



**UNIVERSITÉ
DE LORRAINE**

**BIBLIOTHÈQUES
UNIVERSITAIRES**

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Simulations atomistiques de la réaction d'acétylation d'amines et de l'inhibition covalente de l'enzyme Phosphoinositide 3-kinase (PI3K)

*Atomistic simulations of the amine acetylation
reaction and the covalent inhibition of the
enzyme Phosphoinositide 3-kinase (PI3K)*

THÈSE

présentée et soutenue publiquement le 14 decembre 2022

pour l'obtention du

Doctorat de l'Université de Lorraine

(Mention Chimie)

par

Volkan Fındık

Composition du jury

Président : Prof. Fethiye Aylin Sungur, Istanbul Technical University, Istanbul, Turquie

Rapporteurs : Prof. Fethiye Aylin Sungur, Istanbul Technical University, Istanbul, Turquie
Dr. Hélène Jamet, Université Grenoble Alpes, Grenoble, France

Examineurs : Dr. Alessandro Genoni, Université de Lorraine, Nancy, France
Prof. Ozkan Danis, Marmara University, Istanbul, Turquie

Encadrants : Dr. Manuel F. Ruiz-Lopez, Université de Lorraine, Nancy, France
Prof. Safiye Erdem, Marmara University, Istanbul, Turquie

Mis en page avec la classe thesul.

Contents

Acknowledgments	v
Abbreviations	vii
Résumé	1
1 Introduction	7
1.1 Amine Acetylation and Ester Aminolysis Reactions from Organic Chemistry Perspective	7
1.2 Amine Acetylation and Ester Aminolysis Reactions in Biochemistry	8
1.3 Targeting Covalent Inhibitors (TCIs)	10
1.4 Phosphoinositide 3-Kinase (PI3K) Enzyme	15
2 Objective and scope	19
3 Theoretical Background	23
3.1 Quantum Chemical Methods	23
3.1.1 The Schrödinger Equation	23
3.1.2 Hartree-Fock Theory	24
3.1.3 The Post-HF Methods	25
3.1.4 Semi Empirical Quantum Mechanics	25
3.1.5 Density Functional Theory	26
3.1.6 Basis Sets	27
3.2 Solvation Models	27
3.3 Molecular Mechanics	28
3.4 Molecular Dynamics	30
3.5 The QM/MM Method	31
3.6 Free Energy Calculation Techniques	32

3.6.1	Umbrella Sampling	32
3.6.2	Dual-level Approach	33
4	Ester aminolysis mechanism of model systems via quantum mechanical approach	35
4.1	Introduction	35
4.2	Computational Details	35
4.3	Results and Discussion	37
4.4	Conclusion	46
5	Ester aminolysis mechanism of model systems via QM/MM molecular dynamics simulations	47
5.1	Introduction	47
5.2	Computational Details	48
5.3	Results and Discussion	48
5.3.1	Benchmark calculations	48
5.3.2	Study of the reaction between FPA or PNPA with methylamine	50
5.4	Conclusion	57
6	Classical molecular dynamics simulations of the PI3Kδ enzyme	59
6.1	Introduction	59
6.2	Computational Details	59
6.3	Results and Discussion	61
6.4	Conclusion	65
7	Lysine-targeted covalent inhibition mechanism of PI3Kδ via ONIOM QM:QM approach	69
7.1	Introduction	69
7.2	Computational Details	71
7.3	Results and Discussion	74
7.3.1	Formation of an activated reactant complex	74
7.3.2	Reaction Mechanism: Covalent vs Noncovalent Inhibitors	76
7.3.3	Stereochemical Aspects of Tetrahedral Zwitterion Intermediate	78
7.3.4	Reaction Mechanism: Most Favorable Path Involving Covalent Inhibitors	82
7.3.5	Reaction Mechanism: Electron-Withdrawing vs Electron- Donating Covalent Inhibitors	83

7.4	Conclusion	89
8	Lysine-targeted covalent inhibition mechanism of PI3Kδ via dual-level QM/MM molecular dynamics simulations	95
8.1	Introduction	95
8.2	Computational Details	96
8.3	Results and Discussion	97
8.3.1	Free Energy Profiles via QM/MM MD Simulations	97
8.3.2	Energy Refinement with FEP Correction	103
8.4	Conclusion	104
	Concluding Remarks	107
	Appendix	125

Acknowledgments

I express my deepest gratitude to my thesis supervisors Prof. Safiye Sağ Erdem and Prof. Manuel F. Ruiz-López for their scientific guidance, attention, support and understanding. I am eternally grateful to Safiye Sağ Erdem for the endless support and guidance she has provided me in achieving my academic goals. I am extremely grateful to Manuel F. Ruiz-lópez for giving me the opportunity to work in a highly motivated environment and providing invaluable guidance throughout my PhD studies in Nancy. This will always be an honor for me doing my PhD under supervision of him.

I wish to extend my appreciation to my committee members, Assoc. Prof. Alessandro Genoni, Prof. Özkan Daniş, Prof. F. Aylin Sungur and Assist. Prof. Helène Jamet, thank you for your time, scientific guidance and helpful suggestions. This dissertation is a “cotutelle” thesis between Lorraine University and Marmara University based on a collaborative project in French-Turkish PHC Bosphorous program, supported by Project number 44738WL and TUBITAK-PIA Project no: 119N133. The numerical calculations reported here were partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources). I would like to thank the French Embassy in Ankara, Turkey for “cotutelle PhD scholarship” and TUBITAK for the project scholarship.

I wish to thank all the members of Chemistry Department in Marmara University, who have been with me since the day I started working as a research assistant, for accompanying me on this PhD journey. I would especially like to thank Assoc. Prof. Burcu Oktay, Prof. Emrah Çakmakçı, Assist. Prof. Hatice Birtane and Assist. Prof. Efe Baturhan Orman for making the coffee breaks more enjoyable. I would like to thank the members of the Laboratoire de Physique and Chimie Théoriques at Université de Lorraine for the warm atmosphere. I wish to thank Dr. Marilia Martins-Costa for her endless patience and friendship. Moreover, I would like to thank Prof. Antonio Monari and my office mate Dr. Raul Losantos for scientific discussions, friendship and making our days enjoyable in quarantine in Nancy due to the pandemic. I would also like to thank my beloved office mate Mohamed Houssein for his friendship.

I express my gratitude to my beloved mentor Prof. Viktorya Aviyente, who has supported me throughout my academic life until today and whose support I have always felt behind me. Moreover, I would like to thank Dr. Gamze Tanrıver, Dr. Deniz Akgül, Dr. Z. Pinar Haşlak and İpek Munar from Boğaziçi University for being with me throughout my PhD journey. I would like to extend my gratitude to my dear friends, Hakan Arpacı and Ali Günaydin, for their invaluable support, friendship. I am very lucky to have you as my close friends.

Last but not least, my sincere gratitude goes to my whole family, especially to my wife Başak Fındık, who endured all my hardships, motivated me every day and never complained for a second, and to my fluffy friend Loki, my "white leon". I dedicate my thesis to both of you.

Abbreviations

ω B97xD	Long-Range Corrected Hybrid Functional with Dispersion Correction
3D	Three Dimensional
Ac-CoA	Acetyl-Coenzyme A
AKT	Protein Kinase B
AL	Alkoxide
AL ⁻	Alkoxide Anion
AM	Amide
AM1	Austin Model 1
AM ⁺ _N	N-Protonated Amide
AM ⁺ _O	O-Protonated Amide
Asp	Aspartic Acid
ATP	Adenosine Triphosphate
aug	Augmented
B3LYP	Becke-3-Parameter Lee-Yang-Par functional
BTK	Bruton's Tyrosine Kinase
cc-PVTZ	correlation consistent polarized valence triple zeta
CNDO	Complete Neglect of Differential Overlap
Cys	Cysteine
D3	Grimme D3 correction
DFT	Density Functional Theory
DFTB2	Density Functional Based Tight Binding 2

DFTB3	Density Functional Based Tight Binding 3
DFTBA	Density Functional Based Tight Binding with Analytical Expressions
DNA	Deoxyribonucleic Acid
EDG	Electron Donating Group
ESI	Electronic Supporting Information
EWG	Electron Withdrawing Group
FDA	US Food and Drug Administration
FEP	Free Energy Perturbation
ff14SB	Force Field 14 Stony Brook
FPA	<i>p</i> -F-phenylacetate
GAFF	Generalized Amber Force Field
GGA	Generalized Gradient Approximation
Glu	Glutamic Acid
GPCR	G-Protein Coupled Receptors
GTO	Gaussian-Type Orbitals
HAT	Histone Acetyl Transferase
HF	Hartree-Fock Theory
HL	High Level
Ile	Isoleucine
INDO	Intermediate Neglect of Differential Overlap
IO	Ionic Intermediate
IRC	Intrinsic Reaction Coordinate
K _a	Acid Dissociation Constant
k _{inact}	Rate of Inactivation
K _I	Binding Constant
KAT	Lysine Acetyl Transferase
KDAC	Lysine Deacetylase

KS	Kohn-Sham
LC	Long-range Corrected
LCAO	Linear Combination of Atomic Orbitals
LCMS	Liquid Chromatography Mass Spectrometry
LDA	Local Density Approximation
Leu	Leucine
LL	Low Level
LSDA	Local Spin Density Approximation
Lys	Lysine
M062X	Global Hybrid functional with 54% HF Exchange
MA	Methyl Acetate
MD	Molecular Dynamics
MM/GBSA	Molecular Mechanics with Generalized Born and Surface Area
MNDO	Modified Neglect of Diatomic Overlap
MP2	Møller–Plesset Perturbation Theory Second
NAT	N-Terminal Acetyltransferase
NDDO	Neglect of Diatomic Differential Overlap
NMR	Nuclear Magnetic Resonance
NPT	Isothermal Isobaric Ensemble
NT	Neutral Intermediate
NVT	Canonical Ensemble
ONIOM	Own N-layer Integrated molecular Orbital Molecular Mechanics
OPLS	Optimized Potentials for Liquid Simulations Force Field
PA	Phenyl Acetate
PBC	Periodic Boundary Conditions
PC	Product Complex
PCM	Polarizable Continuum Model

PDB	Protein Data Bank
PDDG	Pairwise Distance Directed Gaussian
PDK1	Pyruvate Dehydrogenase Lipoamide Kinase Isozyme 1
PES	Potential Energy Surface
pH	Potential of Hydrogen
Phe	Phenyl Alanine
PI3K	Phosphoinositide 3-Kinase
PIP2	Phosphatidylinositol-4,5-Bisphosphate
PIP3	Phosphatidylinositol-3,4,5-Triphosphate
PM3	Parametric Method 3
PM3MM	Parametric Method 3 with Molecular Mechanic Correction
PM6	Parametrization Method 6
PMF	Potential Mean Force
PNPA	<i>para</i> -nitrophenyl Acetate
PPI	Protein Pump Inhibitor
PTEN	Phosphatase and Tensin Homolog
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
QSAR	Quantitative Structure-Activity Relationship
RBD	Receptor Binding Domain
RC	Reactant Complex
RDF	Radial Distribution Function
RESP	Restricted Electrostatic Potential
Rg	Radius of Gyration
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RTK	Receptor Tyrosine Kinase

SCF	Self-consistent field
Ser	Serine
STO	Slater-Type Orbitals
$T^{\pm}$	Zwitterionic Tetrahedral Intermediate
T^{-}	Anionic Tetrahedral Intermediate
T^0	Neutral Tetrahedral Intermediate
TCI	Targeting Covalent Inhibitors
TIP3P	Transferable Intermolecular Potential with 3 Points
TS	Transition State
US	Umbrella Sampling
vdW	van der Waals
WHAM	Weighted Histogram Analysis Method
XC	Exchange-Correlation
ZW	Zwitterionic Intermediate

Résumé

Contexte et Objectifs

Les inhibiteurs covalents ciblés (ou TCI de l'anglais *Targeted Covalent Inhibitors*) sont très prometteurs pour la recherche de nouveaux médicaments. Les inhibiteurs de kinases sont les exemples les plus courants de TCI. Les enzymes kinases catalysent le transfert d'un groupe phosphate de l'ATP à leurs substrats. Elles fonctionnent comme des biomolécules de signalisation et sont impliquées dans une grande variété de maladies telles que le cancer et les maladies dégénératives, cardiovasculaires, immunologiques, métaboliques, inflammatoires ou infectieuses.[1] Ainsi, la modulation des fonctions des kinases offre de larges opportunités thérapeutiques.

Ces inhibiteurs offrent un certain nombre d'avantages potentiels par rapport aux inhibiteurs réversibles traditionnels, tels qu'un temps de résidence prolongé, une puissance accrue et la possibilité d'effectuer des modifications pour un design efficace.[1, 2] La plupart des études ciblent généralement les résidus cystéine (Cys) en formant un adduit covalent qui rend l'enzyme inactive. [3] Cependant, les mutations des chaînes latérales Cys sont un mécanisme de résistance courant dans le cancer. C'est pourquoi, au cours des dernières années, les résidus lysine (Lys) ont suscité un intérêt croissant en tant qu'alternative pour l'inhibition covalente ciblée.[4, 5] La Lys est largement présente dans de nombreux sites fonctionnels, notamment les sites actifs des enzymes et les interfaces protéine-protéine.[1, 2, 5] Elle offre donc la diversité qui fait défaut à la Cys. En outre, une mutation ponctuelle ne peut pas se produire sur un résidu Lys fonctionnel. Cependant, la formation d'une liaison covalente avec la Lys est en principe peu favorable du fait qu'elle est protonée et chargée donc positivement à un pH physiologique de 7.4. Récemment, Liu *et al.* (2019)[6] ont cherché à savoir si les lysines catalytiques nucléophiles sont communes dans le kinome humain. Ils ont effectué des simulations de dynamique moléculaire à pH constant continu (CpHMD) pour calculer les valeurs de pK_a et ont identifié plusieurs lysines catalytiques réactives dans les kinases humaines. Les lysines catalytiques des kinases sont connues pour stabiliser les groupes phosphates chargés négativement de l'ATP[7] et les inhibiteurs de kinases entrent généralement en compétition avec l'ATP pour se lier au site actif.[8]

Les enzymes phosphoinositide 3-kinase (PI3K) sont des cibles importantes pour les médicaments en oncologie car elles sont impliquées dans la voie de signalisation pour de nombreuses fonctions cellulaires telles que le contrôle de la croissance, le métabolisme et l'initiation de la traduction. L'inhibition de différentes isoformes de PI3K telles que α , β , γ et δ par de petites molécules est un domaine de recherche actif pour le traitement de divers cancers. Plusieurs inhibiteurs de la PI3K font actuellement l'objet d'essais cliniques pour le traitement de la leucémie lymphocytaire, du lymphome folliculaire, du cancer du sein, de la thyroïde, du can-

cer colorectal et du cancer du poumon[9, 10, 11] ainsi que pour le traitement d'affections inflammatoires telles que l'asthme, la bronchopneumopathie chronique obstructive et la polyarthrite rhumatoïde.[1, 12] L'antibiotique fongique wortmannine, ses dérivés[13] et plusieurs autres composés sont connus pour former des adduits covalents avec les lysines catalytiques des enzymes PI3K, mais ils sont peu sélectifs.[14] Récemment, Dalton *et al.*[12] ont développé les premiers inhibiteurs sélectifs et irréversibles de l'enzyme phosphoinositide 3-kinase delta (PI3K δ) qui ciblent la Lys catalytique (Figure 1). En substituant le groupe sulfonamide dans la structure d'un médicament clinique connu 1 (GSK2292767), ils ont conçu de nouveaux inhibiteurs de PI3K δ ayant des groupes esters comme tête électrophile (Figure 1) qui ciblent la Lys catalytique conservée Lys779. L'analyse par chromatographie liquide-spectrométrie de masse (LCMS) a révélé que les composés 2-7 inactiveraient PI3K δ de manière covalente par la formation d'une liaison amide avec Lys779. De plus, la structure cristalline de PI3K δ inhibée par le dérivé *p*-F 4 obtenue par rayons X a clairement confirmé la présence d'un adduit lié de manière covalente. En revanche, un complexe non-covalent enzyme-inhibiteur a été observé par rayons X pour l'ester méthylique 8. D'autres travaux de Fournier *et al.*[15] ont confirmé que la simple acétylation de Lys779 peut inhiber sélectivement l'activité de PI3K δ .

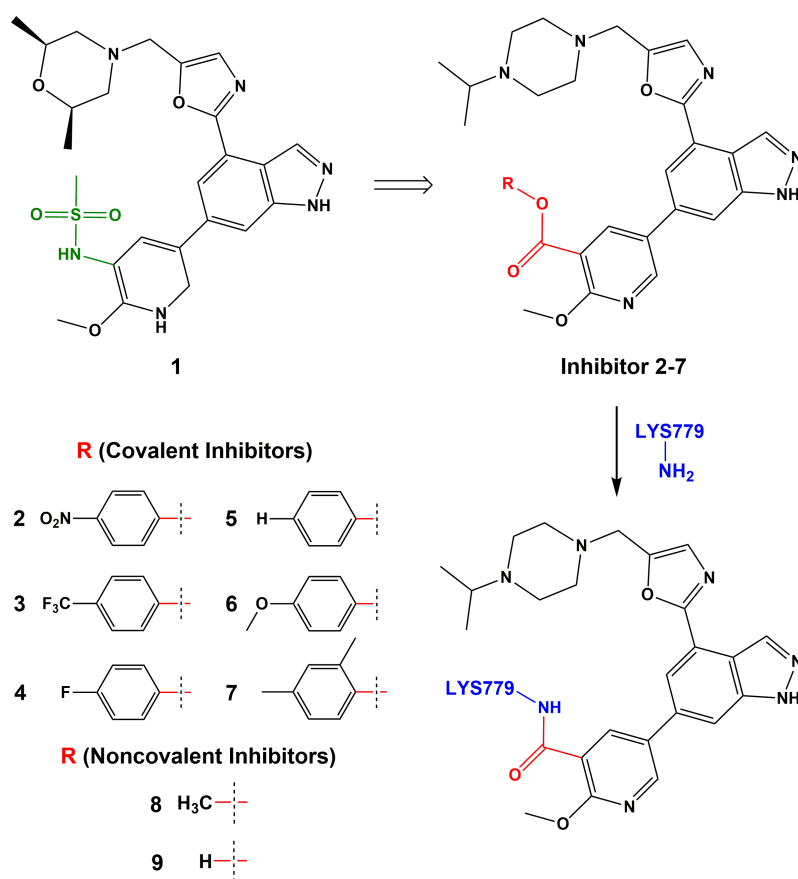


Figure 1: Inhibiteurs covalents de l'enzyme PI3K δ proposés dans la littérature.[12]

Les mesures cinétiques des vitesses d'inactivation (k_{inact}) des inhibiteurs covalents 2-7 (Figure 1) suggèrent que les propriétés électroniques et les réactivités chimiques attendues des esters électrophiles n'ont presque aucun effet sur la vitesse de formation des liaisons covalentes.[12]

L'efficacité relative observée de ces esters envers PI3K δ a donc été attribuée aux étapes initiales de la formation de la liaison réversible des inhibiteurs. Puisque k_{inact} était presque constant et indépendant de l'électrophilie de l'ester, on a supposé que l'étape déterminante de la vitesse n'était pas la première, l'attaque nucléophile au groupe amino de Lys779 (formation d'un zwitterion), mais qu'il pouvait y avoir une autre étape dans le mécanisme. Ainsi, le mécanisme montré dans la Figure 2 a été proposé en supposant que l'étape déterminante de la vitesse est la perte de proton du zwitterion tétraédrique, produisant un intermédiaire anionique. Néanmoins, le mécanisme exact nécessite une clarification car de nombreuses questions se posent : Comment le proton est-il transféré dans l'intermédiaire zwitterionique ? Y a-t-il un résidu basique impliqué dans la perte du proton du zwitterion ? Quels sont les rôles des résidus du site actif ? La formation d'un intermédiaire neutre tétraédrique est-elle également possible ? Les molécules d'eau du site actif sont-elles impliquées à un stade quelconque ? Une question cruciale est de savoir pourquoi le composé 4 (ainsi que les composés 2-7) forment une liaison covalente avec Lys779 alors que le composé 8 ne le fait pas. Aussi, comment la stabilité des intermédiaires tétraédriques ou les énergies d'activation avant/arrière influencent-elles le devenir du ligand (liaison covalente/non-covalente) ? En fin, la Lys779 est-elle suffisamment nucléophile pour réagir avec le groupe ester de l'inhibiteur ?

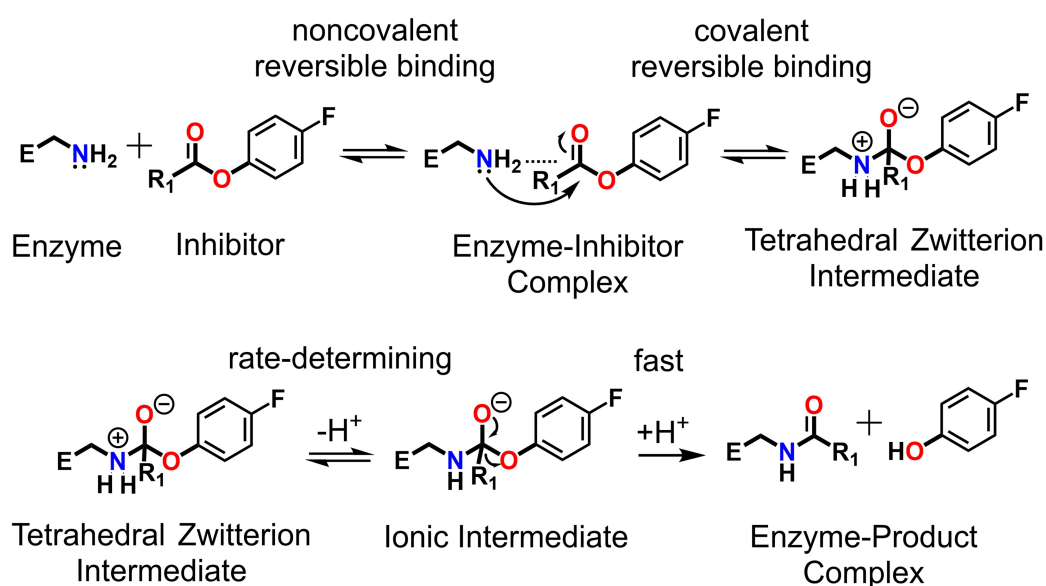


Figure 2: Mécanisme proposé par Dalton *et al.* [12] pour l'inhibition covalente de PI3K δ .

A notre connaissance, il n'existe pas dans la littérature d'étude théorique sur le mécanisme de l'inactivation covalente des enzymes PI3K. Plusieurs études *in silico* dans la littérature ne couvrent que des études de docking moléculaire, des relations structure-activité (QSAR) et de criblage virtuel, [14, 16, 17] et des simulations de dynamique moléculaire (MD) de PI3K α et des isoformes de PI3K γ , [18, 19] ainsi que du domaine catalytique de PI3K δ . [20] D'autre part, plusieurs travaux théoriques se sont concentrés sur le mécanisme des réactions de formation d'amide entre les esters et les amines, [21, 22, 23, 24, 25, 26] ou sur la réaction d'hydrolyse des amides, qui y est étroitement liée. [21, 22, 23] Certaines études étaient motivées par la compréhension de la formation de la liaison peptidique dans des conditions prébiotiques, [21] mais

d'autres étaient liées à des processus synthétiques organiques[27, 28] ou enzymatiques.[29, 30, 31] Tous ces travaux ont contribué à la compréhension du mécanisme de formation de la liaison amide et de la cinétique associée, qui peut changer de manière significative en fonction des dérivés spécifiques et de l'environnement moléculaire.

L'objectif de cette thèse est de déterminer le mécanisme d'inhibition covalente ciblée de PI3K δ visant la lysine. Nous chercherons à apporter des réponses aux questions mentionnées ci-dessus, et à approfondir les connaissances de cette chimie dans le but d'aider à la conception d'inhibiteurs covalents plus efficaces.

Résultats

Dans un premier temps, pour clarifier le mécanisme réactionnel, nous avons effectué une étude computationnelle de la réaction modèle entre la méthylamine et les acétates de méthyle, de phényle et de *p*-NO₂ phényle en solution aqueuse. Notre étude a mis l'accent sur le rôle des groupes aromatiques et aliphatiques partants et sur l'effet de leur capacité à accepter ou donner des électrons.[32] D'abord, des calculs *ab initio* et DFT ont été menés en utilisant un modèle statique du milieu aqueux. Concrètement, nous avons utilisé un modèle discret-continu avec 5 molécules d'eau dans lequel l'ensemble de molécules est entouré d'un environnement diélectrique continu (modèle PCM : Polarizable Continuum Model). Ces mêmes systèmes modèles ont ensuite été étudiés par une approche de dynamique moléculaire. Dans ce cas, nous avons opté pour une approche QM/MM dans laquelle les réactifs sont décrits par une approche quantique (QM) et l'environnement aqueux est décrit par des molécules d'eau explicites traitées par un champ de force de mécanique moléculaire (MM). Pour réduire le coût de calcul, nous avons choisi la technique dite de « dual-level » développée à Nancy. Dans cette approche, l'échantillonnage est obtenu par une méthode QM/MM avec un niveau QM simple (low-level). Les structures obtenues sont ensuite utilisées pour obtenir des corrections perturbatives à l'énergie libre, cette fois avec une approche QM/MM de haut-niveau (high-level). Après quelques tests préliminaires, notre choix s'est porté sur une méthode semiempirique pour le bas niveau (PM3) et une méthode DFT pour le haut niveau (M062X/6-311+G(d,p)).

L'ensemble des travaux sur les systèmes modèles a permis de mettre en évidence les points suivants. Nos calculs indiquent que pour des dérivés suffisamment électrophiles, comme le dérivé *p*-NO₂, le chemin de réaction le plus favorable après la formation de l'intermédiaire tétraédrique zwitterionique n'est pas le transfert de proton, comme proposé pour les inhibiteurs de PI3K α dans la littérature (Figure 2), mais la dissociation de la liaison C-OR, conduisant à la formation d'une paire d'ions. L'apparition d'une importante séparation de charge fournit alors la force motrice pour le transfert de proton et la formation des produits finaux, qui a lieu ensuite, le long d'une surface d'énergie libre plutôt plate. Notre étude indique également que l'efficacité des inhibiteurs ayant des groupes partants phénoliques devrait augmenter lorsque les dérivés ont des substituants attracteurs d'électrons, qui activent l'ester favorisant la formation de l'intermédiaire zwitterionique initial et facilitant les étapes de réaction suivantes. Nous trouvons que l'efficacité de l'inhibiteur observée expérimentalement est bien corrélée avec l'électrophilie calculée de l'ester, et avec la stabilité du zwitterion correspondant. Par conséquent, le calcul de ces propriétés semble être un outil intéressant pour évaluer a priori l'efficacité d'autres candidats inhibiteurs.

Évidemment, les interactions avec l'environnement enzymatique peuvent modifier les mécanismes décrits pour les systèmes modèles en solution aqueuse. Dans la suite de nos travaux, on s'est donc intéressé à ce problème en regardant les aspects mécanistiques dans différents modèles de l'enzyme. Comme pour les systèmes en solution aqueuse, nous avons d'abord réalisé une étude statique.[33] Nous avons construit un modèle du site actif grâce à des simulations de dynamique moléculaire classique et une analyse des trajectoires obtenues. La Figure 3 montre le modèle utilisé pour le site actif de l'enzyme, qui est constitué de quelques résidus de la protéine, plusieurs molécules d'eau et l'inhibiteur. La méthode de calcul dans ce cas est une méthode QM/QM de type ONIOM (M062X/6-311+G(d,p)/PM6). Ces calculs ont permis d'affiner les mécanismes de réaction, qui globalement confirment les étapes obtenues pour les systèmes en solution. Nous avons enfin utilisé ces informations pour aborder une étude dynamique de type QM/MM sur un modèle élaboré de l'enzyme présenté dans la Figure 4. Une approche QM/MM dual-level a également été utilisée au niveau PM3 (bas niveau) et (M062X/6-311+G(d,p)) (haut niveau). Ces calculs nous ont permis d'obtenir le profil d'énergie libre de la réaction d'inhibition pour le composé 2 (dérivé *p*-NO₂). La barrière calculée est en bon accord avec les données cinétiques expérimentales disponibles ce qui nous permet de valider l'approche théorique proposée et les mécanismes obtenus.

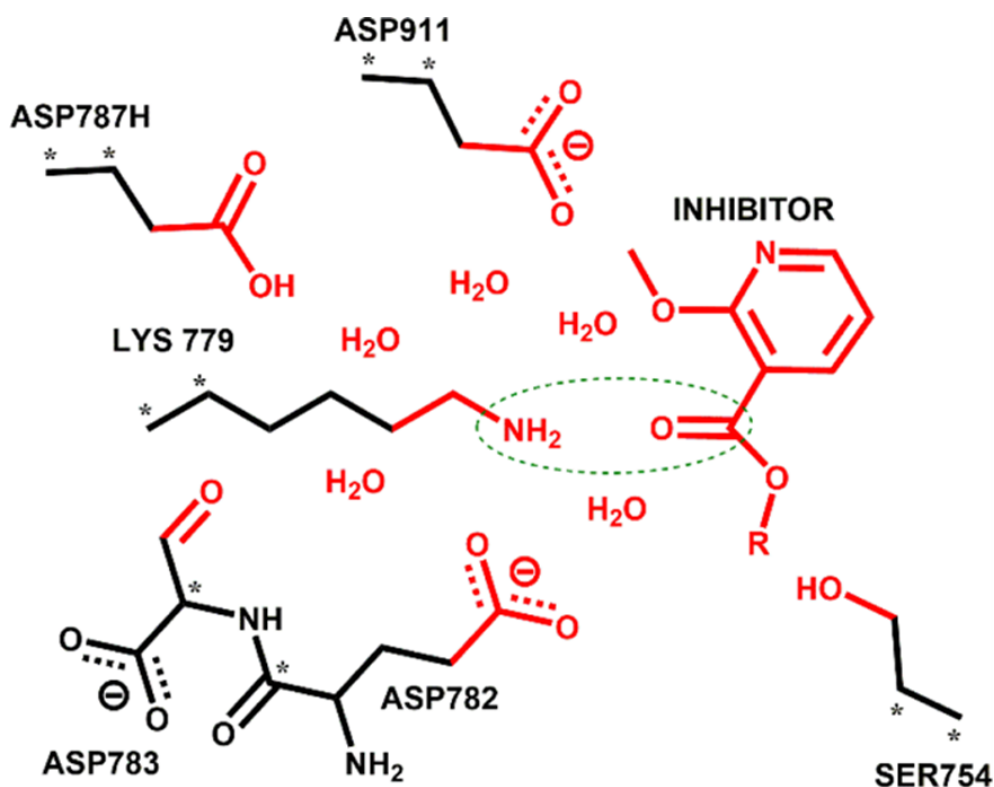


Figure 3: Modèle ONIOM utilisé dans cette étude pour la réaction entre l'enzyme PI3K δ et l'inhibiteur. Les parties rouges représentent le sous-système haut-niveau (M062X/6-31+G(d,p)) et les parties noires le bas niveau (PM6). Les atomes indiqués par un astérisque sont fixés dans l'espace dans la procédure d'optimisation des géométries.

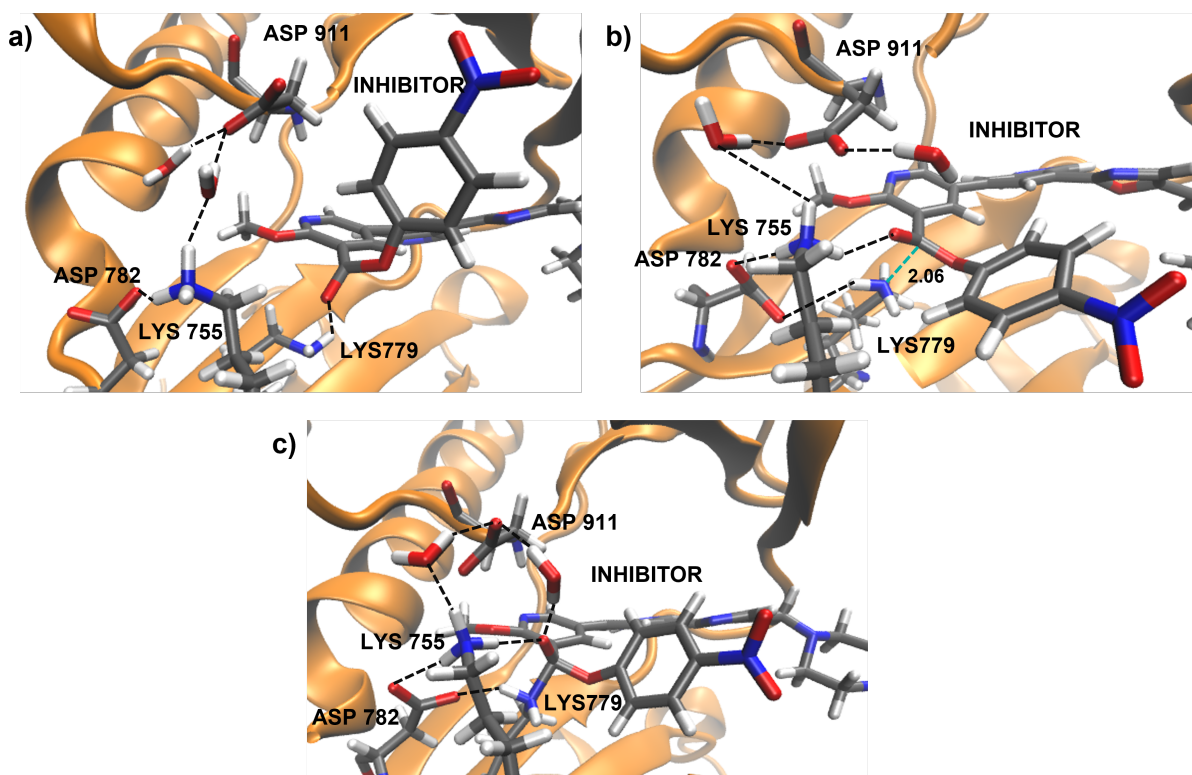


Figure 4: Exemples de structures du complexe réactif (a), de l'état transition state (b) et de l'intermédiaire zwitterionique (c) obtenus dans les simulations de dynamique moléculaire QM/MM de l'enzyme PI3K δ et de l'inhibiteur 2 (dérivé *p*-NO₂).

Conclusions

En résumé, l'ensemble des calculs réalisés pendant cette thèse montrent un mécanisme d'inhibition de l'enzyme PI3K δ qui est un peu plus complexe que celui proposé dans la littérature à partir des données cinétiques expérimentales. Pour l'ensemble d'inhibiteurs étudiés, la réaction commence par une activation du site actif, après la formation du complexe enzyme-substrat, qui se produit par le transfert de proton de la Lys catalytique protonée vers un résidu Asp non-protoné proche. La Lys ainsi activée réagit avec le groupe ester de l'inhibiteur formant un intermédiaire zwitterionique. La suite du mécanisme est cependant très dépendante des propriétés électroniques de l'inhibiteur. Pour les inhibiteurs suffisamment électrophiles, comme le dérivé *p*-NO₂, qui présentent la meilleure activité, la réaction se poursuit par une dissociation du zwitterion en forme de pair d'ions, puis par un transfert de proton qui conduit à la formation des produits attendus (formation d'une liaison amide dans l'enzyme, puis un alcool partant). Les calculs montrent aussi que les intermédiaires zwitterioniques n'ont pas une grande stabilité (dissociation en arrière vers de formes neutres rapide) et que l'étape limitante du processus d'inhibition est celle qui conduit à la dissociation ionique de ces intermédiaires. Ces résultats expliquent enfin l'apparente indépendance des données expérimentales de l'étape d'inactivation (k_{inact}) avec l'électrophilie de l'inhibiteur, qui peut s'expliquer par la présence d'un mécanisme plus complexe que le mécanisme traditionnel d'inactivation en deux étapes. En outre, ils ouvrent de nouvelles perspectives pour le développement d'inhibiteurs plus efficaces.

Chapter 1

Introduction

This work deals with the irreversible inhibition of the enzyme phosphoinositide 3-kinase delta (PI3K δ) by activated esters. The process is relevant for the treatment of various diseases and is especially interesting in oncology. From the chemical point of view, it consists in the formation of a covalent bond between the ester inhibitor and the targeted lysine. It is therefore an ester aminolysis, a process which is well-known in organic chemistry. Amine acetylation reactions are a special case of ester aminolysis and they are frequently used in general organic chemistry to create an amide bond. Numerous uses of the acetyl functional group are found in drug development, pharmaceutical production, polymer chemistry, and naturally occurring reactions within the human body. It can be used in various chemical processes as a protecting group and in the synthesis of peptides.[34, 35, 36, 37, 38] For this reason, in this introduction, the amine acetylation and ester aminolysis reactions will be briefly explained in the first sections in terms of organic chemistry and biochemistry. Then, the enzyme inhibition process will be discussed.

1.1 Amine Acetylation and Ester Aminolysis Reactions from Organic Chemistry Perspective

In organic chemistry, the ester aminolysis reaction is the process in which an ester reacts with a primary or secondary amine to give an amide. When the ester derivative is an acetate, the reaction is also called amine acetylation (the acetyl group CH_3CO is transferred to the amine). An example of ester aminolysis is given in Figure 1.1. In a general overview, this reaction involves nucleophilic addition and elimination steps, occurring at the carbonyl carbon atom where the substitution takes place. Before the elimination reaction, a tetrahedral intermediate forms, and later it transforms into an amide. As shown, the leaving group of the ester (R^3O^- in the Figure 1.1) abstracts one of the protons on the nitrogen atom. Hence, the $\text{p}K_a$ of the formed alcohol is a property playing an important role in the process. For instance, it is known that the $\text{p}K_a$ of the phenol (10.0) is significantly lower than that of methanol (15.5), and therefore, the phenolate ion is a much weaker base than the methoxide ion. So, the reactivity of the corresponding esters is expected to present notable differences, as we will discuss in the chapters below.

The mechanism of amide bond formation through ester aminolysis has been the subject of many theoretical studies in the past using different models and approaches, and possible

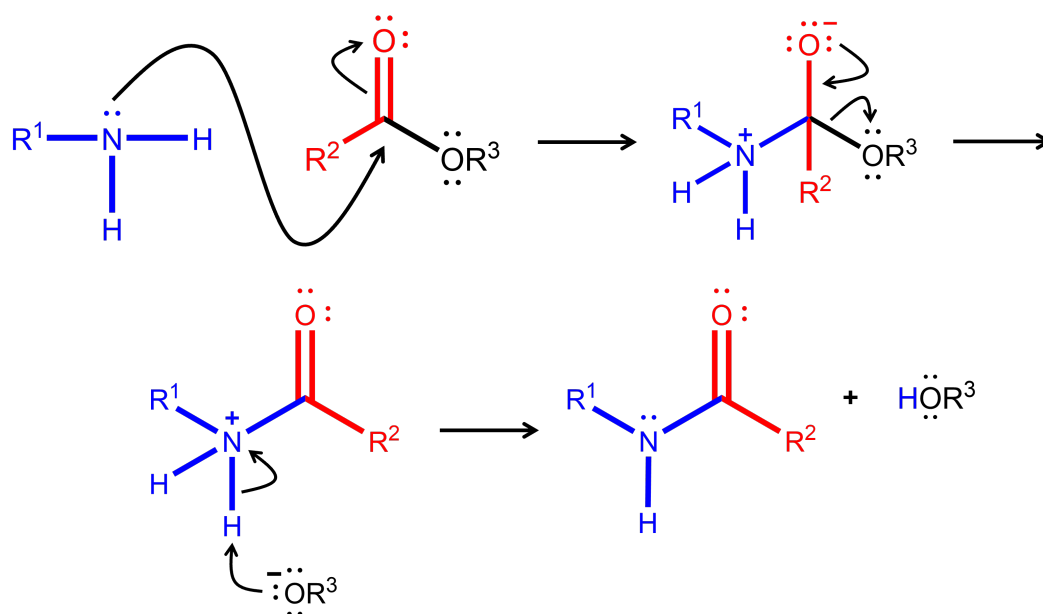


Figure 1.1: Amide formation through a reaction of an amine with an ester.

pathways have been analyzed.[25, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51] However despite some experimental studies [52, 53, 54, 55, 56, 57], the mechanism of the aminolysis of activated esters in aqueous solution is not yet fully understood. Some theoretical calculations are available in acetonitrile solution,[42, 47, 48] but the reaction mechanism is arguably different in this case since the formation of zwitterionic intermediates was ruled out under these conditions, whereas it is expected to play a central role in aqueous solution. However, modeling zwitterionic and ion-pair species in aqueous environments is quite more challenging because of technical difficulties in their theoretical characterization.

1.2 Amine Acetylation and Ester Aminolysis Reactions in Biochemistry

In biological systems, the acetylation of primary amines in proteins, widely known as N-acetylation reactions (also N-acetylation of glucosamine), has a considerable impact on gene expression and metabolism. It can be observed in two different ways.[58] One of the routes is the acetylation at the α -amino group of N-terminal amino acids, catalyzed by the N-terminal acetyltransferase (NAT) enzymes.[59] In this route, the acetyl group of acetyl-coenzyme A (Ac-CoA) is transferred to the α -amino group of the first amino acid residue of the peptide chain. Note that, acetylation of the N-terminal is a major covalent modification (Figure 1.2).

The other route is the acetylation of ϵ -amino group of lysine residues, this is also known as lysine-acetylation. It can occur both enzymatically or non-enzymatically, and for instance, lysine residues are spontaneously acetylated in mitochondria,[60] as illustrated in Figure 1.3.

On the other hand, enzymatically occurring lysine acetylation reactions are catalyzed by Lysine Acetyl Transferases (KATs). They have a similar mechanism to the first aforementioned

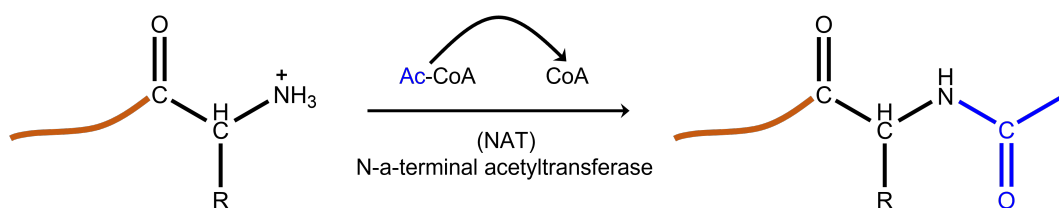


Figure 1.2: Representative scheme of N- terminal acetylation catalyzed by NAT enzymes.[59]

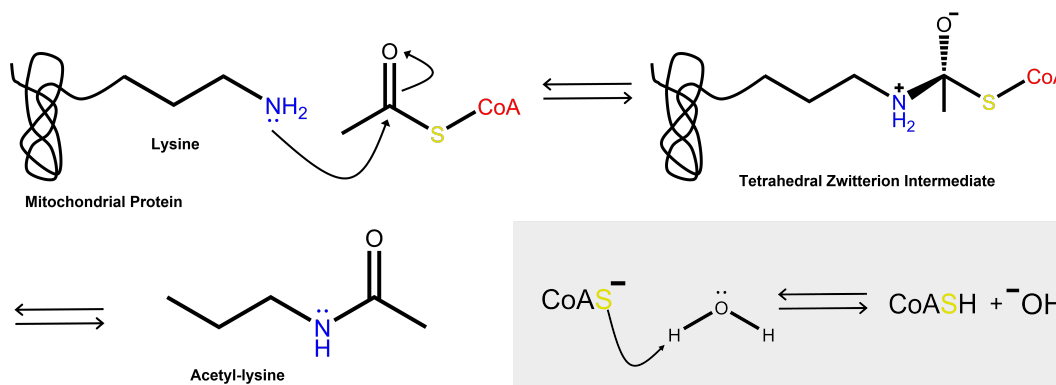


Figure 1.3: Representative scheme of non-enzymatic lysine acetylation mechanism occurring in mitochondria.[60]

reaction, in which, the lysine amino group is deprotonated and an acetyl group is transferred from Ac-CoA. However, in this case, the acetylation is reversible.[61] Up to date, most of the KAT studies are based on the Histone Acetyl Transferases (HATs) catalyzing the same reaction at the N-terminal tail of histone proteins.[62] Note that, this terminal is very important in DNA-templated epigenetic processes.

Histones have positively charged terminal lysine. Therefore, they are wrapped around the DNA which has negatively charged phosphate groups within its backbone. As a result of the acetylation of histones terminal lysine, the condensed chromatin undergoes a structural change into a more relaxed state and DNA binding proteins can engage with exposed regions to stimulate gene transcription. KAT (HAT) activity has been found to be disrupted in a variety of cancers. Deacetylation performed by Lysine Deacetylase (KDAC) enzymes has the opposite effect (Figure 1.4).[63]

DNA is tighten on the deacetylated histone cores, which makes it more difficult for transcription factors to attach to DNA. As a result, the levels of gene expression is suppressed, which is known as gene silencing. This is used to produce therapeutics to combat cancer and other diseases.

