F. L.; Frank, F.; Marks, G. E.; Suzuki, M.; Douglas, K. T.; Bryce, R. A., Molecular dynamics study of chemically engineered green fluorescent protein mutants: comparison of intramolecular fluorescence resonance energy transfer rate. *Proteins* **2009,** *75*, 28-39.

86. Suzuki, M.; Husimi, Y.; Komatsu, H.; Suzuki, K.; Douglas, K. T., Quantum Dot FRET Biosensors that Respond to pH, to Proteolytic or Nucleolytic Cleavage, to DNA Synthesis, or to a Multiplexing Combination. *Journal of the American Chemical Society* **2008,** *130*, 5720-5725.

87. Kolb, H. C.; Finn, M. G.; Sharpless, K. B., Click chemistry: Diverse chemical function from a few good reactions. *Angewandte Chemie-International Edition* **2001,** *40*, 2004-2021.

88. Hoyle, C. E.; Lowe, A. B.; Bowman, C. N., Thiol-click chemistry: a multifaceted toolbox for small molecule and polymer synthesis. *Chemical Society Reviews* **2010,** *39*, 1355-1387.

89. Heck, R. F.; Nolley, J. P., Palladium-catalyzed vinylic hydrogen substitution reactions with aryl, benzyl, and styryl halides. *Journal of Organic Chemistry* **1972,** *37*, 2320-2322.

90. Xu, H. J.; Zhao, Y. Q.; Zhou, X. F., Palladium-Catalyzed Heck Reaction of Aryl Chlorides under Mild Conditions Promoted by Organic Ionic Bases. *Journal of Organic Chemistry* **2011,** *76*, 8036-8041.

91. Hoyle, C. E.; Bowman, C. N., Thiol-Ene Click Chemistry. *Angewandte Chemie-International Edition* **2010,** *49*, 1540-1573.

92. Chan, J. W.; Hoyle, C. E.; Lowe, A. B.; Bowman, M., Nucleophile-Initiated Thiol-Michael Reactions: Effect of Organocatalyst, Thiol, and Ene. *Macromolecules* **2010,** *43*, 6381-6388.

93. Anderson, G. W.; McGregor, A. C., tert-Butoxycarbonylamino acids and their use in peptide synthesis. *Journal of the American Chemical Society* **1957,** *79*, 6180-6183.

94. Flynn, D. L.; Zelle, R. E.; Grieco, P. A., A mild two-step method for the hydrolysis of lactams and secondary amides. *Journal of Organic Chemistry* **1983,** *48*, 2424-2426.

95. Grehn, L.; Gunnarsson, K.; Ragnarsson, U., Removal of formyl, acetyl, and benzoyl groups from amides with conversion into the corresponding tert-butyl carbamates. *Journal of the Chemical Society, Chemical Communications* **1985**, 1317-1318.

96. Bender, M. L., Mechanisms of catalysis of nucleophilic reactions of carboxylic acid derivatives. *Chemical Reviews* **1960,** *60*, 53-113.

97. Satterthwait, A. C.; Jencks, W. P., The mechanism of the aminolysis of acetate esters. *Journal of the American Chemical Society* **1974,** *96*, 7018-7031.

98. Blackburn, G. M.; Jencks, W. P., The mechanism of the aminolysis of methyl formate. *Journal of the American Chemical Society* **1968,** *90*, 2638-2645.

99. Braude, E. A.; Nachod, F. C.; Editors, *Determination of Organic Structures by Physical Methods*. 1955.

100. Smith, M. B.; March, J., *March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*. 5th ed.; Wiley: 2000.

101. Hah, J. M.; Martasek, P.; Roman, L. J.; Silverman, R. B., Aromatic Reduced Amide Bond Peptidomimetics as Selective Inhibitors of Neuronal Nitric Oxide Synthase. *Journal of Medicinal Chemistry* **2003,** *46*, 1661-1669.

102. Dannhardt, G.; Gruchalla, M.; Kohl, B. K.; Parsons, C. G., A novel series of 2-carboxytetrahydroquinolines provides new insights into the eastern region of glycine site NMDA antagonists. *Archiv der Pharmazie (Weinheim, Germany)* **2000,** *333*, 267-274.