Amine acetylation reactions are also employed as a covalent modification for pharmaceutical purposes, such as interruption of enzyme activity. As an example, last year, Fournier *et al.*[15] proposed a series of modified dihydroisobenzofuran (DHB) ester molecules that selectively lead to lysine acetylation resulting in kinase inactivation. Protein liquid chromatography-mass spectrometry (LCMS) experiments confirmed the covalent modification of the catalytic lysine. Likewise, Dalton *et al.* proposed new ester derivatives by modifying the structure of a clinical drug (GSK2292767). They designed new PI3K δ inhibitors constituting ester groups

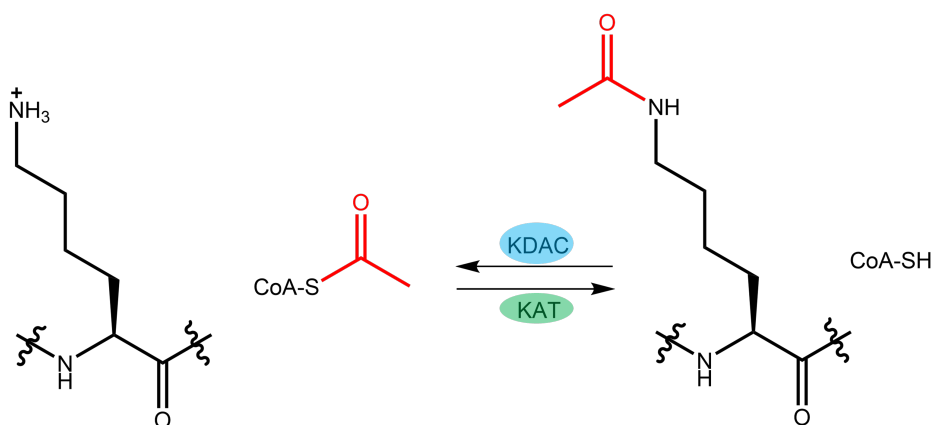


Figure 1.4: Lysine acetylation and deacetylation catalyzed by KAT and KDAC enzymes.[63]

as electrophilic warheads to target the conserved catalytic Lys779 (Figure 1.5) The covalent inhibition approach for rational drug design will be explained in detail, in the following section.

In summary, understanding the details of the chemical mechanisms of amine acetylation appears as a crucial goal for the rational design of new therapeutics, since these reactions are involved in gene expression and metabolic processes.

1.3 Targeting Covalent Inhibitors (TCIs)

Over the last 100 years, covalent targeting of biological structures has drawn attention in drug discovery. As is evident from its name, these drugs attach covalently to their target and they are conceptually different from conventional non-covalent drugs. The first covalent inhibitor is a widely known drug, aspirin, having an acetylsalicylic acid structure, and has been on the shelves of pharmacies since the late 1800s. The acetyl group of aspirin inactivates the cyclooxygenase activity of the enzyme.[64]

According to a survey on the covalent drugs approved by the US Food and Drug Administration (FDA), covalent drugs are mainly used in infections, cancer diseases, gastrointestinal disorders, and cardiovascular indications. The majority of these drugs generally target the enzymes to interrupt their activities. For example, penicillin antibiotics inhibit D-alanyl-D-alanine carboxypeptidase to catalyze a key step in cell-wall synthesis by binding the active site serine of bacterial DD-transpeptidase. Moreover, proton pump inhibitors (PPI) are widely used by millions of patients of reflux disease. Acidic environment of the stomach activates the PPI drug which becomes a covalent modifier of the proton pump and this action minimizes the exposure to acidic reflux.[65]

In the covalent inhibition approach, the regulation of the biological system, such as enzymes, takes place in two steps. [12] It starts with the attachment of a high-affinity ligand (inhibitor (I)) to enzyme target (E) forming a reversible E-I complex. The binding constant, K_I , defines the potency of this complex. The covalent bond formation occurs in the second step, as illustrated in Figure 1.6. The rate of the second step, also known as the rate of inactivation k_{inact} , depends on the concentration of E-I complex. The overall inhibition kinetics is defined by the ratio k_{inact}/K_I . This action results in a chemical modification causing a change in the activity or in

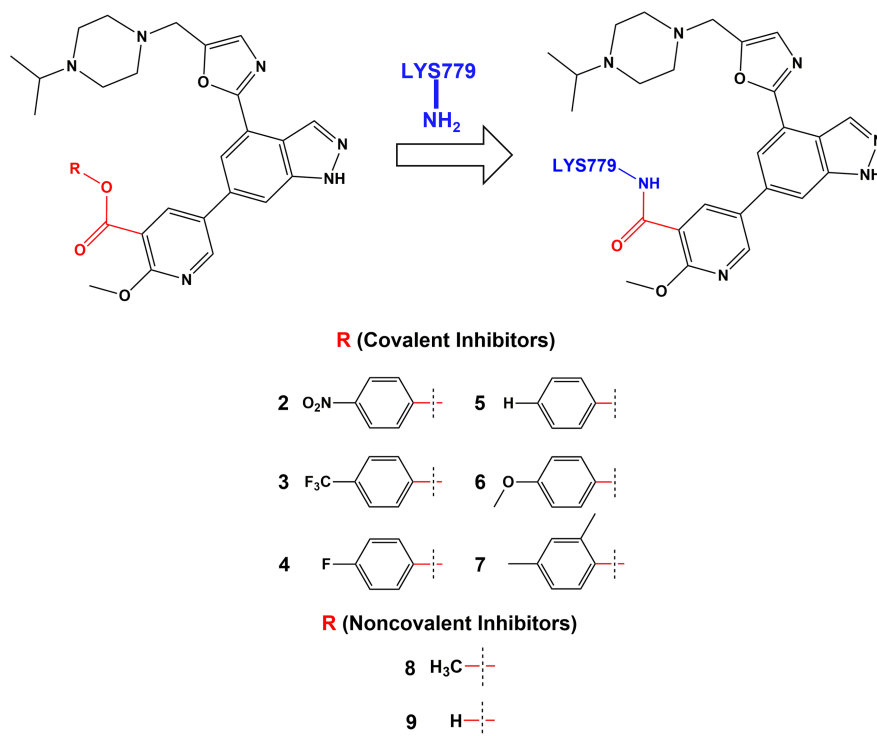


Figure 1.5: Lysine-targeted covalent inhibitors of PI3K δ forming an amide bond.[12]

the structure.[66]



Figure 1.6: Two step binding mechanism between inhibitor and enzyme.

During the last decade, the term "targeting covalent inhibitors (TCIs)" has been introduced since they provide specificity in targeting non-catalytic nucleophiles using structure-based design. They are discrete from traditional reversible inhibitors which are in equilibrium with the target. TCI drugs have several advantages over conventional reversible drugs. These drugs are equipped with electrophilic functionality which provides specificity. The frequency of dosage used in the treatment is much lower than the conventional ones, because of the prolongation of the course of action. Also, they provide increased potency and enable modifications for rational design.[1] They can be used in cancer treatment and other serious diseases because they succeed in targeting protein-protein interactions, signaling pathways, and other biochemical networks. Additionally, kinase inhibitors, which are responsible for signaling biomolecules, are the most common examples of TCIs.

Unlike the conventional ones, TCIs provide irreversible and specific binding to the target resulting in a permanent bond. As aforementioned, these drugs involve electrophilic functional groups, called warhead, which are able to form a covalent bond with the target. Nonetheless,

TCIs are mostly designed based on the traditional reversible drugs, attaching the warheads providing them being a covalent irreversible drugs.[67] The difference in their course of action is illustrated in Figure 1.7. First, the reversible drug-protein complex forms which is represented in the first step of Figure 1.6. In the second step, reaction between the functional group on the drug and the targeted residue spontaneously occurs, which results in a covalent modification of the residue.[66]



Figure 1.7: Schematic representation of the difference between traditional reversible inhibitors and TCIs.

There are a variety of chemical reactions, such as alkylations, and Michael addition reactions, which result in acylation, hemiacetal formation, and disulfide formation, to inactivate the enzymes irreversibly.[68, 69] Therefore, the choice of reacting functionalities, or in other words, the choice of warhead and design of the inhibitor structure should be considered carefully, in order to design a successful TCI.[70] In the enzyme environment, the solvent is highly effective in determining the residue pK_a . Non-conserved residues are exposed to the highly polar environment, which enables species to be charged, while the conserved residues are in a hydrophobic environment, which stabilizes the neutral side chains. Therefore, it is important to take into consideration the local environment when designing a warhead for a specific residue. In the literature, there are several types of warheads specific to targeted amino acid residues such as cysteine,[71] serine,[72] tyrosine[73] and lysine.[12]; they are summarized in Table 1.1.

Cysteine residues are most commonly targeted for TCIs because they bring forth natural selectivity thanks to their high nucleophilicity, exposed location and noncatalytic activity.[6] They can also mediate redox reactions thanks to different oxidization states of the sulfur atom. Especially, an oxidation state having the sulfenic acid functionality figures into the cellular signaling.[71] Acrylamide functionality is known to be the gold standard for warheads on covalent inhibitors. In particular, one of the successful irreversible covalent kinase inhibitors, ibrutinib, has an acrylamide group as a warhead to covalently bond the Cys481 on the BTK.[15] However, a limited number of cysteine residues have been located in the binding pocket of human kinases. In addition to this, the point mutation of the modified cysteine appears as a common resistance mechanism to cancer.[94]

In recent days, lysine residues attracted attention as an alternative approach to cysteine. In contrast to cysteine, lysine is abundant and widely distributed in many functional sites including

Table 1.1: Different types of warhead examples

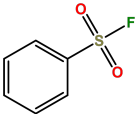
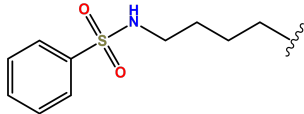
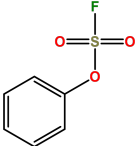
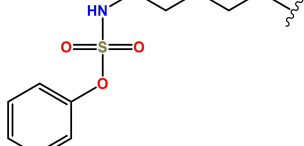
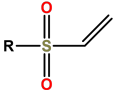
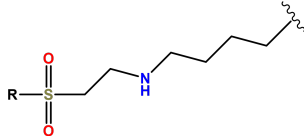
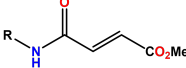
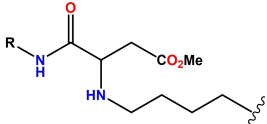
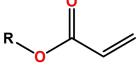
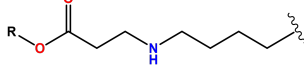
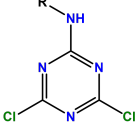
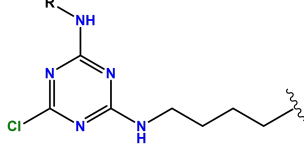
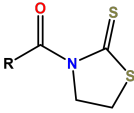
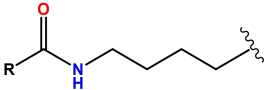
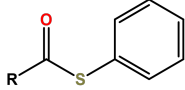
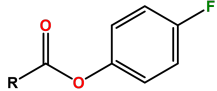
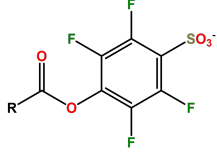
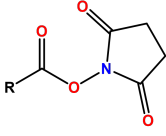
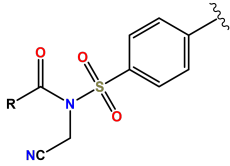
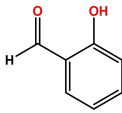
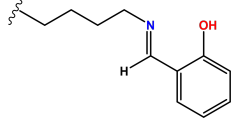
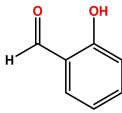
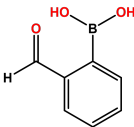
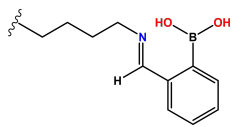
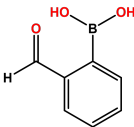
Name	Electrophile	Lysine Adduct	Nucleophile
Aryl sulfonyl fluoride [8, 74, 75, 76]			Lysine Tyrosine Serine
Aryl fluorosulfate [77, 78]			Lysine Tyrosine Serine
Vinyl sulfone/sulfonamide [79, 80]			Lysine Cysteine
Methyl fumaramide [81]			Lysine Cysteine
Acrylate[82]			Lysine Cysteine
Dichlorotriazine[83]			Lysine

Table 1.1: Different types of warhead examples

Name	Electrophile	Lysine Adduct	Nucleophile
Activated ester/amide [5, 12, 84, 85, 86, 87, 88]			Lysine
			
			
			
			
			
Salicylaldehyde [89, 90, 91]			Lysine (N terminus)
			
2-Formyl phenyl boronic acid [92, 93]			Lysine
			

enzyme active sites and protein-protein interfaces.[1, 5] The pK_a of the ϵ -amino group of Lys located on the exposed location of the protein, is around 10.0-10.5 at physiologic pH. In these conditions, the amino group is entirely in the protonated state, which makes the Lys residue non-nucleophilic. It has been shown that the pK_a value of lysines buried inside the protein is however around 5.0. Hence, lysine can adopt the neutral state when the environment is less polar than water.[95]

Targeting lysine has some advantages over cysteine, because of the diversity and lack of point mutation. Aryl sulfonyl fluorides, aryl fluorosulfates, Michael acceptors, dichlorotriazines, activated esters/amines, and aryl aldehydes are functional groups able to form covalent adducts with lysine residues. However, up to date, the covalent modification of lysine residue is a less common strategy, possibly because this residue is often protonated at physiological pH. In 2019, Liu *et al.*[6] inquired if nucleophilic catalytic lysines are common in the X-Ray structures of the human kinome. They performed continuous constant pH molecular dynamics simulations to calculate pK_a values and identified reactive catalytic lysines in human kinases. Catalytic lysines in kinases are known to stabilize the negatively charged phosphate groups of ATP[7] and kinase inhibitors usually compete with ATP for binding to the active site.[8]

1.4 Phosphoinositide 3-Kinase (PI3K) Enzyme

In biochemistry, a kinase is an enzyme that catalyzes the modification of the substrate by covalently adding phosphate, which is known as phosphorylation. The kinase enzymes abstract phosphate from ATP and transfer it to the hydroxyl group of the substrate. They are mainly classified into 3 main groups, namely protein kinases, lipid kinases, and carbohydrate kinases.[96] The human genome contains about 500 protein kinase genes and about 2% of all human genes are composed of them.[97]

Phosphoinositide 3-Kinase (PI3K) enzyme is a plasma membrane-associated lipid kinase contributing to several vital events in cells such as cycle regulation, cell growth, proliferation, cellular metabolism, survival, and angiogenesis.[98] They play an important role in the signaling pathway through the Protein kinase B, also known as AKT. The activation of this pathway leads to the neoplasm progression and resistance to anticancer therapies,[99] and based on the conducted research over years, the dysregulation of this signaling pathway has been shown in almost all human cancers. Therefore, it can be said that targeting this pathway, more specifically the PI3K enzyme is crucial for cancer treatment (Figure 1.8).

PI3K involves three subunits, p85 and p55 as regulatory subunits and p110 as catalytic subunit.[100] Moreover, in human cells, PI3K has 3 main classes: Class I, Class II, and Class III. They are different from each other, regarding their structural characteristics and substrate specificities. Among these, Class I is divided into two subclasses: Class IA and Class IB. Class IA PI3K is a frequently encountered and studied type involved in cancer diseases.[101, 102]

Class IA involves p110 α , p110 β and p110 δ . Whilst p110 α and p110 β are present ubiquitously almost in all tissues, p110 δ is restricted to hematopoietic cells in which it has an important role in B-cell homeostasis and functioning. These three incorporate the inputs from activated receptor tyrosine kinases (RTKs) and G-protein coupled receptors (GPCRs) into the signaling pathway.

PI3K δ enzyme consists of several domains as reported in Figure 1.9. As seen, the receptor

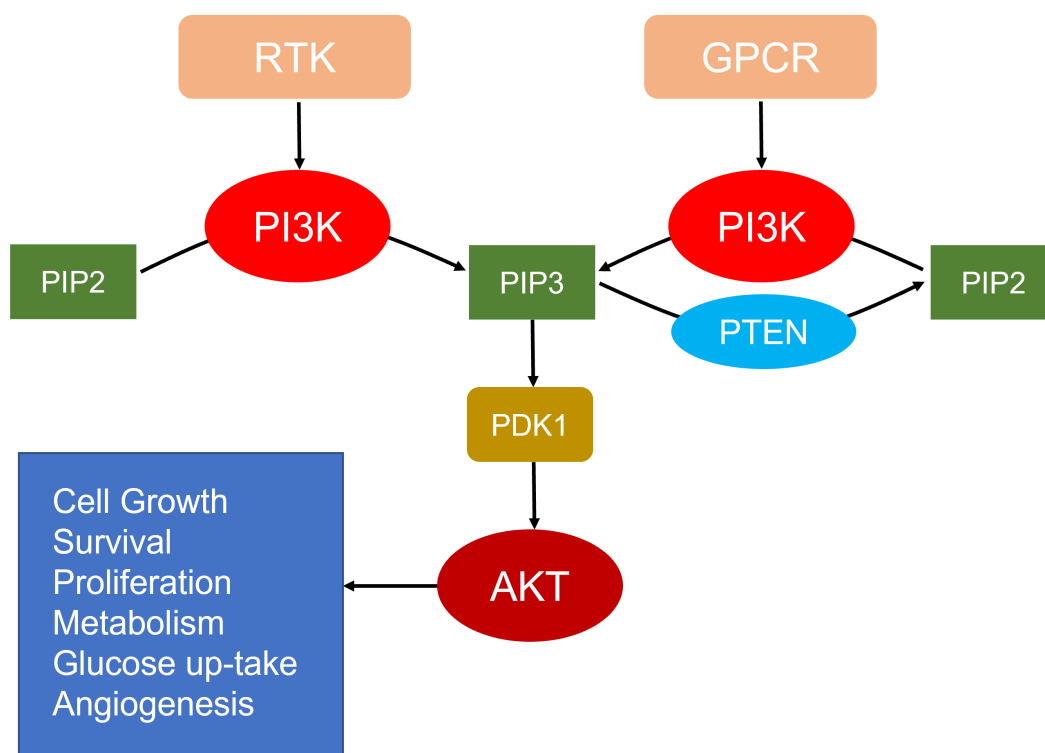


Figure 1.8: The overview of the signaling pathway of PI3K/AKT.

binding domain (RBD) lies between residues 189 and 278, and the C2 domain, responsible for phospholipid binding, lies in the region between residues 324-474. The kinase domain has two parts as N-Lobe (675-830) and C-Lobe (830-1044), which is the catalytic function of the enzyme. Moreover, the kinase and helical domains (in between 500-675 residues) have a broad and tightly packed interface between each other.[20]

More specifically, the transducing signals start from the cell membrane receptors, RTK and GPCR, through the cytoplasm, and meanwhile, PI3K is activated. Upon activation, phosphatidylinositol - 4, 5 - bisphosphate (PIP2) is phosphorylated to produce phosphatidylinositol -3, 4, 5- triphosphate (PIP3), which is catalyzed by PI3K. (Figure 1.10). Note that, Phosphatase and tensin homolog (PTEN) catalyzes the reverse path which is the dephosphorylation of PIP3 to PIP2, and it is responsible for helping suppress the growth of tumor cells. However, in tumor cells, it is seen that PTEN is mostly mutated.[103] Subsequently, a series of vital events is activated when the signaling proteins, such as Pyruvate dehydrogenase lipoamide kinase isozyme 1 (PDK1) and AKT, are bound to lipid products of PI3K. An overview of the signaling pathway is summarized in Figure 1.8. Therefore, it is not surprising that the activity of PI3K signaling is strongly related to malignant cell progression. In light of this information, it is believed that the PI3K enzyme is the key target for cancer treatment and there has been a tremendous effort for the development of drugs targeting PI3K signaling. In July 2014, Idelalisib (the commercial name is Zydelig) was approved by US FDA for patients with Indolent Non-Hodgkin Lymphoma.[104] Idelalisib is a reversible PI3K δ inhibitor. Moreover, Wortmannin, an irre-

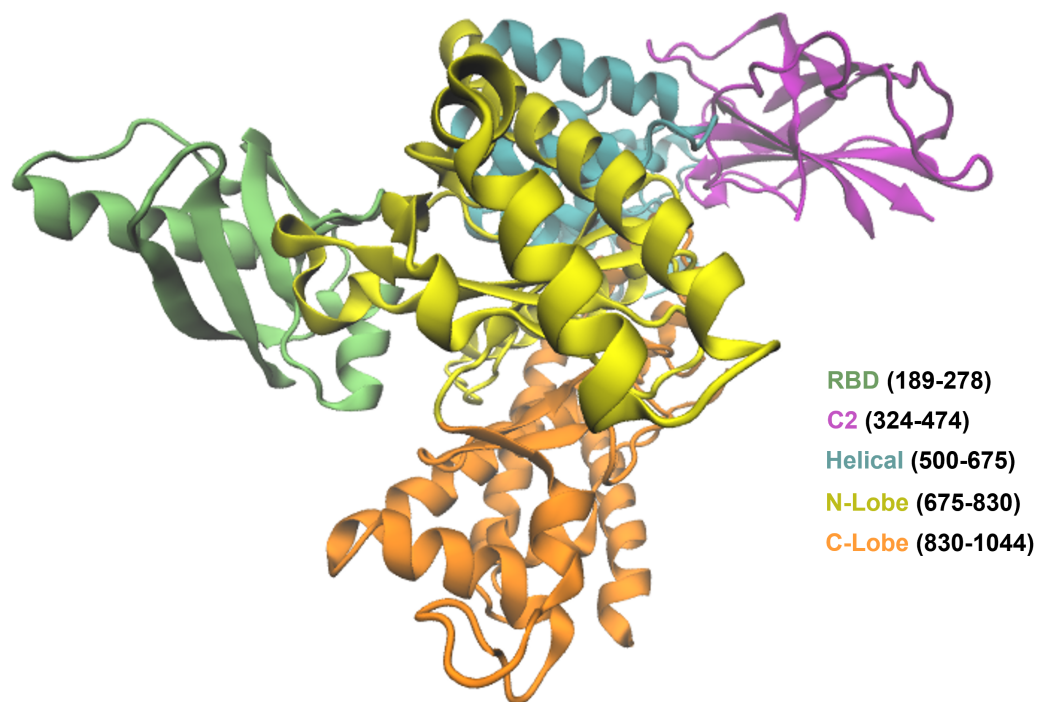


Figure 1.9: Domains of PI3K δ enzyme from crystal structure (PDB ID: 6ez6).[20]

versible covalent inhibitor for PI3K/AKT pathway has been investigated, though it suffers from poor selectivity.[105]

As explained in Section 1.2, in 2018, Dalton *et al.* firstly developed irreversible PI3K δ inhibitors which can selectively react with buried Lys residue (Lys 779 in PI3K δ numbering system), located in the catalytic domain. Their work has prompted us to carry out a theoretical study on these systems, which constitute the central subject of this PhD thesis.

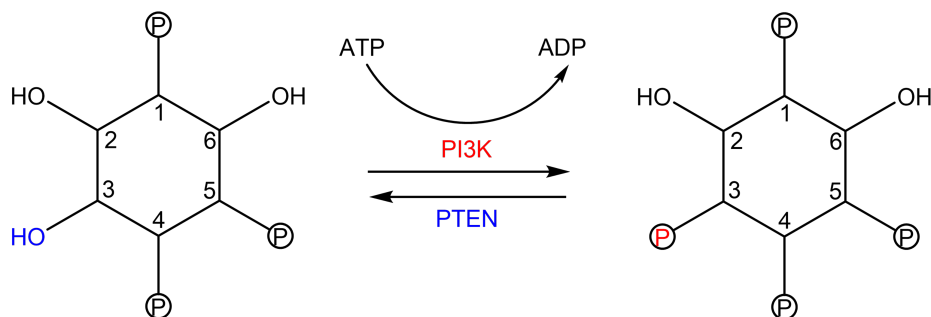


Figure 1.10: The production of PIP3 from PIP2 by phosphorylation catalyzed by PI3K enzyme.

Chapter 2

Objective and scope

Obtaining a more in-depth understanding of the underlying biochemical mechanism in PI3K δ covalent inhibition is an important goal that can help optimize inhibitor efficiency. In this respect, advanced molecular modeling techniques represent a unique tool to elucidate the microscopic details of the reaction mechanism, and the combination of theoretical calculations with experimental kinetic data should certainly be key for rapid progress in the field.

Protein liquid chromatography mass spectrometry (LCMS) analysis revealed that the designed compounds proposed by Dalton *et al.*[12] covalently inactivate PI3K δ by amide bond formation at Lys779. Furthermore, X-Ray crystal structure of PI3K δ inhibited by p-F derivative clearly confirmed a covalently bound adduct, but a noncovalent enzyme-inhibitor complex was observed in the X-Ray structure of methyl ester. Further work by Fournier *et al.*[15] has confirmed that simple acetylation of Lys779 can selectively inhibit PI3K δ activity.

Additionally, the multi-step mechanism indicated in Figure 2.1 has been proposed,[12] which is slightly more complicated than the traditional two-step mechanism for targeted covalent inhibitors (TCI). More specifically, the first step consists of the usual reversible, non-covalent binding of the inhibitor to the enzyme active site. Then, the ester adds to the kinase Lys779 forming a covalent tetrahedral zwitterionic intermediate. This reaction step, assumed also to be reversible, is expected to be affected by the nature (e.g. pK_a) of the leaving phenolic alcohol. The subsequent two steps involve (1) proton loss to form an anionic unstable intermediate, assumed to be the rate-determining step, and (2) irreversible, fast breakdown of this intermediate to form the products. The authors concluded that K_I encompasses the first two steps, while k_{inact} combines the last two. The measured K_I has been found to depend on the leaving group, while the measured k_{inact} has been shown to be little affected by the nature of the ester. Nevertheless, the exact mechanism requires clarification since many questions await answering:

- How is the proton removed from the zwitterionic intermediate?
- Is there any basic residue involved in the proton loss from the zwitterion?
- What are the roles of active site residues? Is the formation of a tetrahedral neutral intermediate possible as well?
- Are the active site water molecules involved at any stage?

- Why does compound 4 (as well as compounds 2-7) form a covalent bond with Lys779 while compound 8 does not? How does the stability of tetrahedral intermediates or the forward/reverse activation energies influence the fate (covalent/noncovalent binding) of the ligand?
- Is Lys779 nucleophilic enough to attack the ester group of the inhibitor?

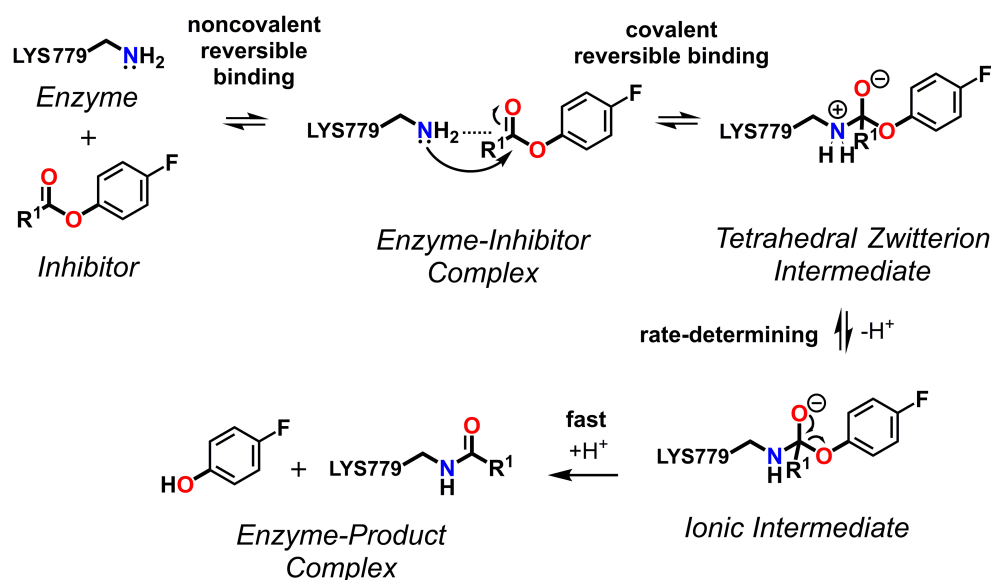


Figure 2.1: The mechanism proposed by Dalton et al.[12] for lysine-targeted covalent inhibition of PI3K δ .

Because of the significance of ester aminolysis in biological processes as well as in designing targeted covalent inhibitors for potential therapeutics, the goal of this study is to enlighten the mechanisms of ester aminolysis reactions via multi-scale computational techniques. For this aim, the overall study has been divided into two major parts involving calculations of simple model systems and the calculations within the enzyme system (Figure 2.2). The model system is used to scrutinize the possible reaction mechanisms and energy profiles of phenolic ester derivatives (Figure 2.3), corresponding to the inhibitors proposed by Dalton *et. al.*, with methyl amine, representing Lys 779.

The information gathered from the model system should shed light on the complex system involving the enzyme. Therefore the main objectives are two-fold:

- The first main objective is to determine with high precision the reaction mechanisms and free energy profiles of the model system in aqueous media.
- The second is to enlighten ester aminolysis involving the catalytic lysine residue within the PI3K δ enzyme for the purpose of targeted covalent inhibitor design.

In the next chapter, the theories behind the computational techniques used within this thesis are explained. Specific details on the methodology can also be found in the corresponding chapters.

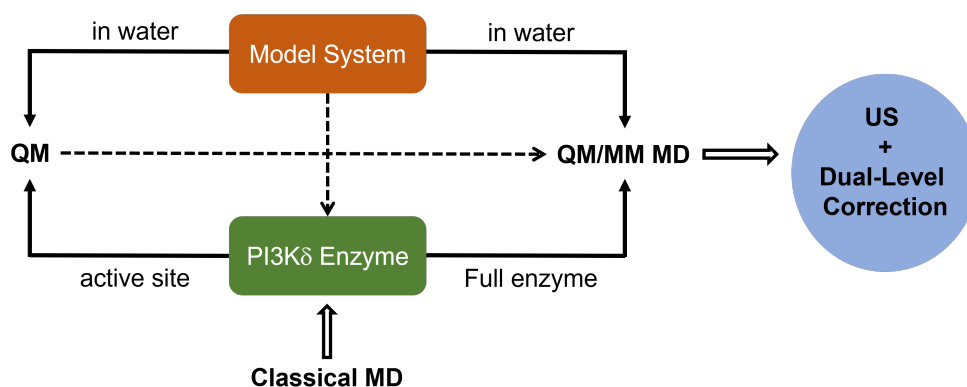


Figure 2.2: Schematic scope of the thesis.

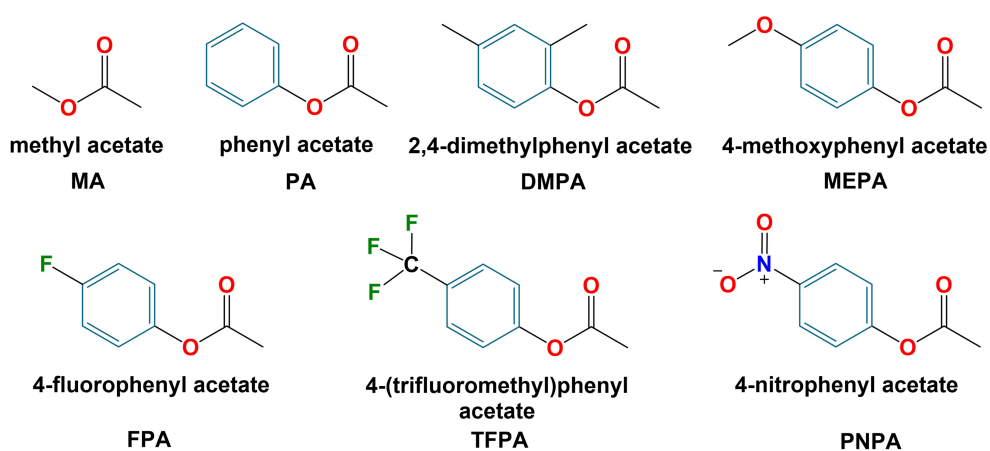


Figure 2.3: Ester derivatives considered for model system calculations.

The results are described in the following five chapters. In Chapter 4 and Chapter 5, the model system is considered. Quantum mechanical results are presented in Chapter 4 using simple dielectric models. They are then used in Chapter 5, in which the most plausible mechanisms are studied with QM/MM MD simulations to get accurate free energy profiles.

In chapter 6, 7 and 8, similar reaction mechanisms are studied in the PI3K δ enzyme environment with Classical MD, ONIOM and QM/MM MD methodologies, respectively. Finally, some concluding remarks will be presented.

Chapter 3

Theoretical Background

3.1 Quantum Chemical Methods

3.1.1 The Schrödinger Equation

Quantum mechanics was developed at the beginning of the 20th century to explain the behavior of particles at the atomic scale, such as electrons, because classical mechanics was inadequate to describe them. Quantum mechanics assumes the wave-particle duality of matter, and the classical concept of trajectory is replaced by that of a wave function (ψ). The wave function contains all the information about the properties of a particle. In 1926, Schrodinger, published the fundamental equation of quantum mechanics, known as the time-dependent Schrodinger equation, which allows to find the future state of a quantum mechanical system if the wave function at time t_0 is known. In the case of conservative systems for which the potential energy is time-independent, the time-dependent Schrodinger equation simplifies, and it is possible to write a time-independent form of this equation that can be used to obtain the energy of atoms and molecules. (Equation 3.1)

$$H\psi = E\psi \quad (3.1)$$

H in the Equation 3.1 is the Hamiltonian operator, which corresponds to the sum of the kinetic and potential energy operators of the system. ψ represents the wave function describing the system and E stands for the energy of the system. Although Equation 3.1 may seem only as an eigenvalue-eigenvector problem, it can only be solved exactly for the hydrogen atom. For multielectron atoms and molecules this equation becomes too complicated and cannot be solved analytically.

An estimation of E can be done using the variation principle. According to this principle, the computed E value with an approximated wave function, Equation 3.2 can only be higher than or equal to the exact ground-state energy (E_0).

$$E = \frac{\int \psi^* \hat{H} \psi d\tau}{\int \psi^* \psi d\tau} \quad (3.2)$$

Moreover, several approximations can be employed to simplify the solution of Schrödinger equation. The first one is known as the Born-Oppenheimer approximation, which separates

the motion of electrons and nuclei, since the latter move much slower than the former. For this reason, the kinetic energy of the nuclei is ignored and only the potential energy term is considered in the equation. The total energy operator for a diatomic molecule in the Born-Oppenheimer approximation can be written as in equation 3.3

$$\hat{H} = -\frac{1}{2} \sum_i^{elec} \left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \right) - \sum_i^{elec} \sum_s^{nuc} \left(\frac{Z_s}{r_{is}} \right) + \sum_i^{elec} \sum_j^{elec} \left(\frac{1}{r_{ij}} \right) + \sum_s^{nuc} \sum_t^{nuc} \left(\frac{Z_s Z_t}{R_{st}} \right) \quad (3.3)$$

where the first term is the kinetic energy of the electrons (elec). The following terms are potential energies, namely Coulombic attraction between the electrons and nuclei (nuc) (2nd term), the electron-electron repulsion (3rd term) and the nuclear-nuclear repulsion (last term), which is a constant due to the assumption of fixed positions of nuclei.

The Schrödinger equation can be solved for H_2^+ within the Born-Oppenheimer approximation, but for more complex systems, other approximations are required. The Hartree-Fock self-consistent field (HF-SCF) approach described hereafter is the main technique used to study molecular systems.

3.1.2 Hartree-Fock Theory

The Hartree-Fock (HF) method is based on the treatment of electrons as independent particles moving in the average field created by the other electrons and the nuclei, and serves as the foundation for all ab initio quantum chemistry methods. The wave function is approximated by a single Slater determinant:

$$\psi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \dots & \chi_{N-1}(1) & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \dots & \chi_{N-1}(2) & \chi_N(2) \\ \dots & \dots & \dots & \dots & \dots \\ \chi_1(N-1) & \chi_2(N-1) & \dots & \chi_{N-1}(N-1) & \chi_N(N-1) \\ \chi_1(N) & \chi_2(N) & \dots & \chi_{N-1}(N) & \chi_N(N) \end{vmatrix} \quad (3.4)$$

which is constructed from a set of N single-electron wave functions (N being the total number of electrons in the molecule and χ a spin-orbital equal to the spin function multiplied by the spatial wave function).

Using the variational principle, it is possible to determine the optimal spin orbitals as those minimizing the energy E of the system. The resulting equations are known as Hartree-Fock equations,

$$f_i \chi_i = \epsilon_i \chi_i \quad (3.5)$$

where f_i is the Fock operator, χ_i is an eigenfunction and ϵ_i the corresponding orbital energy (eigenvalue). The one electron Fock-operator is defined as

$$f_i = -\frac{1}{2} \nabla_i^2 - \sum_k^{nuclei} \frac{Z_k}{r_{ik}} + V_{HF}(i) \quad (3.6)$$

where the Hartree-Fock potential $V_{HF}(i)$ represents a measure of the average repulsive potential that each electron experiences in presence of the other electrons, and therefore depends on the molecular orbitals (for a precise definition of this operator, see for instance Ref [106]).

The Hartree-Fock approach, which is an example of the self-consistent field (SCF) method, involves an iterative process. It begins with an initial assumption regarding the molecular orbitals for a specific molecular geometry, which allows obtaining a first approximation for the HF potential. Then, the eigenvalue problem is solved and a new set of spin-orbitals are calculated leading to a new approximation for the HF potential. The cycle is continued in this manner until energy convergence is obtained.

The weakest point of HF theory is its insufficient description of electron-electron correlations, which may cause significant errors in chemical information. More elaborated methods have been developed to account for electron correlation effects in different ways, although such techniques are in general much more computationally expensive. We briefly describe below the so-called post-HF methods.

3.1.3 The Post-HF Methods

They are aimed to overcome the limitations of the HF method in which the wavefunction is described as a single Slater determinant. The starting point is the HF solutions, and the exact energy is written as:

$$E_{\text{exact}} = E_{\text{Hartree-Fock}} + E_{\text{correlation}} \quad (3.7)$$

The correlation energy can be computed in different ways. The Configuration Interaction (CI) approach is based on the variational principle assuming that the wave function is described as a linear combination of Slater determinants. The approach is quite expensive and most post-HF calculations are based on perturbation theory, namely the Moller-Plesset theory[107]. In this case, the Hamiltonian of the system is written as the sum of the Fock Hamiltonian ($\hat{F}$) (coming from the HF method) and a perturbation operator ($\hat{V}$) which contains the correlation information:

$$\hat{H} = \hat{F} + \hat{V} \quad (3.8)$$

In general, the perturbation is limited to the second order and the method is known as MP2,[108] though methods up to the fourth order, namely MP4,[109] have been developed.

3.1.4 Semi Empirical Quantum Mechanics

To reduce the computation time needed to calculating and manipulating integrals in ab initio procedures, semiempirical methods were developed. In this approach, only the valence electrons are treated explicitly, as they are the only electrons involved in chemical properties; the core electrons are merged with the nuclear charge. In addition, simplifications lead to the neglect of some integrals and their compensation by empirical parameters, which directly influences the accuracy of the methods. Several semi-empirical techniques are available.[110, 111, 112, 113, 114, 115, 116, 117, 118, 119]

Pople and colleagues introduced the first methods: complete neglect of differential overlap (CNDO),[110] intermediate neglect of differential overlap (INDO),[111] and neglect of diatomic differential overlap (NDDO).[112] Although these methods are not commonly used in modern computational chemistry, they constituted the groundwork for other semiempirical

approaches. Dewar et al. developed the modified neglect of differential overlap in 1977 (MNDO).[115] As a result of this achievement, all modern semiempirical methods are derived from MNDO. The most popular semiempirical methods in contemporary computational studies are AM1,[116], PM3,[117] and PM6.[119]

3.1.5 Density Functional Theory

Density functional theory (DFT)[120] was built based on a theorem proven by Hohenberg and Kohn[121], and states that the energy of a system can be calculated if the electronic density of the system is known. From the original theorems, a variational principle can be established since the ground state density minimizes the total energy.

The electron density is defined as shown in Equation 3.9 where r_i corresponds the coordinates of the electrons.

$$\rho(r) = N \int \dots \int |\Psi(r_1, r_2, \dots r_n)|^2 dr_1 dr_2 \dots dr_n \quad (3.9)$$

The electronic energy as a function of the electron density is:

$$E[\rho(r)] = \int V(r)\rho(r)dr + T[\rho(r)] + V_{ee}[\rho(r)] \quad (3.10)$$

Equation 3.10 can be rewritten following the approach developed by Kohn and Sham[122] for a system of non-interacting electrons generating the same total density as the real system. The electronic energy is obtained as

$$E[\rho(r)] = \int V(r)\rho(r)d(r) + T_{ni}[\rho(r)] + J[\rho(r)] + E_{XC}[\rho(r)] \quad (3.11)$$

in which ($J[\rho]$) is the coulomb energy, ($T_{ni}[\rho]$) is the kinetic energy of the non-interacting electrons, and ($E_X[\rho]$) is the exchange-correlation energy functional, which accounts for the remaining energy terms (those not included in the non-interacting kinetic and electrostatic terms)[123]. The exchange-correlation energy is written:

$$E_{XC}[\rho(r)] = \int \rho(r)\epsilon_{XC}(\rho(r))dr \quad (3.12)$$

The Kohn-Sham orbitals ψ_i are established by solving the Kohn-Sham equations in which h_{KS} is the hamiltonian operator

$$h_{KS} = -\frac{\nabla^2}{2} - \sum_{A=1}^N \frac{Z_A}{|r - R_A|} + \int \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 + V_{XC}[\rho(r)] \quad (3.13)$$

where $V_{XC}[\rho(r)]$ is the exchange-correlation potential associated to the exchange-correlation energy. The exact form of the exchange-correlation functional is not known and approximated functionals have to be used.

The Local Density Approximation (LDA) is the groundwork for approximating exchange-correlation functionals. In this approach it is assumed that the exchange-correlation energy depends only on the electron density at each point in space (the local density). Further improvements are obtained by considering the Generalized gradient approximation (GGA)[124], which

takes into account the inhomogeneous nature of electron density by considering the exchange and correlation energies dependence on the density and its gradient. An alternative approach is the combination of GGA with exact HF exchange, which leads to functionals known as hybrid functionals. Among those functionals, B3LYP [125, 126, 127] is one of the most popular ones. Another widely used functional is M062X, developed by Zhao *et al.*[128] and based on meta-GGA, the more accurate version of GGA.

3.1.6 Basis Sets

In quantum chemical calculations, the molecular orbitals within a system are expanded as a combination of functions called the basis set.[129] The basis set is usually constituted by an ensemble of atomic orbitals, which can be defined through Slater-type orbitals (STOs) or Gaussian-type orbitals (GTOs),[130] though the latter type is more broadly used in modern *ab initio* and DFT calculations.

The basis sets are hierarchically ordered according to their size, which provides a controlled way to obtain more accurate solutions, while resulting in higher computational time. For a minimal basis set, only one basis function is used to describe each atomic orbital, while several functions are used in extended basis sets called double-zeta, triple-zeta, quadruple-zeta, etc.

Split-valence basis sets developed by Pople *et al.* treat differently the core and valence orbitals. Only one basis function is applied to core atomic orbitals, while several functions are used for the valence atomic orbitals (eg: 3-21G, 6-31G, 6-31+G, 6-311G). Moreover, it is possible to modify the basis sets with the addition of polarization and/or diffuse functions, which provide higher flexibility to describe the electronic cloud, especially in polarizable and charged systems, and leads to a huge improvement in the electronic energy calculation.

Another type of basis functions is the so-called correlation consistent basis sets. They are constructed by adding shells of functions to a core set of atomic Hartree-Fock functions,[131] each function providing a similar contribution to the correlation energy.

3.2 Solvation Models

A solvent model is a technique used in computational chemistry to take into consideration the behavior of molecules in solution.[132] By using appropriate solvation models, thermodynamic calculations of reactions that take place in solution are possible.

The effect of the solvent can be taken into consideration using implicit or explicit solvent models. In explicit solvation, the molecular details of each solvent molecule are included. They are explicitly added to the system as discrete molecules. Nevertheless, in quantum chemical calculations, including many solvent molecules explicitly is a challenging task due to the high computational cost and the difficulties to deal with a high number of degrees of freedom.

Implicit solvent models were developed in order to avoid costly computations. They represent the most cost-effective way to include condensed-phase effects into quantum chemical calculations. They are also known as “continuum solvation models” since in general the solvent is described as a polarizable dielectric continuum medium around a cavity occupied by the solute molecule. Implicit models include different physical effects which contribute to the overall solvation process. The total solvation free energy is expressed as

$$\Delta G_{solvation} = \Delta G_{cavity} + \Delta G_{dispersion} + \Delta G_{electrostatic} + \Delta G_{repulsion} \quad (3.14)$$

In this equation, ΔG_{cavity} is a positive energy required to create the cavity in the medium, and the other terms represent the different types of interaction between the solute and the solvent. The cavity is generally defined by interlocking van der Waals spheres centered at atomic positions, possibly with corrected atomic radii.

The original method was developed in Nancy,[133, 134, 135] but the Polarizable Continuum Model (PCM) developed by Tomasi *et al.*,[136] is the most popular one nowadays, using either the Conductor-like approach (C-PCM), or the Integral Equation Formalism (IEF-PCM).[137]

For moderate size molecules, it is also possible to employ explicit and implicit solvation approaches together (discrete-continuum solvation models). In this case, a number of solvent molecules are placed around the solute molecule (which is usually kept at a minimum to minimize the computational cost) and a dielectric continuum is assumed to surround the solute-solvent molecular cluster.

The study of solutes in solution can also be done using explicit models in the framework of molecular mechanics,[138] and the combination of models in which part of the system is described quantum mechanically and the other part is described with molecular mechanics leads to the combined QM/MM methods. We briefly describe these techniques hereafter.

3.3 Molecular Mechanics

Molecular Mechanics (MM) utilizes classical mechanics to describe complex molecular systems such as systems in solution or biomolecules. Chemical bonds between atoms are treated as springs and are typically described as harmonic oscillators. The other interactions are also simplified using the concept of force field.[139]

In a force-field, the total potential energy of the system is divided into bonded and non-bonded interactions (Figure 3.1, Equation 3.15). Bonded interactions include bond stretching, angle bending and torsions. Non-bonded interactions are present for every pair of atoms in the system and include electrostatic and non-electrostatic (dispersion and short-range repulsion) terms. The latter are usually treated by a Lennard-Jones potential.

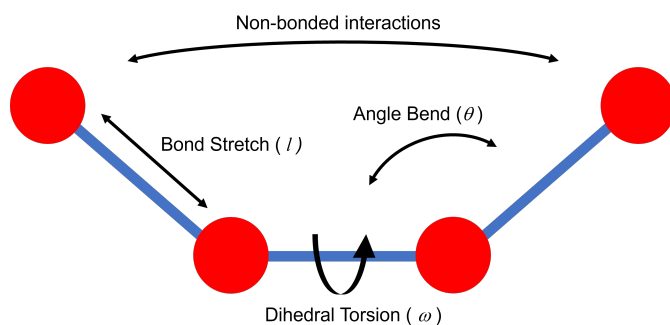


Figure 3.1: Intramolecular terms of force-fields.

$$V_{Total} = \sum V_{bonded} + \sum V_{angle} + \sum V_{torsion} + \sum V_{improper} + \sum V_{el} + \sum V_{vdW} \quad (3.15)$$

Bond stretching and angle bending can be expressed using simple harmonic motion:

$$V_{bonded} = \sum_{bond} \frac{1}{2} k_{bond} (l_0 - l)^2 \quad (3.16)$$

$$V_{angle} = \sum_{angle} \frac{1}{2} k_{angle} (\theta_0 - \theta)^2 \quad (3.17)$$

in which the force constants are denoted as k_{bond} and k_{angle} , for bonds and angles, respectively. The torsional term is typically described by a periodic function as follows:

$$V_{torsion} = \sum_{n=0}^N \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] \quad (3.18)$$

where V_n , n and γ hold for amplitudes, periodicity and phase factor, respectively.

Improper torsions are used to define the out-of-plane bending motions of groups of four atoms that are not sequentially linked. The mathematical description can be as follows:

$$V_{improper} = \frac{1}{2} k_{\omega} [1 - \cos 2\omega] \quad (3.19)$$

Atoms of different molecules, or in the same molecule separated by more than two bonds, involve non-bonded interactions.

In non-bonded interactions, for the electrostatic contribution, atoms are usually treated as point charges. The electrostatic interaction between the particles is calculated using Coulomb's law,

$$V_{el} = \sum_i \sum_j \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (3.20)$$

where q_i and q_j are the point charges on the atoms, while the distance between them is denoted as r_{ij} .

The non-bonded non-electrostatic, or van der Waals (vdW) interactions are generally weaker than the electrostatic ones and are relevant for particles in close contact. They are defined commonly by a Lennard-Jones function:

$$V_{vdW} = 4\epsilon_{AB} \left[\left(\frac{\sigma_{AB}}{r} \right)^{12} - \left(\frac{\sigma_{AB}}{r} \right)^6 \right] \quad (3.21)$$

where ϵ_{AB} is the well depth of the interaction between atoms A and B, σ_{AB} (called the collision diameter) represents the distance at which the intermolecular potential is zero. The attractive part (the R^{-6} term) and the repulsive part (the R^{-12} term) are related to dispersion and short-range repulsion interactions, respectively.

Many different force-field parameterizations are available in the literature. Force fields widely used for describing biomolecular systems include AMBER,[140] CHARMM,[141] and OPLS.[142]

3.4 Molecular Dynamics

Molecular Dynamics (MD) is a computational simulation technique to describe the time-dependent behavior of a chemical system. In an MD simulation, the particles are allowed to interact with each other and their movements are described by numerically solving Newton's equations of motion:

$$f_i(t) = m_i a_i(t) = -\frac{\partial V(x(t))}{\partial r_i(t)} \quad (3.22)$$

where the net force acting on the atom i at time (t) is denoted as $f_i(t)$, $a_i(t)$ is the acceleration and m_i represents the mass. The vector $r(t)$ represents the position of the N interacting atoms. In MD simulations, molecular mechanics force fields are often used, but quantum mechanical potentials can also be assumed.

The motion of the system is described by the classical Hamiltonian ($H(p_i(t), r_i(t))$):

$$H(p_i(t), r_i(t)) = \sum_{i=1}^N \frac{1}{2m_i} p_i^2 + V(r_i) \quad (3.23)$$

in which the momentum vector is denoted as $p_i(t)$ and $V(r_i)$ is the potential energy. The equations of motion are solved by using the partial derivation of Hamiltonian

$$-\frac{\partial H}{\partial r_i} = F \quad (3.24)$$

and Newton's 2nd law:

$$\frac{\partial^2 r_i}{\partial t^2} = \frac{F}{m_i} \quad (3.25)$$

There are many methods available to integrate the equations of motion, which is done by assuming a discrete time interval called the time step, ∂t . In the *Velocity Verlet* algorithm[143] the velocity of a particle (v) and its position (r) at time $(t + \partial t)$ are defined as:

$$r(t + \partial t) = r(t) + \partial t v t + \frac{1}{2} \partial t^2 m^{-1} F(t) \quad (3.26)$$

$$v(t + \partial t) = v(t) + \frac{1}{2m} \partial t (F(t) + F(t + \partial t)) \quad (3.27)$$

Initial coordinates, $r(0)$, for the system can be obtained from various sources (eg X-Ray crystallography studies or NMR structures for biomolecules), whereas initial velocities, $v(0)$ can be derived from the Maxwell-Boltzmann distribution.

In MD simulations, *Periodic Boundary Conditions* (PBC) are often used, which allow to describe extended systems such as condensed phases. Most often, non-bonded interactions are limited to a cutoff radius, and in that case, the Ewald sum method can be used to account for the neglected long-range electrostatic terms.[144]

Simulations can be carried out using different statistical ensembles in which the temperature (T), the volume (V), the number of particles (N) and/or the pressure (P) can be maintained constants:

- NVE: microcanonical ensemble - has constant N, V and E;
- NVT: canonical ensemble - has constant N, V and T;
- NPT: isothermal isobaric ensemble - has constant N, P and T

To control the temperature or the pressure, algorithms behaving as thermostats or barostats have been developed.

3.5 The QM/MM Method

The hybrid quantum mechanics/molecular mechanics (QM/MM) approach was firstly introduced by Warshel and Levitt in 1976. It allows studying chemical reactions in systems having a high number of atoms.[145, 146, 147] In this approach, two separated parts denoted as QM region and MM region are defined and allowed to interact with each other, as illustrated in Figure 3.2. The interaction between each region can be treated through two main approaches, namely, additive and subtractive schemes.

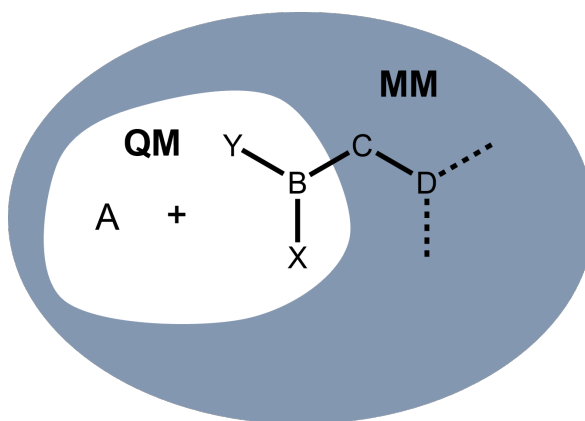


Figure 3.2: The system partition as two separated parts namely, QM and MM region.

In the additive scheme, the Hamiltonian is defined by the QM region contribution (H_{QM}), the MM contribution (H_{MM}) and their interaction ($H_{QM/MM}$):

$$H = H_{QM} + H_{MM} + H_{QM/MM} \quad (3.28)$$

in which the last term contains electrostatic and non-electrostatic terms:

$$H_{QM/MM} = \sum_k \frac{q_M Z_k}{R_{kM}} - \sum_{iM} \frac{q_M}{r_{iM}} + V_{QM/MM}^{vdw} \quad (3.29)$$

where q_M is the MM point charge, the nuclear charges in QM are denoted as Z_k , R_{kM} is the distance between an atom k in the QM part and an atom M in the MM region, and r_{iM} is the separation between QM electrons and MM atoms.

In the subtractive scheme, the interaction between QM and MM region is taken into account in a different manner. In this scheme (Equation 3.30), the total energy is calculated in three

steps. It involves 1) the calculation of the whole system energy at the MM level, 2) the calculation of the QM layer energy at the MM level, and 3) the calculation of the QM layer energy at the QM level. This scheme is used in ONIOM calculations (our Own N-layer Integrated molecular Orbital Molecular Mechanics), developed by Morokuma.[148, 149, 150, 151, 152] ONIOM has a unique feature over conventional QM/MM methods since it is based on a simple extrapolation procedure that allows it to be extended to any n-layer n-level scheme, such as two QM-layers ONIOM (QM1:QM2) or the three-layer ONIOM (QM1:QM2:MM). Thanks to the subtractive scheme, there is no requirement for communication between high and low layers, which provides simplicity of usage.

$$E_{ONIOM(QM:MM)}^{Total} = E_{(QM)}^{QM} + E_{(QM+MM)}^{MM} - E_{(QM)}^{MM} \quad (3.30)$$

The approach is available in different version that allow to account for electrostatic embedding or simply for mechanical embedding.

The treatment of frontier bonds in QM/MM methods deserves some comments (frontier bonds refer to chemical bonds in which one atom lies in the QM region while the other lies in the MM region). To treat technical issues related to the presence of such bonds, a general approach named “link atom method” can be used. In this method, a monovalent atom, mostly hydrogen, is placed aligned along the frontier bond, and allowed to interact with both regions. It can be used with both additive and subtractive schemes.

3.6 Free Energy Calculation Techniques

3.6.1 Umbrella Sampling

The free energy ($G(\xi)$) of a molecular system with respect to a reaction coordinate (ξ) is related to the probability ($P(\xi)$) of finding the system in a state corresponding to ξ :

$$G(\xi) = -k_B T \ln(P(\xi)) \quad (3.31)$$

in which Boltzmann constant is denoted as k_B and T is the temperature. Unfortunately, when the free energy changes are significantly larger than $k_B T$ the probabilities are very low, and it becomes impossible to obtain them from direct MD simulations. Biased techniques have been developed to solve this problem.

The umbrella sampling[153] is a technique in which the space is discretized into windows. A schematic representation of the method is given in Figure 3.3. The system is enforced to remain in a given state by adding a bias potential V_i centered on each window (i) to the unbiased energy (E^0), resulting in the biased energy (E^*). It can be summarized as

$$E^*(\xi) = E^0(\xi) + \sum_{windows} V_i(\xi) \quad (3.32)$$

An MD simulation performed using $E^*(\xi)$ yields the biased probability P_i^* in each window. The biased free energy is expressed from Equation 3.31 with P_i^* and the unbiased free energy can be obtained from the equation:

$$G_i(\xi) = -k_B T \ln(P_i^*(\xi)) - V_i(\xi) - F_i \quad (3.33)$$

where F_i is a constant associated with a given simulation. Therefore, the main issue in umbrella sampling is to find a suitable constant F_i for each window in order to rebuild the full free energy surface.

To overcome this issue, the weighted histogram analysis method (WHAM)[154, 155] is used. The free energy between two successive windows should be identical in a series of umbrella sampling simulations, if the probabilities of these windows overlap. Thus, if sufficient overlap is obtained between two successive windows, the F constant is optimized to minimize the difference in free energy. To sum up, umbrella sampling calculations require the next fulfillments:

- a reaction coordinate (ξ) should be chosen.
- a satisfactory overlap between two neighboring windows should be obtained, which depends on the width ($\delta\xi$) of each window, as well as on the parameter(s) defining the bias potential.
- the maximum of the biased probability distribution must be centered in the middle of the corresponding window.

3.6.2 Dual-level Approach

The dual-level approach has been developed to increase the accuracy in free energy calculations using the QM/MM partition while keeping the computational time within affordable limits.

The basic idea is to combine a low-level (LL) calculation of the free energy using for instance the umbrella sampling technique, and a free-energy perturbation approach to correct this free energy by using a high-level (HL) method on selected points of the profile.[156]

The free energy profile obtained at LL can be corrected assuming that the change from the LL to the HL Hamiltonian is a perturbation[157]. Then, the free energy at HL can be estimated through the equation[158]:

$$W_{HL}(\xi) - W_{LL}(\xi) = -\frac{1}{\beta} \ln \langle e^{-\beta \Delta U(\xi)} \rangle_{LL} \quad (3.34)$$

where

$$\Delta U^i(\xi) = U_{HL}^i(\xi) - U_{LL}^i(\xi) \quad (3.35)$$

represents the potential energy difference between the HL and LL for configuration i and β is the inverse temperature $(k_B T)^{-1}$. The average is calculated using a set of snapshots from the LL sampling selected at regular time intervals in the simulation and displaying a particular value of the reaction coordinate.

This is a key benefit of the strategy because it dramatically reduces the computational work required, as high-level calculations are only required at a few selected points of the reaction

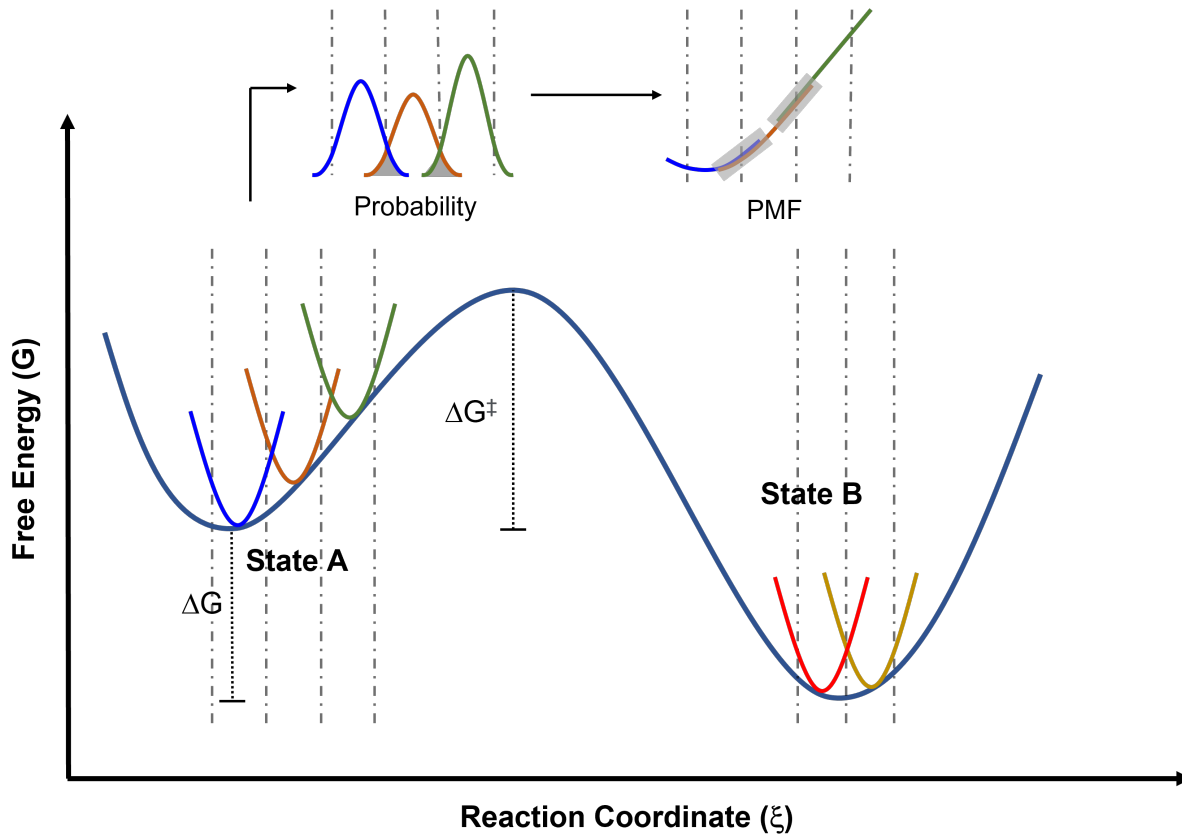


Figure 3.3: Representation of the umbrella sampling method used to get the potential of mean force (PMF) for a given the biased free energy profile with the computation of the biased probability in each window.

coordinate rather than on the entire domain. Rewriting the equation above using the fluctuations of the potential energy difference

$$\delta\Delta U^i(\xi) = U_{HL}^i(\xi) - U_{LL}^i(\xi) - \langle\Delta U(\xi)\rangle_{LL} \quad (3.36)$$

leads to the useful expression:

$$W_{HL}(\xi) - W_{LL}(\xi) = \langle\Delta U(\xi)\rangle_{LL} - \frac{1}{\beta} \ln \langle e^{-\beta\delta\Delta U(\xi)} \rangle_{LL} \quad (3.37)$$

Here, the first term represents a free-energy correction due to average differences on the potential energy, which in general is expected to provide the largest contribution. The second term, which contains the fluctuations with respect to the average, is connected to thermal corrections.

Chapter 4

Ester aminolysis mechanism of model systems via quantum mechanical approach

4.1 Introduction

In this chapter, the aim is to model the aminolysis mechanism of simple esters in aqueous solution to get insights into the aminolysis of PI3K δ by covalent inhibitors, since Lys779 (the residue targeted by the inhibitors) is located near a water-accessible surface.[12] Though the reaction in water is expected to present variations with respect to the reaction in the PI3K δ enzyme, the qualitative trends found for the series of ester derivatives (Figure 2.3 on Page 21) should provide useful information about the important factors that govern the inhibition process. Therefore, the reaction mechanism of some model systems, namely, methylamine (CH_3NH_2) with phenyl acetate (PA) and 4-nitrophenyl acetate (or *p*-nitrophenyl acetate PNPA) were studied. Methyl acetate (MA) has also been considered for comparison, and as a reference to evaluate the efficiency of phenyl derivatives in PI3K δ experiments.[12] A combined discrete-continuum model is assumed for the water solvent similar to that used before by Gorb *et al.*[26] for a related fundamental reaction (formamide hydrolysis), which reported results in excellent agreement with experimental kinetics data. Similar models have been used by other authors for other aminolysis processes.[45, 46, 50, 159, 160] Then, the results are analyzed in terms of the electronic properties of the leaving group to check the potential correlation with the experimental data for the series of model esters. Finally, the potential implications of our results[32] will be discussed. This chapter is based on a published paper "Mechanistic insights into lysine-targeting covalent inhibition through a theoretical study of ester aminolysis" in *Org. Biomol. Chem.*, 2021, 19, 9996-10004.

4.2 Computational Details

All calculations have been carried out by means of the Gaussian09 software package.[161] The default integration grid for DFT-based computations in Gaussian 09 (fine grid) has been used.

In this part of the study, a combined explicit-implicit, or discrete-continuum solvation model has been considered. Specifically, five explicit water molecules were included in the chemical system, while using the polarizable continuum solvation model (PCM, assuming the static di-

electric constant of water) to account for long-range solvation effects. The explicit-implicit solvation model has been chosen after some preliminary test computations and is similar to that used in related previous studies.[26] The inclusion of explicit water molecules is compulsory to correctly describe the mechanisms involving zwitterionic and charged species. Previous studies have shown that when a pure continuum solvation model is used, only neutral species-based mechanisms are found,[47, 51, 162] and the same applies for explicit-implicit models using a very small number of explicit waters. Some detailed studies on zwitterion stability as a function of explicit water molecules number have been reported.[46] Explicit consideration of five water molecules appear to be a good compromise between accuracy and computational cost.[50] Moreover, the initial position of the water molecules has been chosen based on the most plausible locations to form hydrogen bonds with the reactants, as well as using data reported in previous studies for related systems.[25, 46, 50] In most cases, different orientations have been explored, in particular, to describe proton transfer elementary steps. In these cases, the possibility of direct and water-assisted reactions has been contemplated.

Geometry optimizations, frequency, and intrinsic reaction coordinate[163] (IRC) calculations have been performed by means of density functional theory (DFT) based methods at B3LYP/6-311+G(d,p) level of theory. Electronic energies have been improved through single-point energy computations at the MP2/aug-cc-pVTZ levels on B3LYP/6-311+G(d,p) optimized geometries. Zero-point energy and thermal contributions calculated at the B3LYP/6-311+G(d,p) level were then added to these refined electronic energies to obtain total free energies (assuming a reference state of 1 M at T = 298 K). The nature of the stationary points, minima or transition structures (TS), was checked by a frequency analysis (zero or one imaginary frequency, respectively). The minima joined by each transition structure were determined through the IRC computations followed by full geometry optimization.

The discrete-continuum solvation model does not allow describing proton transfers to bulk water in detail but the relative energy of the process can be estimated by using the experimental free energy of the hydrated proton, as described in the supporting information of Derbel *et al.* ($G_{aq}(H^+) = -270.28 \text{ kcal mol}^{-1}$, reference state 1 M).[164] In addition, we have computed the free energy profile for the explicit proton transfer to water molecules in the first shell, forming a tetrahedral ion pair $T^{\cdot-} \dots (H_3O)^+$, which can be considered to be the first step in the proton transfer to the bulk water solvent.

Electrophilicity indices ω for the various ester molecules have been calculated at the B3LYP/6-311+G(d,p) level in gas phase as[165]:

$$\omega = \frac{\mu^2}{\eta} \quad (4.1)$$

where the chemical potential μ , and the chemical hardness, η , are approximated by:

$$\mu = \frac{1}{2}(\epsilon_{HOMO} + \epsilon_{LUMO}) \quad (4.2)$$

$$\eta = (\epsilon_{LUMO} - \epsilon_{HOMO}) \quad (4.3)$$

4.3 Results and Discussion

First, the mechanisms for the reaction of methylamine with MA, PA, and PNPA esters have been explored through their potential energy surfaces, trying to locate all the relevant stationary points to determine the possible pathways. These calculations have been guided by the results of previous studies on different esters.[25, 41, 47, 51, 52, 53, 54, 166] The different pathways found are shown in Figure 4.1. The nomenclature used thorough the paper for the different species is defined in the Figure caption.

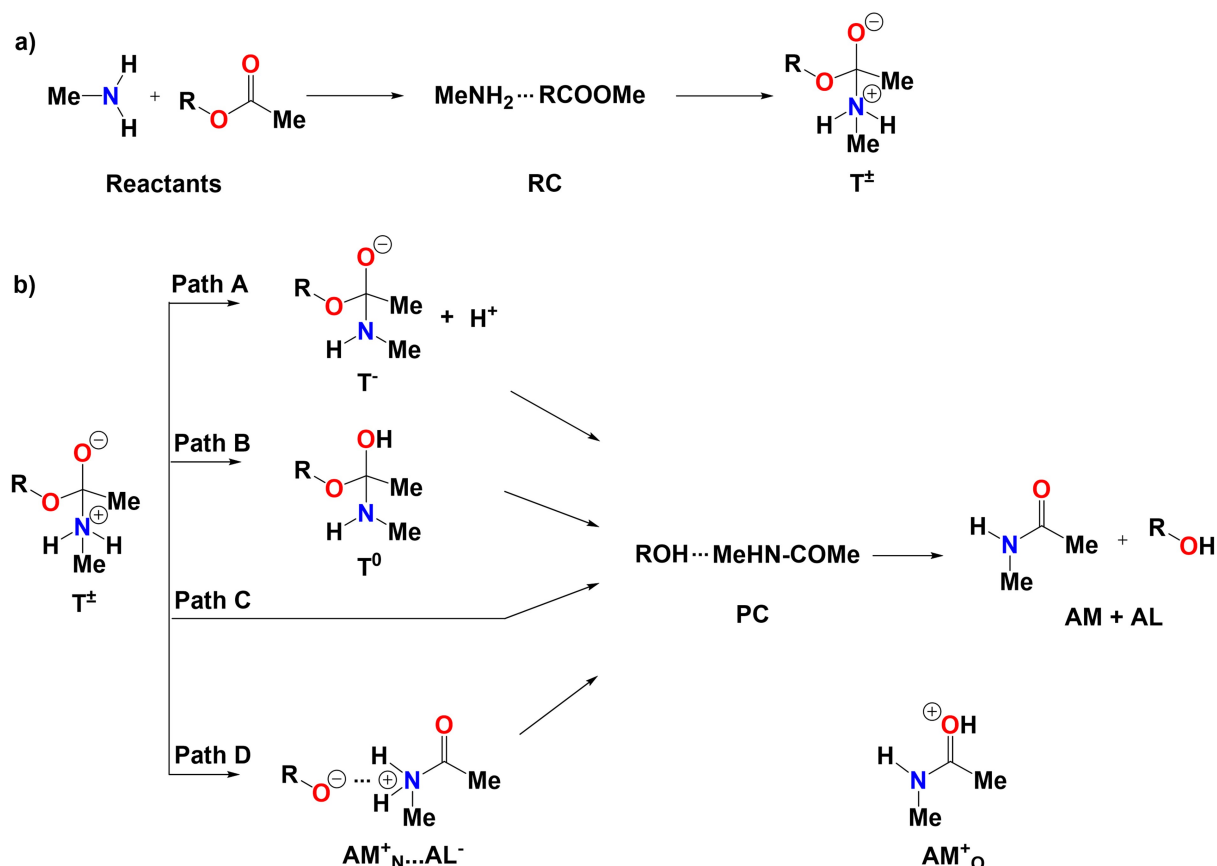


Figure 4.1: Reaction mechanisms for the ester aminolysis in water solution described in this chapter: (a) Formation of the pre-reactive complex RC and the zwitterionic tetrahedral intermediate $T^{\pm}$, (b) Mechanisms A-D leading from $T^{\pm}$ to the products. R is either the methyl, phenyl, or p-nitrophenyl group. The following acronyms are used throughout the chapter: RC = reactant complex, $T^{\pm}$ = zwitterionic tetrahedral intermediate, T^0 = neutral tetrahedral intermediate, T^{-} = anionic intermediate, AM^{+}_N = N-protonated amide, AM^{+}_O = O-protonated amide, AL^{-} = alkoxide intermediate, PC = product complex, AM = amide product, AL = alcohol product. AM^{+}_O (shown in the bottom right part of the figure) is a structure appearing only in some variants of mechanism B (see text).

All mechanisms start with the formation of a non-covalent reactant complex (RC), which then evolves into the characteristic covalent tetrahedral intermediate exhibiting a zwitterionic structure, $T^{\pm}$ (Figure 4.1a). This $T^{\pm}$ species results from the formation of a covalent bond

between the N atom of the amine and the C atom of the ester COO group. We have also explored the potential energy surface corresponding to the direct addition of the amine N-H group to the ester C=O or C-OR bonds, leading respectively to the formation of a neutral tetrahedral intermediate or the final products, but the computations for the three esters considered failed to locate any transition states for these hypothetical reaction pathways. From $T^\pm$, four main mechanisms (A–D) have been found, with some variants and specificities depending on the ester derivative, as described in detail below. Figure 4.1b presents them schematically. Mechanisms A–C corresponds to those usually expected for the aminolysis of non-activated esters, which all start with a proton transfer from the N atom but differ in the proton acceptor. The latter can be the water solvent (mechanism A), the negative O^- atom (mechanism B), or the leaving OR group in $T^\pm$ (mechanism C). In mechanism A (proton transfer to water), an anionic form of the tetrahedral intermediate (T^-) is formed, which then dissociates to the amide (AM) and to the alkoxide anion (AL^-). Finally, the alkoxide captures a proton from the medium to form the alcohol (AL). Mechanism A parallels that proposed in the experimental study of the PI3K δ inhibitors.[12] In mechanism B, the proton transfer leads to the formation of a neutral form of the tetrahedral intermediate (T^0), from which the system evolves to the final amide and alcohol products by subsequent proton transfer, from OH to OR group. In mechanism C, the proton transfer to OR leads in a single step to the amide (AM) and alcohol (AL) final products.

In addition to these pathways, in this study, another mechanism, called hereafter mechanism D, has been found in the case of activated phenolic esters. In contrast to the mechanisms (A–C), which start with a proton transfer, mechanism D starts by the heterolytic dissociation of the C–OR bond in the $T^\pm$ species forming an ion pair between the N-protonated amide (AM^+) and the leaving alkoxide (AL^-). As the incipient ion pair is being formed, the reaction proceeds (in a more or less concerted way depending on the derivative) by the proton transfer between the two moieties, either directly or through water molecules, which allows reaching the amide (AM) and alcohol (AL) products.

The free energy profiles for the studied reaction with MA, PA, and PNPA, are presented in Figures 4.2-4.4, respectively. The most relevant structures are drawn in Figures 4.5-4.7. In the description of the mechanisms, numbers are used as subscripts to differentiate between different conformations of a particular structure (e.g. T_1^0 , T_2^0 , $T_1^\pm$, $T_2^\pm$, etc.), or between different transition structures (e.g. TS_1 , TS_2 , etc.). $T_1^\pm$ will always represent the tetrahedral intermediate structure reached from RC, while $TS(RC-T_1^\pm)$ is the transition structure joining them. Note that, in mechanism A, the TS structures for proton transfer to/from bulk water cannot be located in the model, and for this reason, the corresponding reaction paths will be indicated as dashed lines in the Figures below. Likewise, for mechanism B in the case of PA and PNPA, the TSs leading to the formation of PC from ion-pair intermediates could not be located, and dashed lines are used in those cases too. Nevertheless, as shown below, mechanisms A and B have high barriers in other steps of the overall reaction pathway, and thus can be excluded as favorable mechanisms. For simplicity, TSs structures between different configurations and conformations of the same intermediate or product have not been searched. Note finally that, strictly speaking, in RC (reactant complex) and PC (product complex) there is not always a direct interaction between the ester and amine reagents, or between the amide and alcohol products.

For MA (Figure 4.2), mechanisms of types A, B, and C have been found. Two possibilities have been considered for mechanism A, depending on the location of the proton, first solvation shell or bulk water. In the first case, the ion pair $T^- \dots H_3O^+$ can evolve to form the neutral

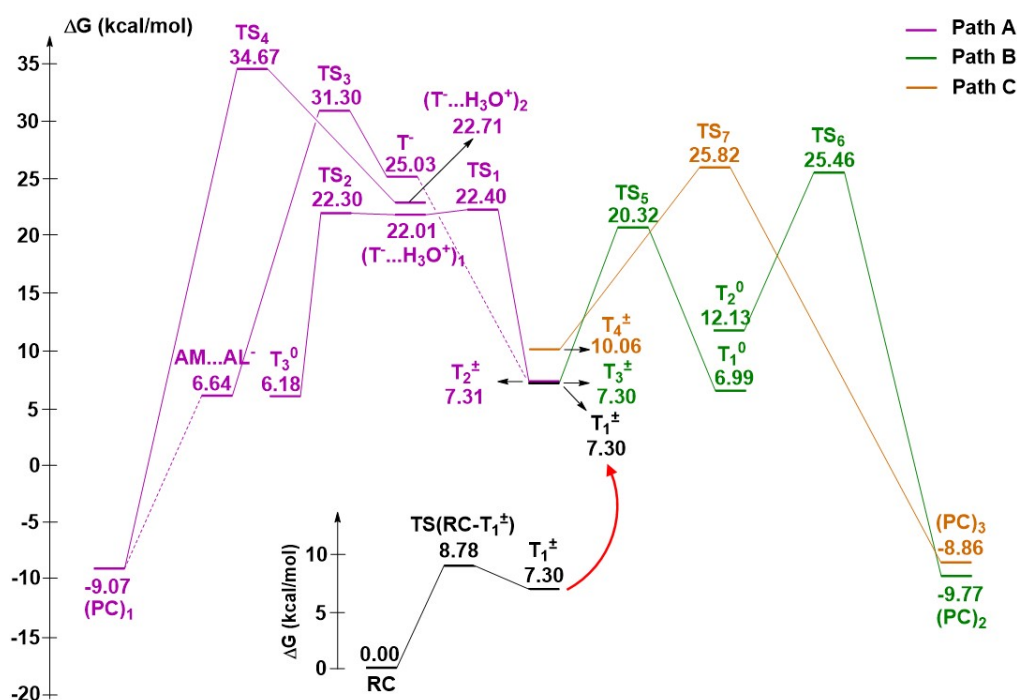


Figure 4.2: Free energy profile obtained for the methylamine + MA reaction in water at the MP2/aug-cc-pVTZ//B3LYP/6-311+G(d,p) level. For simplicity, the whole reaction path has been separated into two graphs, from RC to $T^\pm$, and from $T^\pm$ to the products.

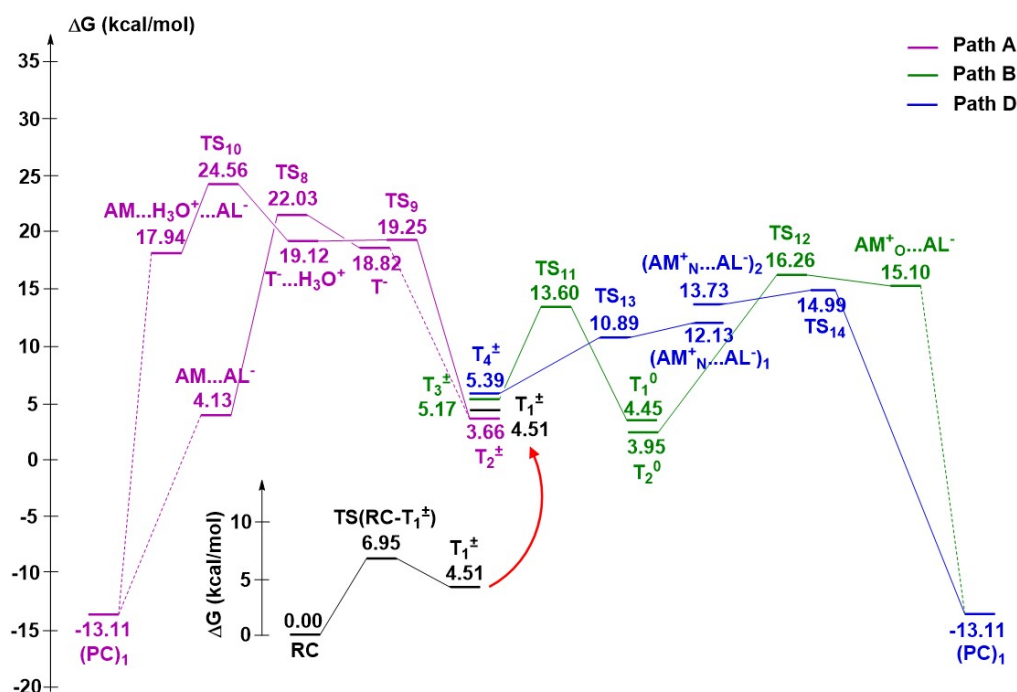


Figure 4.3: Free energy profile obtained for the methylamine + PA reaction in water at the MP2/aug-cc-pVTZ//B3LYP/6-311+G(d,p) level. For simplicity, the whole reaction path has been separated into two graphs, from RC to $T^\pm$, and from $T^\pm$ to the products.

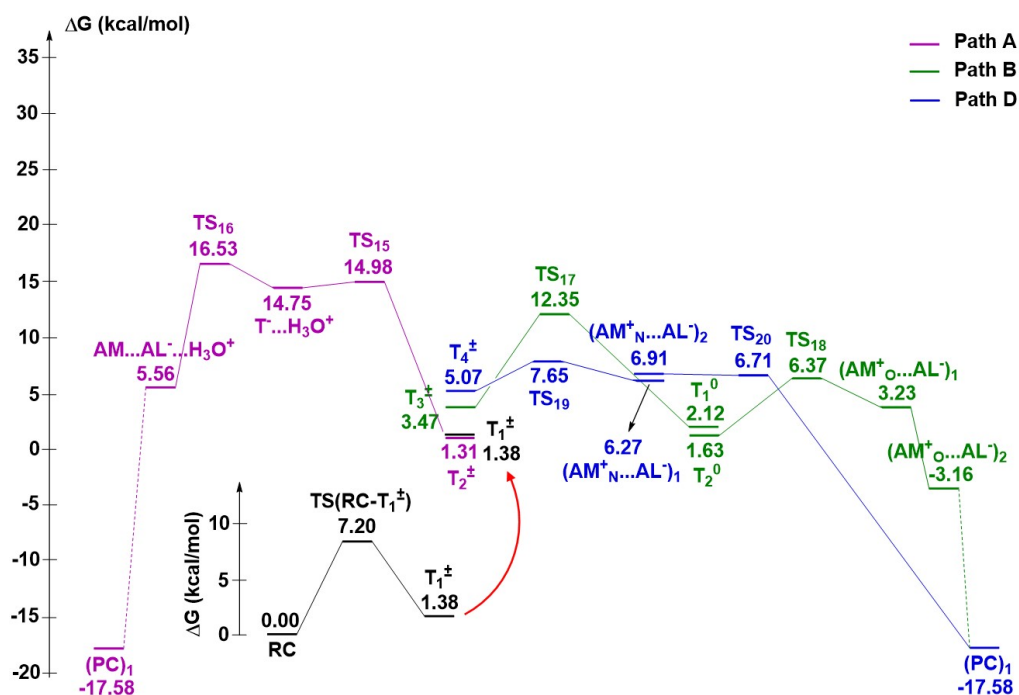


Figure 4.4: Free energy profile obtained for the methylamine + PNPA reaction in water at the MP2/aug-cc-pVTZ//B3LYP/6-311+G(d,p) level. For simplicity, the whole reaction path has been separated into two graphs, from RC to $T_1^\pm$, and from $T_1^\pm$ to the products.

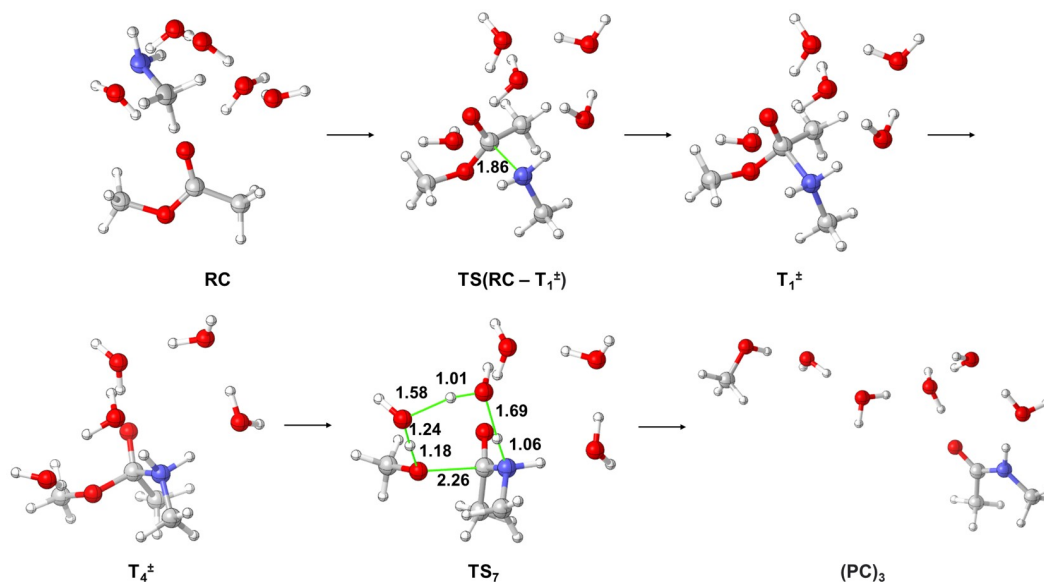


Figure 4.5: Main stationary structures along path C of MA aminolysis in water optimized at the B3LYP/6-311+G(d,p) level of theory.

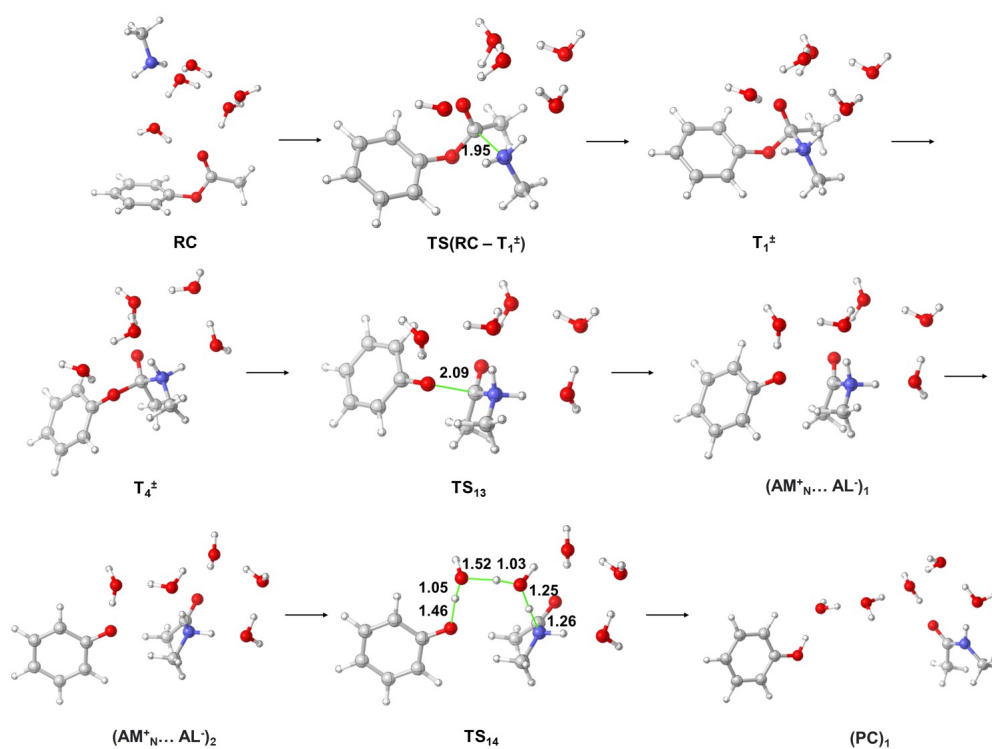


Figure 4.6: Main stationary structures along path D of PA aminolysis in water optimized at the B3LYP/6-311+G(d,p) level of theory.

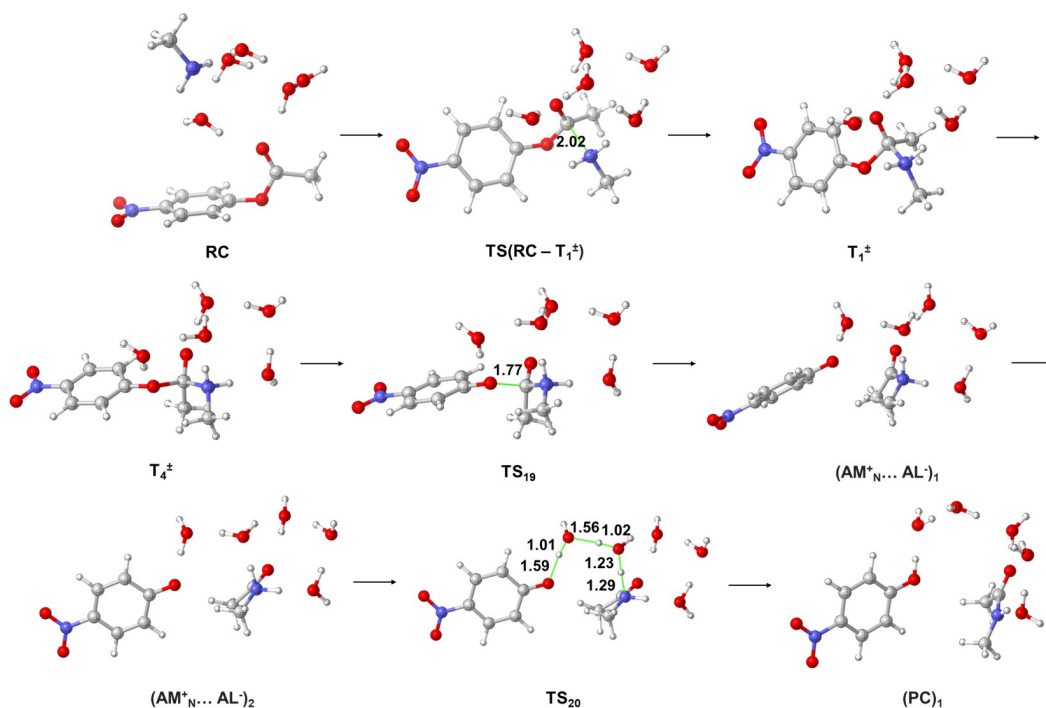


Figure 4.7: Main stationary structures along path D of PNPA aminolysis in water optimized at the B3LYP/6-311+G(d,p) level of theory.

intermediate (T_3^0) and from this point, the reaction is like mechanism B. In the second case, the proton has been fully transferred to the bulk and the reaction proceeds through dissociation of the remaining T^- intermediate, forming the amide product (AM) and the alkoxide anion (AL^-), which then captures the proton to form the alcohol (AL). Despite the possible errors introduced by combining MP2 and experimental energies for estimating the stability of the T^- structure (see the methodology section), it seems reasonable to conclude that mechanism A is more likely to occur via T_3^0 : this mechanism can be described as a variant of mechanism B in which the proton transfer from N to O^- occurs in a stepwise, water-assisted way. In all mechanisms A, B, and C, transition structures displaying high-energy (25 kcal mol⁻¹ or larger) are present somewhere along the reaction paths, suggesting a low reaction rate for the aminolysis of MA, as experimentally found.[167]

The reactions with the activated phenolic esters PA (Figure 4.3) or PNPA (Figure 4.4) displays a significantly different landscape, compared to MA. Mechanism C joining $T^\pm$ to the products directly has not been found in these cases. A variant of Mechanism B has been found for these esters where the dissociation of T^0 to AM...AL is not concerted (as in MA) but proceeds via an ion-pair intermediate ($AM^+O^- \dots AL^-$) involving the O-protonated amide and the alkoxide. It displays significantly lower activation energies compared to MA. Mechanism A has also been found, but as for MA, it does not appear to be a favorable pathway. An important result obtained in our study is that the intermediate T^- is not stable for the activated PNPA ester. Indeed, when the zwitterionic tetrahedral intermediate $T^\pm$ is deprotonated, the system spontaneously dissociates to form the alkoxide and the amide. A $T^- \dots H_3O^+$ ion pair intermediate has been located, however.