103. Lee, J.; Lee, J.; Kang, M.; Shin, M.; Kim, J. M.; Kang, S. U.; Lim, J. O.; Choi, H. K.; Suh, Y. G.; Park, H. G.; Oh, U.; Kim, H. D.; Park, Y. H.; Ha, H. J.; Kim, Y. H.; Toth, A.; Wang, Y.; Tran, R.; Pearce, L. V.; Lundberg, D. J.; Blumberg, P. M., N-(3-Acyloxy-2-benzylpropyl)-N'-[4-(methylsulfonylamino)benzyl]thiourea Analogues: Novel Potent and High Affinity Antagonists and Partial Antagonists of the Vanilloid Receptor. *Journal of Medicinal Chemistry* **2003,** *46*, 3116-3126.

104. Campaigne, E., Thiones and thials. *Chemical Reviews* **1946,** *39*, 1-77.

105. Jesberger, M.; Davis, T. P.; Barner, L., Applications of Lawesson's reagent in organic and organometallic syntheses. *Synthesis* **2003**, 1929-1958.

106. Ozturk, T.; Ertas, E.; Mert, O., Use of Lawesson's Reagent in Organic Syntheses. *Chemical Reviews* **2007,** *107*, 5210-5278.

107. Cava, M. P.; Levinson, M. I., Thionation reactions of Lawesson's reagents. *Tetrahedron* **1985,** *41*, 5061-5087.

108. Pedersen, B. S.; Scheibye, S.; Clausen, K.; Lawesson, S. O., Studies on organophosphorus compounds. XXII. The dimer of p-methoxyphenyl-thionophosphine sulfide as thiation reagent. A new route to O-substituted thioesters and dithioesters. *Bulletin des Societes Chimiques Belges* **1978,** *87*, 293-297.

109. Pedersen, B. S.; Scheibye, S.; Nilsson, N. H.; Lawesson, S. O., Studies on organophosphorus compounds. XX. Syntheses of thioketones. *Bulletin des Societes Chimiques Belges* **1978,** *87*, 223-228.

110. Scheibye, S.; Pedersen, B. S.; Lawesson, S. O., Studies on organophosphorus compounds. XXIII. Synthesis of salicylthioamides and 4H-1,3,2-benzoxaza-phosphorine derivatives from the dimer of p-methoxyphenylthionophosphine sulfide and salicylamides. *Bulletin des Societes Chimiques Belges* **1978,** *87*, 299-306.

111. Cherkasov, R. A.; Kutyrev, G. A.; Pudovik, A. N., Organothiophosphorus reagents in organic synthesis. *Tetrahedron* **1985,** *41*, 2567-2624.

112. Rauchfuss, T. B.; Zank, G. A., Mechanistic studies on the thiation of carbonyls by Lawesson's reagent: the role of a 3-coordinate phosphorus(V) species. *Tetrahedron Letters* **1986,** *27*, 3445-3448.

113. Scheibye, S.; Pedersen, B. S.; Lawesson, S. O., Studies on organophosphorus compounds. XXI. The dimer of p-methoxyphenylthionophosphine sulfide as thiation reagent. A new route to thiocarboxamides. *Bulletin des Societes Chimiques Belges* **1978,** *87*, 229-238.

114. Oshida, H.; Ishii, A.; Nakayama, J., First isolation of 1,3,2-dithiaphosphetane 2-sulfide. *Tetrahedron Letters* **2004,** *45*, 1331-1334.

115. Emsley, J.; Hall, D., *The Chemistry of Phosphorus*. 1976.

116. Csomos, P.; Fodor, L.; Mandity, I.; Bernath, G., An efficient route for the synthesis of 2-arylthiazino[5,6-b]indole derivatives. *Tetrahedron* **2007,** *63*, 4983-4989.

117. Klyne, W.; Prelog, V., Description of steric relationships across single bonds *Experientia* **1960,** *16*, 521-523.

118. Cahn, R. S.; Ingold, C. K.; Prelog, V., The specification of asymmetric configuration in organic chemistry. *Experientia* **1956,** *12*, 81-94.

119. Rogers, M. T.; Woodbrey, J. C., A proton magnetic resonance study of hindered internal rotation in some substituted N,N-dimethylamides *Journal of Physical Chemistry* **1962,** *66*, 540-546.

120. Neuman, R. C.; Young, L. B., Hindered Rotation and Carbon-13-Hydrogen Coupling Constants in Amides, Thioamides, and Amidines. *Journal of Physical Chemistry* **1965,** *69*, 2570-2576.