Mechanism D, which was not found for MA, is a stepwise analog of Mechanism C and appears to be the most favorable pathway for the phenyl derivatives, especially for PNPA. The highest TS in this mechanism lies at about 15 kcal mol⁻¹ (TS₁₄) for PA and only at about 7.7 kcal mol⁻¹ (TS₁₉) for PNPA.

Figure 4.8 summarizes and compares the most favorable mechanisms found for the aminolysis of the three esters. For PA and PNPA, mechanism D is the most favorable one, as mentioned. In the case of MA, mechanisms B and C should present a comparable kinetics. We have chosen to represent mechanism C in this Figure since it is the concerted analog of mechanism D. This comparison enlightens several important facts:

- the aminolysis reaction is considerably faster for the phenyl esters, compared to the alkyl ester,
- the presence of the nitro group in the phenyl ring considerably enhances the process,
- though the same mechanism type holds for PA and PNPA, the bottleneck in the two free energy profiles is different, being final proton transfer for PA (TS₁₄), and zwitterion formation (TS(RC- $T^\pm$)) or C-O bond dissociation (TS₁₉) for PNPA,
- the stability of the zwitterionic intermediates follow the order MA < PA < PNPA, and
- the reactions are found to be exergonic, with higher reaction energy (in absolute value) in the same order as MA < PA < PNPA.

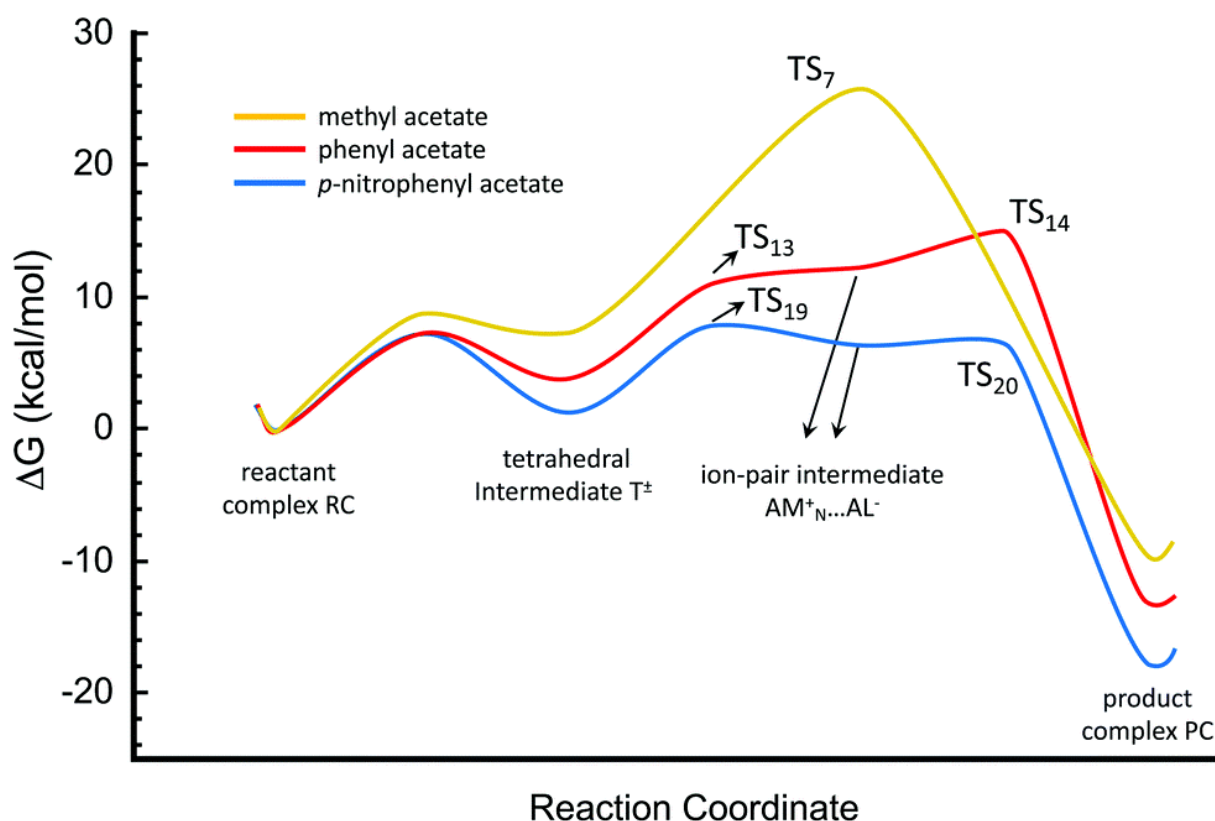


Figure 4.8: Schematic representation of the most favorable reaction mechanisms for the aminolysis of MA, PA, and PNPA esters in water, as predicted by MP2/aug-cc-pVTZ//B3LYP/6-311+G(d,p) calculations. In this schematic representation, we have systematically considered the most stable conformation/configuration of the reaction intermediates and product complexes.

Experimentally, activated phenolic esters have shown promising properties to be used as PI3K δ inhibitors, with the *p*-nitrophenyl derivative displaying the most interesting properties among the different esters shown in Figure 1.5. The findings confirm the high reactivity of this ester and also explain the strong affinity to covalently bind Lys residues in the enzyme. They also reveal that different reaction mechanisms might be at play for different inhibitors. Mechanisms of type D should be involved in the case of the most efficient covalent inhibitors.

Since mechanism D involves the formation of an alkoxide intermediate, the acidity of the corresponding alcohol is expected to be key in determining the type of the aminolysis reaction pathway. Maude *et al.*[28] have reported an experimental study of the alcohol pK_a influence on aminolysis reactions in acetonitrile, and found that a change of mechanism intervenes at pK_a around 8-9. Our computations confirm a change of mechanism in water too, possibly at a slightly higher pK_a value, since we found different pathways for MA ($pK_a = 15.2$) and PA ($pK_a = 9.95$) or PNPA ($pK_a = 7.15$). Dalton *et al.*[12] tried to relate the inhibitor activity and the alcohol pK_a and reported a subtle relationship. Roughly, they found that K_I is dependent on the leaving group pK_a , while k_{inact} is basically independent. This finding was interpreted as the fact that the zwitterionic intermediate will proceed rapidly to the products in an irreversible way.

Table 4.1: Comparison of calculated and experimental properties for selected esters. Calculated electrophilicity index in gas phase ω in eV (B3LYP/6-311+G(d,p) level), zwitterion intermediate formation energy ΔG in kcal mol⁻¹ (MP2/aug-cc-pVTZ//B3LYP/6-311+G(d,p) level), and experimental pK_a values[168, 169] apply for the model esters shown in Figure 2.3 (on page 21). Kinetic parameters[12] K_I (M), k_{inact} (s⁻¹) and k_{inact}/K_I (s⁻¹M⁻¹) correspond to esters shown in Figure 1.5 (on page 11).

R	ω	ΔG	pK_a	K_I	k_{inact}	k_{inact}/K_I
<i>o,p</i> -DiMePh	2.10	5.20	10.60	7.8×10^{-6}	0.0066	8.46×10^2
<i>p</i> -MeOPh	2.12	4.67	10.05	2.2×10^{-6}	0.0068	3.09×10^3
Ph	2.36	4.51	9.95	3.5×10^{-7}	0.0058	1.66×10^4
<i>p</i> -FPh	2.65	3.31	9.89	2.6×10^{-7}	0.0055	2.12×10^4
<i>p</i> -CF ₃ Ph	3.30	3.15	9.39	2.5×10^{-7}	0.0070	2.80×10^4
<i>p</i> -NO ₂ Ph	5.88	1.38	7.15	4.0×10^{-8}	0.0075	1.88×10^5

To get a deeper insight into the relationship between ester properties and reactivity, two main properties have been computed for the esters as models for the inhibitor candidates. These properties are the electrophilicity index ω of the ester in the gas phase (B3LYP/6-311+G(d,p)) and the free energy difference ΔG between RC and the zwitterionic intermediate T[±] in water (MP2/aug-cc-pVTZ//B3LYP/6-311+G(d,p)). The same water model as in the calculations presented above is used in these calculations. Figure 4.9 displays the correlations between calculated and experimental properties. Since the experimental k_{inact} appears to be very similar for the different esters, the correlations have been done using the second order rate constant k_{inact}/K_I , which is used to characterize irreversible inhibitors and should capture the global reactivity. As shown in Figure 4.9, and despite the limited number of available experimental data, we found that the whole set of inhibitor kinetic data and calculated ester properties separate into two different groups corresponding to phenolic derivatives bearing electron-donating (*o,p*-diMePh, *p*-MeOPh) or electron-withdrawing substituents (*p*-FPh, *p*-CF₃Ph, *p*-NO₂Ph), with phenol being in the crossing point. It is worth noting that the correlations of $\ln(k_{\text{inact}}/K_I)$ with either ω or ΔG are quite similar, and indeed a comparable correlation is obtained when pK_a is used instead of ω or ΔG , which is not surprising considering the strong interdependency between all these properties. Electron-donating groups usually have lower electrophilicity and higher pK_a compared to electron-withdrawing groups. The split correlation found here (two straight lines intersecting at one point) is quite similar to the correlations found by Maude *et al.*[28] between the measured rate constants of the aminolysis reactions in acetonitrile and the esters pK_a . As in this latter case, our study suggests that there is a change in mechanism between the phenolic esters bearing electron-donating groups and those carrying electron-withdrawing groups, which strongly increases the inhibition kinetics.

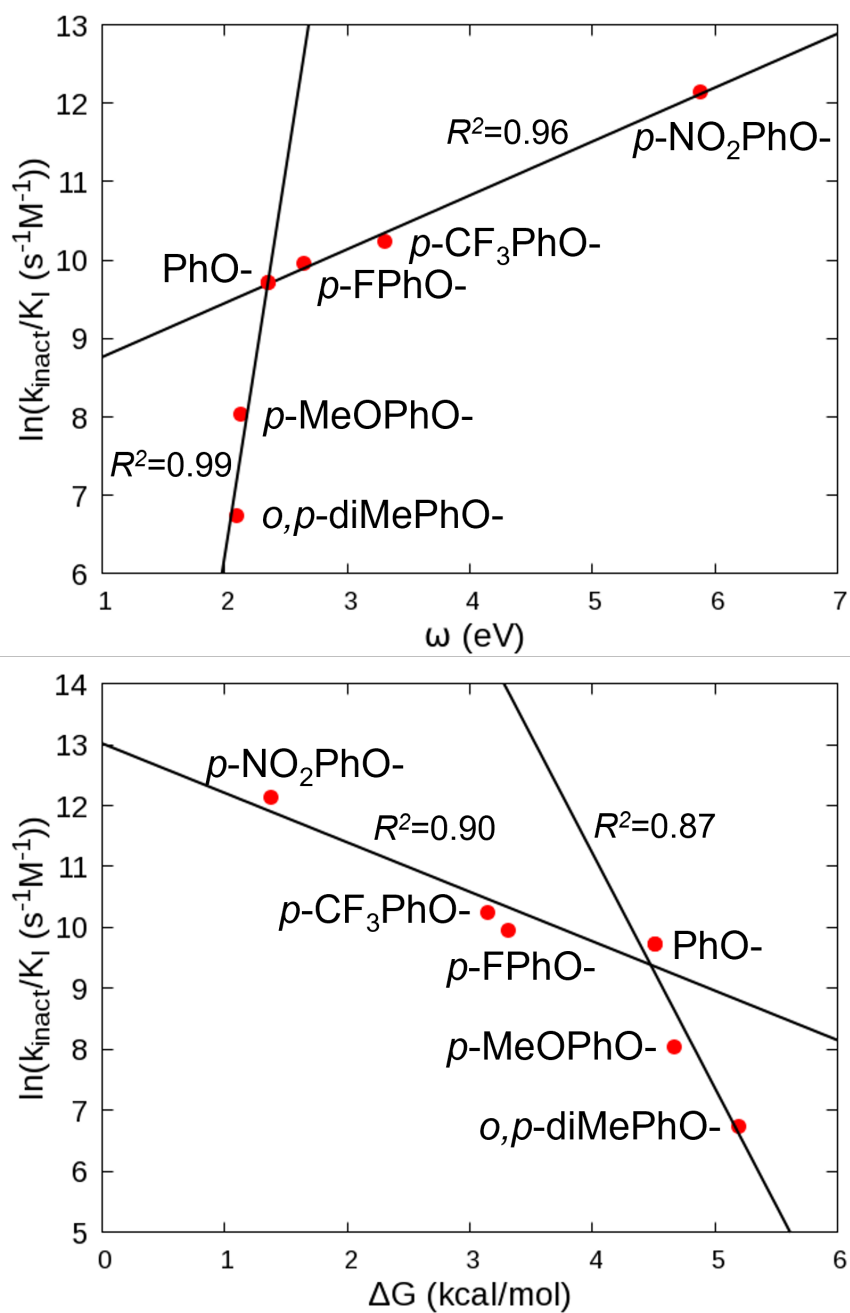


Figure 4.9: Correlations between calculated and experimental ester properties collected in Table 4.1

4.4 Conclusion

The analysis of the different aminolysis pathways of activated phenyl esters studied in our work reveals that the mechanism schematized in Figure 2.1 (on page 20) for PI3K δ inhibitors needs to be revised. Our calculations indicate that for sufficiently high electrophilic derivatives, such as PNPA, the main reaction coordinate following the formation of the zwitterionic tetrahedral intermediate is not proton transfer, as in mechanism A proposed before for PI3K δ inhibitors, but C-OR bond dissociation, leading to the formation of an ion pair $AM^+ \dots AL^-$. The development of this important charge separation then provides the driving force for proton transfer and the formation of the final products, which takes place afterward, along a rather flat free energy surface. This mechanism is in line with previous experimental studies in acetonitrile, which showed that a change of mechanism occurs for leaving groups with a sufficiently low pK_a . This study indicates that the most efficient inhibitors with phenolic leaving groups should correspond to derivatives bearing electron-withdrawing substituents, which activate the ester favoring the formation of the starting zwitterionic intermediate and rendering the subsequent reaction steps easier. Overall, the experimental inhibitor efficiency has been found to be nicely correlated with the calculated electrophilicity of the ester, and with the stability of the zwitterion. Therefore, the calculation of these properties should be very useful to evaluate a priori the efficacy of other inhibitor candidates.

Chapter 5

Ester aminolysis mechanism of model systems via QM/MM molecular dynamics simulations

5.1 Introduction

In Chapter 4, plausible aminolysis mechanisms for ester derivatives have been explored involving five explicit water molecules within the polarizable continuum solvation model. However, in order to gain insight into these mechanisms and obtain the associated free energies more accurately, an investigation involving dynamic effects of the environment should be performed. Full ab initio molecular dynamics represents the most elaborated approach to achieve this goal but the associated computational cost is excessively large and approximated methods have to be used in general.

The pivotal method to be used in our work is based on MD simulations with combined quantum mechanics and molecular mechanics (QM/MM) partitions. In the QM/MM dual-level version used here, the idea is to combine different QM theoretical levels (low and high levels). The low-level is used in the sampling step, while the high-level is used afterward to calculate properties using perturbation theory. Within this computational scheme, it is possible to get free energy profiles at high levels, which is often required to obtain chemical accuracy in the calculation of kinetic properties.[156, 158] A typical calculation thus involves the calculation of free energy profiles by the umbrella sampling (US) technique using a simple QM/MM approach, followed by the calculation of free energy corrections using Free-Energy Perturbation theory (FEP) and a high-level QM/MM method to reach the required accuracy. The low QM level is typically a semiempirical method, while the high QM level is typically a DFT or MP2 method.

In this chapter, we have carried out a study of the reaction between *p*-F-phenyl acetate and *p*-NO₂-phenyl acetate (FPA and PNPA, respectively) with methyl amine in water solution. The purpose of this study is to obtain accurate free energies for this process and to serve as a benchmark study to select the appropriate computational strategy for the enzymatic reaction, which will be presented in Chapter 8.

The direct dissociation pathway, denoted as mechanism D, has been considered, as it has been found to be the most plausible pathway. As a reminder, mechanism D involves two steps:

the first one is the nucleophilic attack of the N atom of the amine to the carbonyl carbon atom of the ester, followed by the direct dissociation of the phenolate leaving group from the zwitterionic structure, as depicted in Figure 4.1 on page 37. In our QM/MM calculations, the ester derivative and the methyl amine have been chosen as the QM subsystem, while water molecules lie in the MM part. Other computational details are presented in the following section.

5.2 Computational Details

The Amber16 software[170] is used for the QM/MM MD simulations. The reaction profiles of each step were obtained using the Umbrella Sampling (US) method. The reaction coordinate is defined as follows. For the first step of the process, the distance between the carbon atom of the ester group and the nitrogen atom of the amine has been chosen. The US incremental distance is 0.05 Å, and the force constant between two atoms is 500 kcal/mol/Å². For the second step, the US constraint is the distance between the carbonyl carbon atom and the oxygen atom of the leaving group. The same incremental distance and force constant have been used. As an indication, note that these calculations lasted approximately 18 CPU hours using 12 cores for each US window.

To thermalize the system, 500 minimization steps and 50 ps heating simulation are first applied to the initial system, which is followed by 200 ps of equilibration. Then, 100 ps of production are performed for each constrained distance in the NPT ensemble. We use a 0.2 fs time step. As mentioned above, the QM layer consists of the ester and the amine, while the MM region consists of the surrounding water molecules, which have been described here by the TIP3P model.

In order to select the most efficient dual-level QM/MM approach several benchmark calculations have been carried out. The details will be given below in the Results and Discussion section. Finally, PM3 has been chosen for the low-level QM calculations in the US sampling simulations and the M062X DFT-based method for the high-level QM calculations in the free energy corrections. Free energy corrections have been obtained for several points along the reaction coordinate. For each point, 1000 snapshots are extracted from the corresponding US trajectory keeping the water molecules lying within 10 Å from the ZW, as illustrated in Figure 5.1. The high-level computations are then performed and the free-energy corrections are calculated according to the dual-level approach described in the Theoretical Background chapter. These QM/MM calculations at DFT level were done using the conventional AMBER/Gaussian interface and they have lasted approximately 2 minutes of CPU in 12 cores for each snapshot (total 1000 snapshots) to obtain the corrected energy of one point on the reaction coordinate.

5.3 Results and Discussion

5.3.1 Benchmark calculations

For the QM method to be used in the US simulation, different semi-empirical methods (DFTB3, DFTB2, DFTBA, PM6, PM3, PM3MM, PDDG, and AM1) have been tested. For this purpose, relaxed scans in the presence of 5 explicit water molecules has been carried out for the reaction

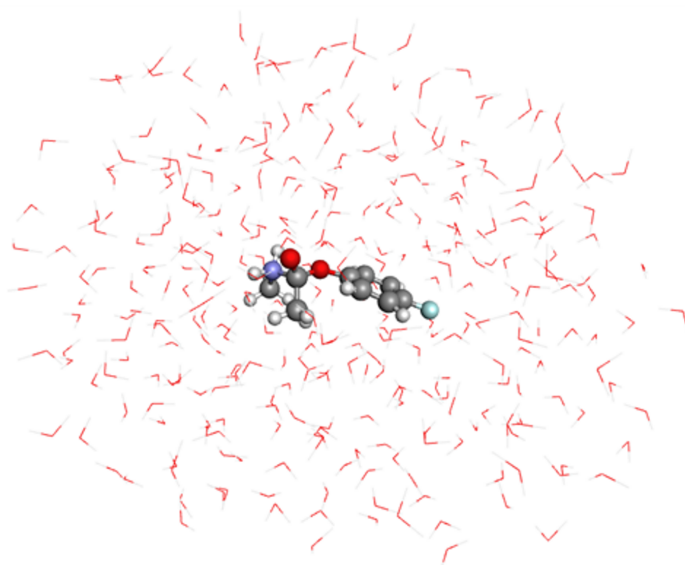


Figure 5.1: A representative snapshot containing the zwitterionic (ZW) structure surrounded by water molecules within 10 Å. The illustrated structure belongs to the reaction of methylamine and *p*-F-phenylacetate (FPA).

of FPA with methylamine. In these “static computations”, we have not used the PCM solvation model (in contrast to Chapter 4) since in Gaussian Software it is not available for DFTB methods. Hence for consistency, all calculations were conducted in vacuum. In this benchmark study, the performance of the semi-empirical methods listed above has been compared to the B3LYP/6-311+G(d,p) level of theory, since this method was used in our preceding study. The results for the zwitterion formation step are reported in Figure 5.2. As shown, most DFTB methods cannot locate the minimum energy point for the ZW structure. AM1 method locates the ZW structure but the calculated reaction barrier is excessively high. In the case of PM6 and DFTBA, the equilibrium CN distance in ZW is much longer than in B3LYP. In contrast, PM3 and PDDG (a variant of PM3) are successful to locate ZW and the potential energy surfaces are comparable to that of B3LYP. An additional study has been carried out to check the appropriateness of the PM3 level of theory. From the relaxed scan pathways obtained with PM3, a single point energy calculation has been performed with B3LYP/6-311+G(d,p). The resulting energy profile is plotted in Figure 5.3, and as shown, it displays a quite good agreement with the relaxed scan calculated at the B3LYP/6-311+G(d,p) level. Based on all these findings, it can be said that the PM3 method appears as a convenient technique to be used as the low-level QM region in QM/MM MD simulations.

Next, in order to decide on the high-level QM method for the dual-level corrections, the performance of various DFT methods (B3LYP, B3LYP-D3, M062X, and ω B97XD with 6311+G(d,p) basis set) has been assessed. In this case, the reaction for methyl acetate with methyl amine has been considered. The MP2 method is chosen as reference here. Using higher reference levels would be computationally too expensive due to the presence of five explicit water molecules and the necessity to explore the complex potential energy surface associated to such a multimolec-

ular cluster. The energy differences between the reactant complex (RC) and the zwitterionic intermediate (ZW) are compared in Table 5.1, and the energy barriers for the rate-determining step, i.e. the leaving group dissociation, are tabulated in Table 5.2. Based on these data, the M062X functional, which gives the lowest relative error in the rate-determining step kinetics, has finally been chosen.

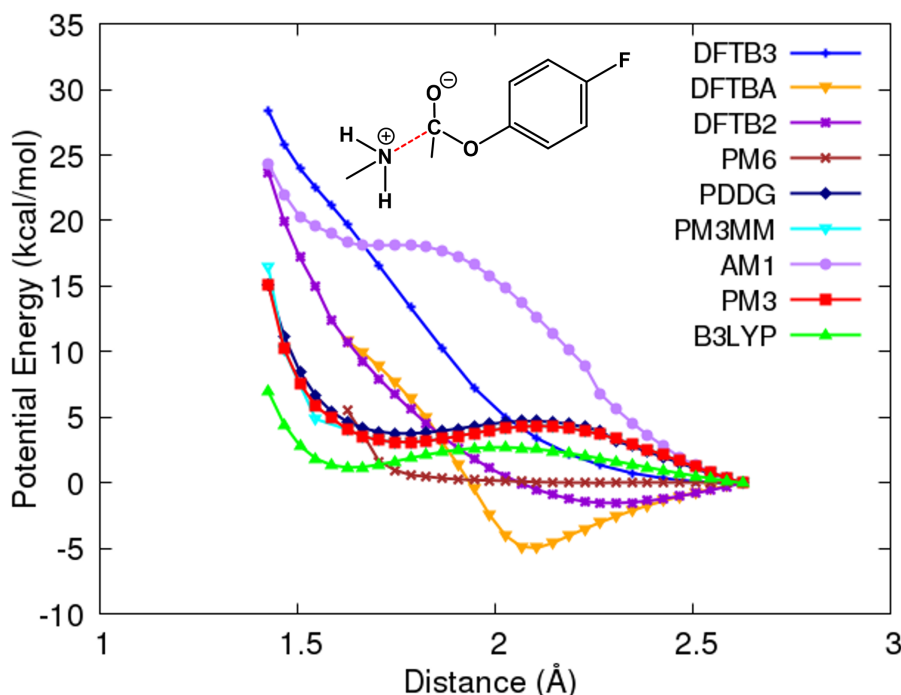


Figure 5.2: Relaxed scan results as a function of the C-N distance using different QM methods with 5 explicit water molecules in vacuum.

To complete this section devoted to benchmark calculations, and for the sake of comparison with the results presented below, we report in Figure 5.4 the energy profiles for the reactions of FPA and PNPA with methylamine at the M062X/6-311+G(d,p) level using the discrete continuum solvation model (as before, with 5 water molecules). Note that the reaction with PNPA was studied in Chapter 4 at the combined MP2//B3LYP level and the corresponding profile was presented in Figure 4.8 on page 43. As shown, the agreement between the two approaches is satisfying.

5.3.2 Study of the reaction between FPA or PNPA with methylamine

Free Energy Profiles via PM3/MM MD Simulations

First step: ZW formation. The free energy profiles for *p*-F-phenyl acetate (FPA) and *p*-NO₂ phenyl acetate (PNPA) are reported in Figure 5.5 for the ZW formation step. As shown, in both cases the transition state is located at a C-N distance around 2 Å and the zwitterionic tetrahedral structure at around 1.75 Å. The QM/MM simulations do not predict a free energy minimum for the reactant complex (RC) in explicit water solution but the rest of the free energy profile is

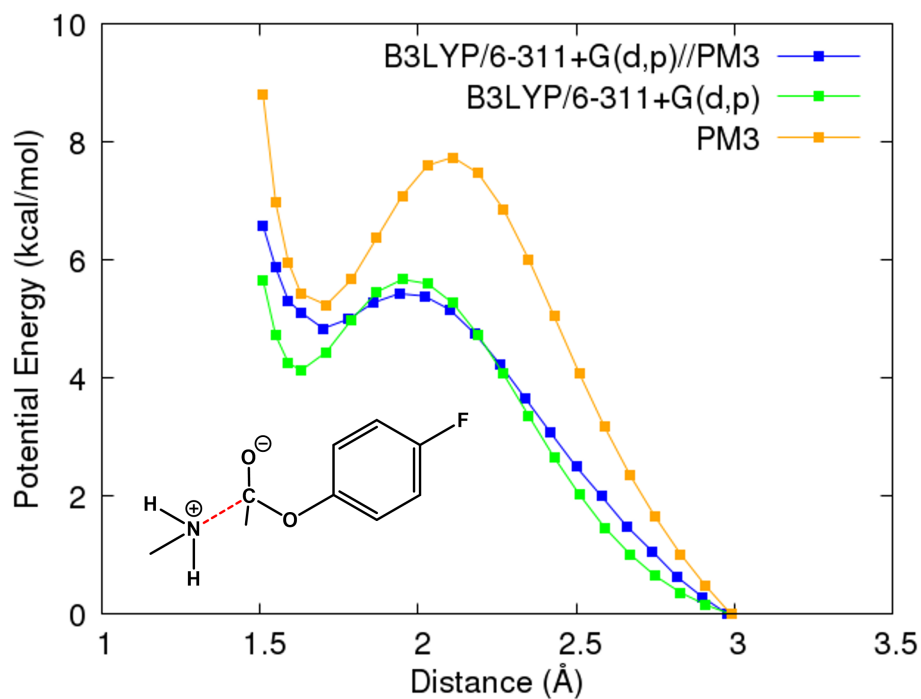


Figure 5.3: Relaxed PES as a function of the C-N distance with discrete continuum solvation model at different level of theories.

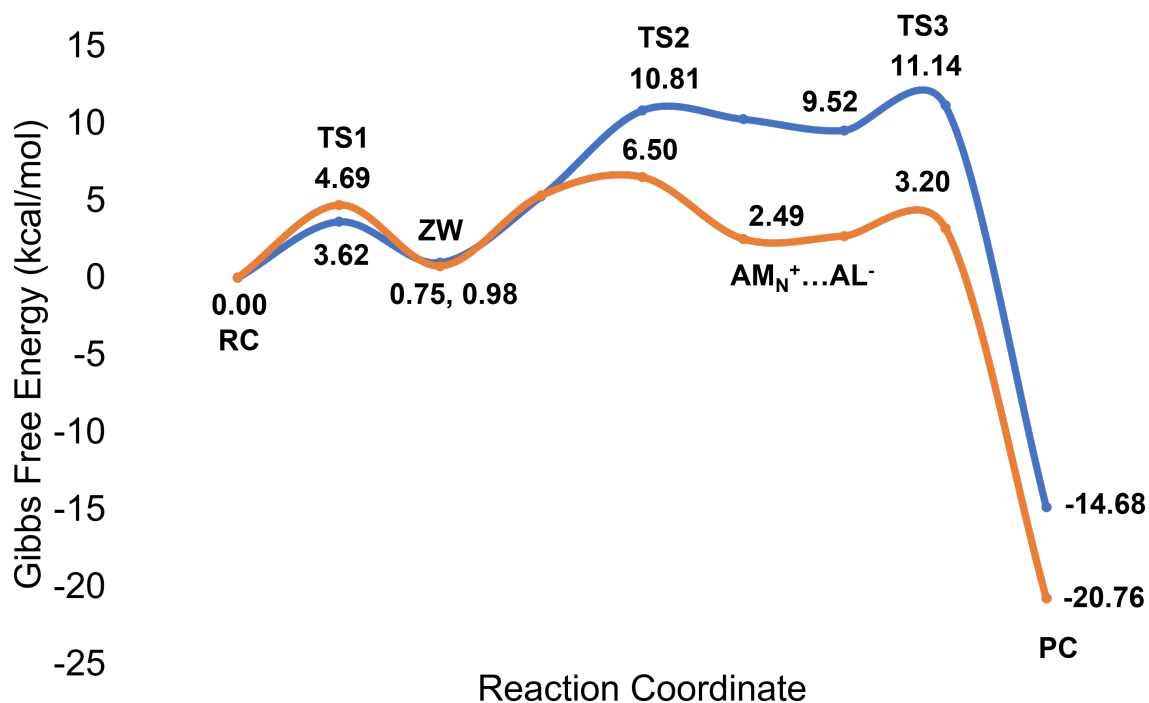


Figure 5.4: Reaction profiles for FPA (blue) and PNPA (orange) with methylamine at M062X/6-311+G(d,p) level of theory using a discrete continuum solvation model with 5 water molecules.

Table 5.1: Energy differences (kcal/mol) between the zwitterionic intermediate (ZW) and the reactant complex (RC) for the reaction of methyl acetate with methyl amine in the presence of 5 explicit water molecules without PCM model, calculated with different DFT functionals and the 6-31+G(d,p) basis set. Errors with respect to MP2 calculations are tabulated.

Method	Energy Difference		Relative Error	
	ΔE_{ZW-RC}	ΔG_{ZW-RC}	ΔE_{ZW-RC}	ΔG_{ZW-RC}
MP2	2.30	8.80	0.00	0.00
B3LYP	9.30	15.80	7.00	7.00
B3LYP-D3	5.86	10.93	3.55	2.13
M062X	1.59	5.54	-0.72	-3.26
ω B97XD	3.91	8.71	1.61	-0.10

found to be similar to the profile at the QM level in the implicit-explicit model system (Figure 5.4). Representative snapshots of ZW structures are reported in Figures 5.6 and 5.7, for FPA and PNPA, respectively. One may note that the activation barrier is slightly lower for the nitro derivative (18.09 kcal/mol and 17.32 kcal/mol for FPA and PNPA, respectively). Likewise, the ZW energies are lower in the case of the nitro compound (17.6 kcal/mol and 16.25 kcal/mol for FPA and PNPA, respectively). In other words, the presence of NO₂ on para position appears to favor the process at this computational level.

The radial distribution functions (RDF) of the carbonyl oxygen atom of each ester species with water H-atoms have been calculated and are reported in Figure 5.8. In the ZW structure, the oxygen atom becomes formally negatively charged and the interaction with water is strong, as revealed by the RDFs. The slightly lower magnitude of the RDFs first and second peaks in the case of PNPA can be explained by a larger delocalization of the charge in this case due to the stronger electron-withdrawing capacity of the nitro group. Figure 5.9 displays the RDF of two H-atoms on nitrogen atom with oxygen atoms of surrounding water molecules. There is no significant difference between two systems. The reason for it can be attributed to the positive charge on H-atoms, which is not delocalized over the system as in the case of the negative charge.

Second step: leaving group dissociation. The second step involves the dissociation of the phenolate leaving group from the ZW, which we study by increasing the C...O distance in the US simulations. The range of available CO distances is limited however by the spontaneous occurrence of “chemical events” at specific CO values that bring the system to a different potential energy surface, complicating the interpretation of the US sampling in terms of free energy differences. Two different issues appeared. At short CO distances close to the equilibrium value in ZW, US simulations led to dissociation of the amine, i. e. the system tends to form the reactants, which can be explained by the small reverse barrier found in the ZW formation step (see above). At long CO distances, when the phenolate anion is formed, a spontaneous proton transfer from the N atom is obtained in the case of the FPA derivative (but not in the

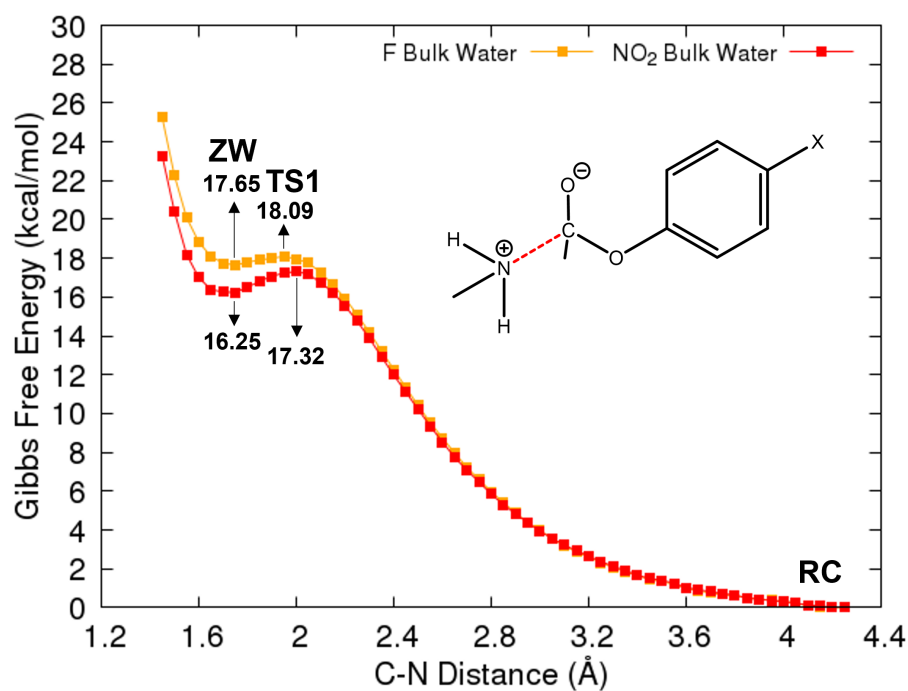


Figure 5.5: Gibbs Free Energy profiles of the first step of the aminolysis reaction of *p*-F-phenyl acetate (yellow) and *p*-NO₂ phenyl acetate (red) by means of US PM3/TIP3P calculations in water solution.

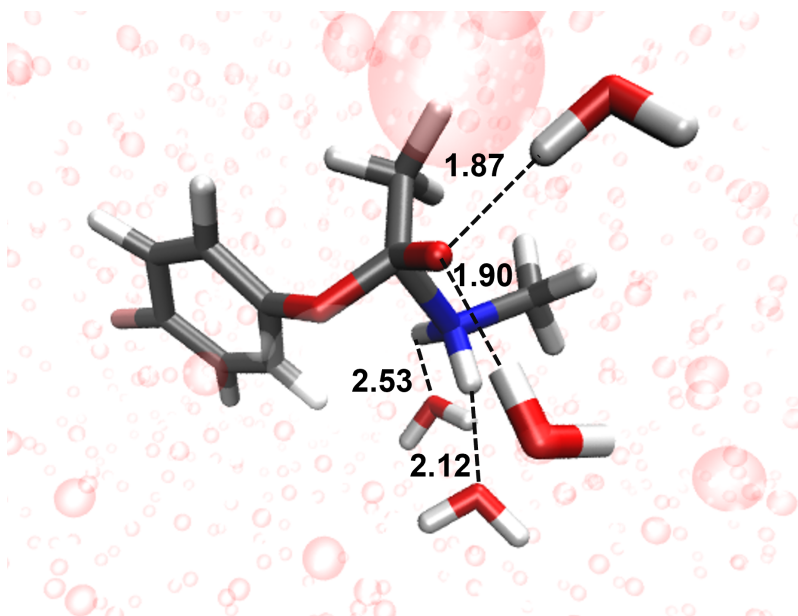


Figure 5.6: Representative MD snapshot from the ZW structure formation of FPA by means of US PM3/TIP3P calculations in water, which is illustrated as bubbles.

Table 5.2: Activation energies (kcal/mol) for the rate-determining step of the reaction of methyl acetate with methyl amine in the presence of 5 explicit water molecules without PCM model, calculated with different DFT functionals and the 6-31+G(d,p) basis set. Errors with respect to MP2 calculations are tabulated.

Method	Activation Energy		Relative Error	
	$\Delta E^\ddagger$	$\Delta G^\ddagger$	$\Delta E^\ddagger$	$\Delta G^\ddagger$
MP2	20.72	16.92	0.00	0.00
B3LYP	16.94	13.97	-3.78	-2.95
B3LYP-D3	16.95	12.77	-3.77	-4.15
M062X	21.53	17.33	0.81	0.42
ω B97XD	21.45	16.25	0.73	-0.67

case of the PNPA derivative). In principle, to solve these issues, a multiple constraint could be used, for instance using a combination of C...O, C...N and N...H distances in the reaction coordinate, but the calculation of such a free-energy hypersurface would be too costly. To avoid this problem, as said above, the range of C...O distance variation was limited. The starting and final C...O distance values were 1.7 Å and 2.75 Å for FPA, and 1.65 Å and 3.70 Å for PNPA, respectively. Final values were chosen so that the ion pair is already formed and the proton transfer to form the final products has not yet started. The obtained reaction profiles are reported in Figure 5.10.

The fact that the ion-pair in the case of PNPA derivative appears to be more stable in the case of the PNPA derivative, compared to FPA (i.e. we do not observe a proton transfer to form the alcohol in PNPA even for large C...O distances) is consistent with the results reported in the previous chapter and with the lower pK_a of the nitro phenol derivative compared to the fluoro phenol derivative, as reported in Table 4.1 on page 44.

Radial distribution functions for the interaction between the oxygen atom on the leaving group and water hydrogens are analyzed in Figure 5.11 for both FPA and PNPA. We use the simulations for a reaction coordinate equal 2.75 Å, just before proton transfer occurs in FPA. As shown, the first peak for FPA is significantly higher than the first-peak in PNPA suggesting that the phenolate O atom carries a much larger negative charge in this case, which facilitates the proton abstraction observed at 2.80 Å.

Corrected Free Energy Profiles via Dual-Level Approach

The refined free energy profiles using the dual-level FEP method at the M062X/6-311+G(d,p) level of theory are shown in Figures 5.12 and 5.13 for FPA and PNPA, respectively. Globally, the activation energies for ZW formation decrease significantly at the high-level of theory. For instance, for the FPA derivative, the first step reaction barrier is 18.09 kcal/mol at the PM3 level and 11.31 kcal/mol after correction at the M062X level. As predicted at the PM3 level, the

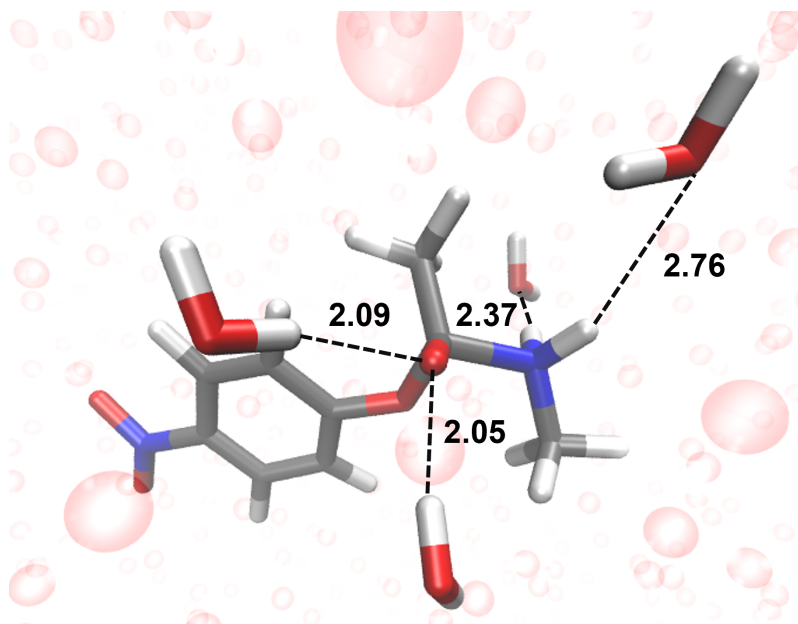


Figure 5.7: Representative MD snapshot from the ZW structure formation of PNPA by means of US PM3/TIP3P calculations in water, which is illustrated as bubbles.

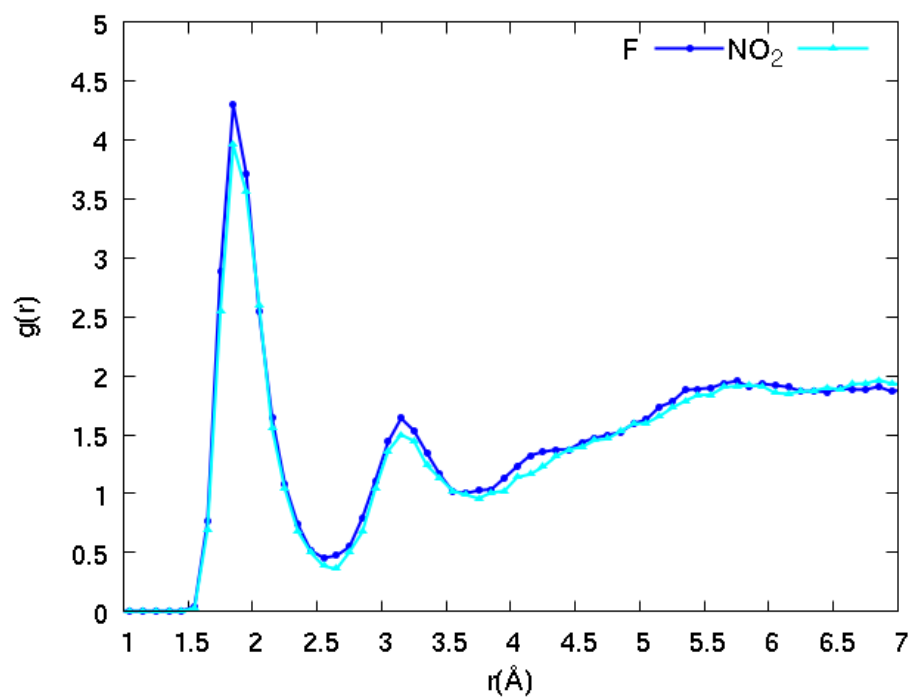


Figure 5.8: Radial Distribution Function of carbonyl oxygen of FPA and PNPA with water H-atoms for ZW structure.

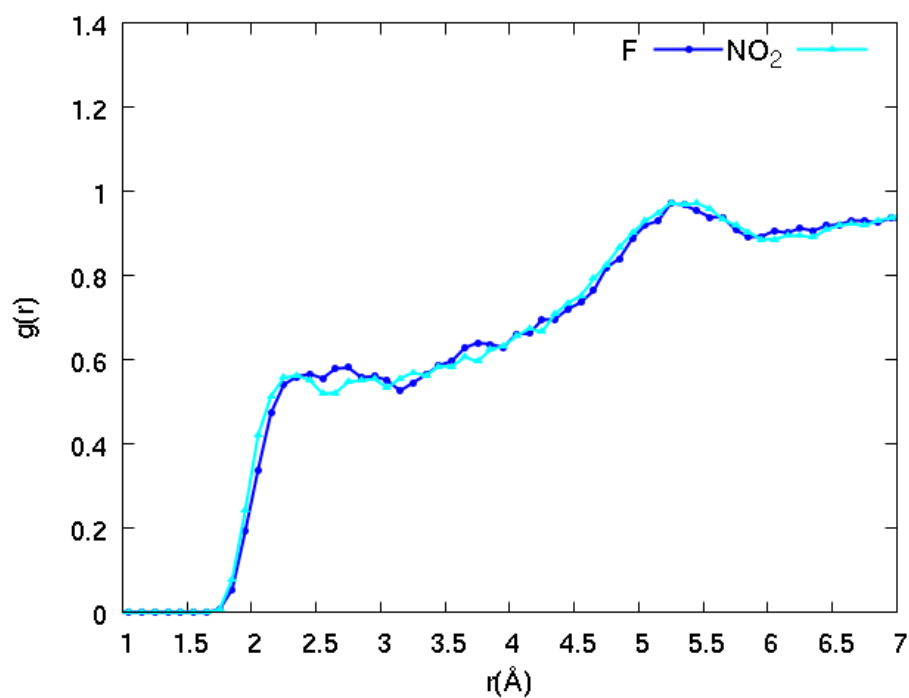


Figure 5.9: Radial Distribution Function of H-atoms on nitrogen atom of FPA and PNPA with the O atoms of water for ZW structure.

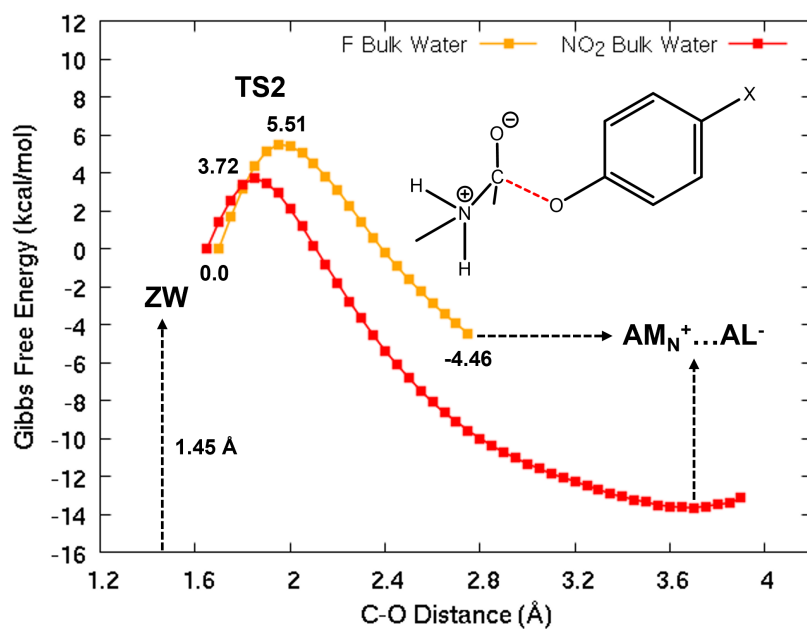


Figure 5.10: Gibbs Free Energy profiles of the second step of the aminolysis reaction of FPA (yellow) and PNPA (red) by means of US PM3/TIP3P in bulk water.

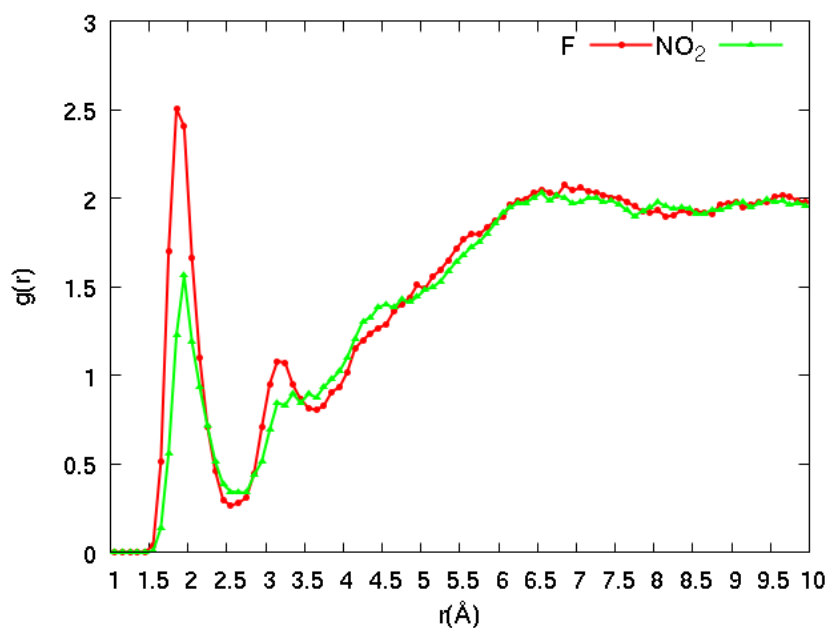


Figure 5.11: Radial Distribution Function of phenolate oxygen of FPA and PNPA with water H-atoms obtained when the C...O reaction coordinate is equal to 2.75 Å (ion pairs are formed).

reaction is significantly more favorable with PNPA, as the two barriers (for the first and second steps) are lower than for FPA.

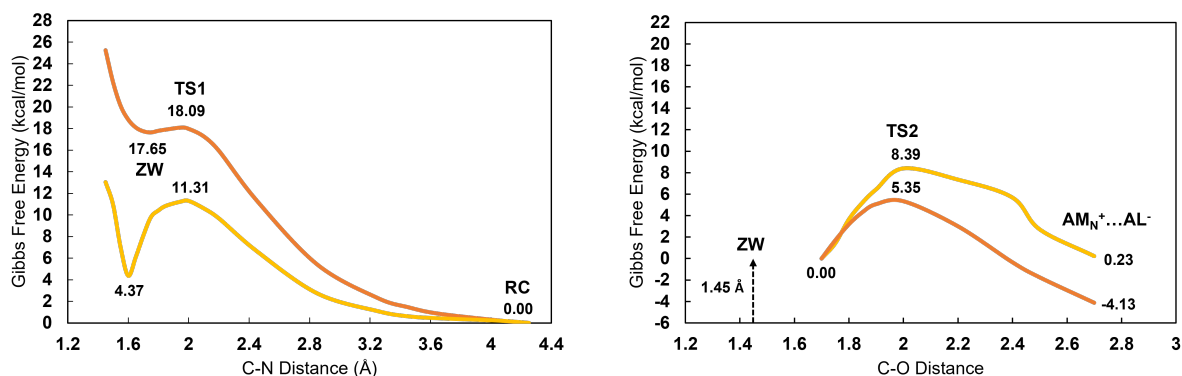


Figure 5.12: The comparative free energy profiles of 1st (left) and 2nd (right) steps of the reaction for FPA derivative. Orange line represents the US simulation done by PM3, yellow line is FEP corrected results.

5.4 Conclusion

An efficient protocol for dual level QM/MM MD simulations of ester aminolysis reactions has been established in this chapter. According to our benchmark studies, a combination of

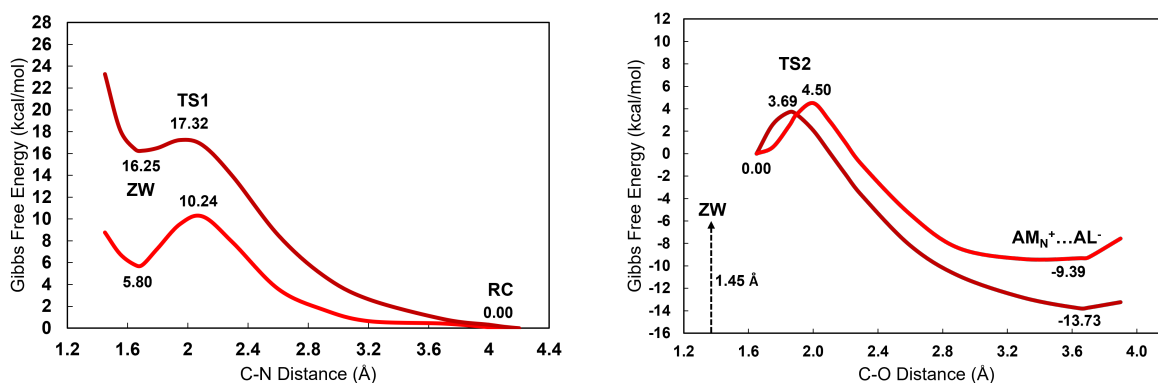


Figure 5.13: The comparative free energy profiles of 1st (left) and 2nd (right) steps of the reaction for PNPA derivative. Dark red line represents the US simulation done by PM3, red line is FEP corrected results.

the semiempirical method PM3 for the low-level, and the DFT-based M062X/6-311+G(d,p) method for the high-level has provided satisfactory results.

Reactions of PNPA and FPA with methylamine have been studied with this protocol. According to the free energy diagrams obtained, the zwitterion formation barrier is found to be more favorable for PNPA than for FPA. Likewise, the leaving group dissociation barrier for PNPA appears to be lower than for FPA, though the latter calculations are complicated by the fact that the range of variation of the C...O distances is limited due to the spontaneous occurrence of reactive events in some cases. Specific solution to get the complete free energy profile will be proposed in Chapter 8. The phenolate anion of PNPA is found to be significantly stable, which can be explained by the delocalization of the electronic charge into the nitro group. In the case of FPA, as the phenolate forms, fast abstraction of the proton from the amino group has been predicted.

Chapter 6

Classical molecular dynamics simulations of the PI3K δ enzyme

6.1 Introduction

In the literature, computational studies on PI3K enzymes cover several molecular docking, QSAR and virtual screening applications,[14, 16, 17] as well as MD simulations for the PI3K α and PI3K γ isoforms[18, 19] and the catalytic domain of PI3K δ . [20] In 2019, Zhu *et al.* studied the binding modes of an inhibitor and four PI3K isoforms and the results revealed the key residues of PI3K γ for selective inhibition.[171] Sabbah *et al.*[18] employed molecular mechanics/generalized born surface area (MM/GBSA) calculations to get the binding affinities of clinical drugs, namely wortmannin and LY294002 for PI3K α and PI3K γ , respectively. They identified the specific residues which can be used as targets for ligand design.

In this chapter, we have carried out classical MD simulations to analyze the interaction between the proposed inhibitors and the PI3K δ enzymatic environment. The X-ray structure of the complex formed by the PI3K δ enzyme with methyl 5-(4-(5-((4isopropylpiperazin-1-yl)methyl)oxazol-2-yl)-1H-indazol-6-yl)-2-methoxynicotinate (inhibitor 8 in Figure 1.5 on page 11) has been used as the starting point. The crystal structure was taken from Protein Data Bank (PDB Number: 6EZ6).[12] The main goal is to select an appropriate inhibitor/enzyme initial structure for the calculations of the reaction mechanisms that will be discussed in the next two chapters (ONIOM and QM/MM MD calculations), and to analyze the underlying non-covalent interactions.

6.2 Computational Details

In order to predict the secondary structure of missing loop parts, homology modeling was applied by the Swiss model.[172] The obtained complete structure of the enzyme and its crystal structure were superimposed and %99.5 similarity was found between the two structures (Figure 6.1). PI3K δ is a relatively large enzyme with a structure weight of 121.34 kDa and consists of 1051 residues. Therefore, for exploring the reaction center, we focused on its kinase domain only. However, to make sure that dealing with the kinase domain instead of the full enzyme does not change much its structural features, snapshots after 100 ns MD simulations of the full

enzyme and of the kinase domain were superimposed. As seen in Figure 6.2, RMSD between the two structures is very small. Hence, we used the MD results of the catalytic domain for further investigation, which permitted us to avoid expensive computations.

The catalytic kinase domain of the enzyme (residues between 675-1032) was truncated and classical MD simulation was applied to this structure. A solvent box was created (85.23 x 90.42 x 89.41 Å³) with TIP3P water models. For the inhibitor, a non-standard residue was parametrized by the RESP protocol. The GAFF force field[173] was used for the inhibitor, while the ff14SB force field[140] was selected for the rest of the enzyme during the simulations. The thermalization of the system was carried out as follows: first, 20000 minimization steps were performed to the complex, after which a heating trajectory was applied during 500 ps followed by an equilibration trajectory of 500 ps. A 100 ns production simulation was carried out using a time step of 2 fs in the NPT ensemble (at 298K). Amber16 Software[170] was used for MD simulations. The pK_a analysis was done using the propkatraj.py utility in the MDAnalysis software.[174] This python script is based on PROPKA 3 method [175] and allows to estimate pK_a values of selected protein residues within an MD trajectory.

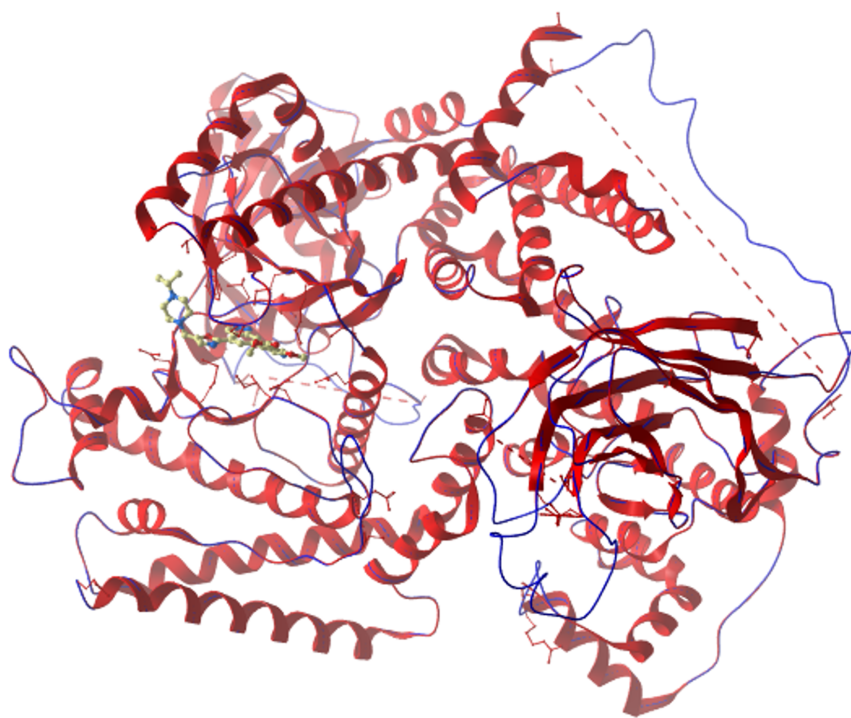


Figure 6.1: Alignment of experimental crystal structure (PDB ID:6EZ6) (red) and full enzyme model obtained from Swiss model (blue) (3318 of 3452 atoms superimposed, similarity= 99.5%).

6.3 Results and Discussion

The MD simulation results were examined by standard analysis tools considering the average values of electrostatic energy, root-mean-squared deviation (RMSD) and root-mean-squared fluctuation (RMSF), and radius of gyration (Rg) profiles throughout the trajectories. RMSD and RMSF are reported in Figure 6.3. As shown, RMSD stabilizes after 25 ns and remains below 3

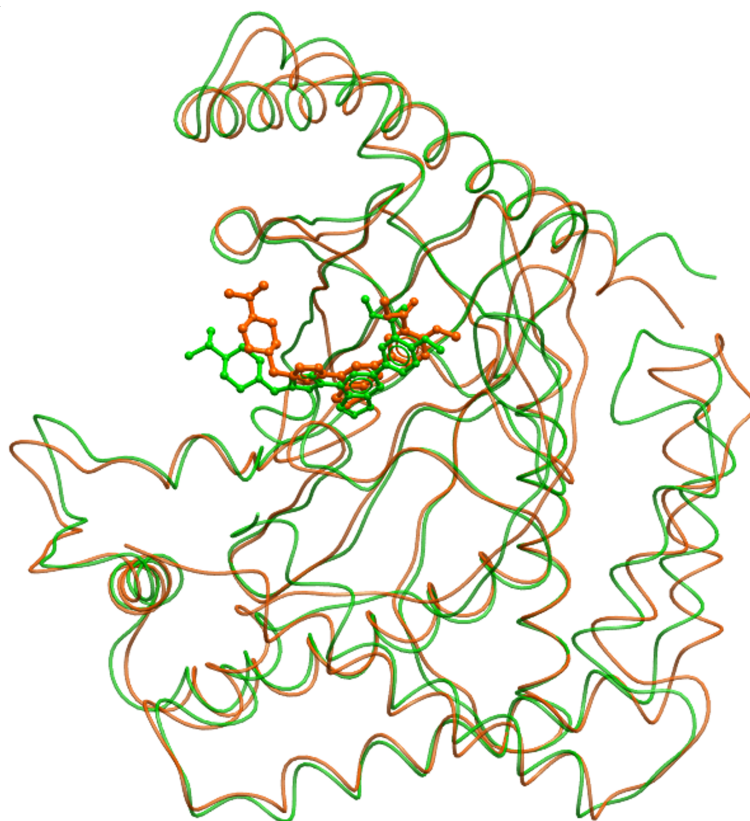


Figure 6.2: Alignment of kinase domains after 100 ns simulation of full enzyme and of catalytic kinase domain model (976 of 1432 atoms superimposed, rmsd_full_weighted=0.591).

Å until the end of the 100 ns simulation. This result shows that the enzyme-inhibitor complex gains its internal stability after 25 ns. RMSF plot indicates that several residues fluctuate more than other parts of the enzyme. Such high fluctuations are expected for the disordered loop regions in enzymes and this enzyme contains several loops in residue chains Glu767, Phe922, Leu987 which are consistent with RMSF plot. Moreover, the electrostatic interaction energy between the inhibitor and the enzyme is studied in Figure 6.4. The fluctuation of interaction energies is low and the average value is -27.05 kcal/mol, even if the complex is relaxed during the simulation. The radius of gyration (Rg) of a protein is a measure of its compactness. It remains stabilized between 22.1-22.6 Å until the end of the 100 ns simulation, suggesting stability of its folding over time.

We also analyzed the distance between the atoms involved in the covalent bond formation between the inhibitor ester group and the targeted Lys779 residue of the enzyme. Thus, the

distance between the nitrogen of side chain of Lys779 and the carbonyl carbon of the inhibitor was calculated during the simulation (Figure 6.5). Interestingly, this distance is around 4.0-4.5 Å in most of the frames, revealing the stability of the complex and the high frequency of proximity between reacting atoms, which is the first requirement for covalent inhibition of the enzyme to take place.

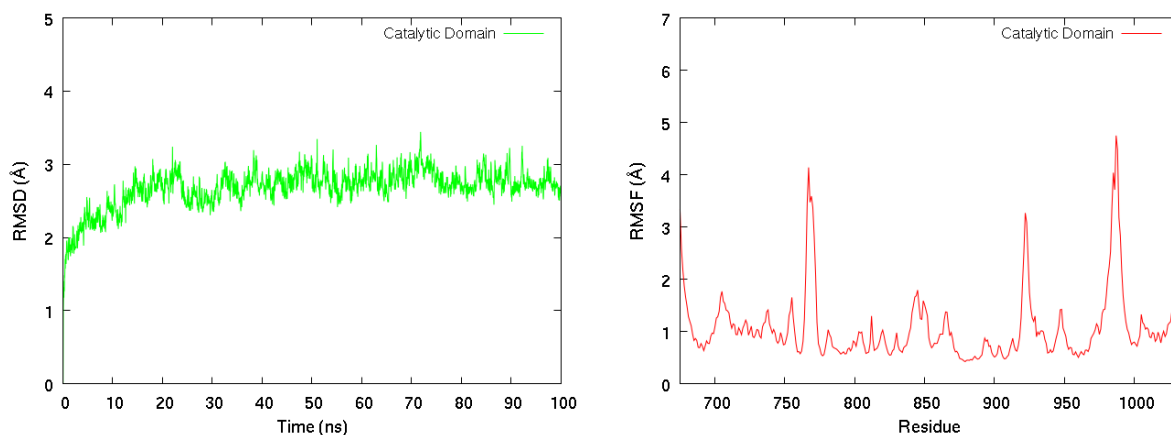


Figure 6.3: RMSD and RMSF results of the catalytic domain for 100 ns MD simulation.

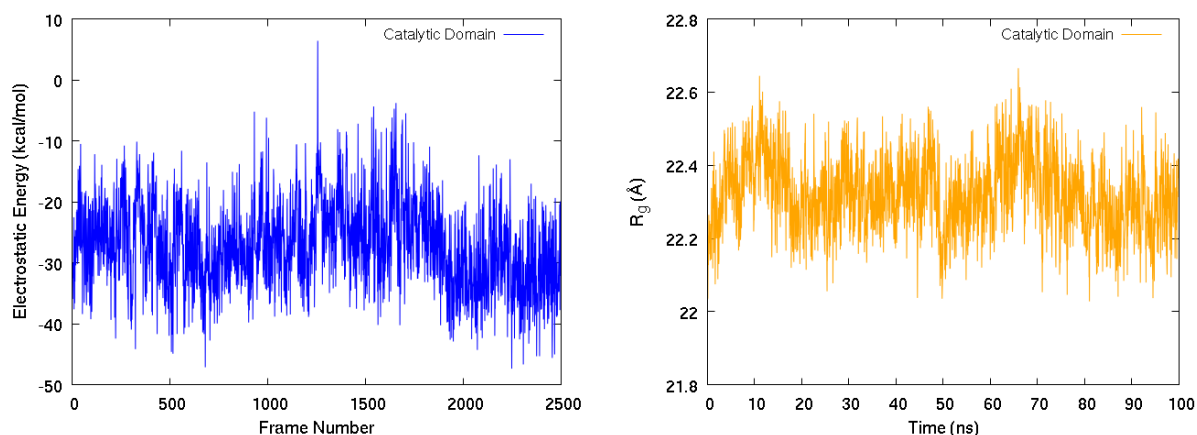


Figure 6.4: Electrostatic interaction energy between ligand and catalytic domain for 100 ns MD simulation (2500 frame total). Radius of gyration results for non-H atoms of catalytic domain during 100 ns MD simulation.

In the vicinity of the reaction center, apart from Lys779, there are several Asp residues which may be involved in the reaction. Knowledge of their pK_a and protonation state is of course an important issue to understand the reactivity of the system. In order to understand whether Asp residues are protonated or not we have estimated the pK_a values for all Asp and Lys residues during the simulation in the catalytic domain (Figures 6.6 and 6.7). The pK_a analysis of catalytic domain trajectory showed that all Asp residues are deprotonated and Lys residues are protonated. However, the pK_a of Asp787 is higher than for the other Asp residues, therefore it is more prone to be protonated. The side chain pK_a of Lys779 is found around

11, which is a similar value as in physiological environment.[176] Since Lys779 is the residue responsible for covalent inhibition by attacking as a nucleophile, it needs to lose its proton prior to covalent bond formation.

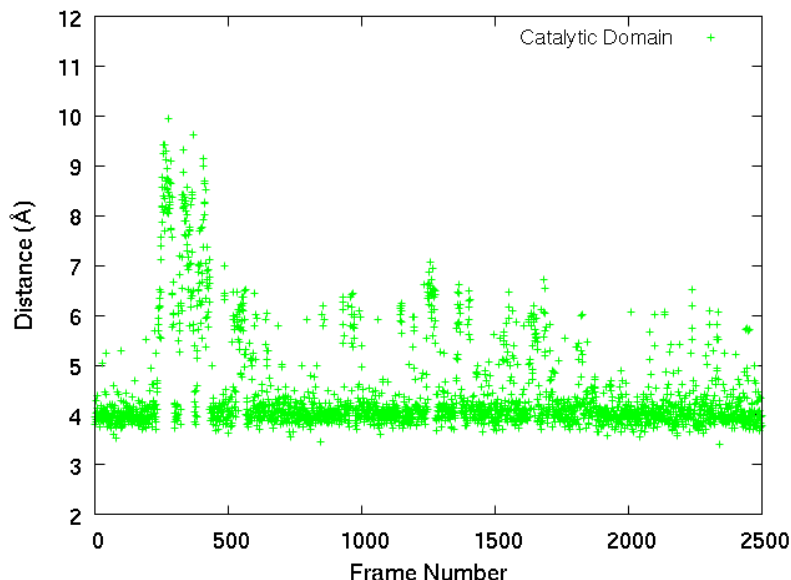


Figure 6.5: Distance between C atom of ligand (ester carbonyl carbon) and the side chain N atom of Lys779 along the 100 ns MD simulation (2500 frames).

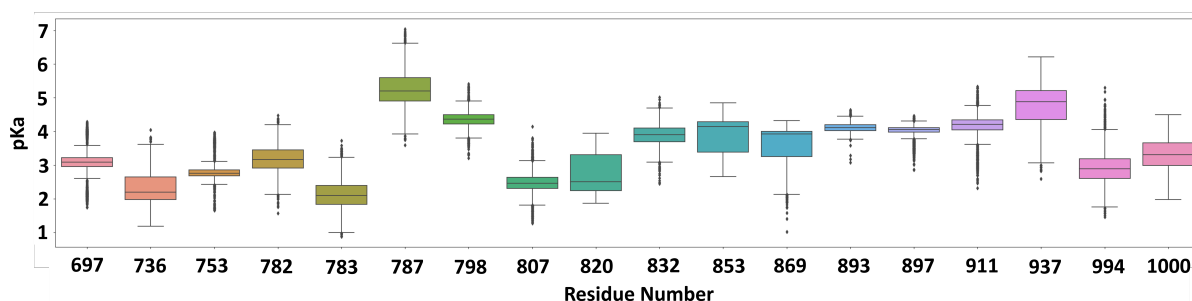


Figure 6.6: Change in pK_a 's of Asp residues in catalytic domain of enzyme for 100 ns simulation.

To get a deeper insight into this question, MD simulations of two kinase domain models M03 and M04 have been performed to check the stabilities of the different protonation states. In M03, we adjust only Asp787 as protonated in the kinase domain while M04 is adjusted to have protonated Asp787 and deprotonated Lys779. Structural stabilities of these systems are compared based on their RMSD profiles during the simulations and the results are reported in Figure 6.8. RMSD profiles of M01, M02, M03 resemble each other such that after about 20 ns, they reach stability with almost constant RMSD value. On the other hand, for enzyme model M04 RMSD keeps increasing slightly along the simulation indicating a less stable enzyme. This reveals that deprotonation of Lys779 generates more fluctuations which may facilitate the attach

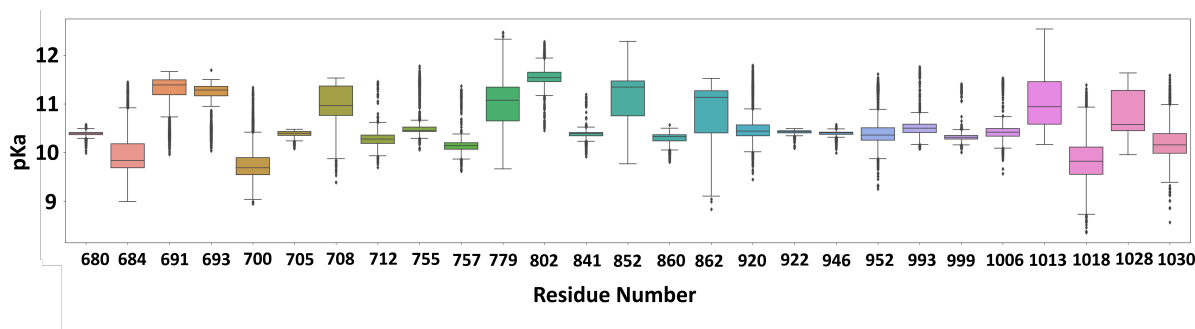


Figure 6.7: Change in pK_a 's of Lys residues in catalytic domain of enzyme for 100 ns simulation.

of Lys779 to the inhibitor. Among four RMSD profiles M03 (orange) appears to be the most stable system indicating the tendency of Asp787 for protonation. The next stable system with similar RMSD profile as M03 is the native catalytic domain M02 (green). Since M02 is the native enzyme obtained from experimental x-ray structure, we preferred to use this structure for further studies.

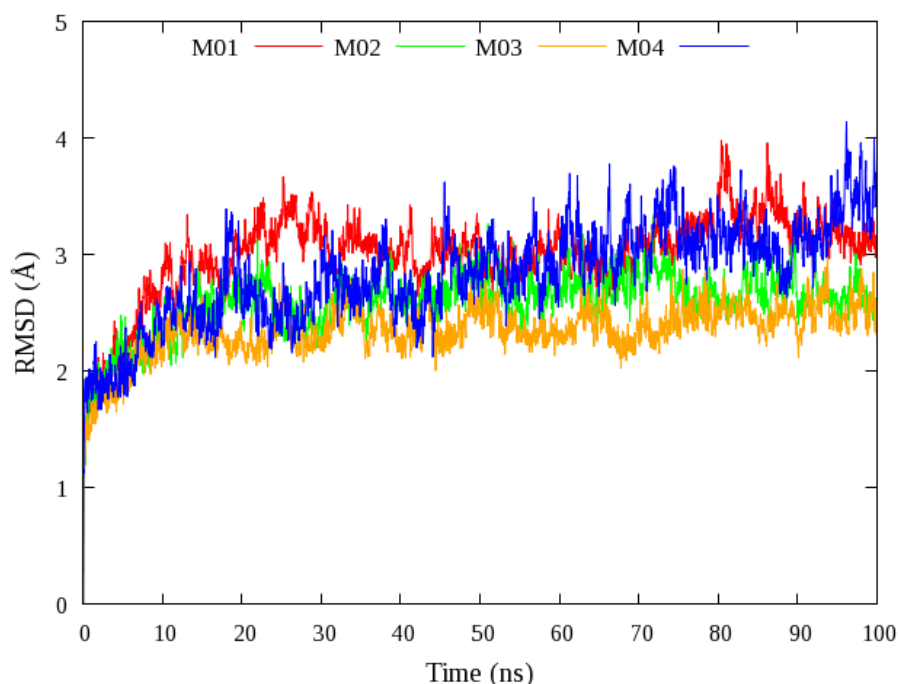


Figure 6.8: RMSD for 100 ns molecular dynamic simulations of full enzyme (M01), kinase domain (M02), protonated Asp787 model of the kinase domain (M03), protonated Asp787 and deprotonated Lys779 model of the kinase domain (M04).

Protonated Lys779 does not have sufficient nucleophilicity to attack inhibitor ester group. Therefore, we assume that it will give its proton to one of the aspartates prior to covalent bond formation with ester. Although our findings so far indicate that Asp787 is the residue more susceptible for proton abstraction, it is 6 Å away from Lys779 which makes it difficult to abstract

the proton of Lys779. On the other hand, another aspartate, Asp782 is in close contact with Lys779 by means of a H-bond. Therefore, we also want to consider the possibility of Asp782 abstracting the proton of Lys779 in the process of selecting an initial structure for ONIOM calculations. For that purpose, we examined the distance between each O atom of Asp782 and N atom of Lys779 during full trajectory. The distribution of distances shown in Figure 6.9 will be used to select a snapshot for the further study.

In order to study the inhibition reaction by ONIOM calculations in the next chapter,[177] we have selected an initial structure from the MD trajectory fulfilling two conditions, C-N lower than 4 Å and N-O lower than 3 Å. The frame taken at 78.8 ns of the simulation (1970 in Figure 6.10) contains one of the smallest distances for these important parameters and this frame was therefore selected. The quantum cluster shown in Figure 6.11 shows the inhibitor and the active site residues of the catalytic domain with water molecules in this frame.

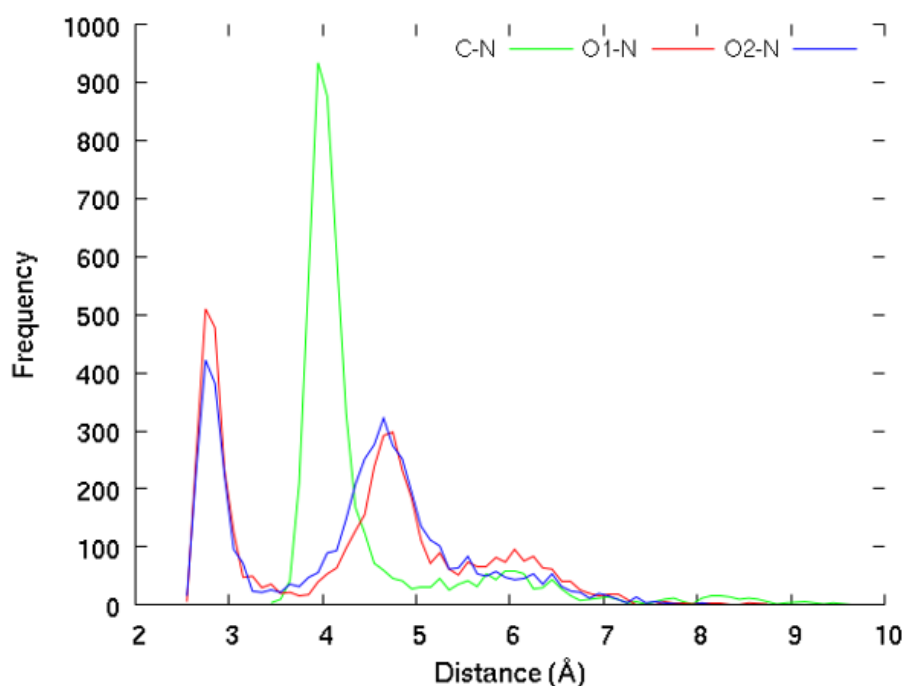


Figure 6.9: Frequency of distances between C atom of the ligand and N atom of Lys779 (green), O atoms of Asp782 and N atom of Lys779 (red and blue) for the catalytic domain of PDB:6ez6.

6.4 Conclusion

In this chapter, we have carried out a MD simulation of the PI3K δ enzyme to study the inhibitor/enzyme structure and select an appropriate structure of the active site for the reaction calculations in the forthcoming chapters. Starting from an X-ray structure, the missing residues were completed by using homology modeling. A classical MD simulation was carried out with the full enzyme and the catalytic domain, and the similarity of these two structures was found to consist of 976 atoms of a total of 1432 atoms. A standard analysis was then done for the

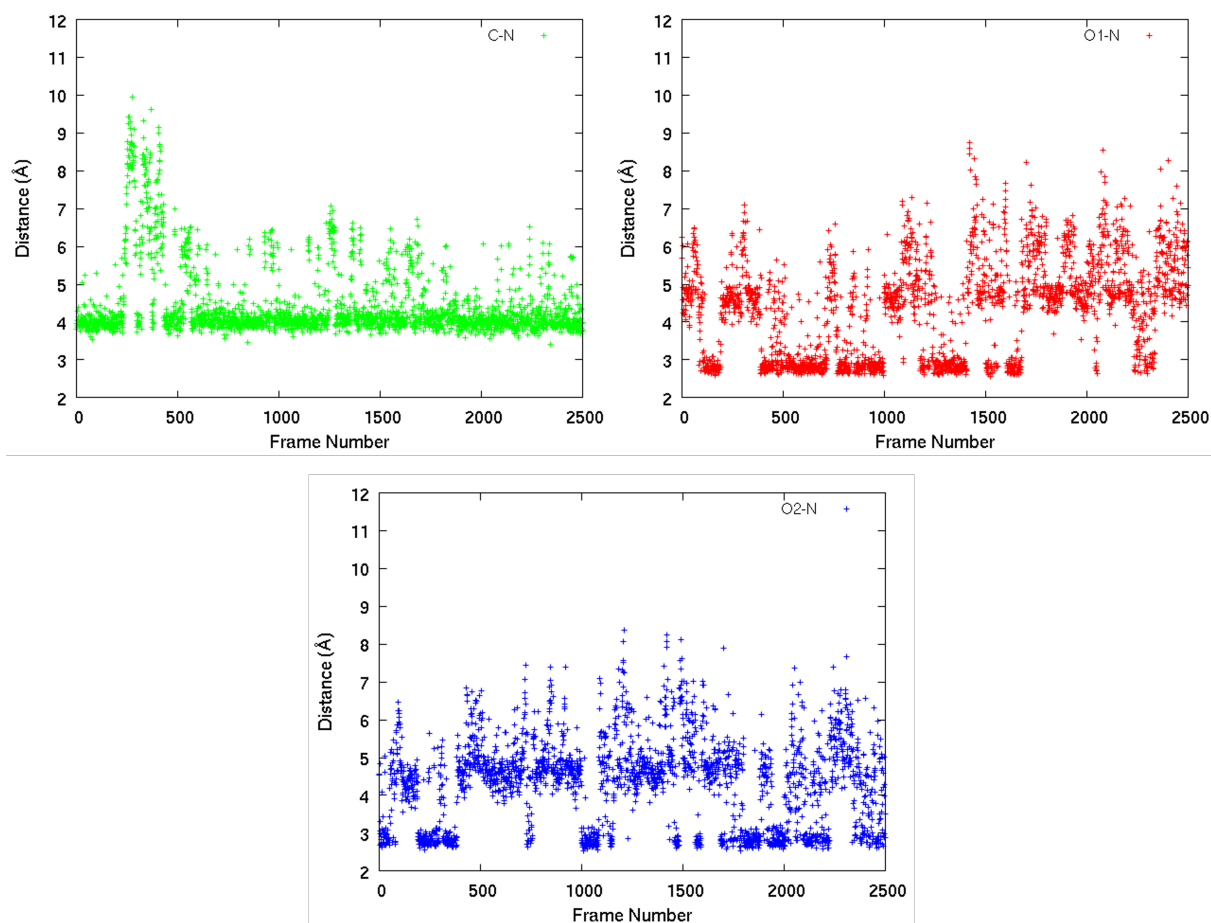


Figure 6.10: Distances between the carbonyl C atom of the inhibitor and the N atom of Lys779 (green), and between the O atoms of Asp782 and the N atom of Lys779 (red and blue). These distances were extracted from 100 ns MD simulation of the catalytic domain. For each 40ps, a frame from the simulation is taken (total number 2500).

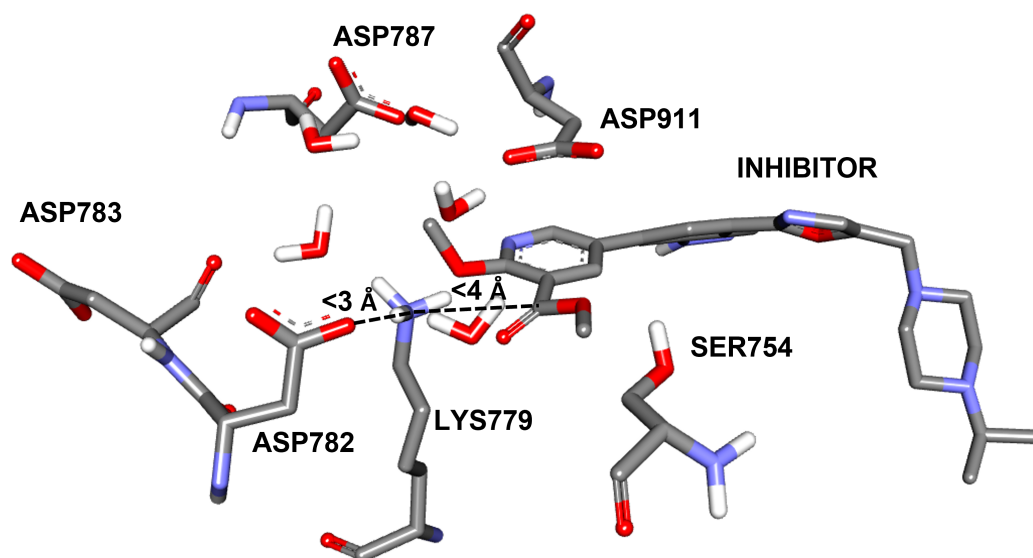


Figure 6.11: The inhibitor and the active side residues of catalytic domain with water molecules from the 78.8 ns of 100 ns MD simulation.

catalytic domain. The pK_a of Lys and Asp residues within the catalytic domain were studied by carrying out 3 different MD simulations in which different protonation states of Lys779 and Asp787 were assumed. Based on this analysis and the distances between the nitrogen atom of Lys779 and the oxygen and the carbonyl carbon atoms of the inhibitor, a frame has been selected to be used in the ONIOM model that we present in the next chapter.

Chapter 7

Lysine-targeted covalent inhibition mechanism of PI3K δ via ONIOM QM:QM approach

7.1 Introduction

In this chapter, quantum mechanical (QM) calculations using realistic enzyme models were carried out to enlighten the lysine-targeted covalent inhibition mechanism proposed by Dalton *et al.* (Figure 2.1) and gain chemical insights allowing to design more effective covalent PI3K inhibitors. We assume a static enzyme model using an ONIOM QM:QM approach, which is described in the next section. A more elaborated enzyme model together with a dynamic approach will be presented and discussed in the next chapter, using in that case molecular dynamics simulations with a QM/MM partition. We have explored the mechanisms of compounds 2, 4, 5, 6, and 8, which have been selected as representatives of covalent and noncovalent inhibitors studied by Dalton *et al.* and are reminded in Figure 7.1.

The ONIOM calculations presented in this chapter were conducted in collaboration with two master students. After I modeled four different mechanistic pathways of the ester aminolysis reaction between *p*-NO₂ phenyl ester (inhibitor 2 in Figure 7.1) and the targeted Lys779 residue in the active site of the enzyme, the same mechanisms were applied to three other covalent inhibitors (*p*-F, *p*-H and *p*-OCH₃ phenyl ester derivatives, compounds 4, 5, 6 in Figure 7.1) by Betül Tuba Varınca Gerçik in order to explore the influence of the para substituent. In addition, the similar mechanistic calculations were performed on the noncovalent methyl ester inhibitor by Öykü Sinek for comparison purpose. We published an article "Mechanistic Investigation of Lysine-Targeted Covalent Inhibition of PI3K via ONIOM QM:QM Computations" in *J. Chem. Inf. Model*, 2022., together with all these data. [33] The text below will be based on this article. For the sake of the completeness of the work and not to lose the synergy of this collaborative work in elucidating the important aspects of the mechanism, all the results and discussion are collectively included in this chapter.

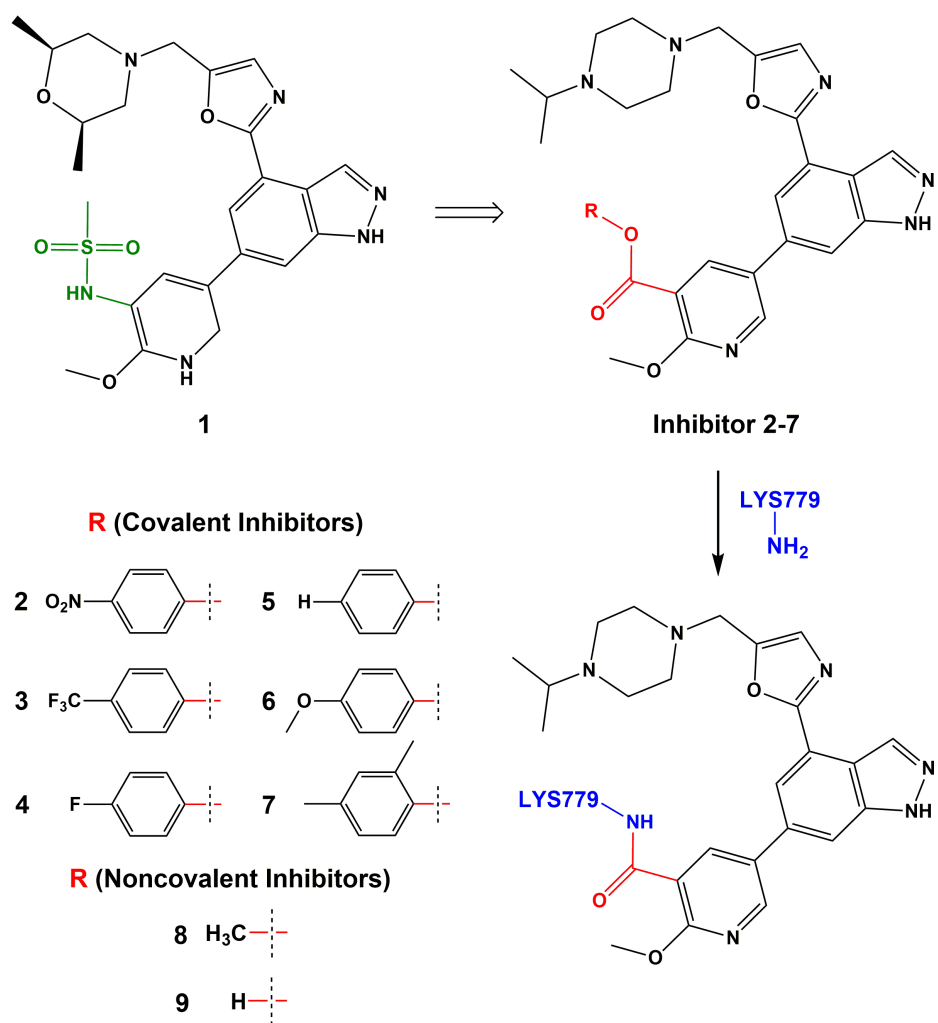


Figure 7.1: Lysine-Targeted Covalent Inhibitors of PI3K δ designed from clinical drug 1.[12]

7.2 Computational Details

The combined QM:QM ONIOM method of Morokuma and coworkers, [177] together with a cluster model, have been used to study the reaction of several ester inhibitors with Lys779 residue in the catalytic site of the PI3K δ enzyme. Potential covalent inhibition mechanisms have been investigated by modeling alternative pathways of the reaction.

The ONIOM method[177] is a hybrid approach dividing the enzyme (or a big molecular system in general) into two fragments. The most significant atoms in the active site of the enzyme are described at a high theoretical level (“high layer”, M062X/6-31+G(d,p) method[178] here) and the rest of the atoms with a less accurate method (“low layer”, semi-empirical PM6 method[119, 179] here). The ONIOM approach displays a good compromise between accuracy and cost and has been proven to be advantageous in locating all stationary points along the potential energy surface of large molecular systems.[177] The M062X/6-31+G(d,p):PM6 ONIOM scheme has been chosen because it was successfully applied in our previous studies on monoamine oxidase,[180, 181] and many similar ONIOM (M062X/6-31+G(d,p):PM6) combinations are reported in the literature.[181, 182, 183, 184] The M062X functional accounts for dispersion and long-range interactions, and it is highly recommended for the calculation of activation energy barriers.[178] Additionally, our benchmark study (Table 7.1 and Figure 7.2) using small model molecules revealed that M062X/6-31+G(d,p) has similar performance as the post-Hartree-Fock MP2/6-31+G(d,p) method (signed error in Gibbs free energy of activation is only 0.42 kcal/mol). Our previous model system (methyl amine reacting with methyl acetate representing the most simple case) in chapter 4, with five explicit water was used for this benchmark.

Table 7.1: Calculated potential and free activation energies (kcal/mol) for the leaving group dissociation step of small model system by different methods with 6-31+G(d,p) basis set in presence of five explicit water molecules.

	Forward Barrier		Reverse Barrier		Reaction Energy	
Method	ΔE	ΔG	ΔE	ΔG	ΔE	ΔG
MP2	20.72	16.92	31.51	33.79	-10.79	-16.87
B3LYP	16.94	13.97	38.80	41.23	-21.86	-27.27
B3LYP-D3	16.95	12.77	29.98	31.79	-13.03	-19.20
M062X	21.53	17.33	31.26	31.18	-9.74	-13.84
WB97XD	21.45	16.25	33.19	34.72	-11.75	-18.47

The PI3K δ -inhibitor cluster model to be studied with ONIOM is based on the MD simulations carried out in chapter 6. We summarize hereafter those calculations and the way the active site model has been chosen. First, a 100 ns Molecular Dynamics (MD) simulation was carried out for an enzyme-inhibitor system starting from the corresponding X-ray crystal structure.