121. Sandstrom, J., Barriers to internal rotation in thioamides. Experimental results and molecular orbital calculations. *Journal of Physical Chemistry* **1967,** *71*, 2318-2325.

122. Stewart, W. E.; Siddall, T. H., Nuclear Magnetic Resonance studies of amides. *Chemical Reviews* **1970,** *70*, 517-551.

123. Wiberg, K. B., The interaction of carbonyl groups with substituents. *Accounts of Chemical Research* **1999,** *32*, 922-929.

124. Wiberg, K. B.; Rush, D. J., Solvent effects on the thioamide rotational barrier: An experimental and theoretical study. *Journal of the American Chemical Society* **2001,** *123*, 2038-2046.

125. Wiberg, K. B.; Rablen, P. R., Why does thioformamide have a larger rotational barrier than formamide? *Journal of the American Chemical Society* **1995,** *117*, 2201-2209.

126. Rao, C. N. R.; Rao, K. G.; Goel, A.; Balasubramanian, D., Configuration of secondary amides and thioamides: spectroscopic and theoretical studies. *Journal of the Chemical Society A: Inorganic, Physical, Theoretical* **1971,** 3077-3083.

127. Varma, R. S.; Kumar, D., Microwave-Accelerated Solvent-Free Synthesis of Thioketones, Thiolactones, Thioamides, Thionoesters, and Thioflavonoids. *Organic Letters* **1999,** *1*, 697-700.

128. Mase, T.; Itoh, T., General and practical synthesis of benzothiazoles. *Pure and Applied Chemistry* **2008,** *80*, 707-715.

129. Malik, J. K.; Manvi, F. V.; Nanjwade, B. K.; Singh, S.; Purohit, P., Review of the 2-Amino Substituted Benzothiazoles: Different Methods of the Synthesis. *Der Pharmacia Lettre* **2010,** *2*, 347-359.

130. Bondock, S.; Fadaly, W.; Metwally, M. A., Recent trends in the chemistry of 2-aminobenzothiazoles. *Journal of Sulfur Chemistry* **2009,** *30*, 74-107.

131. Gupta, A.; Rawat, S., Synthesis and Cyclization of Benzothiazole: Review. *Journal of Current Pharmaceutical Research* **2010,** *3*, 13-23.

132. El Meligy, M. S. A.; Mohamed, S. A., Plumbophosphates as oxidizing agents in the preparation of benzimidazoles, benzothiazoles, and benzoxazoles. *Journal für Praktische Chemie* **1974,** *316*, 154-158.

133. Alajarín, M.; Vidal, A.; Tovar, F.; Jones, P. G.; Bautistac, D., A novel benzothiazole synthesis by cyclization of ketenimines bearing sulfenylimine fragments. Unexpected sulfur to carbon migration of an imino group. *Archive for Organic Chemistry* **2005**, 35-50.

134. Jackson, Y. A.; Lyon, M. A.; Townsend, N.; Bellabe, K.; Soltanik, F., Reactions of some N-(2,5-dimethoxyaryl)thiobenzamides: en route to an analogue of kuanoniamine A. *Journal of the Chemical Society, Perkin Transactions 1* **2000**, 205-210.

135. Hutchinson, I.; Chua, M. S.; Browne, H. L.; Trapani, V.; Bradshaw, T. D.; Westwell, A. D.; Stevens, M. F. G., Antitumor Benzothiazoles. 14. Synthesis and in Vitro Biological Properties of Fluorinated 2-(4-Aminophenyl)benzothiazoles. *Journal of Medicinal Chemistry* **2001,** *44*, 1446-1455.

136. Downer, N. K.; Jackson, Y. A., Synthesis of benzothiazoles via ipso substitution of ortho-methoxythiobenzamides. *Organic & Biomolecular Chemistry* **2004,** *2*, 3039-3043.

137. Mortimer, C. G.; Wells, G.; Crochard, J. P.; Stone, E. L.; Bradshaw, T. D.; Stevens, M. F. G.; Westwell, A. D., Antitumor benzothiazoles. 26. 2-(3,4-Dimethoxyphenyl)-5-fluorobenzothiazole (GW 610, NSC 721648), a simple fluorinated 2-arylbenzothiazole, shows potent and selective inhibitory activity against lung, colon, and breast cancer cell lines. *Journal of Medicinal Chemistry* **2006,** *49*, 179-185.