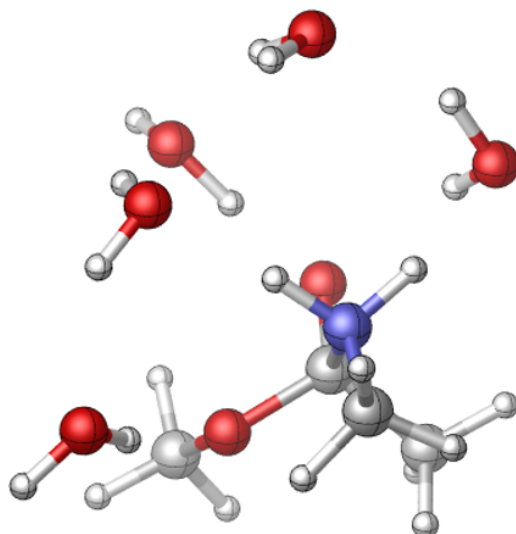


Figure 7.2: Small model system used in the benchmark study: Optimized zwitterion geometry by M062X/6-31+G(d,p) in the presence of five explicit water molecules.

The crystal structure (pdb:6ez6) for the noncovalent inhibitor (methyl substituted inhibitor) complexed with PI3K δ was chosen since it provides the full initial picture before the reaction. The crystal structure with covalent inhibitor (*p*-F-phenyl substituted inhibitor) (pdb:6eyz) was not preferred because the leaving group of the ester is missing in this structure (presumably, it already dissociated from the enzyme after amide bond formation). The MD simulations also allowed us to explore the pK_a values of the important amino acid residues, which is essential to describe the complex in the initial stage of the reaction (see below). The simulation procedure and structural analysis are explained in the previous chapter in detail. Then, using some MD snapshots, we selected a cluster structure containing the important chemical groups, i.e. the active site residues surrounding the reaction center, the solvating water molecules (five in our case) and the inhibitor. For the sake of limiting the computational cost, the structure of the inhibitor was truncated at the pyridine ring bearing the ester functional group. The cluster model used is shown in Figure 7.3.

The cluster structure was divided into two layers for the ONIOM computations: the functional groups of Lys779, Ser754, Asp782, Asp787, and Asp911 residues, the inhibitor model, and the water molecules were chosen as high layer. Asp783, backbone of Leu784, backbones of Asp782, Asp911 and Ser754, backbone nitrogen of Gly781 and the four carbons in the alkyl chain of Lys779 were chosen as low layer. H atoms were used as link atoms to saturate the chemical bonds in the QM:QM frontier between the two layers.[149, 177, 185] According to the well-established protocol of the quantum chemical cluster approach[185] and our experience from previous studies,[181, 182] two carbon atoms of amino acid backbones, shown as asterisk in Figure 7.3, were frozen to avoid unnatural changes of the atomic distances in the enzyme structure during optimizations. Fixing these coordinates also accounts for the steric restrictions imposed by the enzyme environment. The remaining geometric parameters were fully optimized. From the relaxed cluster structure with inhibitor 8, the cluster structure for

other inhibitors (2, 4, 5 and 6 in the Figure 7.1) was built and fully optimized afterwards.

In order to test the effect of using larger QM size on the reaction mechanism, we performed some preliminary test by including almost full size of the inhibitor in the high layer and repeating the calculations for the *p*-NO₂ phenyl ester inhibitor (Figure 7.4). Pyridine ring of the inhibitor is connected to an aromatic π -electron system and a saturated ring at the end. Dalton *et al.*[12] emphasized that the main role of these cyclic moieties is to fix the position of the inhibitor in the binding pocket of the enzyme by means of many attractive forces with the surrounding residues. It is evident from the experimental K_I values that, noncovalent binding to form enzyme-inhibitor complex is quite strong justifying that the conformations of these ring systems (the tail of the inhibitor) are almost locked at this pocket. Therefore, the dynamics of the full inhibitor are assumed to have negligible influence on the covalent inhibition reaction. On the other hand, if there is resonance delocalization between pyridine and benzene rings, it may influence the reaction energetics. In fact, in the crystal structure and the MD simulations, these aromatic rings are not planar (the torsional angle between pyridine and benzene moieties is about 40°) indicating that electron delocalization is interrupted at least partially. Nevertheless, the main contribution of this tail part is expected to come from the π -system. Therefore, in order to estimate the contribution of this π -system we enlarged the high QM layer of our ONIOM cluster model including the aromatic rings in the inhibitor structure as shown in Figure 7.4. For simplicity and to save computing time, we compared the energy profiles for Path A and Path C described below. The results showed that that the contribution of the π -electron delocalization from this tail part of the inhibitor is quite limited and does not alter the mechanistic preference of the inhibitor. Hence, the simplified high QM layer can be considered robust with respect to the larger high QM layer model.

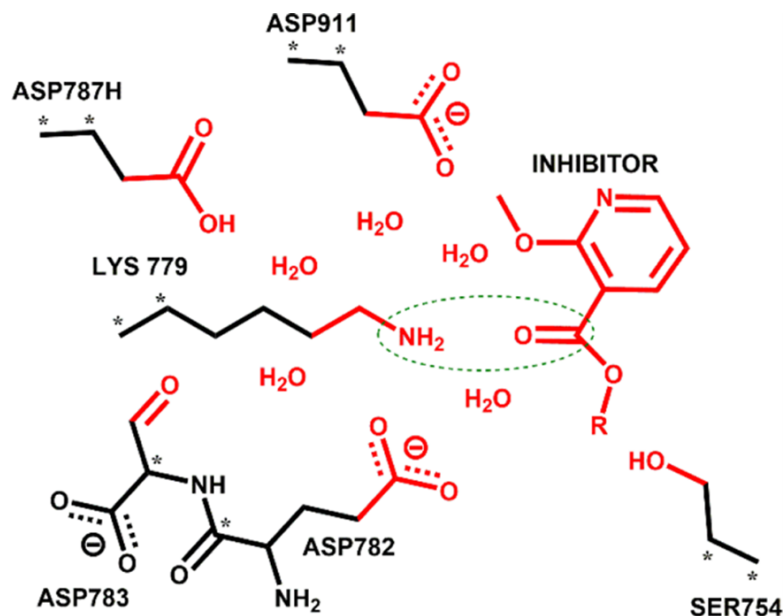


Figure 7.3: The quantum chemical cluster structure used in this study to model the reaction between PI3K δ and the inhibitor. Red parts represent the high layer (M062X/6-31+G(d,p)) and black parts the low layer (PM6) in the ONIOM partitioning scheme. Frozen atoms are shown with asterisk.

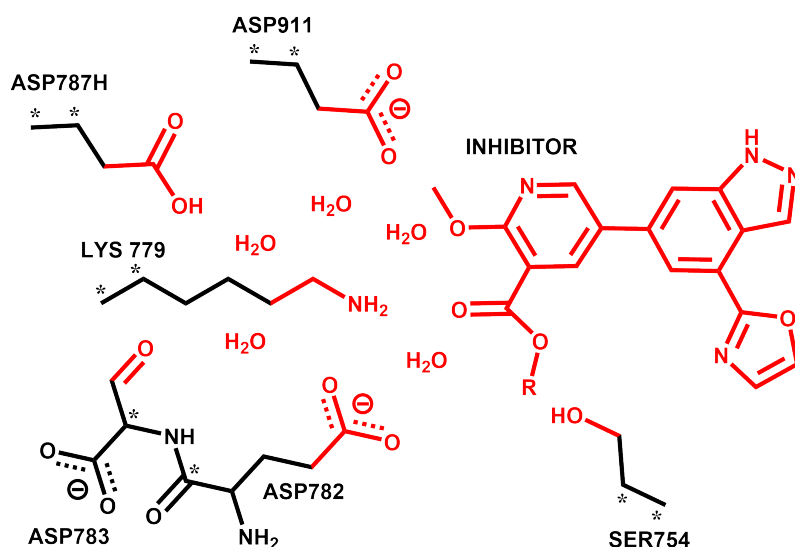


Figure 7.4: The big inhibitor structure included in the high layer QM part of the chemical cluster. Red parts represent the high layer (M062X/6-31+G(d,p)), and black parts the low layer (PM6) in the ONIOM partitioning scheme. R = *p*-NO₂ phenyl.

The reaction mechanisms explored are described in the next section. Transition state structures were characterized by only one imaginary frequency in normal mode analysis and further supported by intrinsic reaction coordinate (IRC)[163] calculations. The remaining stationary structures (reactant complexes, intermediates, and product complexes) were obtained by optimizing the final structures of the IRC output and were characterized by frequency analysis. Thermodynamic corrections and Gibbs free energies were calculated at 25 °C and 1 atm. All calculations were performed using Gaussian 09 revision B.01.[161]

7.3 Results and Discussion

7.3.1 Formation of an activated reactant complex

As indicated in chapter 6, Lys residues, which are expected to be positively charged at physiological pH, must be deprotonated in the earliest stages of the covalent inhibition process. At the active site of PI3K δ , there are four aspartic acid (Asp) residues: Asp782, Asp783, Asp787, and Asp911. Our MD simulations show that three of them, Asp782, Asp787, and Asp911 form an H-bond network with the water molecules slightly above the reaction center (Figure 6.11 on page 67).

Although the intrinsic pK_a values of lysine and aspartic acid side chains are known to be 10.4 and 4.0, respectively, their apparent pK_a values in proteins can be shifted depending on the protein microenvironment.[186] Therefore, we calculated the pK_a values of Asp and Lys residues in our model from the MD simulation with the noncovalent inhibitor (Figure 6.6 on page 63 and Figure 6.7 on page 64). During the simulation, the pK_a of Lys779 changes roughly between 10.5 and 11.5 indicating that (as expected) its amino group is protonated (Lys side chains are usually considered deprotonated when their $pK_a < 7.8$ [6]). The pK_a values of Asp782, Asp787

and Asp911 residues are found to be approximately ~ 3 , ~ 5 and ~ 4 , respectively. Because an aspartate residue binds a proton at a pH close to its pK_a , these three aspartates should be in the deprotonated carboxylate form at physiological pH. However, since Asp787 exhibits the largest pK_a , it appears more prone to abstract a proton in comparison to the other Asp residues. In fact, using the ONIOM M062X/6-31+G(d,p):PM6 method, we calculated the free energy change for Lys779 deprotonation (i.e. proton transfer to one Asp residue) to be smaller for Asp787 (4.8 kcal/mol) than for Asp911 (7.8 kcal/mol) for inhibitor 8.[33] Besides, all our efforts to calculate the energy of the systems with a protonated Asp782 residue ended up with the transfer of the proton back to Lys779, suggesting a relative basicity order Asp787>Asp911>Asp782. Thus, in order to gain the required nucleophilicity and react with the ester group of the inhibitor, Lys779 should preferentially donate its proton to Asp787. The inhibition mechanism proposed by Dalton *et al.*[12] in Figure 2.1 (on page 20) does not take into account the requirement for Lys deprotonation, although it is a well-known issue in lysine-targeted covalent inhibition of the enzymes.[82, 187] Liu *et al.*[6] state that, to display nucleophilic reactivity, Lys residues in kinases should have $pK_a < 7.8$. We propose that Lys779 is protonated in the initial step of the reaction of PI3K δ with the inhibitor, and that proton transfer to Asp787 is a prerequisite for the reaction to take place, which requires about 4.8 kcal/mol. The two water molecules situated between Asp787 and Lys779 can act as proton shuttles to assist Lys deprotonation, as illustrated in Figure 7.5. The same deprotonation step was also examined for two covalent inhibitors to elucidate their influence. For this purpose, methyl group (R) was replaced with *p*-NO₂ phenyl (inhibitor 2) and *p*-F phenyl (inhibitor 4) groups. The free energy change from the ionized configuration (Lys779NH₃⁺, Asp787COO⁻) to the neutral one (Lys779NH₂, Asp787COOH) was found to be 5.0 and 5.5 kcal/mol energy for 2 and 4[33], respectively, suggesting that different R substituents have small or moderate effect on the deprotonation step of Lys779 residue.

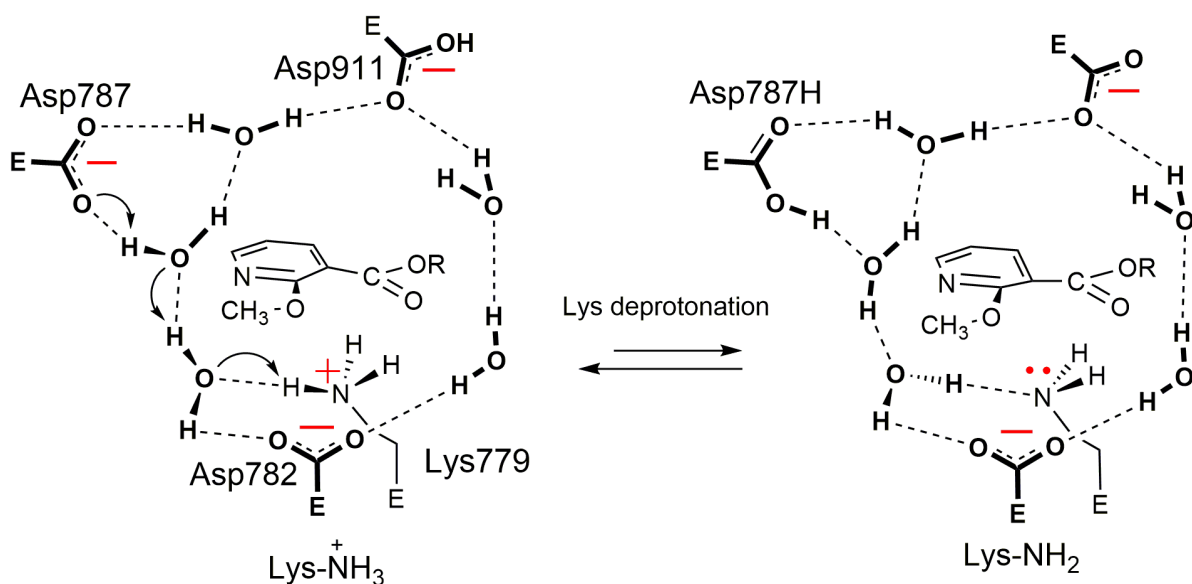


Figure 7.5: Proposed mechanism for the proton transfer between Lys779 and Asp787 via two water molecules. The inhibitor and Lys779 are located slightly below the H-bonding network.

This endothermic process leads to an activated enzyme-inhibitor complex, which is expected

to provoke several conformational alterations along with a reorganization of the H-bond network in the initial reactant complex (RC). Indeed, our calculations show a Lys conformational change in RC so that the nitrogen electron lone pair is now approximately directed towards the ester group. In addition, the ester group undergoes a small alteration with the carbonyl oxygen slightly pushed out of plane, which promotes the establishment of a new H-bond of this oxygen atom with a water molecule located between Asp787H and Asp911 (RC1 in Figure 7.6). These two changes are likely to promote the next stage of the reaction.

7.3.2 Reaction Mechanism: Covalent vs Noncovalent Inhibitors

As suggested by calculations with simple model systems,[32] four main mechanistic scenarios can be considered for the ester aminolysis reaction starting from RC. They are schematically represented in Figure 7.6 (Path A, Path B) and Figure 7.7 (Path C, Path D).

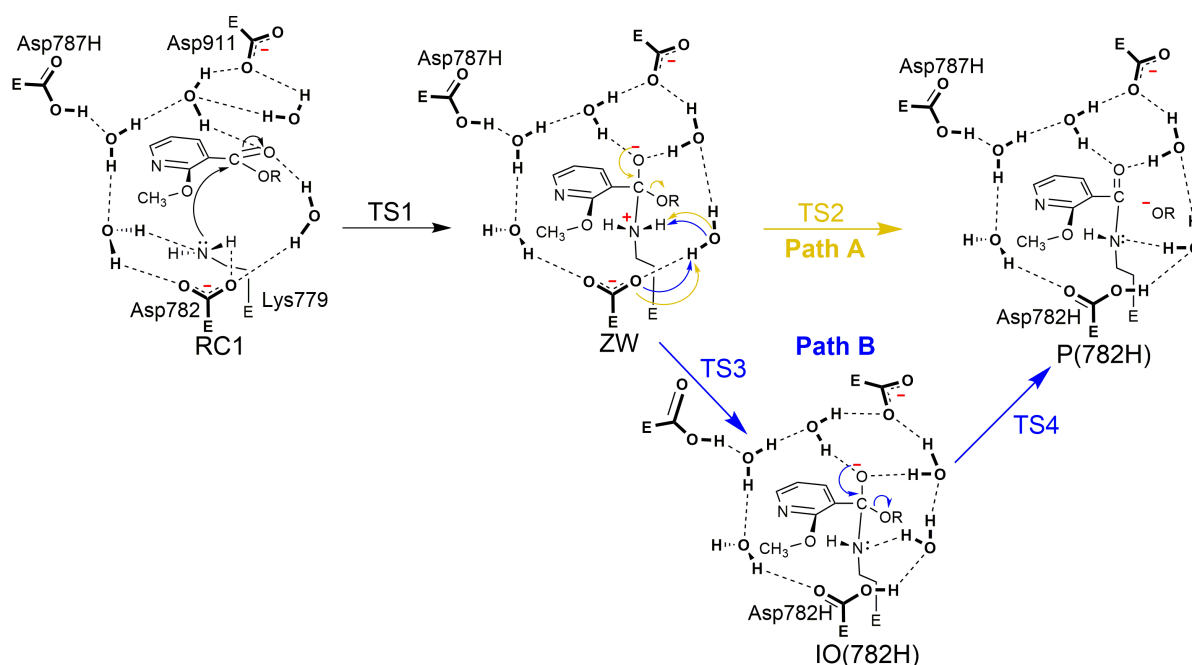


Figure 7.6: Mechanisms (Path A and B) modeled for covalent inhibitors (E=Enzyme). Path A (yellow): leaving group dissociation from ZW occurs in concerted manner with proton transfer to Asp782. Path B (blue): this is a stepwise version of Path A in which proton transfer occurs first forming an ionic intermediate IO(782H).

The first step common to all mechanisms is the attack of the Lys-NH₂ group to the ester carbonyl carbon, giving rise to a tetrahedral zwitterionic intermediate (ZW). Although the final product of ester aminolysis is an achiral amide, ZW formed in the first step is a chiral tetrahedral intermediate. Detailed information on the stereochemistry of ZW formation is given in the following section. From ZW, the amidation occurs in a concerted manner in Path A (Figure 7.6); as the leaving group departs, Asp782 abstracts a proton from amine nitrogen via one water molecule acting as a proton shuttle. Therefore, Asp782 is protonated in the product state denoted as P(782H). Path B (Figure 7.6) is the stepwise version of Path A; abstraction of

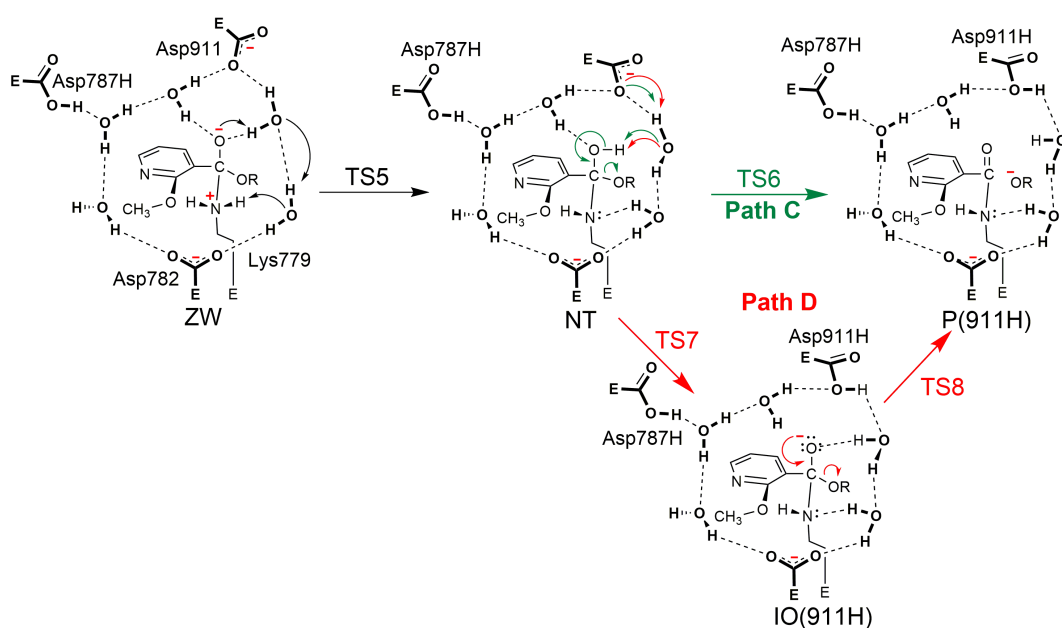


Figure 7.7: Mechanisms (Path C and D) modeled for covalent inhibitors (E=Enzyme). Path C (green): dissociation of the neutral (NT) intermediate occurs in a concerted manner with proton transfer to Asp911. Mechanism D (red): this is a stepwise version of Path C in which Asp911 takes a proton from NT and ionic intermediate IO(911H) is formed.

ZW amine proton by Asp782 takes place before the leaving group departure. Such a stepwise process affords the formation of a tetrahedral anionic intermediate IO(782H) (similar to the one proposed by Dalton *et al.* in Figure 2.1 on page 20) where Asp782 is protonated. In Path C (Figure 7.7), the amine proton of ZW migrates to the negatively charged oxygen by means of two water molecules acting as shuttle, resulting in a neutral hemiaminal intermediate (NT). Then, the hemiaminal OH proton is transferred to Asp911 residue as the leaving group departs. Path D (Figure 7.7), is similar to Path C but the last step occurs in a stepwise manner; an ionic intermediate IO(911H) (as proposed by Dalton *et al.* in Figure 2.1 on page 20) forms before the departure of the leaving group. We attempted to model similar mechanisms to Path C/D but with Asp782 abstracting the NT proton, and to Path A/B with Asp911 abstracting the ZW proton, but the related transition structures of these cases could not be located for the systems described below, and such alternative pathways are therefore discarded.

We first present the results obtained for two covalent inhibitors, *p*-NO₂ phenyl ester (2) and *p*-F phenyl ester (4), and for the noncovalent inhibitor methyl ester (8). Note that, although methyl ester acts as a reversible noncovalent inhibitor of PI3K δ [12], we have still investigated its covalent bond formation mechanism in order to compare with the mechanism of the covalent inhibitors, and to understand why this ester inhibitor does not favor aminolysis reaction with Lys779. Such a comparative study will allow us to have a general overview of the ester inhibitors and to determine the important factors in designing new covalent/noncovalent inhibitors. Our calculations confirm that the four mechanisms A-D described above apply for the covalent inhibitors 2 and 4, whereas in the case of the noncovalent inhibitor 8, path D could not be located, and the last steps of Paths A-C are slightly different (Figure 7.8).[33] Indeed, a

neutral leaving group is found in this case (CH_3OH) that is formed by capturing a proton during dissociation; otherwise, the leaving group (CH_3O^-) would be a highly unstable anion. This result is in line with our previous computational study on amine acetylation reactions in a small model system.[32] The Gibbs free energies are given in Tables 7.2, 7.3 and 7.4, and they are illustrated in Figures 7.9, 7.10, 7.11, for inhibitors 2, 4, and 8, respectively. In these plots, RC2 represents the prereactive complex obtained from reverse IRC calculations, where the amine lone pair electrons are close to and directed toward the ester carbonyl carbon, whereas RC1 is a more stable RC structure obtained through full optimization after elongating the N(Lys)—C(carbonyl) distance to 4.0 Å. In all cases, relative free energies are given with respect to this most stable RC1 structure to ease the comparison.

The relative energies of TS1 and ZW1 for *p*-NO₂ phenyl, *p*-F phenyl, and methyl ester inhibitors reveal that their stabilities decline as the electron-withdrawing character of the leaving groups decreases. Thus, methyl ester exhibits the least stable TS1 and ZW. In addition, the last steps for all three pathways for this ester have quite unstable transition states (TS2, TS4, TS6), which will prohibit the forward reaction (Figure 7.11). This result is consistent with the experimentally observed noncovalent behavior of this inhibitor candidate. In contrast, the reaction free energy profiles of *p*-NO₂ phenyl (Figure 7.9) and *p*-F phenyl esters (Figure 7.10) exhibit more stable transitions states (at least by 16 kcal/mol) and greater exergonicity in general for the four mechanisms with respect to the methyl ester free energy profile (Figure 7.11). According to these calculations, the aminolysis reaction with the phenyl esters (compounds 2 and 4) is likely to occur through either path A, B, C, or D in physiological conditions, and the theoretical results are therefore consistent with the experimental data showing that these esters can behave as covalent inhibitors.

7.3.3 Stereochemical Aspects of Tetrahedral Zwitterion Intermediate

The reported mechanism was chosen based on the following results. For the tetrahedral ZW intermediate, we screened through different conformations of the phenyl moiety generated from the rotation of the bond between phenyl carbon and oxygen, and adopted the most stable one. It is stabilized by the favorable interaction between phenyl hydrogen and Asp911 oxygen as well as H-bonding interaction between the *p*-substituent and Ser754 residue. This conformation is similar in all structures. The other conformations of the inhibitor are restricted due to the constraints exerted by the protein environment.

For the stereochemistry of the ZW formation step, we considered the approach of Lys to two prochiral faces of ester plane as follows: The upper side of the ester plane (si-face) is more hydrophilic and hindered for the attack of Lys since this face is occupied by two water molecules and Asp911. Instead, bottom side (re-face) attack is feasible because there is enough room (a hydrophobic pocket) below the ester plane (re-face) and lysine can occupy this space as it approaches to the ester carbon, which leads to tetrahedral zwitterion in S-configuration. The bottom side attack of Lys orients the carbonyl oxygen to upward position where it benefits from H-bonding interactions with the water molecules. Alternatively, conformational change via rotation of the bond connecting α -C and the carbonyl C by 180° alters the re-face to the si-face which can be approached from the bottom side as well. Nevertheless, such a rotation is restricted under the constraints of the protein surrounding.

Table 7.2: Total energies (E), Gibbs free energies (G) and relative Gibbs free energies (ΔG) for all stationary points optimized at M062X/6-31+G(d,p):PM6 level for *p*-NO₂ phenyl ester (inhibitor 2).

<i>p</i> -NO ₂	E (a.u)	G (a.u)	ΔG (kcal/mol)
RC1	-2420.168411	-2419.270616	0.0
RC2	-2420.160551	-2419.259959	6.7
TS1	-2420.155917	-2419.252852	11.1
ZW1	-2420.165466	-2419.259673	6.9
PATH A			
ZW2	-2420.165932	-2419.261810	5.5
TS2	-2420.154485	-2419.252838	11.2
P1(782H)	-2420.187493	-2419.291779	-13.3
PC-A	-2420.192024	-2419.294161	-14.8
PATH B			
ZW3	-2420.164201	-2419.257563	8.2
TS3	-2420.150797	-2419.250012	12.9
IO(782H)	-2420.151066	-2419.249317	13.4
TS4	-2420.152911	-2419.250257	12.8
P2(782H)	-2420.190034	-2419.289288	-11.7
PATH C			
ZW4	-2420.162137	-2419.253783	10.6
TS5	-2420.150558	-2419.245479	15.8
NT1	-2420.162863	-2419.255143	9.7
NT2	-2420.162891	-2419.255478	9.5
TS6	-2420.15501	-2419.253257	10.9
P1(911H)	-2420.199565	-2419.300527	-18.8
PC-C	-2420.191184	-2419.292181	-13.5
PATH D			
NT3	-2420.162938	-2419.256759	8.7
TS7	-2420.158838	-2419.255444	9.5
IO(911H)	-2420.160179	-2419.254192	10.3
TS8	-2420.158278	-2419.254257	10.3
PC(911H)	-2420.195271	-2419.295769	-15.8

Table 7.3: Total energies (E), Gibbs free energies (G) and relative Gibbs free energies (ΔG) for all stationary points optimized at M062X/6-31+G(d,p):PM6 level for *p*-F phenyl ester (inhibitor 4).[33]

<i>p</i> -F	E (a.u)	G (a.u)	ΔG (kcal/mol)
RC1	-2314.928435	-2314.037168	0.0
RC2	-2314.922424	-2314.030988	3.9
TS1	-2314.914366	-2314.018527	11.7
ZW1	-2314.921593	-2314.025541	7.3
PATH A			
ZW2	-2314.925866	-2314.029736	4.7
TS2	-2314.904216	-2314.012253	15.6
P1(782H)	-2314.927616	-2314.034293	1.8
PC-A	-2314.948843	-2314.058015	-13.8
PATH B			
ZW3	-2314.922093	-2314.023740	8.4
TS3	-2314.90967	-2314.019078	11.4
IO(782H)	-2314.910241	-2314.017746	12.2
TS4	-2314.902187	-2314.012877	15.2
P2(782H)	-2314.919769	-2314.031702	3.4
PC-B	-2314.955427	-2314.064197	-17.0
PATH C			
ZW4	-2314.922311	-2314.024309	8.1
TS5	-2314.910341	-2314.017931	12.1
NT1	-2314.930124	-2314.031616	3.5
NT2	-2314.924572	-2314.028155	5.7
TS6	-2314.905610	-2314.021672	9.7
P1(911H)	-2314.968368	-2314.082043	-28.2
PC-C	-2314.973269	-2314.082692	-28.6
PATH D			
NT3	-2314.930152	-2314.032098	3.2
TS7	-2314.921133	-2314.026616	6.6
IO(911H)	-2314.921161	-2314.026716	6.6
TS8	-2314.912289	-2314.020472	10.5
PC(911H)	-2314.968368	-2314.082043	-28.2
PC-D	-2314.973269	-2314.082692	-28.6

Table 7.4: Total energies (E), Gibbs free energies (G) and relative Gibbs free energies (ΔG) for all stationary points optimized at M062X/6-31+G(d,p):PM6 level for methyl ester (inhibitor 8).[33]

Me	E (a.u)	G (a.u)	ΔG (kcal/mol)
RC1	-2024.04499805	-2023.19270000	0.0
RC2	-2024.04074674	-2023.19052300	1.4
TS1	-2024.02298298	-2023.16679700	16.3
ZW1	-2024.02600445	-2023.16693600	16.2
PATH A			
ZW2	-2024.02922877	-2023.17503900	11.1
TS2	-2023.98122454	-2023.13209800	38.0
PC-A	-2024.05658988	-2023.20840900	-9.9
PATH B			
TS3	-2024.00985648	-2023.16068600	20.1
IO(782H)	-2024.01019941	-2023.15875700	21.3
TS4	-2023.99231755	-2023.14241100	31.6
PC-B	-2024.04554564	-2023.19562700	-1.8
PATH C			
TS5	-2024.00909898	-2023.15831600	21.6
NT1	-2024.03342322	-2023.17901200	8.6
NT2	-2024.02790743	-2023.17388800	11.8
TS6	-2023.99077483	-2023.13891400	33.8
PC-C	-2024.04924613	-2023.19843500	-3.6

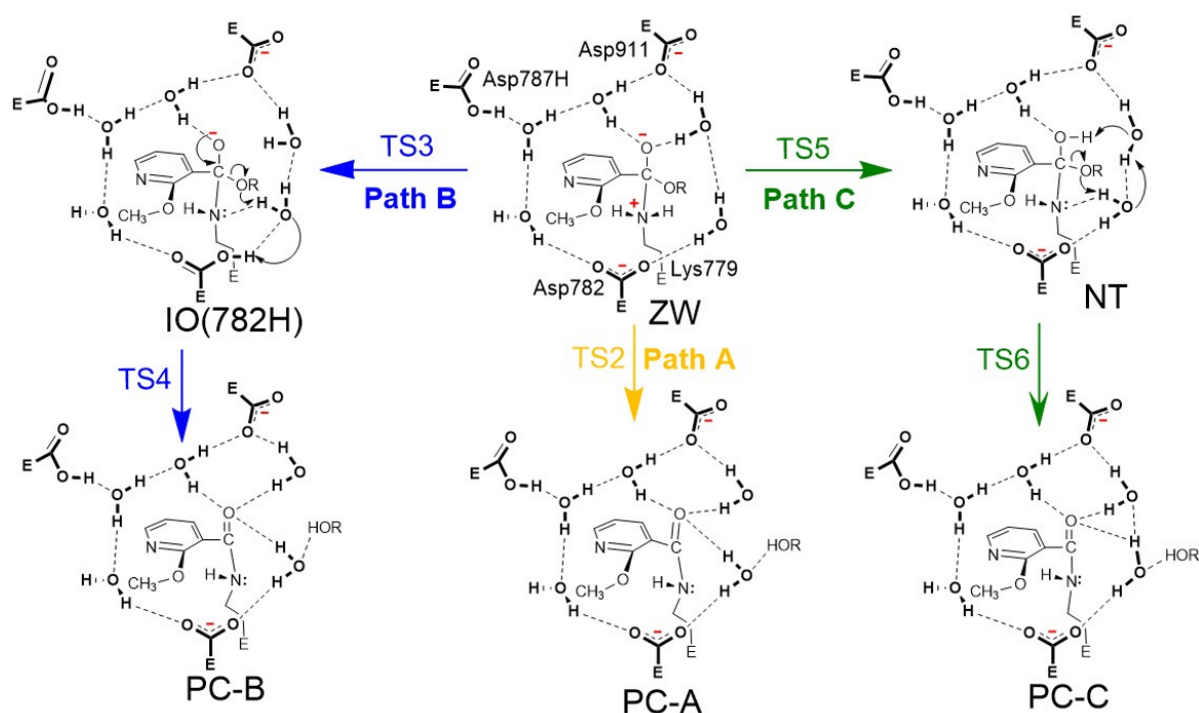


Figure 7.8: Mechanisms explored for methyl ester inhibitor in order to understand why it does not favor the formation of a covalent bond with Lys779 ($R=CH_3$, $E=Enzyme$): In Paths A, B, and C, the leaving group dissociates in protonated form (HOR) and the product complexes PC-A, PC-B and PC-C are produced, respectively.[33]

7.3.4 Reaction Mechanism: Most Favorable Path Involving Covalent Inhibitors

Inspection of the p -NO₂ phenyl ester energy profiles in Figure 7.9 reveals that Path A (yellow) appears as the most plausible one, as it exhibits the lowest energy transition state (TS2, 11.2 kcal/mol). The optimized structures along this pathway (RC1, TS1, ZW1, TS2, P1(782H), PC-A) are shown in Figure 7.12. Comparison of the different paths clearly shows that after ZW forms, direct dissociation of the leaving group to the product (Path A, and to a lower extent Path B) is preferred instead of forming a NT intermediate (Path C). This is in line with our previous study[32] and quite reasonable since the leaving group is remarkably stabilized by the strong electron-withdrawing character of the p -NO₂ substituent, which triggers the direct dissociation of ZW.

In the case of p -F phenyl ester, the profile in Figure 7.10 shows that Path C (green), passing through TS5 (12.1 kcal/mol), which connects ZW to the NT intermediate, and TS6 (9.7 kcal/mol), appears as the most plausible one. 3-Dimensional views of the optimized structures along this path are given in Figure 7.13.[33] The distinct behavior between p -F and p -NO₂ phenyl esters can be attributed to the much lower stability of the p -F phenolate leaving group relative to p -NO₂ phenolate resulting from the direct dissociation of ZW in Path A.

Therefore, p -F phenyl ester favors a different route (Path C) along which the ZW proton is first transferred to the carbonyl oxygen to form the NT intermediate; subsequently, the leaving

group dissociates and Asp911 abstracts this proton (TS6) as depicted in Figure 7.13.[33] The benefit of the formation of NT intermediate is that the migrating proton is now closer to the more basic Asp911 (Figure 7.7) which promotes the abstraction of this proton along with the departure of the *p*-F phenolate leaving group. In Paths A/B, the migrating proton is abstracted by the less basic Asp782 during leaving group dissociation; in the case of *p*-NO₂ phenyl ester, the lower basicity of Asp782 is compensated by the strong electron-withdrawing stabilization of *p*-NO₂ phenolate.

7.3.5 Reaction Mechanism: Electron-Withdrawing vs Electron- Donating Covalent Inhibitors

The discrepancies in the most favorable paths for *p*-NO₂ (path A) and *p*-F (path C) derivatives indicate that different mechanisms can be in play for covalent inhibitors depending on their electronic properties. Both, the nitro and fluorine substituents can be considered as electron-withdrawing groups although the former is a much stronger one than the latter. These results prompted us to explore the free energy profiles for an electron-donating group, namely *p*-OCH₃ phenyl ester 6. For simplicity, only Path A and Path C have been considered in this case (Table 7.5).[33]

As illustrated in Figure 7.14, Path C (green) involving the NT intermediate is indeed predicted to be more favorable than Path A. Note, however, that *p*-OCH₃ phenyl ester acts in a slightly different manner than *p*-F phenyl ester: in this case the highest energy transition state along Path C is TS6 (14.2 kcal/mol) (instead of TS5 for *p*-F phenyl ester), which can be explained by the fact that TS6 is associated with the departure of the phenolate leaving group, and this anion is significantly destabilized by the electron-donating *p*-OCH₃ group.

Finally, we wanted to check which mechanism dominates for compound 5, which will be interesting since it is a special moderate case with unsubstituted phenyl ester. The Gibbs free energy profile of compound 5 for Path A and Path C in Figure 7.15 demonstrates that Path A is favored over Path C but by a smaller extent (by only 1.3 kcal/mol) when compared to the *p*-NO₂ case (Table 7.6).[33] Ultimately, it behaves more like the *p*-NO₂ phenyl than the *p*-F phenyl case. Since there is no substituent in the phenyl group, this compound may illustrate the general tendency in phenolate esters. Thus, it is reasonable to assume that direct dissociation (Path A) is the preferred covalent inhibition mechanism in general as in our previous model study.[32] Nevertheless, since Asp residues in the vicinity of the reaction center administer a delicate balance between Path A and Path C (as explained in a previous section), the preferred mechanism may switch to Path C depending on the electronic nature of the substituent. Our results disclose that electron-donating substituents with π -conjugation (*p*-OCH₃ and *p*-F) to the phenyl ring drive the mechanism to Path C.

Based on this theoretical study, an attempt can be made to classify the covalent inhibitors shown in Figure 7.1 into three classes.

- Class 1: Phenyl esters having strong electron-withdrawing groups. They should proceed through Path A (fast dissociation of the leaving group) and display the lowest activation energies and most favorable kinetics. The *p*-NO₂ ester derivative (compound 2) can be considered as a prototypical example of this class.

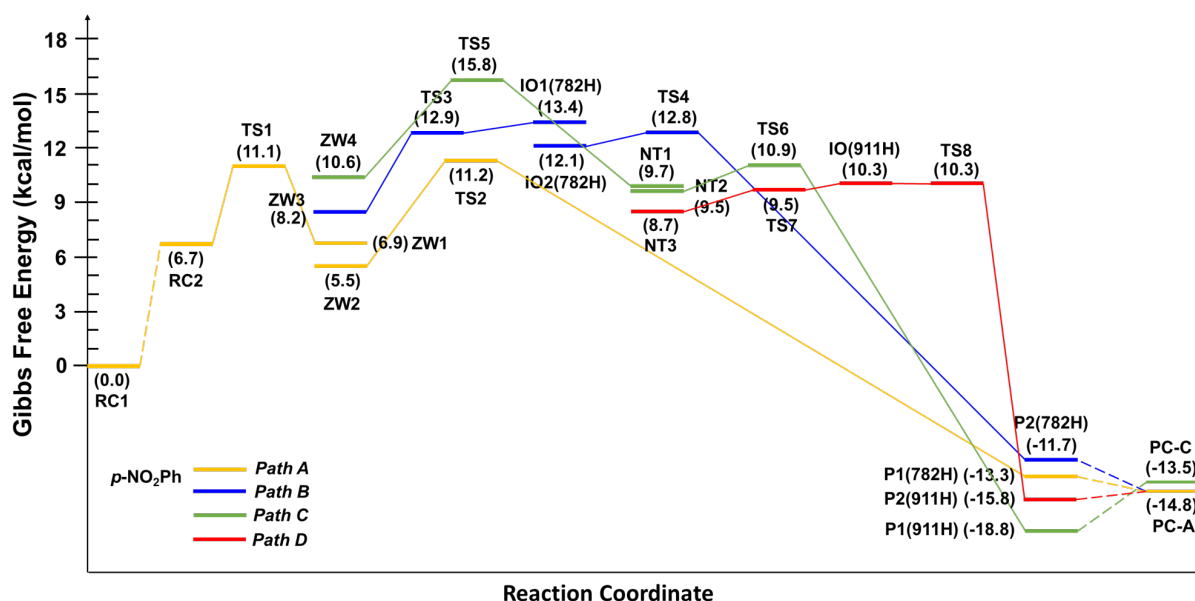


Figure 7.9: Gibbs free energy profile for the reaction of *p*-NO₂ phenyl ester (compound 2) with Lys779 in the active site of PI3K δ calculated with the ONIOM (M062X/6-31+G(d,p):PM6) method for four different pathways. Energies are relative to the most stable reactant complex RC1. The first common step (ZW1 formation) is shown by the same color as the most plausible Path A (yellow). ZW1, ZW2, ZW3, ZW4 represent different conformations of the zwitterion intermediate obtained through IRC calculations from TS1, TS2, TS3, and TS5, respectively. NT1, NT2, NT3 represent different conformations of the neutral intermediate obtained through IRC calculations from TS5, TS6, TS7, respectively. IO1(782H) and IO2(782H) are anionic intermediates (Asp782 is protonated) obtained through IRC calculations from TS3 and TS4, respectively. P1(782H) and P2(782H) represent amide products (Asp782 is protonated), complexed with *p*-NO₂ phenolate ion, obtained through IRC calculations from TS2 and TS4. P1(911H) and P2(911H) represent amide products (Asp911 is protonated), complexed with *p*-NO₂ phenolate, obtained through IRC calculations from TS6 and TS8. PC-A and PC-C are the final product complex structures (amide + *p*-NO₂ phenol) of Path A/B and Path C/D.

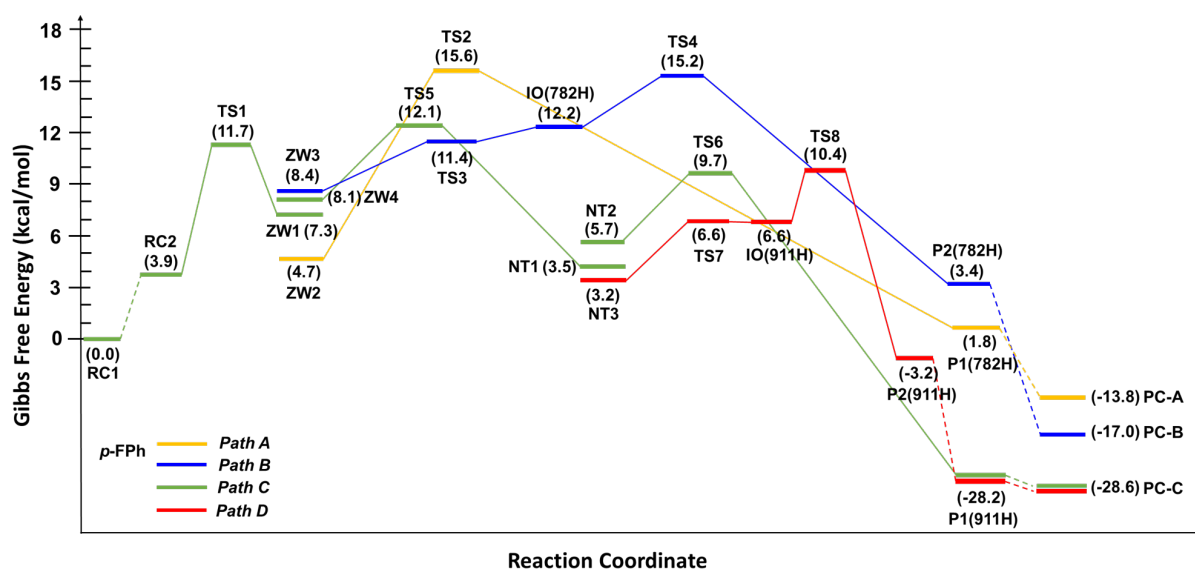


Figure 7.10: Gibbs free energy profile for the reaction of *p*-F phenyl ester (compound 4) with Lys779 in the active site of PI3K δ calculated with the ONIOM (M062X/6-31+G(d,p):PM6) method for four different pathways. Energies are relative to the most stable reactant complex RC1. The first common step (ZW1 formation) is shown by the same color as the most plausible Path C (green). ZW2, ZW3, ZW4 represent different conformations of the zwitterion intermediate obtained through reverse IRC calculations from TS2, TS3, TS5, respectively. NT1, NT2, NT3 represent different conformations of the neutral intermediate obtained through IRC calculations from TS5, TS6, TS7, respectively. IO(782H) is the anionic intermediate where Asp782 is protonated. IO(911H) is the anionic intermediate where Asp911 is protonated. P1(782H) and P2(782H) represent amide products (Asp782 is protonated), complexed with *p*-F phenolate ion, obtained through IRC calculations from TS2 and TS4. P1(911H) and P2(911H) represent amide products (Asp911 is protonated), complexed with *p*-F phenolate, obtained through IRC calculations from TS6 and TS8. PC-A, PC-B and PC-C are the final product complex structures (amide + *p*-F phenol) of Path A, B and Path C/D.[33]

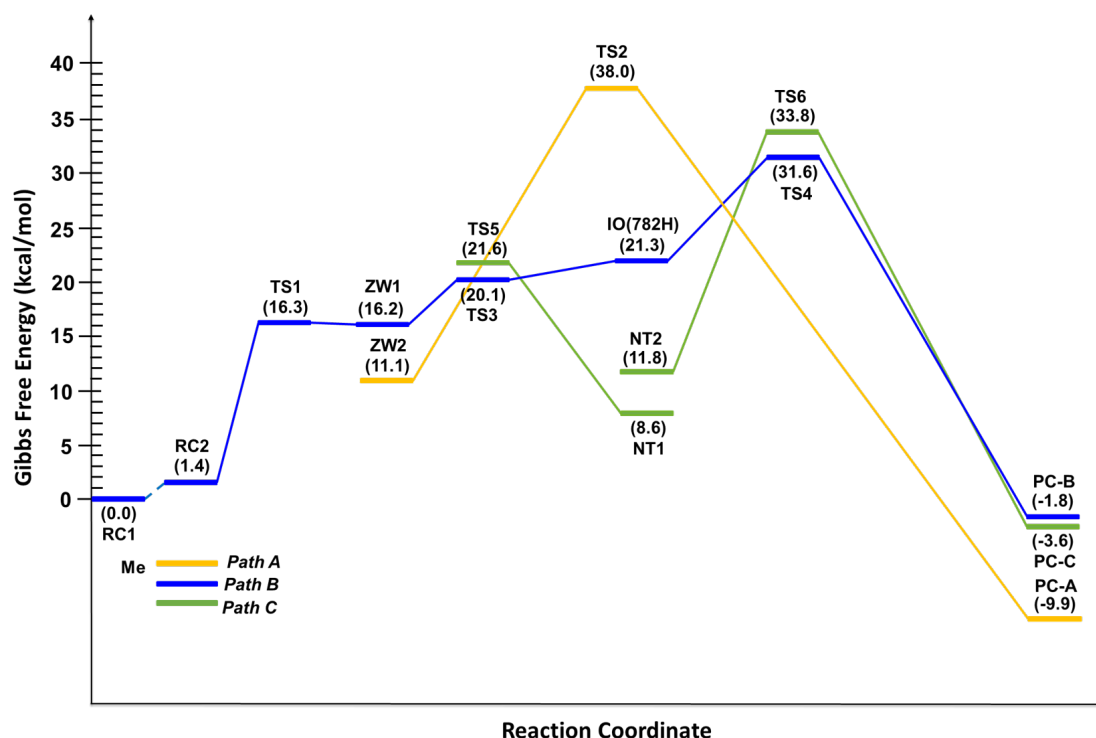


Figure 7.11: Gibbs free energy profile for the reaction of methyl ester (compound 8) with Lys779 in the active site of PI3K δ calculated with the ONIOM (M062X/6-31+G(d,p):PM6) method for four different pathways. Energies are relative to the most stable reactant complex RC1. The first common step (ZW1 formation) is shown by the same color as the most plausible Path B (blue). ZW2, NT1, NT2 and IO(782H) represent different conformations of the zwitterion/neutral/anionic intermediates obtained through IRC calculations from TS2, TS5, TS6 and TS4. PC-A, PC-B and PC-C are the final product complex structures (amide + methanol) of Path A, B and C, respectively.[33]

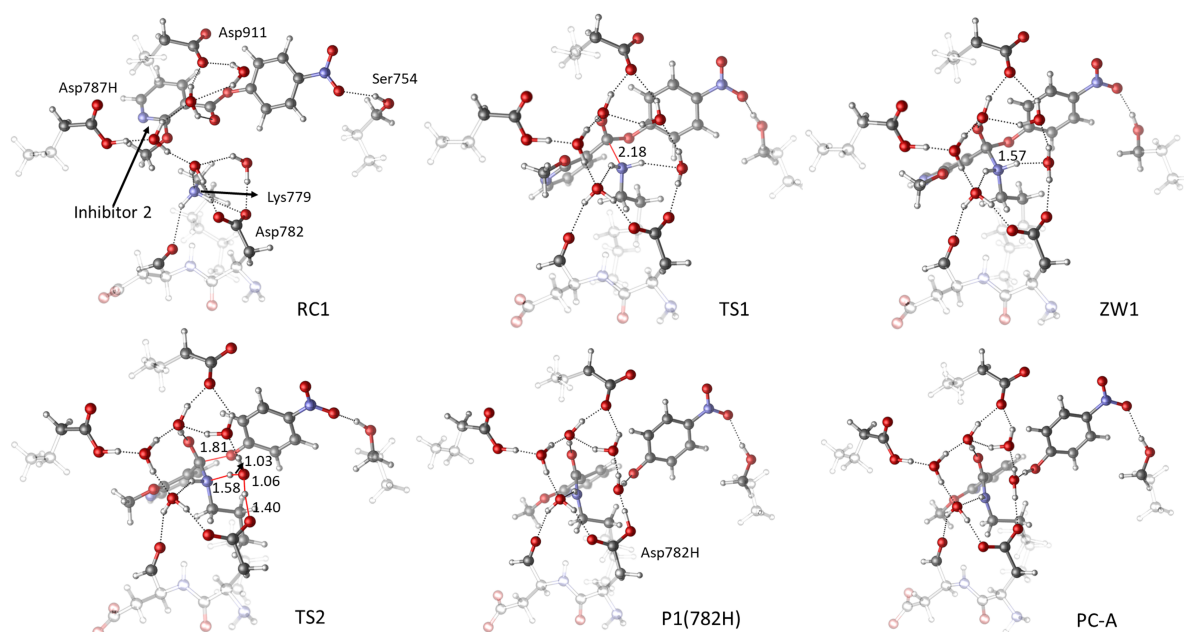


Figure 7.12: 3-Dimensional views of the optimized structures along Path A of *p*-NO₂ phenyl ester reaction with Lys779 including the ZW formation step. The inhibitor and Lys779 are blurred because they are situated behind. Transparent parts are in low layer.

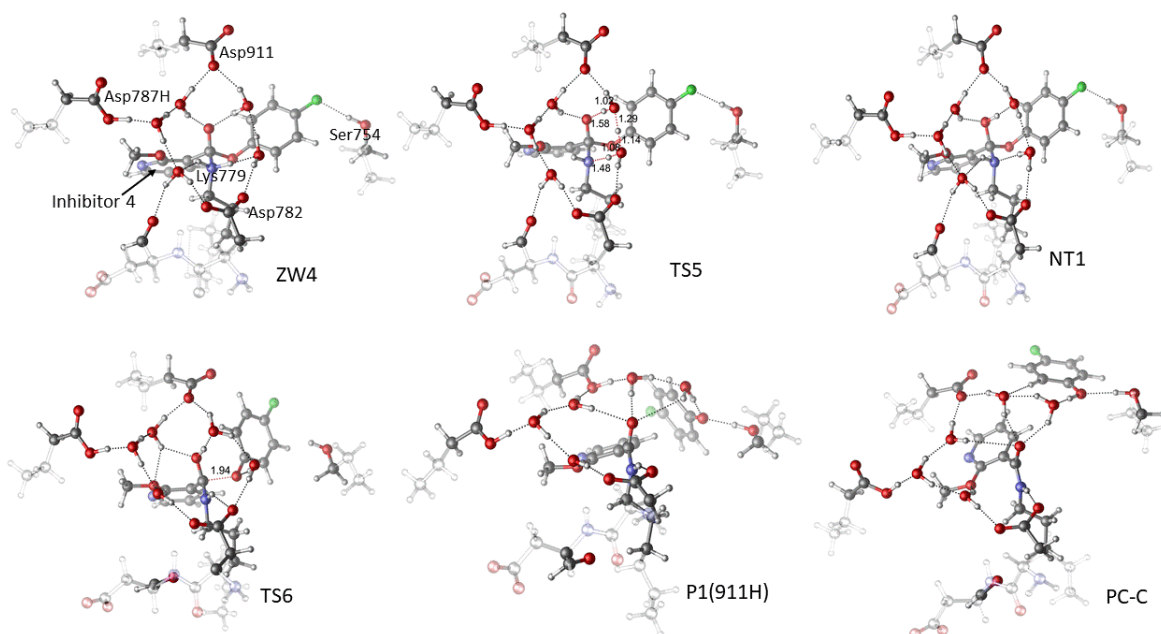


Figure 7.13: 3-Dimensional views of the optimized structures along Path C of *p*-F phenyl ester reaction with Lys779. The inhibitor and Lys779 are blurred because they are situated behind. Transparent parts are in the low layer.[33]

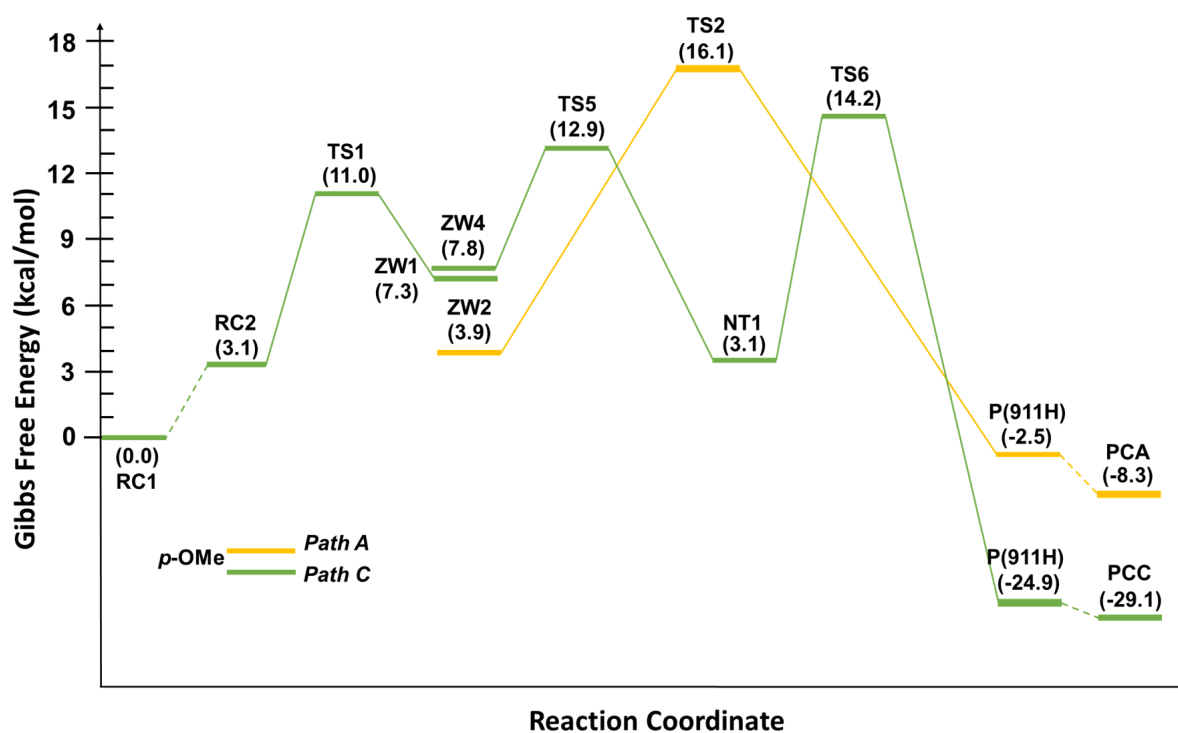


Figure 7.14: Gibbs free energy profile for the reaction of *p*-OCH₃ phenyl ester (compound 6) with Lys779 in the active site of PI3K δ calculated with the ONIOM (M062X/6-31+G(d,p):PM6) method. Energies are relative to the most stable reactant complex RC1. Path A via Asp782 protonation and Path C via Asp911 protonation are compared. The first common step (ZW1 formation) is shown by the same color as the most plausible Path C (green).[33]

- Class 2: Phenyl esters having weak electron-withdrawing/ donating groups or unsubstituted phenyl. These esters should exhibit a moderate dissociation tendency of the leaving group and could proceed through Path A or C depending on the actual substituent. They should display intermediate activation energies and reaction kinetics. Compounds 5 (*p*-H) and 4 (*p*-F) belong to this class. The fluorine atom is usually considered as a weak electron-withdrawing substituent with some extent of electron-donating ability via π -conjugation effect. Anomalous behavior of fluorine is documented in the literature with studies on photoelectron spectroscopy, proton affinities,[188] and reactivity toward electrophilic aromatic substitution reactions.[189] This unexpected behavior of fluorine was ascribed to the fact that, at the para position, the π -electron donating character of fluorine dominates its electron withdrawing character. Accordingly, our results confirm its π -electron donating character.
- Class 3: Phenyl esters bearing strong electron-donating groups. They should clearly prefer Path C because of an unfavorable dissociation step, and would exhibit higher activation barriers and lower kinetics than Class 1 and Class 2 inhibitors. The *p*-OCH₃ (compound 6) is an illustrative example of this class.

Though the whole enzymatic mechanism is certainly more complex than the reaction studied here, our results for *p*-NO₂, *p*-F, *p*-H, and *p*-OCH₃ qualitatively agree with the kinetic study reported by Dalton et al.[12] for PI3K δ .

Thus, the reactivity order predicted by the calculations *p*-NO₂ > *p*-F > *p*-H > *p*-OCH₃ (based on the rate-determining TS of the most preferred path) is in perfect agreement with the experimental k_{inact}/K_I values reported by Dalton et al.[12] for *p*-NO₂ ($1.9 \times 10^5 \text{ s}^{-1} \text{ M}^{-1}$), *p*-F ($2.1 \times 10^4 \text{ s}^{-1} \text{ M}^{-1}$), *p*-H ($1.7 \times 10^4 \text{ s}^{-1} \text{ M}^{-1}$), and *p*-OCH₃ ($3.1 \times 10^3 \text{ s}^{-1} \text{ M}^{-1}$). Note that Dalton et al. proposed that K_I involves the first reversible steps in Figure 2.1 on page 20 (including ZW formation) while k_{inact} is associated with the remaining steps (presumably with the rate-determining one), so that k_{inact}/K_I represents the overall inhibition process.

7.4 Conclusion

Lysine-targeted covalent inactivation of PI3K δ enzyme by the ester inhibitors in Figure 7.1 is a more complex mechanism than it appears in Figure 2.1 on page 20. First of all, no single mechanism is valid for all the studied inhibitors. Instead, the most plausible covalent inhibition pathway and/or the rate-determining transition state alter according to the electron withdrawing/donating nature of the para substituent in the phenolate leaving groups because the enzyme active site offers alternative paths with slightly more stable transition states for different *p*-substituted phenyl ester inhibitors. The fact that the energy difference between the TSs of these paths is smaller than 3 kcal/mol (i.e., TS2 of *p*-NO₂, TS2 of *p*-H, TS5 of *p*-F and TS6 of *p*-OCH₃) may account for the similar k_{inact} values observed in the kinetic studies.[12]

Second, our results reveal that Asp787 plays active role in the deprotonation of Lys779 while the slightly different basicity of Asp782 and Asp911 residues are particularly important in tuning the direction of the covalent inhibition paths by abstracting the proton either from the zwitterion in Paths A/B or from neutral intermediate in Paths C/D. The anionic intermediates IO(782H) in Path B and IO(911H) in Path D may account for the experimentally proposed ionic

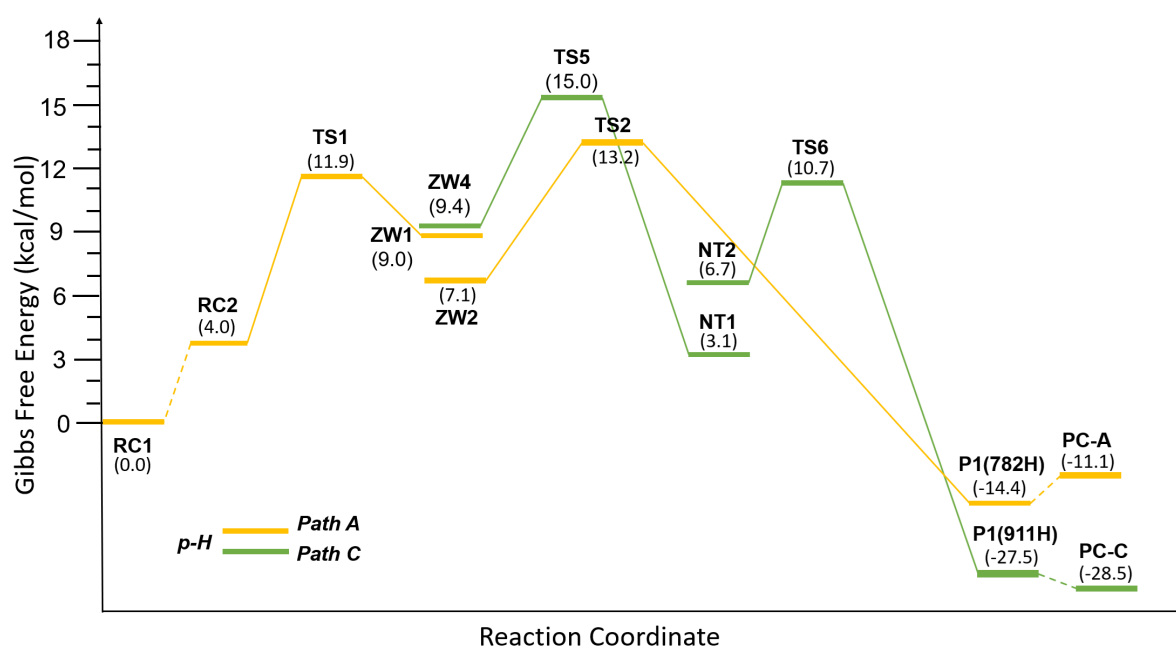


Figure 7.15: Gibbs free energy profile for the reaction of *p*-H phenyl ester (compound 5) with Lys779 in the active site of PI3K δ calculated with the ONIOM (M062X/6-31+G(d,p):PM6) method. Energies are relative to the most stable reactant complex RC1. Path A via Asp782 protonation and Path C via Asp911 protonation are compared. The first common step (ZW1 formation) is shown by the same color as the most plausible Path A (orange).[33]

intermediate in Figure 2.1 on page 20, but the mechanisms involving them appear to be less plausible. As a result, we propose that inhibitors with sufficiently strong electron-withdrawing group proceed with direct departure of the leaving group, whereas inhibitors having an electron-donating substituent prefer the formation of the NT intermediate before dissociation.

Table 7.5: Total energies (E), Gibbs free energies (G) and relative Gibbs free energies (ΔG) for all stationary points optimized at M062X/6-31+G(d,p):PM6 level for *p*-OCH₃ phenyl ester (inhibitor 6).[33]

<i>p</i> -OCH ₃	E (a.u)	G (a.u)	ΔG (kcal/mol)
RC1	-2330.198681	-2329.268863	0.00
RC2	-2330.195405	-2329.263922	3.10
TS1	-2330.185955	-2329.251330	11.00
ZW1	-2330.194585	-2329.257229	7.30
PATH A			
ZW2	-2330.200066	-2329.262614	3.9
TS2	-2330.176781	-2329.243163	16.1
P1(782H)	-2330.205185	-2329.272862	-2.5
PC-A	-2330.213121	-2329.282028	-8.2
PATH C			
ZW4	-2330.194816	-2329.256425	7.8
TS5	-2330.182522	-2329.248391	12.9
NT1	-2330.202077	-2329.263860	3.1
NT2	-2330.190184	-2329.254248	9.2
TS6	-2330.179956	-2329.246235	14.2
P1(911H)	-2330.234966	-2329.308501	-24.9
PC-C	-2330.246013	-2329.315293	-29.1

Third, we propose that, prior to the covalent bond formation, water-assisted proton transfer from Lys779 to Asp787 should occur to activate Lys for nucleophilic attack (Figure 7.5). Since targeting the lysine779 in PI3K δ is a relatively new approach, there is a limited amount of experimental/theoretical studies, and many aspects of such a covalent inhibition mechanism remain unexplored. In particular, there is no experimental support for the role of Asp787 in PI3K δ . From this point of view, we think that our computational results will make a novel contribution to the literature. In most enzymes/proteins, Lys residues are protonated and cannot act as a nucleophile. However, there are several examples of enzymatic reactions emphasizing the requirement for Lys deprotonation,[3, 190, 191] and the role of aspartates and glutamates in those processes has been revealed.[191, 192, 193] Moreover, deprotonation of Lys via water

Table 7.6: Total energies (E), Gibbs free energies (G) and relative Gibbs free energies (ΔG) for all stationary points optimized at M062X/6-31+G(d,p):PM6 level for phenyl ester (inhibitor 5).[33]

<i>p</i> -H	E (a.u)	G (a.u)	ΔG (kcal/mol)
RC1	-2214.629947	-2214.813469	0.00
RC2	-2214.626191	-2214.807152	4.0
TS1	-2214.618581	-2214.794550	11.9
ZW1	-2214.623608	-2214.799616	8.7
PATH A			
ZW2	-2214.624491	-2214.802234	7.1
TS2	-2214.616592	-2214.792432	13.2
P1(782H)	-2214.652227	-2214.836336	-14.4
PC-A	-2214.651707	-2214.831090	-11.1
PATH C			
ZW4	-2214.623537	-2214.798445	9.4
TS5	-2214.619231	-2214.790356	14.5
NT1	-2214.632695	-2214.808490	3.1
NT2	-2214.626815	-2214.802803	6.7
TS6	-2214.621044	-2214.796400	10.7
P1(911H)	-2214.674514	-2214.857224	-27.5
PC-C	-2214.676251	-2214.858892	-28.5

molecules in the water-accessible surface of some enzymes has been reported as well. Similar to our water-assisted mechanism, there are various examples of long-range proton transfers from protonated Lys/Lys-Schiff bases mediated by H-bonded network of water molecules and/or amino acid residues with hydroxyl groups.[19, 194, 195, 196, 197, 198] Especially, Lys-water- Glu motif of histone-Lys modifying enzymes is analogous to our case.[190] Therefore, we think that the water-assisted proton transfer from Lys779 to Asp787 supported by our calculations is a reasonable one. Experimentally, the role of Asp787 could be verified by site-directed mutagenesis of this residue (for example to alanine).

It should be kept in mind that the quantum cluster approach used here employs the simplified description of the enzyme involving the residues and water molecules about 10 Å from the ester functional group in the active site. Besides, the description of the process is based on a static vision neglecting, the full dynamics of the protein system on the actual potential energy hypersurface. It is expected that the structural and conformational flexibility of enzymes also influences the calculated energies.[199, 200] Although QM/MM MD simulations account for the enzyme dynamics and achieve quite accurate energetics, they are very expensive and prohibitive for the complex systems requiring the examination of many alternative mechanisms. On the other hand, the quantum chemical cluster approach is a known powerful tool in modeling enzymatic reactions since it allows optimization of the large number of intermediates and transition structures along reaction coordinates of various alternative mechanisms. General trends in their calculated energies can discriminate between different reaction pathways and administer relatively fast evaluation of the most plausible mechanisms.[185]

Thus, our computational results provide a clear overview of four alternative reaction mechanisms of activated phenyl esters for the lysine-targeted covalent inactivation of PI3K δ within its active site. Calculations can correctly discriminate the covalent and noncovalent inhibitors which might be a useful strategy for the screening of new potential PI3K δ inhibitors. Refinement of the most favorable pathways from quantum cluster calculations will also facilitate the highly accurate QM/MM MD simulations which is our next goal.

Chapter 8

Lysine-targeted covalent inhibition mechanism of PI3K δ via dual-level QM/MM molecular dynamics simulations

8.1 Introduction

It is obvious that in order to obtain accurate free energy profiles of large complex systems such as enzymes, statistical sampling and dynamical effects must be included by employing state-of-the-art MD simulations based on first-principles. Therefore, the main task of this chapter is to elucidate fully the free energy profile of the inhibition mechanism of PI3K δ enzyme by applying dual level QM/MM MD simulations. However, despite significant advances in computer technology and theoretical approaches, such simulations demand high computational costs. In this context, the information gathered from the simpler molecular systems reported in previous chapters sheds light on this part of the study, as follows:

- (i) Based on the QM studies applied to small models both in the aqueous medium (Chapter 4) and within the enzyme active site (Chapter 7), the most plausible inhibition pathway has been identified for *p*-NO₂ phenyl derivative of the ester inhibitor (inhibitor 2). Therefore, such mechanism (the direct dissociation mechanism) was selected to employ QM/MM MD simulations. This chapter presents QM/MM MD simulations on the direct dissociation mechanism of covalent inhibition by *p*-NO₂ derivative and partly by *p*-F phenyl derivative of the ester inhibitor.
- (ii) Reaction coordinates of direct dissociation mechanism were established in the light of Chapter 4 and Chapter 7. The first reaction coordinate is the C-N distance representing the Lys779 amine nitrogen approaching the carbonyl carbon of the inhibitor, producing the ZW intermediate. The second coordinate is the C-O distance representing the dissociation of the phenolate anion from the ester inhibitor.
- (iii) ONIOM computations in Chapter 7 let us understand that prior to the nucleophilic attack on ester, Lys779 should be deprotonated by Asp787 which is the best candidate to abstract this proton. The starting structural setup is decided in light of this information.

- (iv) The dual-level protocol applied to the model system in bulk water was established in Chapter 5 and it was adopted for the simulations modeling in the enzyme. The computational protocol has mainly two steps: finding the free energy profiles by US QM/MM MD simulations with a low QM level (semi-empirical PM3 method), next, applying the FEP correction on each point of the US simulation to refine the energies with high QM level (M062X/6-311+G(d,p)).

8.2 Computational Details

According to our findings from ONIOM calculations (chapter 7 section 7.3.1), Lys779 should donate its proton to Asp787 in order to gain the required nucleophilicity for the ester aminolysis reaction. Therefore, prior to US simulations, classical MD simulation of this activated form of the enzyme (deprotonated Lys779, protonated Asp787) was performed on the catalytic domain of the x-ray crystal structure of the enzyme (PDB ID: 6ez6) in complex with the methyl ester derivative (inhibitor 8 in Figure 7.1). The details of the MD simulation procedure are given in chapter 6. The initial structure for the US simulations was obtained from the last frame of the 100 ns MD simulation after substituting the methyl group of the inhibitor 8 with the *p*-NO₂ phenyl group and then 200 ps equilibration of this newly created structure.

In order to perform US simulations on direct dissociation mechanism (Path A in Chapter 7) QM and MM layer partitioning was performed as follows: Similar to ONIOM layer separations, the inhibitor (its pyridine ring and the ester group), the amino group of Lys779 and the carboxylate group of Asp782 were selected as QM layer; all remaining atoms of the protein, other heterocyclic groups of the inhibitor and water molecules were represented in MM layer (Figure 8.1). Asp782 was included in QM region because, according to ONIOM calculations, this residue behaves as a base in the second step of the reaction. On the other hand, Asp911 and Asp787 are not directly involved in the mechanism. Thus, they were not included in the QM region to avoid high computational cost. Unfortunately, it was not possible to include any water molecule in the QM region because of the complications arising from the interchange of the water molecules with the other water molecules in the MM region during US simulations. For the QM/MM boundary, hydrogen was used as link atom as described in section 3.5.

For the inhibitor, the General Amber Force Field (GAFF)[173] was used to determine the atomic types and bond, angle, and dihedral angle parameters. Point charges were obtained by using the Restricted Electrostatic Potential (RESP) protocol by means of the Antechamber module of AMBER 16.[170] As aforementioned, PM3 method[117] was utilized for QM region, while enzyme structure and water molecules placed in MM region were defined by ff14SB[140] and TIP3P, respectively.

Each US point was subjected to 20 ps equilibration followed by 100 ps production run. The last snapshot of the 4th window was used as the starting structure for the next 4 windows. The Weighted histogram analysis method (WHAM)[154, 155] was used to generate the energy profile with respect to the reaction coordinate. Berendsen barostat was used for pressure control while Langevin thermostat was employed to maintain the constant temperature with a collision frequency of 2ps⁻¹. For the first step of the reaction, the reaction coordinate was selected as the C-N distance between carbonyl carbon of the inhibitor and nitrogen of the Lys779. The windows were spaced with 0.05Å intervals along the reaction coordinate, which was varied

from 4.00Å to 1.50Å. For the second step, the C-O distance between the ester carbonyl carbon and the leaving group oxygen atom located on the inhibitor backbone was selected as reaction coordinate. Similar to the first step, the windows were created for each 0.05Å interval starting from 1.55Å until 3.60Å C-O distance. However, the choice of the C-O coordinate intervals was limited due to the spontaneous occurrence of “chemical events” belonging to a different potential energy surface, as discussed below. As an indication, note that these calculations lasted approximately 240 CPU hours using 12 cores for each US window.

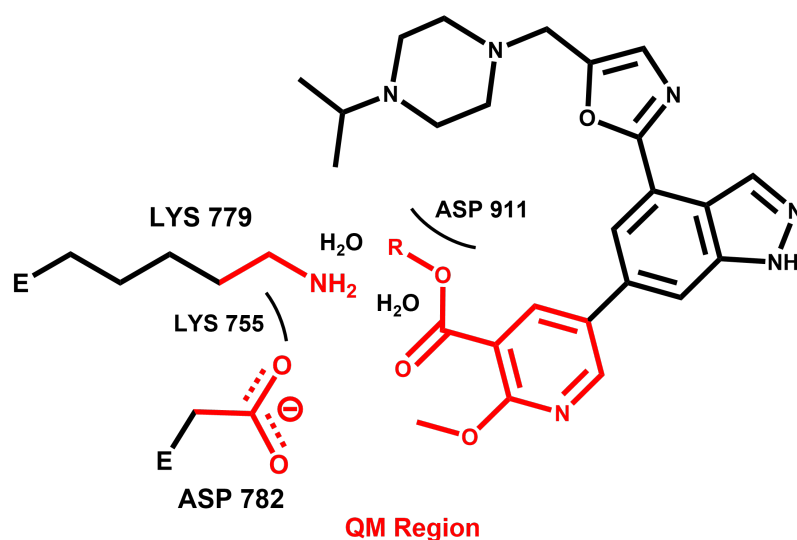


Figure 8.1: Schematic representation of QM region (red) and its nearest surrounding in MM region (black) used in QM/MM MD simulations. Full catalytic domain is included in the simulations but the remaining residues and water molecules are not shown for simplicity.

Free energy corrections were calculated for many points along the reaction coordinate. For each point, 1000 snapshots were extracted from the corresponding US trajectory including the inhibitor, enzyme and the water molecules. Same QM/MM scheme was selected as applied in US simulations described above. 10Å cutoff distance is used (both Coulombic and van der Waals interactions) for single point calculations. The high-level computations by M062X/6-311+G(d,p) were performed for QM layer to obtain the free-energy corrections in terms of the dual-level approach described in the Theoretical Background chapter. These QM/MM calculations at DFT level were done using the conventional AMBER/Gaussian interface and they have lasted approximately 8 minutes of CPU in 12 cores for each snapshot (total 1000 snapshots) to obtain the corrected energy of one point on the reaction coordinate.

8.3 Results and Discussion

8.3.1 Free Energy Profiles via QM/MM MD Simulations

First step: ZW formation. The free energy surface generated from US is reported in Figure 8.2. It should be noted that the reactant complex (RC) is formed at a C-N distance equal 3.90Å

and the reported energies are relative to this RC. The transition state is found at 2.10Å, having 15.43 kcal/mol of free energy barrier. The ZW formation is observed around 1.70Å with 12.55 kcal/mol energy relative to RC. Thus, the first step is endergonic and has a small reverse activation barrier (about 3 kcal/mol), confirming that ZW formation is a reversible step.

The 3D structures of typical snapshots corresponding to the minima and saddle points obtained in the US simulations are shown in Figure 8.3. Apart from Lys779, the residues Asp782, Lys755, Asp911 and 2 water molecules are found in the vicinity of the reaction center. Several hydrogen bonding interactions are responsible for the stability of the reactant complex at 3.90Å C-N distance. One hydrogen of Lys779 interacts with the carbonyl oxygen of the inhibitor. Asp 782 exhibits hydrogen bonding with the protonated Lys755, and Lys755 has also indirect interaction with Asp 911 through a water molecule. On the other hand, in the TS structure, positively charged Lys755 stabilizes the negative charge developing on the carbonyl oxygen. At the same time, it interacts with Asp782 through hydrogen bonding. And accordingly, the positive charge developing on the nitrogen of Lys779 is stabilized with the negatively charged Asp782. Moreover, hydrogen bond interaction has been observed between water molecules and residues, Asp911 and Lys755. Lastly, the ZW structure is located at the 1.70Å distance between the nitrogen atom of Lys 779 and carbonyl carbon of the inhibitor. The positive and negative charges on the ZW structure have been stabilized by the surrounding residues. Hydrogens attached to the positively charged nitrogen atom have interactions with Asp782, while Lys755 and a water molecule stabilize the negative charge on the carbonyl oxygen through hydrogen bonding. Concurrently, Asp911 and Lys755 indirectly interact with the inhibitor via water molecules.

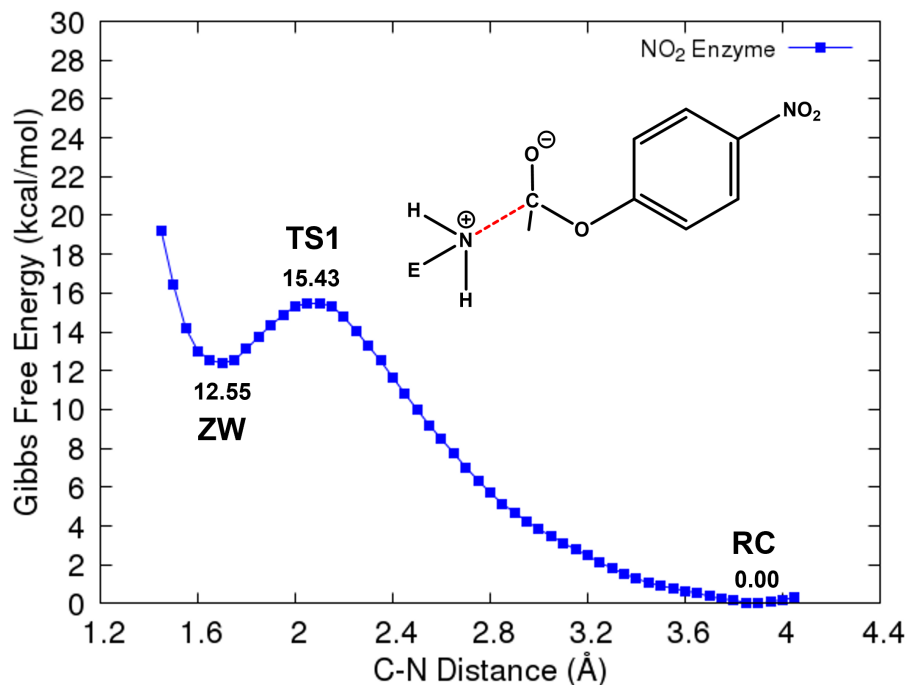


Figure 8.2: Free energy profile of ZW formation step within PI3K δ enzyme obtained from US PM3/ff14SB/TIP3P simulations.

Second step: leaving group dissociation. The QM/MM MD simulations of this step were

applied to both *p*-NO₂ and *p*-F phenyl ester inhibitors. Our US simulations of model system (chapter 5, section 5.3.2) revealed that *p*-F phenyl ester displays slightly different behavior than *p*-NO₂ phenyl ester, but only for the leaving group dissociation step. Thus, in order to explore the influence of ester inhibitor substitution in the second step of the reaction, the fluorine derivative was considered within the enzyme environment as well.

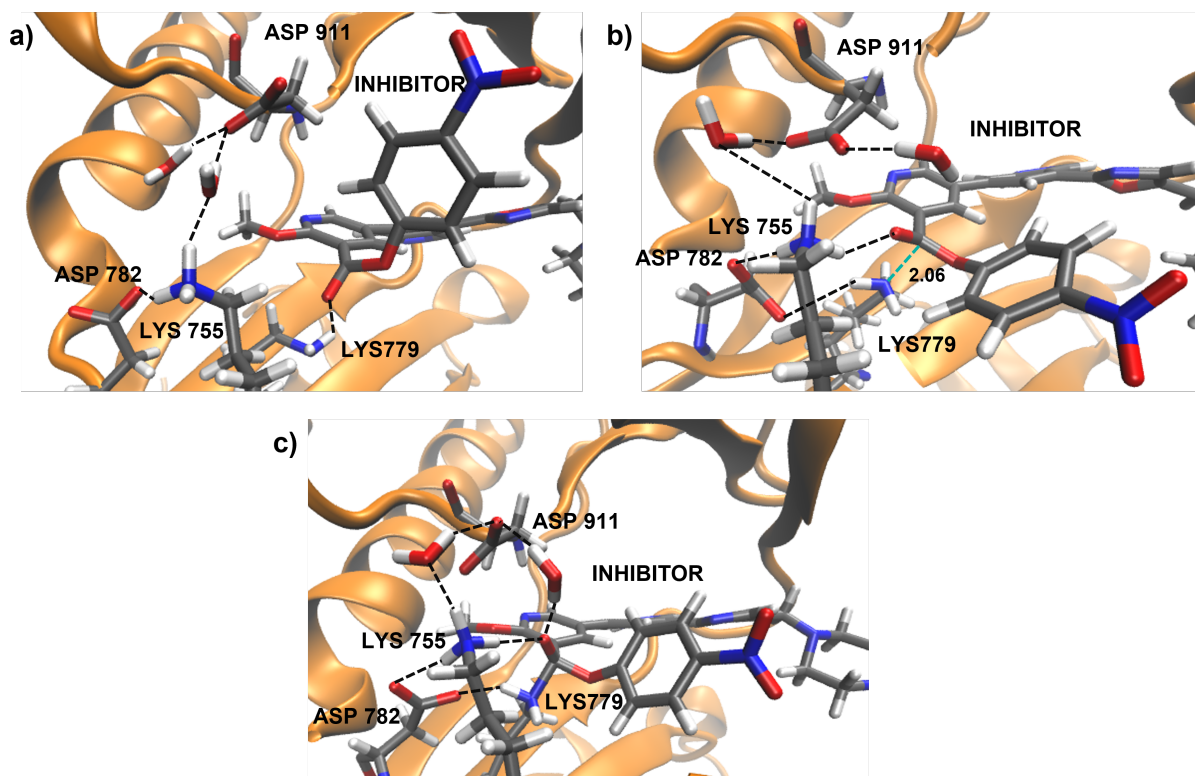


Figure 8.3: 3D structures of a) reactant complex (at 3.90Å C-N distance) b) transition state (at 2.06Å C-N distance) c) zwitterionic intermediate (at 1.70Å C-N distance), obtained from US PM3/ff14SB/TIP3P simulations.

The second step of the reaction starts with the ZW intermediate and the dissociation of the C-O bond of the ester results in the separation of phenolate leaving group from the ZW. Therefore, we study the elongation of the C-O distance in the US simulations. The free energy profiles obtained for both inhibitors are given in Figure 8.4. Unfortunately, similar inconvenience as described in chapter 5 (section 5.3.2) is encountered here as well. The first problem appears when we start US at short C-O distances (1.45Å) close to the equilibrium value in ZW; US simulations lead to dissociation of the amine group, the system returning back to the reactants because of the small activation energy for the reverse reaction. For this reason, the US simulations and the free energy profile in Figure 8.4 starts with a structure at 1.65Å C-O distance instead of the fully relaxed ZW structure. However, in order to complete the energy profile, we were able to estimate the energy change between the ZW and the structure at 1.65Å using the probability distribution of the C-O distance in the direct MD simulation of ZW (i.e. the simulation in the US process for the first step and for a C-N distance corresponding to the ZW structure at the free energy minimum). The C-O probability distribution $P(\text{C-O})$ is transformed into a free

energy change $-RT\ln(P(\text{CO}))$ around the correct ZW geometry, which is fitted by a 2nd order polynomial, then extrapolated to C-O 1.65Å. This part of the free energy profile is displayed as a blue dotted line in Figure 8.4. As shown, the free energy change from C-O 1.45Å to 1.65Å is estimated as 5 kcal/mol. The second problem appears only for the *p*-F phenyl ester inhibitor at long C-O distances: After the *p*-F phenolate anion is formed at 2.65Å, a spontaneous proton transfer from the Lys779 N atom to the *p*-F phenolate takes place leading to the formation of *p*-F phenol at C-O distances > 2.65Å.

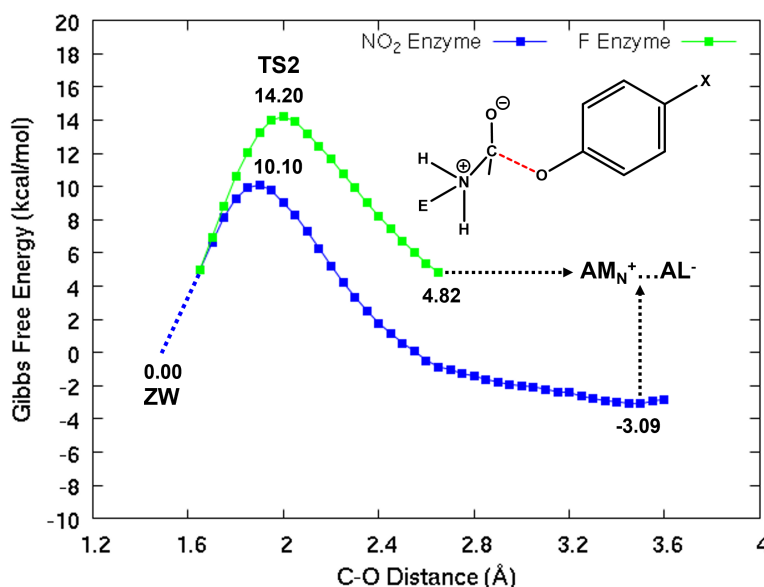


Figure 8.4: Free energy profiles of the leaving group dissociation steps for the reactions of *p*-NO₂ and *p*-F phenyl ester inhibitors with PI3K δ enzyme, obtained from US PM3/ff14SB/TIP3P simulations. ($\text{AM}_\text{N}^+ \cdots \text{AL}^-$) represents the ion-pair product (a complex between a protonated amide and the substituted phenolate anion)

As a result of these drawbacks, the range of available C-O distances along the reaction coordinate is limited because of the spontaneous occurrence of “chemical events” at certain C-O distances. At these specific distances (C-O > 2.65Å for *p*-F phenyl ester and C-O < 1.65Å for both esters), the system is on a different potential energy surface.

For *p*-NO₂ phenyl ester inhibitor, activation energy of the second step is 10.10 kcal/mol with respect to ZW. This step is exergonic forming the amide with a positively charged nitrogen and the *p*-NO₂ phenolate anion which we refer as ion-pair product ($\text{AM}_\text{N}^+ \cdots \text{AL}^-$). This ion-pair is more stable than ZW structure and TS2 by 3.09 kcal/mol and 13.19 kcal/mol, respectively. The 3D structures of TS2 and ion-pair product complex are illustrated in Figure 8.5 with some representative snapshots. Note that the 3D view of the ZW has been already given in Figure 8.3c. In the TS2 structure, the critical distance between C and O is 1.90Å and the oxygen of the *p*-NO₂ phenolate interacts with the protons on the Lys779 nitrogen atom. It is also observed that Lys755 acts as a hydrogen-bond bridge between Asp782 and the ester carbonyl oxygen by donating its protons to these groups. Additionally, in this reaction center, two water molecules exhibit hydrogen bond interactions with each other and with surrounding residues (especially with Asp911 and Lys755). After TS2, the ion-pair product complex is found at 3.50Å C-O

distance. During the formation of ion-pair product, the positively charged Lys755 plays an important role in the reaction center: it stabilizes carbonyl oxygen of the inhibitor, and two negatively charged Asp residues (namely, Asp 782 and Asp 911). Moreover, the protonated amide structure formed at the end is stabilized by negatively charged Asp 782 and *p*-NO₂ phenolate oxygen.

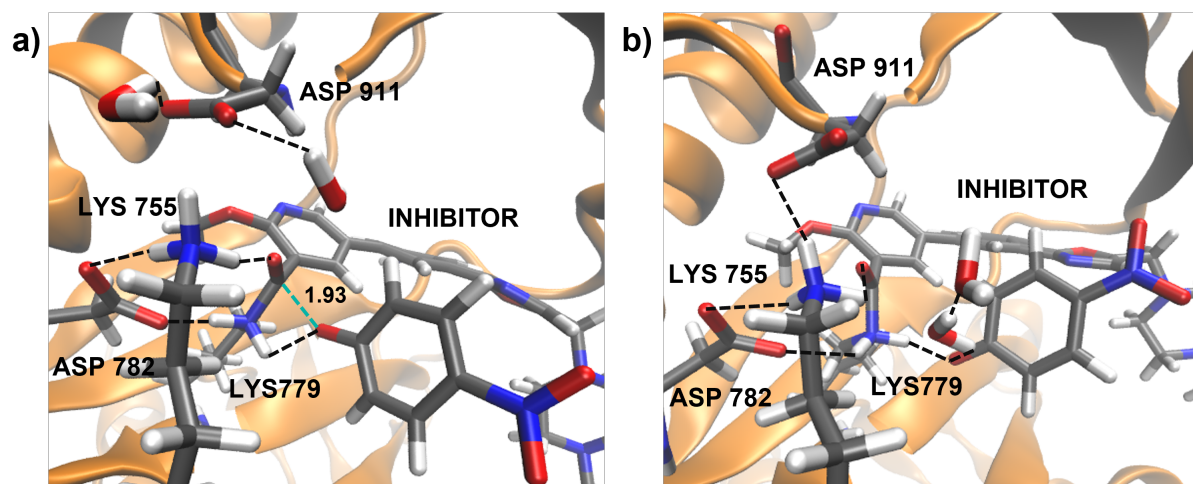


Figure 8.5: 3D structures of a) transition state TS2 (at 1.93Å C-O distance) and b) ion-pair product complex for the *p*-NO₂ phenyl ester inhibitor (at 3.50Å C-O distance) obtained from US PM3/ff14SB/TIP3P simulations.