138. Jacobson, P., Ueber Bildung von Anhydroverbindungen des Orthoamidophenylmercaptans aus Thioanilides. *Berichte der Deutschen Chemischen Gesellschaft* **1886,** *19* 1067-1077.

139. Metzger, J., Thiazoles and their Benzo Derivates. In *Comprehensive Heterocyclic Chemistry: the structure, reactions, synthesis and uses of heterocyclic compounds*, Katritzsky, A. R.; Ress, C. W., Eds. Pergamon Press: Oxford, 1984; Vol. 6.

140. Metzger, J.; Planck, H., Étude dans la série des azoles. VIII (I) Nouvelles synthéses dans la série du benzothiazole. *Bulletin de la Societe Chimique de France* **1956**, 684-689.

141. Stevens, M. F. G.; McCall, C. J.; Lelievald, P.; Alexander, P.; Richter, A.; Davies, D. E., Structural Studies on Bioactive Compounds. 23. Synthesis of Polyhydroxylated 2-Phenylbenzothiazoles and a Comparison of their Cytotoxicities and Pharmacological Properties with Genistein and Quercetin. *Journal of Medicinal Chemistry* **1994,** *37*, 1689-1695.

142. Collum, D. V.; McNeil, A. J.; Ramirez, A., Lithium diisopropylamide: solution kinetics and implications for organic synthesis. *Angewandte Chemie-International Edition* **2007,** *46*, 3002-3017.

143. Wakefield, B. J., *The Chemistry of Organolithium Compounds*. 1974.

144. Williard, P. G.; Salvino, J. M., Synthesis, isolation, and structure of an LDA-THF complex. *Journal of Organic Chemistry* **1993,** *58*, 1-3.

145. Sharghi, H.; Asemani, O., Methanesulfonic Acid/SiO2 as an Efficient Combination for the Synthesis of 2-Substituted Aromatic and Aliphatic Benzothiazoles from Carboxylic Acids. *Synthetic Communications* **2009,** *39*, 860-867.

146. Rudrawar, S.; Kondaskar, A.; Chakraborti, A. K., An efficient acid- and metal-free one-pot synthesis of benzothiazoles from carboxylic acids. *Synthesis* **2005**, 2521-2526.

147. Bogert, M. T.; Naiman, B., Researches on thiazoles XX The synthesis of benzothiazoles from aldehydes and ortho-aminothiophenol - The action of aldehydes upon zinc ortho-aminothiophenolate and upon related thionated aromatic amines. *Journal of the American Chemical Society* **1935,** *57*, 1529-1533.

148. Mosslemin, M. H.; Fazlinia, A., An Efficient Procedure for the Synthesis of Benzoxazole and Benzothiazole Derivatives Using a H2O2/SiO2-FeCl3 System. *Phosphorus Sulfur and Silicon and the Related Elements* **2010,** *185*, 2165-2170.

149. Abayeh, O. J.; Olagbemiro, T. O.; Agho, M. O.; Amupitan, J. O., Facile Synthesis of 2-substituted benzothiazoles via ketenes *Bulletin des Societes Chimiques Belges* **1994,** *103*, 687-689.

150. Chao, R. Y.; Ding, M.-F.; Chen, J.-Y.; Lee, C.-C.; Lin, S.-T., Preparation and Characterization of Substituted 3-Benzothiazol-2-ylcoumarins. *Journal of the Chinese Chemical Society* **2010,** *57*, 213-221.

151. Raulin, F.; Lussiana, J.-P., Prebiotic formation of iminothioesters. II: Addition of thiophenols to malonic nitriles. *Origins of Life and Evolution of Biospheres* **1984,** *14*, 157-162.

152. Saito, K.; Kambe, S.; Nakano, Y.; Sakurai, A.; Midorikawa, H., Synthetic studies using alpha-beta-unsaturated nitriles: A convenient preparation of 1,3-benzothiazole Derivatives. *Synthesis* **1983**, 210-212.

153. Elgemeie, G. H.; Shams, H. Z.; Elkholy, Y. M.; Abbas, N. S., Novel synthesis of pyrido [2,1-b] benzothiazoles and 1,3-benzothiazole derivatives. *Phosphorus Sulfur and Silicon and the Related Elements* **2000,** *165*, 265-272.

154. Landor, S. R.; Landor, P. D.; Tanee Fomum, Z.; Tanyi Mbafor, J.; Mpango, G., W. B., Allenes-41 : The addition of thiols to allenyl- and phenylpropynyl- nitrile and the formation of thiazolines and benzothiazoles. *Tetrahedron* **1984,** *40*, 2141-2149.

155. Baldwin, J. E., Rules for ring closure. *Journal of the Chemical Society, Chemical Communications* **1976**, 734-736.

156. Knoevenagel, E., Condensation von Malonsäure mit aromatiachen Aldehyden durch Ammoniak und Amine. *Berichte der Deutschen Chemischen Gesellschaft* **1898,** *31*, 2596-2619.

157. Bogdal, D., Coumarins: Fast synthesis by Knoevenagel condensation under microwave irradiation. *Journal of Chemical Research* **1998**, 468-469.

158. Shockravi, A.; Shargi, H.; Valizadeh, H.; Heravi, M. M., Solvent Free Synthesis of Coumarins. *Phosphorus Sulfur and Silicon and the Related Elements* **2002,** *177*, 2555-2559.

159. Gardner, P. D.; Brandon, R. L., Preparation and reactions of some aralkyl cyanoacetic esters. *Journal of Organic Chemistry* **1957,** *22*, 1704-1705.

160. Sugino, T.; Tanaka, K., Solvent-free coumarin synthesis. *Chemistry Letters* **2001**, 110-111.

161. Volmajer, J.; Toplak, R.; Leban, I.; Le Marechal, A. M., Synthesis of new iminocoumarins and their transformations into N-chloro and hydrazono compounds. *Tetrahedron* **2005,** *61*, 7012-7021.

162. Benmaktouf, L.; Ammar, H.; Le Bigot, Y.; Abid, S., Synthesis of New Iminocoumarins Bearing Parabanic Moieties. *Synthetic Communications* **2011,** *41*, 1017-1026.

163. Still, W. C.; Kahn, M.; Mitra, A., Rapid chromatographic technique for preparative separations with moderate resolution. *Journal of Organic Chemistry* **1978,** *43*, 2923-2925.

Summary and closing remarks

Molecular recognition is the foundation for the two research projects presented in this thesis: analysis of protein-ligand interactions, and fluorescent indicators design.

In the first part of this thesis, the structural basis of molecular recognition in protein-ligand complexes was studied. Specifically, as a step towards improved understanding of PI3K binding preferences, we examined the basis on which the γ isoform of phosphoinositide 3-kinase (PI3Kγ) distinguishes selective inhibitors **1** and **2**, with a ~30-fold preference for the former. Interestingly, despite the chemical similarity of the two ligands, the X-ray structures for their PI3Kγ complexes exhibit the molecules in different conformers, *s*-cis for inhibitor **1** and *s*-trans for inhibitor **2**. Consequently, our first goal was to determine the conformational preference of inhibitors **1** and **2**. In chapter 3, firstly, we investigated computationally the intrinsic conformational preference of the isolated ligands, and found that for inhibitors **1** and **2** the *s*-cis and *s*-trans conformations are possible from the energetic point of view. Subsequently, we revisited the available crystallographic data and found that for inhibitor **1** the *s*-cis conformation possesses a much higher occupancy (87%) than the originally modelled *s*-trans isomer (13%). For inhibitor **2**, the evidence provided by the electron density maps suggests the *s*-cis conformation is the preferred one. Additionally, the process of manual rebuilding and re-refinement of the PI3Kγ structure resulted in a protein model with improved quality metrics, including R factor, R_{free}, and Ramachandran scores.

Modelling protein plasticity during the ligand binding process is essential to understanding the mechanism of the process. To provide a dynamical insight into the subtle basis of molecular recognition by PI3Kγ, in chapter 4, we performed 140 ns MD simulations of the four PI3Kγ complexes in explicit solvent. This reveals similar conformational flexibility at the active site for all systems. Throughout the four different MD simulations, the hydrogen bond interactions to Lys833, Asp841 and Val882 are maintained very well. Moreover, the cavity lined by the hydrophobic residues is well conserved during the trajectories. Yet, the conformations sampled by the *s*-cis isomers are more consistent with the conformations reported by the X-ray crystal structures. One of the most interesting observations from the MD simulations is the movement of the DFG aspartate Asp964 (located at the beginning of the activation

loop). The available crystallographic evidence supporting this movement has been discussed.

Subsequent energetic analysis was performed incorporating ensemble-averaging and desolvation effects via the Poisson-Boltzmann continuum solvent model (chapter 5). For both inhibitors the *s*-cis isomers are predicted to be the most favourable conformations. Furthermore, *s*-cis inhibitor **1** is predicted to be a stronger binder than *s*-cis inhibitor **2**. The difference in their predicted binding free energies ($\Delta\Delta G_{bind}$) was -2.0 kcal·mol^{-1}, which is consistent with the trend observed experimentally. From the results of the per-residue decomposition analysis it was possible to identify the key residues whose contributions differ the most when comparing the different protein-ligand systems. When comparing the energetic analysis of the PI3Kγ/*s*-cis inhibitor **1** system *versus* the PI3Kγ/*s*-cis inhibitor **2** system, it appears that residues Glu880, Ile881 and Val882 are responsible for the higher affinity of PI3Kγ for *s*-cis inhibitor **1** over *s*-cis inhibitor **2**.

The crucial ligand-Val882 interaction of the PI3Kγ/*s*-cis inhibitor **1** complex and the PI3Kγ/*s*-cis inhibitor **2** complex was examined utilising molecular mechanical and quantum chemical treatments. The results suggests that the *ff99sb* force field underestimates the relative strength of the quinoxaline N···HN(Val882) hydrogen bond of the *s*-cis inhibitor **1** by 0.3 kcal·mol^{-1}; and overestimates the strength of the 1,3-benzodioxole ring O···HN(Val882) hydrogen bond of the *s*-cis inhibitor **2** by 0.7 kcal·mol^{-1}. Therefore, we correct the gas-phase protein-ligand interaction to account for inaccuracies of the force field in modelling the ligand-Val882 hydrogen bond. This correction leads to a preference in total binding energy of 2.9 kcal·mol^{-1} for *s*-cis inhibitor **1**. This study therefore provides structural, dynamical and energetic insights into the subtle basis of molecular recognition by PI3Kγ.

In the second half of this thesis we consider the application of molecular recognition to fluorescent sensor design. Fluorescent probes have evolved into an extremely useful tool for the detection of calcium in biological systems. Benzothiazole derivatives BTC, and its iminocoumarin analogue BTIC, are two low affinity calcium indicators featuring many desirable properties for cellular calcium measurement. In an effort to produce fluorophores that can be chemically conjugated with a screening protein, such as green fluorescent protein (GFP), we aimed to derivatise BTC and BTIC. Two synthetic approaches towards the synthesis of these potential fluorescent probes were outlined.

Firstly, a synthetic route was proposed for the synthesis of potential fluorescent probe **11**, a derivative of BTC. The proposed synthesis involved the condensation of esterified benzothiazole **16** with the versatile precursor salicylaldehyde **15** by a Knoevenagel-type reaction. A five-step synthesis of **16** was designed commencing with commercially available 4-aminophenethylamine. The first four transformations in this five-step synthesis were successfully completed. Unfortunately, however, the difficulties in the last step impeded the successful completion of the synthetic route leading to benzothiazole **16**. This in turn hindered the total synthesis of BTC's derivative **11**.

Secondly, a synthetic route was proposed for the synthesis of potential fluorescent probe **12**, a derivative of BTIC. The proposed synthesis involved the condensation of benzothiazole **20** with the versatile precursor salicylaldehyde **15** by a Knoevenagel-type reaction. Benzothiazole **20** was successfully synthesised in a one-step synthesis. It was then successfully condensed with salicylaldehyde **15**, to produce compound **21**. Deprotection of the tetra ^{t}butyl esters would furnish potential fluorescent probe **12**. The spectral properties and binding affinities for Ca^{2+} could then be determined. However, a key issue that has been made obvious by the studies of this thesis is the unexpected (light-induced) instability of **21**. This in reality precludes its eventual use as a GFP-bound probe and a different fluorophore skeleton should be sought.

Overall, a complex and many-step synthesis was achieved, and indicated that this particular framework (compound **21**) is not suitable for a fluorophore to link to proteins (e.g. GFP mutant uv5GFPC48I229), surfaces (e.g. 96-well plate formats) or other media (e.g. nuclei acids). This however, could not have been assessed in any other way than a trial synthesis. In this respect, this research removes what to all appearances was a likely candidate.