For the *p*-F phenyl ester inhibitor, the TS2 barrier is found to be 14.2 kcal/mol based on the assumption that the dotted-line estimation is also 5 kcal/mol. This barrier is relatively higher than that of *p*-NO₂ substituted case, as expected. As observed in previous calculations on the *p*-F phenyl acetate model in aqueous solution (Chapter 5), a spontaneous proton transfer from amide nitrogen to the oxygen of the *p*-F phenolate leaving group is observed, leading directly from ZW to the neutral PC instead of to an ion-pair intermediate, as in the case of *p*-NO₂ presented above. The snapshots belonging to the ZW-like structure and TS2 are illustrated in Figure 8.6. In the first structure, the carbonyl oxygen of the inhibitor interacts with the proton of Lys755 and of a water molecule. One of the hydrogens attached to the positively charged nitrogen atom of Lys779 is stabilized by the negatively charged Asp782. On the other hand, in transition-state structure, in addition to similar interactions as in the ZW-like structure, the negatively charged oxygen of the leaving group and one of the hydrogen atoms of amino group of Lys779 have an interaction with each other.

In order to illustrate the structural changes caused by the spontaneous proton transfer, 3D views of snapshots at 2.65Å and 2.70Å C-O distances are displayed in Figure 8.7. At 2.65Å separation, the leaving group oxygen is hydrogen-bonded to the hydrogen (around 1.75Å) of positively charged nitrogen of Lys779. Moreover, in this structure, one water molecule is bridging the carbonyl oxygen and Asp911. At 2.70Å C-O distance (Figure 8.7b) the proton transfer from Lys779 nitrogen to *p*-F-phenolate oxygen has taken place and not surprisingly, the hydrogen bonding interaction between Lys779 and Asp782 is lost.

To shed light on the different behavior displayed by *p*-NO₂ and *p*-F phenyl ester inhibitors, the interactions of the leaving group oxygen with surrounding H atoms have been further in-

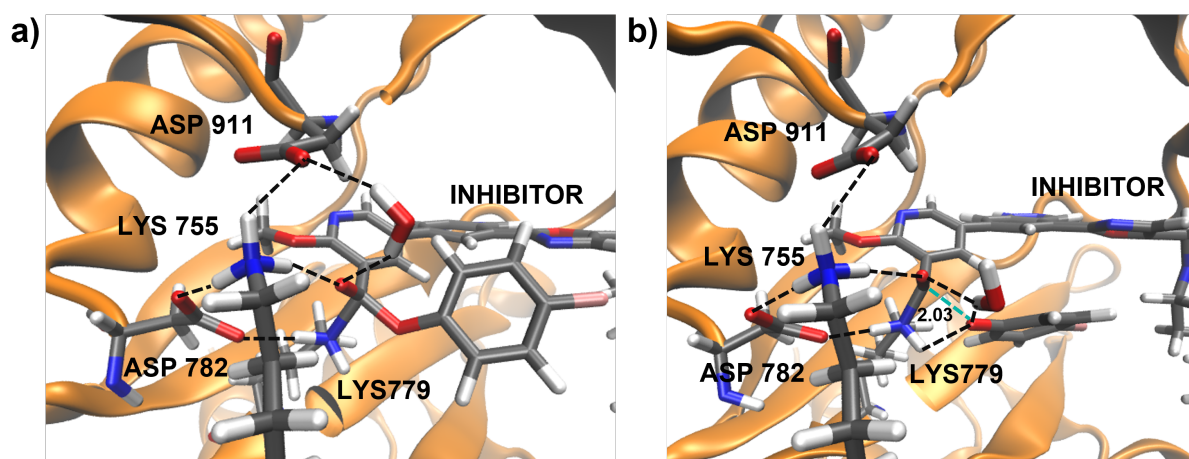


Figure 8.6: 3D structures of a) ZW-like structure (at 1.65Å C-O distance) b) transition state TS2 (at 2.03Å C-O distance) of *p*-F-phenyl ester inhibitor obtained from US PM3/ff14SB/TIP3P simulations.

investigated by looking at some RDFs for a reaction coordinate (C-O distance) equal to 2.65Å (i.e. just before proton transfer occurs in the *p*-F phenyl ester inhibitor case) (Figure 8.8). Similar to the bulk water results in chapter 5, the intensity of the first peak at around 1.75-1.80Å is significantly lower for *p*-NO₂ than for *p*-F, which is an indication that the electron density developing at the phenolate oxygen is more strongly withdrawn by the substituent in the first case, providing a higher stability of the leaving group. Eventually, *p*-NO₂ derivative does not require the proton abstraction during dissociation.

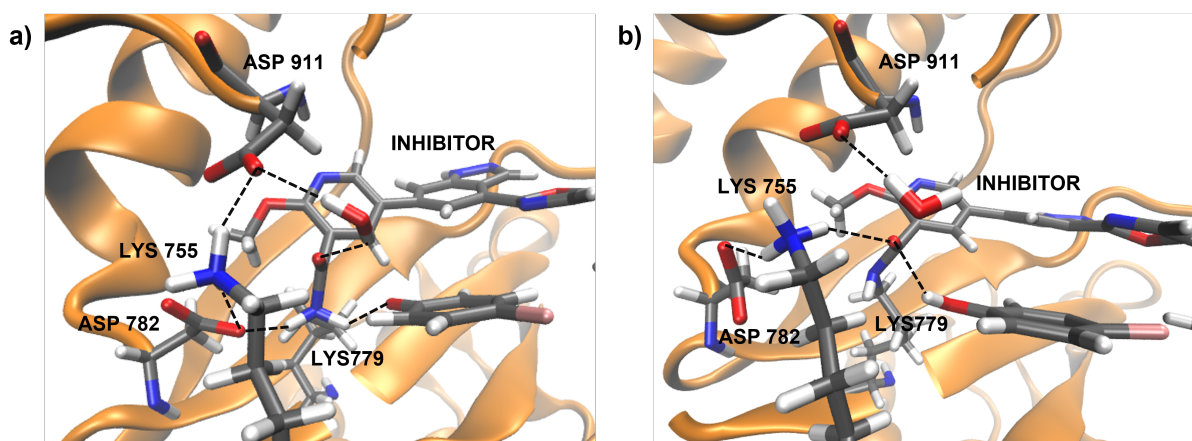


Figure 8.7: Representative snapshots taken from C-O separation around a) 2.65Å and b) 2.70Å of the C-O reaction coordinate of *p*-F-phenyl derivative obtained from US PM3/ff14SB/TIP3P simulations.

8.3.2 Energy Refinement with FEP Correction

High-level free energy corrections for the different reaction steps have been carried out for the *p*-NO₂ phenyl ester inhibitor. The overall free energy profile is given in Figure 8.9. The corrected free energy change between the ZW minimum at a C-O distance 1.45Å and the ZW-like structure at a C-O distance 1.6Å has been approximated by using the PM3 free energy profile after checking that the high-level corrections do not change too much this profile (variations of at most 0.5 kcal/mol have been obtained for the points close to the minimum but an accurate corrected value using free-energy perturbation theory could not be achieved for C-O distances close to 1.6Å due to the lack of a sufficient number of snapshots in the MD trajectory having such distance).

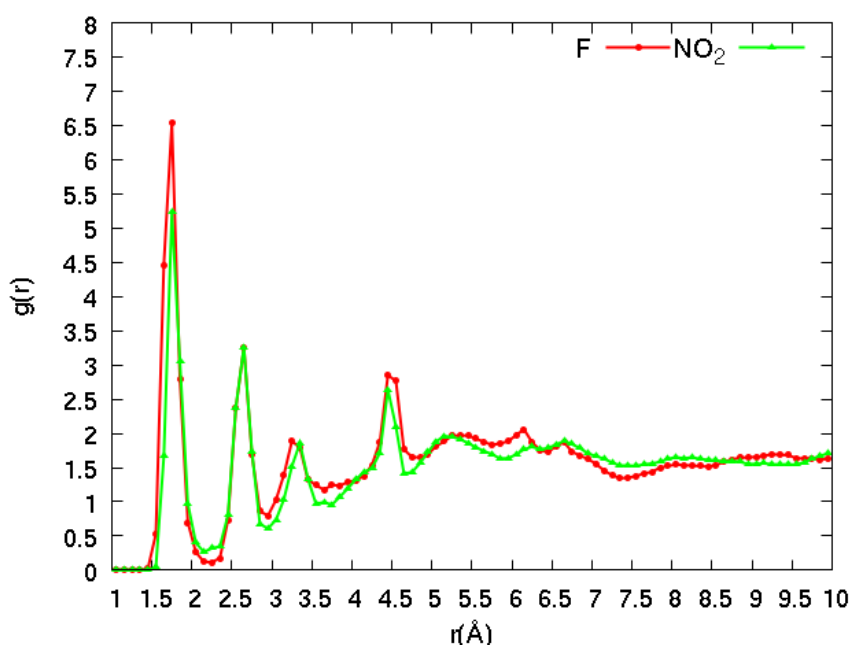


Figure 8.8: Radial Distribution Function of the phenolate oxygen of the *p*-NO₂ (green) and *p*-F (red) phenyl ester inhibitors with surrounding H-atoms.

The energy refinement at the M062X/6-311+G(d,p) level reveals the following important points. A reactant complex free energy minimum is located at a C-N distance 2.60Å, which is consistent with the results found in the model systems, but contrasts with the low-level results in the enzyme. The activation free energy for the formation of the zwitterionic intermediate amounts 5.4 kcal/mol, and this first step is slightly exergonic. In the second step, the activation energy is about 18-19 kcal/mol and an ion pair minimum is obtained at about C-O equal to 2.7Å. This step is endergonic, though we assume that quickly after formation of the ions, a spontaneous proton transfer occurs leading the system to the final products.

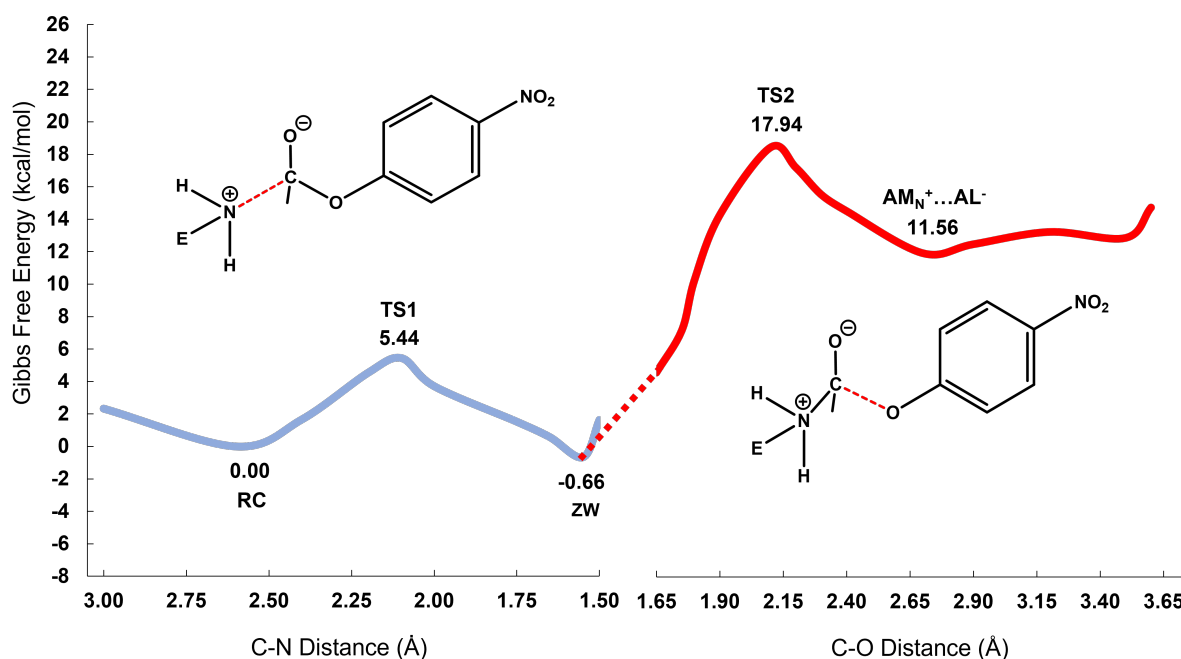


Figure 8.9: Gibbs free energy profile for the inhibition process (first (blue) and second (red) steps) of the PI3K δ enzyme by the *p*-NO₂ phenyl derivative (inhibitor 2), as calculated with the dual-level QM/MM MD approach. For the low and high QM levels, we use the PM3 and M062X/6-311+G(d,p) methods, respectively.

8.4 Conclusion

In this chapter, we employed state-of-the-art QM/MM MD simulations on the chemical mechanism of the ester aminolysis reaction between inhibitor 2 and Lys779 of activated PI3K δ . For the sake of exploring the difference in their behavior, and only for the second step of the reaction, QM/MM MD simulations were employed on inhibitor 4 as well. In line with our previous QM/MM MD simulations on model system in bulk water (chapter 5) and the ONIOM calculations (chapter 7), the dissociation of *p*-NO₂ phenolate ion is more feasible with a smaller energy barrier than that of *p*-F phenolate. Indeed, *p*-F phenolate tends to abstract a proton during its departure because of its lower stability with respect to the *p*-NO₂ phenolate (higher pK_a of the corresponding alcohol).

The high-level computations here predict a free energy difference between TS2 (the transition step for the zwitterion dissociation into ions) and ZW equal about 18-19 kcal/mol. This is an approximated value because of the assumptions made in the calculations, and more specifically on the difficulty to evaluate the free energy change around the ZW structure with the US technique due to its spontaneous dissociation. The comparison with experiment can be made by using the reported k_{inact} value, which is 0.0075 s^{-1} , and from which one deduces an activation free energy equal to 20.4 kcal/mol. The experimental k_{inact} is deduced from the measured k_{obs} at high inhibitor concentrations.[12] It is supposed to describe the activation process from

the non-covalently bounded enzyme-inhibitor complex to the final products with a covalent enzyme-inhibitor bond. The experimental value is close to our calculated value for the ester aminolysis reaction starting from the RC complex. However, one should remind that the RC complex to be formed requires a proton reorganization in the active site, which should imply a proton transfer from the protonated Lys779 to an aspartate group, probably Asp787, when the inhibitor approaches the enzyme active site. The process may occur in a stepwise way or in a concerted mode, i.e. proton transfer occurs concomitantly with the nucleophilic attack. It is estimated to require about 5 kcal/mol according to the study presented in Chapter 7, which should be considered for the comparison with the measured k_{inact} . In the absence of more detailed experimental data, however, it is difficult to conclude if the non-covalently bounded enzyme-inhibitor complex lying behind the measurement of k_{inact} is the structure theoretically found before or after Lys/Asp proton exchange. Further discussion on the different enzyme-ligand binding possibilities associated to k_{inact} measurements can be found in refs.[3, 201]

To elucidate this point in particular, and getting more accurate results in general, one could imagine increasing the size of the QM system in the QM/MM calculations. In particular, the role of Lys555 is important since this residue provides stability for the ZW intermediate by acting as a hydrogen-bond donor to the carbonyl oxygen of the ester and might help preventing the issues concerning its dissociation back to RC in the US simulations. Calculations in this direction are ongoing.

Concluding Remarks

In this dissertation, different computational approaches (QM, MD, ONIOM, QM/MM MD) have been used to get deeper insight into the ester aminolysis mechanism in aqueous solution and in the active site of PI3K δ enzyme targeting the Lys779 amino group by the ester inhibitors. The main objective is to elucidate the mechanism of lysine targeted covalent inhibition of PI3K δ with the aim of helping the design of more effective new covalent inhibitors.

Initially, ester aminolysis mechanism in aqueous solution were explored with small model molecules (methylamine reacting with methyl, phenyl, *p*-nitro phenyl acetates) to explore the fundamental aspects of the mechanism and to develop a computational protocol for the “dual-level” QM/MM MD simulations. First, QM calculations were carried out using a static model at the MP2/aug-cc-pVTZ//B3LYP/6-311+G(d,p) level to find all elementary steps of the alternative mechanistic pathways (Chapter 4). A discrete-continuous model in which 5 explicit water molecules interacting with the reactants surrounded by a continuous dielectric environment (PCM) was employed. These small model systems were then studied using a dual-level QM/MM MD approach where the sampling was obtained by PM3/TIP3P method and the obtained structures were then used to obtain perturbative corrections at free energy with a high-level M062X/6-311+G(d,p)/TIP3P approach (Chapter 5). All of this work on model systems allowed us to highlight the following points: For sufficiently electrophilic ester derivatives, such as the *p*-NO₂ phenyl derivative, the most favorable reaction path after formation of the zwitterionic tetrahedral intermediate is not the proton transfer, as proposed for PI3K δ inhibitors in the literature, but the dissociation of the C-OR bond, leading to the formation of an ion pair. The occurrence of a large charge separation then provides the driving force for proton transfer and the formation of the end products, which then takes place, along a rather flat free energy surface. Our study also indicates that the effectiveness of inhibitors with phenolic leaving groups should increase when the derivatives have electron-withdrawing substituents, which activate the ester promoting the formation of the initial zwitterionic intermediate and facilitating subsequent reaction steps. We find that the efficiency of the inhibitors observed experimentally is well correlated with the calculated electrophilicity of the esters, and with the stability of the corresponding zwitterion. Therefore, the calculation of these properties seems to be an interesting tool to evaluate a priori the efficiency of other candidate inhibitors.

Obviously, interactions with the enzymatic environment can modify the mechanisms described for the model systems in aqueous solution. In the rest of our work, we were therefore interested in this problem by looking at the mechanistic aspects in different models of the enzyme. As for the systems in aqueous solution, we first carried out a static study (Chapter 7). Based on the analysis of the trajectories obtained from the classical molecular dynamics simulations (Chapter 6), we built a cluster model of the active site which consists of important

amino acid residues, 5 water molecules and the inhibitor. The calculation method in this case is a QM:QM method of the ONIOM type (M06-2X/6-31+G(d,p):PM6). These calculations provide an energetic overview of different reaction mechanisms that might be occurring within the enzyme active site. Finally, the refinement of the most feasible mechanistic path allowed us to use this information for the QM/MM molecular dynamics simulations (Chapter 8) on the covalent inhibition mechanism by the *p*-NO₂ phenyl derivative of the inhibitor (the most active inhibitor). Employing the similar “dual-level” QM/MM approach as in the small model system we obtained the free energy profile of the covalent inhibition reaction for the enzyme. The calculated barrier is in good agreement with the experimental kinetic data available, which allows us to validate the proposed theoretical approach and the mechanisms obtained.

Overall, all the calculations carried out during this thesis show that the covalent inhibition of the PI3K enzyme by the ester electrophiles is a more complex mechanism than that proposed in the literature from experimental kinetic data. After the formation of the enzyme-inhibitor complex, the reaction begins with an activation of the active site, which occurs by proton transfer from the protonated catalytic Lys to a nearby non-protonated Asp residue. Then the activated Lys reacts with the ester group of the inhibitor forming a zwitterionic intermediate. The rest of the mechanism is however dependent on the electronic properties of the inhibitor. For sufficiently electrophilic inhibitors, such as the *p*-NO₂ phenyl derivative, the reaction continues by the dissociation of the zwitterion to an ion pair complex, then by a proton transfer leading to the formation of the expected products (formation of an amide bond in the enzyme, then a leaving alcohol). The calculations also show that the zwitterionic intermediate has low stability (tend to dissociate backward towards neutral reactants) and that the limiting step of the inhibition process is the one which leads to the ionic dissociation of the zwitterion, at least for the most active inhibitors such as *p*-NO₂ phenyl derivative case. We believe the results of these calculations open new perspectives for the development of more effective inhibitors. Further QM/MM molecular dynamics studies on the enzyme with a more elaborated description of the QM layer are in progress.

Bibliography

- [1] Ferguson, F. M.; Gray, N. S. Kinase inhibitors: The road ahead. *Nature Reviews Drug Discovery* **2018**, *17*, 353–376.
- [2] Mukherjee, H.; Grimster, N. P. Beyond cysteine: recent developments in the area of targeted covalent inhibition. *Current Opinion in Chemical Biology* **2018**, *44*, 30–38.
- [3] Luo, Y. L. Mechanism-Based and Computational-Driven Covalent Drug Design. *Journal of Chemical Information and Modeling* **2021**, *61*, 5307–5311.
- [4] Jones, L. H. Reactive Chemical Probes: Beyond the Kinase Cysteinome. *Angewandte Chemie - International Edition* **2018**, *57*, 9220–9223.
- [5] Hacker, S. M.; Backus, K. M.; Lazear, M. R.; Forli, S.; Correia, B. E.; Cravatt, B. F. Global profiling of lysine reactivity and ligandability in the human proteome. *Nature Chemistry* **2017**, *9*, 1181–1190.
- [6] Liu, R.; Yue, Z.; Tsai, C. C.; Shen, J. Assessing Lysine and Cysteine Reactivities for Designing Targeted Covalent Kinase Inhibitors. *Journal of the American Chemical Society* **2019**, *141*, 6553–6560.
- [7] Knight, J. D. R.; Qian, B.; Baker, D.; Kothary, R. Conservation, variability and the modeling of active protein kinases. *PLoS ONE* **2007**, *2*, e982.
- [8] Zhao, Q.; Ouyang, X.; Wan, X.; Gajiwala, K. S.; Kath, J. C.; Jones, L. H.; Burlingame, A. L.; Taunton, J. Broad-spectrum kinase profiling in live cells with lysine-targeted sulfonyl fluoride probes. *Journal of the American Chemical Society* **2017**, *139*, 680–685.
- [9] Goulielmaki, E.; Bermudez-Brito, M.; Andreou, M.; Tzenaki, N.; Tzardi, M.; De Bree, E.; Tsenteliero, E.; Makrigiannakis, A.; Papakonstanti, E. A. Pharmacological inactivation of the PI3K p110 δ prevents breast tumour progression by targeting cancer cells and macrophages article. *Cell Death and Disease* **2018**, *9*, 678.
- [10] Hanna, B. S.; Roessner, P. M.; Scheffold, A.; Jebaraj, B. M.; Demerdash, Y.; Öztürk, S.; Lichter, P.; Stilgenbauer, S.; Seiffert, M. PI3K δ inhibition modulates regulatory and effector T-cell differentiation and function in chronic lymphocytic leukemia. *Leukemia* **2018**, *33*, 1427–1438.

- [11] Chen, G.; Kumar, J.; Gregory, A.; Stenzel, M. H. Efficient synthesis of dendrimers via a thiol-yne and esterification process and their potential application in the delivery of platinum anti-cancer drugs. *Chemical Communications* **2009**, 6291–6293.
- [12] Dalton, S. E.; Dittus, L.; Thomas, D. A.; Convery, M. A.; Nunes, J.; Bush, J. T.; Evans, J. P.; Werner, T.; Bantscheff, M.; Murphy, J. A.; Campos, S. Selectively Targeting the Kinome-Conserved Lysine of PI3K δ as a General Approach to Covalent Kinase Inhibition. *Journal of the American Chemical Society* **2018**, *140*, 932–939.
- [13] 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*, 909–19.
- [14] Castro-Falcón, G.; Seiler, G. S.; Demir, Ö.; Rathinaswamy, M. K.; Hamelin, D.; Hoffmann, R. M.; Makowski, S. L.; Letzel, A. C.; Field, S. J.; Burke, J. E.; Amaro, R. E.; Hughes, C. C. Neolymphostin A Is a Covalent Phosphoinositide 3-Kinase (PI3K)/Mammalian Target of Rapamycin (mTOR) Dual Inhibitor That Employs an Unusual Electrophilic Vinylogous Ester. *Journal of Medicinal Chemistry* **2018**, *61*, 10463–10472.
- [15] Fournier, J. C.; Evans, J. P.; Zappacosta, F.; Thomas, D. A.; Patel, V. K.; White, G. V.; Campos, S.; Tomkinson, N. C. Acetylation of the Catalytic Lysine Inhibits Kinase Activity in PI3K. *ACS Chemical Biology* **2021**, *16*, 1644–1653.
- [16] Zhou, Z.; Li, L.; Yan, N.; Du, L.; Sun, C.; Sun, T. Synthesis, crystal structure, absolute configuration and antitumor activity of the enantiomers of 5-bromo-2-chloro-N-(1-phenylethyl) pyridine-3-sulfonamide. *Molecules* **2015**, *20*, 20926–20938.
- [17] Chen, R.; Luo, X.; Liang, G. Theoretical studies on the aminolysis mechanism of propylene carbonate with ammonia. *Theoretical Chemistry Accounts* **2015**, *134*, 32.
- [18] Sabbah, D. A.; Vennerstrom, J. L.; Zhong, H. A. Binding selectivity studies of phosphoinositide 3-kinases using free energy calculations. *Journal of Chemical Information and Modeling* **2012**, *52*, 3213–3224.
- [19] Zhang, X.; Ouyang, S.; Kong, X.; Liang, Z.; Lu, J.; Zhu, K.; Zhao, D.; Zheng, M.; Jiang, H.; Liu, X.; Marmorstein, R.; Luo, C. Catalytic mechanism of histone acetyltransferase p300: From the proton transfer to acetylation reaction. *Journal of Physical Chemistry B* **2014**, *118*, 2009–2019.
- [20] Berndt, A. et al. The p110 δ structure: Mechanisms for selectivity and potency of new PI(3)K inhibitors. *Nature Chemical Biology* **2010**, *6*, 117–124.
- [21] Antonczak, S.; Ruiz-López, M. F.; Rivail, J. L. Ab Initio Analysis of Water-Assisted Reaction Mechanisms in Amide Hydrolysis. *Journal of the American Chemical Society* **1994**, *116*, 3912–3921.

- [22] Antonczak, S.; Ruiz-López, M.; Rivail, J. L. The hydrolysis mechanism of formamide revisited: Comparison between ab initio, semiempirical and DFT results. *Journal of Molecular Modeling* **1997**, *3*, 434–442.
- [23] Antonczak, S.; Monard, G.; Ruiz-López, M. F.; Rivail, J. L. Modeling of peptide hydrolysis by thermolysin. A semiempirical and QM/MM study. *Journal of the American Chemical Society* **1998**, *120*, 8825–8833.
- [24] Antonczak, S.; Monard, G.; Ruiz-López, M. F.; Rivail, J.-L. Insights in the Peptide Hydrolysis Mechanism by Thermolysin: A Theoretical QM/MM study. *Journal of Molecular Modeling* **2004**, *6*, 527–538.
- [25] Chalmet, S.; Harb, W.; Ruiz-López, M. F. Computer simulation of amide bond formation in aqueous solution. *Journal of Physical Chemistry A* **2001**, *105*, 11574–11581.
- [26] Gorb, L.; Asensio, A.; Tuñón, I.; Ruiz-López, M. F. The mechanism of formamide hydrolysis in water from Ab initio calculations and simulations. *Chemistry - A European Journal* **2005**, *11*, 6743–6753.
- [27] Schreiner, E.; Nair, N. N.; Marx, D. Peptide synthesis in aqueous environments: The role of extreme conditions on peptide bond formation and peptide hydrolysis. *Journal of the American Chemical Society* **2009**, *131*, 13668–13675.
- [28] Maude, A. B.; Williams, A. Effective charge development in the transfer of the acetyl group between nucleophiles in acetonitrile solution: Acetolysis and butylaminolysis of substituted phenyl esters. *Journal of the Chemical Society. Perkin Transactions 2* **1997**, 179–183.
- [29] Beringer, M.; Bruell, C.; Xiong, L.; Pfister, P.; Bieling, P.; Katunin, V. I.; Mankin, A. S.; Böttger, E. C.; Rodnina, M. V. Essential mechanisms in the catalysis of peptide bond formation on the ribosome. *Journal of Biological Chemistry* **2005**, *280*, 36065–36072.
- [30] Abe, T.; Hashimoto, Y.; Zhuang, Y.; Ge, Y.; Kumano, T.; Kobayashi, M. Peptide bond synthesis by a mechanism involving an enzymatic reaction and a subsequent chemical reaction. *Journal of Biological Chemistry* **2016**, *291*, 1735–1750.
- [31] Świderek, K.; Marti, S.; Tuñón, I.; Moliner, V.; Bertran, J. Peptide Bond Formation Mechanism Catalyzed by Ribosome. *Journal of the American Chemical Society* **2015**, *137*, 12024–12034.
- [32] Findık, V.; Ruiz-López, M. F.; Erdem, S. S. Mechanistic insights into lysine-targeting covalent inhibition through a theoretical study of ester aminolysis. *Organic Biomolecular Chemistry* **2021**, *19*, 9996–10004.
- [33] Findık, V.; Varınca Gerçik, B. T.; Sinek, ; Erdem, S. S.; Ruiz-López, M. F. Mechanistic Investigation of Lysine-Targeted Covalent Inhibition of PI3K via ONIOM QM:QM Computations. *Journal of Chemical Information and Modeling* **2022**, PMID: 35980989.

- [34] Tajbakhsh, M.; Hosseinzadeh, R.; Tajbakhsh, M., et al. TiO₂ NPs as Catalyst for N-Formylation and N-Acetylation of Amines Under Solvent-Free Conditions. *Letters in Organic Chemistry* **2013**, *10*, 657–663.
- [35] Barravecchia, L.; Neira, I.; Pazos, E.; Peinador, C.; García, M. D. Amino Acid–Viologen Hybrids: Synthesis, Cucurbituril Host–Guest Chemistry, and Implementation on the Production of Peptides. *The Journal of Organic Chemistry* **2022**, *87*, 760–764.
- [36] Hattori, T.; Yamamoto, H. Synthesis of Silacyclic Dipeptides: Peptide Elongation at Both N- and C-Termini of Dipeptide. *Journal of the American Chemical Society* **2022**, *144*, 1758–1765.
- [37] Isidro-Llobet, A.; Álvarez, M.; Albericio, F. Amino Acid-Protecting Groups. *Chemical Reviews* **2009**, *109*, 2455–2504.
- [38] Farina, V.; Reeves, J. T.; Senanayake, C. H.; Song, J. J. Asymmetric Synthesis of Active Pharmaceutical Ingredients. *Chemical Reviews* **2006**, *106*, 2734–2793.
- [39] Zipse, H.; Wang, L.-H.; Houk, K. N. Polyether Catalysis of Ester Aminolysis – A Computational and Experimental Study. *Liebigs Annalen* **1996**, *1996*, 1511–1522.
- [40] Wang, L. H.; Zipse, H. Bifunctional catalysis of ester aminolysis - A computational and experimental study. *Liebigs Annales* **1996**, *1996*, 1501–1509.
- [41] Yi, G. Q.; Zeng, Y.; Xia, X. F.; Xue, Y.; Kim, C. K.; Yan, G. S. The substituent effects of the leaving groups on the aminolysis of phenyl acetates: DFT studies. *Chemical Physics* **2008**, *345*, 73–81.
- [42] Galabov, B.; Ilieva, S.; Hadjieva, B.; Atanasov, Y.; Schaefer, H. F. Predicting reactivities of organic molecules. Theoretical and experimental studies on the aminolysis of phenyl acetates. *Journal of Physical Chemistry A* **2008**, *112*, 6700–6707.
- [43] Rao, H. B.; Wang, Y. Y.; Zeng, X. Y.; Xue, Y.; Li, Z. R. Theoretical study on the aminolysis of p-substituted phenyl acetates with dimeric ammonia in vacuo and acetonitrile. *Computational and Theoretical Chemistry* **2013**, *1008*, 8–14.
- [44] Suárez, D.; Merz, K. M. Quantum chemical study of ester aminolysis catalyzed by a single adenine: A reference reaction for the ribosomal peptide synthesis. *Journal of the American Chemical Society* **2001**, *123*, 7687–7690.
- [45] Yang, W.; Drueckhammer, D. G. Computational studies of the aminolysis of oxoesters and thioesters in aqueous solution. *Organic Letters* **2000**, *2*, 4133–4136.
- [46] Singleton, D. A.; Merrigan, S. R. Resolution of conflicting mechanistic observations in ester aminolysis. A warning on the qualitative prediction of isotope effects for reactive intermediates [16]. *Journal of the American Chemical Society* **2000**, *122*, 11035–11036.
- [47] Galabov, B.; Atanasov, Y.; Ilieva, S.; Schaefer, H. F. Mechanism of the aminolysis of methyl benzoate: A computational study. *Journal of Physical Chemistry A* **2005**, *109*, 11470–11474.

- [48] Ilieva, S.; Atanasov, Y.; Kalcheva, V.; Galabov, B. Computational study of the general base catalysed aminolysis of 2-benzoxazolinone. *Journal of Molecular Structure: THEOCHEM* **2003**, *633*, 49–55.
- [49] Xia, X.; Zhang, C.; Xue, Y.; Kim, C. K.; Yan, G. DFT study and Monte Carlo simulation on the aminolysis of XC(O)OCH_3 ($\text{X} = \text{NH}_2$, H, and CF_3) with monomeric and dimeric ammonias. *Journal of Chemical Theory and Computation* **2008**, *4*, 1643–1653.
- [50] Sung, D. D.; Koo, I. S.; Yang, K.; Lee, I. Effects of atom pairs O and S on the stability of zwitterionic tetrahedral intermediate: A theoretical study. *Chemical Physics Letters* **2006**, *432*, 426–430.
- [51] Ilieva, S.; Atanasov, Y.; Galabov, B. Mechanism of the aminolysis of phenyl acetate: A computational study. *Bulgarian Chemical Communications* **2008**, *40*, 401–408.
- [52] Rajarathnam, D.; Jeyakumar, T.; Nadar, P. A. Structure-reactivity correlation in the aminolysis of 4-fluorophenyl acetate in aqueous medium. *International Journal of Chemical Kinetics* **2002**, *34*, 366–373.
- [53] Rajarathnam, D.; Jeyakumar, T.; Nadar, P. A. Mechanistic change in the reactivity of substituted phenyl acetates over phenyl thiolacetates toward imidazole in aqueous phase. *International Journal of Chemical Kinetics* **2005**, *37*, 211–221.
- [54] Blackburn, G. M.; Jencks, W. P. The Mechanism of the Aminolysis of Methyl Formate. *Journal of the American Chemical Society* **1968**, *90*, 2638–2645.
- [55] Arcelli, A.; Concilio, C. Kinetics and mechanism of aminolysis of phenyl acetates in aqueous solutions of poly(ethylenimine). *Journal of Organic Chemistry* **1996**, *61*, 1682–1688.
- [56] Castro, E. A.; Millan, D.; Aguayo, R.; Campodónico, P. R.; Santos, J. G. Reactions of aryl acetates with secondary alicyclic amines in ethanol/water mixtures: Effect of the solvent composition on the kinetics and mechanism. *International Journal of Chemical Kinetics* **2011**, *43*, 687–693.
- [57] Marlier, J. F.; Haptonstall, B. A.; Johnson, A. J.; Sacksteder, K. A. Heavy-atom isotope effects on the hydrazinolysis of methyl formate. *Journal of the American Chemical Society* **1997**, *119*, 8838–8842.
- [58] Christensen, D. G.; Meyer, J. G.; Baumgartner, J. T.; D'Souza, A. K.; Nelson, W. C.; Payne, S. H.; Kuhn, M. L.; Schilling, B.; Wolfe, A. J. Identification of Novel Protein Lysine Acetyltransferases in *Escherichia coli*. *mBio* **2018**, *9*, e01905–18.
- [59] Starheim, K. K.; Gevaert, K.; Arnesen, T. Protein N-terminal acetyltransferases: When the start matters. *Trends in Biochemical Sciences* **2012**, *37*, 152–161.
- [60] Wagner, G. R.; Payne, R. M. Widespread and enzyme-independent N ϵ -acetylation and N ϵ -succinylation of proteins in the chemical conditions of the mitochondrial matrix. *Journal of Biological Chemistry* **2013**, *288*, 29036–29045.

- [61] Tweedie-Cullen, R. Y.; Brunner, A. M.; Grossmann, J.; Mohanna, S.; Sichau, D.; Nanni, P.; Panse, C.; Mansuy, I. M. Identification of combinatorial patterns of post-translational modifications on individual histones in the mouse brain. *PloS one* **2012**, *7*, e36980.
- [62] Smith, B. C.; Denu, J. M. Chemical mechanisms of histone lysine and arginine modifications. **2009**, *1789*, 45–57.
- [63] He, M.; Han, Z.; Liu, L.; Zheng, Y. G. Chemical Biology Approaches for Investigating the Functions of Lysine Acetyltransferases. **2018**, *57*, 1162–1184.
- [64] Montinari, M. R.; Minelli, S.; De Caterina, R. The first 3500 years of aspirin history from its roots - A concise summary. *Vascular pharmacology* **2019**, *113*, 1–8.
- [65] Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. The resurgence of covalent drugs. *Nature Reviews Drug Discovery* **2011**, *10*, 307–317.
- [66] Baillie, T. A. Targeted Covalent Inhibitors for Drug Design. *Angewandte Chemie - International Edition* **2016**, *55*, 13408–13421.
- [67] Singh, J. The Ascension of Targeted Covalent Inhibitors. *Journal of Medicinal Chemistry* **2022**, *65*, 5886–5901.
- [68] Potashman, M. H.; Duggan, M. E. Covalent modifiers: An orthogonal approach to drug design. *Journal of Medicinal Chemistry* **2009**, *52*, 1231–1246.
- [69] Noe, M. C.; Gilbert, A. M. Targeted Covalent Enzyme Inhibitors. *Annual Reports in Medicinal Chemistry* **2012**, *47*, 413–439.
- [70] Lanning, B. R. et al. A road map to evaluate the proteome-wide selectivity of covalent kinase inhibitors. *Nature Chemical Biology* **2014**, *10*, 760–767.
- [71] Hallenbeck, K.; Turner, D.; Renslo, A.; Arkin, M. Targeting Non-Catalytic Cysteine Residues Through Structure-Guided Drug Discovery. *Current Topics in Medicinal Chemistry* **2016**, *17*, 4–15.
- [72] Fadeyi, O. O.; Hoth, L. R.; Choi, C.; Feng, X.; Gopalsamy, A.; Hett, E. C.; Kyne, R. E.; Robinson, R. P.; Jones, L. H. Covalent Enzyme Inhibition through Fluorosulfate Modification of a Noncatalytic Serine Residue. *ACS Chemical Biology* **2017**, *12*, 2015–2020.
- [73] Bum-Erdene, K.; Liu, D.; Gonzalez-Gutierrez, G.; Ghazayel, M. K.; Xu, D.; Meroueh, S. O. Small-molecule covalent bond formation at tyrosine creates a binding site and inhibits activation of Ral GTPases. *Proceedings of the National Academy of Sciences of the United States of America* **2020**, *117*, 7131–7139.
- [74] Grimster, N. P.; Connelly, S.; Baranczak, A.; Dong, J.; Krasnova, L. B.; Sharpless, K. B.; Powers, E. T.; Wilson, I. A.; Kelly, J. W. Aromatic sulfonyl fluorides covalently kinetically stabilize transthyretin to prevent amyloidogenesis while affording a fluorescent conjugate. *Journal of the American Chemical Society* **2013**, *135*, 5656–5668.

- [75] Hoppmann, C.; Wang, L. Proximity-enabled bioreactivity to generate covalent peptide inhibitors of p53-Mdm4. *Chemical Communications* **2016**, 52, 5140–5143.
- [76] Gushwa, N. N.; Kang, S.; Chen, J.; Taunton, J. Selective targeting of distinct active site nucleophiles by irreversible Src-family kinase inhibitors. *Journal of the American Chemical Society* **2012**, 134, 20214–20217.
- [77] Baranczak, A.; Liu, Y.; Connelly, S.; Du, W. G. H.; Greiner, E. R.; Genereux, J. C.; Wiseman, R. L.; Eisele, Y. S.; Bradbury, N. C.; Dong, J.; Noodleman, L.; Sharpless, K. B.; Wilson, I. A.; Encalada, S. E.; Kelly, J. W. A Fluorogenic Aryl Fluorosulfate for Intraorganellar Transthyretin Imaging in Living Cells and in *Caenorhabditis elegans*. *Journal of the American Chemical Society* **2015**, 137, 7404–7414.
- [78] Mortenson, D. E.; Brighty, G. J.; Plate, L.; Bare, G.; Chen, W.; Li, S.; Wang, H.; Cravatt, B. F.; Forli, S.; Powers, E. T.; Sharpless, K. B.; Wilson, I. A.; Kelly, J. W. "inverse Drug Discovery" Strategy to Identify Proteins That Are Targeted by Latent Electrophiles As Exemplified by Aryl Fluorosulfates. *Journal of the American Chemical Society* **2018**, 140, 200–210.
- [79] Anscombe, E. et al. Identification and Characterization of an Irreversible Inhibitor of CDK2. *Chemistry and Biology* **2015**, 22, 1159–1164.
- [80] Suh, E. H.; Liu, Y.; Connelly, S.; Genereux, J. C.; Wilson, I. A.; Kelly, J. W. Stilbene vinyl sulfonamides as fluorogenic sensors of and traceless covalent kinetic stabilizers of transthyretin that prevent amyloidogenesis. *Journal of the American Chemical Society* **2013**, 135, 17869–17880.
- [81] Manglik, A.; Kruse, A. C.; Kobilka, T. S.; Thian, F. S.; Mathiesen, J. M.; Sunahara, R. K.; Pardo, L.; Weis, W. I.; Kobilka, B. K.; Granier, S. Crystal structure of the μ -opioid receptor bound to a morphinan antagonist. *Nature* **2012**, 485, 321–326.
- [82] Pettinger, J.; Bihan, Y. L.; Widya, M.; van Montfort, R. L. M.; Jones, K.; Cheeseman, M. D. An Irreversible Inhibitor of HSP72 that Unexpectedly Targets Lysine-56. *Angewandte Chemie* **2017**, 129, 3590–3594.
- [83] Shannon, D. A.; Banerjee, R.; Webster, E. R.; Bak, D. W.; Wang, C.; Weerapana, E. Investigating the proteome reactivity and selectivity of aryl halides. *Journal of the American Chemical Society* **2014**, 136, 3330–3333.
- [84] Choi, S.; Connelly, S.; Reixach, N.; Wilson, I. A.; Kelly, J. W. Chemoselective small molecules that covalently modify one lysine in a non-enzyme protein in plasma. *Nature Chemical Biology* **2010**, 6, 133–139.
- [85] Tamura, T.; Hamachi, I. Chemistry for Covalent Modification of Endogenous/Native Proteins: From Test Tubes to Complex Biological Systems. *Journal of the American Chemical Society* **2019**, 141, 2782–2799.

- [86] Yasueda, Y.; Tamura, T.; Fujisawa, A.; Kuwata, K.; Tsukiji, S.; Kiyonaka, S.; Hamachi, I. A Set of Organelle-Localizable Reactive Molecules for Mitochondrial Chemical Proteomics in Living Cells and Brain Tissues. *Journal of the American Chemical Society* **2016**, *138*, 7592–7602.
- [87] Patricelli, M. P.; Szardenings, A. K.; Liyanage, M.; Nomanbhoy, T. K.; Wu, M.; Weisig, H.; Aban, A.; Chun, D.; Tanner, S.; Kozarich, J. W. Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry* **2007**, *46*, 350–358.
- [88] Ward, C. C.; Kleinman, J. I.; Nomura, D. K. NHS-Esters As Versatile Reactivity-Based Probes for Mapping Proteome-Wide Ligandable Hotspots. *ACS Chemical Biology* **2017**, *12*, 1478–1483.
- [89] Metcalf, B. et al. Discovery of GBT440, an Orally Bioavailable R-State Stabilizer of Sick Cell Hemoglobin. *ACS Medicinal Chemistry Letters* **2017**, *8*, 321–326.
- [90] Cross, B. C.; Bond, P. J.; Sadowski, P. G.; Jha, B. K.; Zak, J.; Goodman, J. M.; Silverman, R. H.; Neubert, T. A.; Baxendale, I. R.; Ron, D.; Harding, H. P. The molecular basis for selective inhibition of unconventional mRNA splicing by an IRE1-binding small molecule. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*.
- [91] Sanches, M. et al. Structure and mechanism of action of the hydroxy-aryl-aldehyde class of IRE1 endoribonuclease inhibitors. *Nature Communications* **2014**, *5*.
- [92] Cal, P. M.; Vicente, J. B.; Pires, E.; Coelho, A. V.; Veiros, L. F.; Cordeiro, C.; Gois, P. M. Iminoboronates: A new strategy for reversible protein modification. *Journal of the American Chemical Society* **2012**, *134*, 10299–10305.
- [93] Akcay, G.; Belmonte, M. A.; Aquila, B.; Chuaqui, C.; Hird, A. W.; Lamb, M. L.; Rawlins, P. B.; Su, N.; Tentarelli, S.; Grimster, N. P.; Su, Q. Inhibition of Mcl-1 through covalent modification of a noncatalytic lysine side chain. *Nature Chemical Biology* **2016**, *12*, 931–936.
- [94] Liu, Q.; Sabnis, Y.; Zhao, Z.; Zhang, T.; Buhrlage, S. J.; Jones, L. H.; Gray, N. S. Developing irreversible inhibitors of the protein kinase cysteinome. *Chemistry and Biology* **2013**, *20*, 146–159.
- [95] Isom, D. G.; Castañeda, C. A.; Cannon, B. R.; García-Moreno, B. E. Large shifts in pKa values of lysine residues buried inside a protein. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 5260–5265.
- [96] Fabbro, D.; Cowan-Jacob, S. W.; Moebitz, H. Ten things you should know about protein kinases: IUPHAR Review 14. **2015**, *172*, 2675–2700.
- [97] Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. **2002**, *298*, 1912–1934.

- [98] Katso, R.; Okkenhaug, K.; Ahmadi, K.; White, S.; Timms, J.; Waterfield, M. D. Cellular function of phosphoinositide 3-Kinase: Implications for development, immunity, homeostasis, and cancer. *Annual Review of Cell and Developmental Biology* **2001**, *17*, 615–675.
- [99] Martini, M.; De Santis, M. C.; Braccini, L.; Gulluni, F.; Hirsch, E. PI3K/AKT signaling pathway and cancer: An updated review. *Annals of Medicine* **2014**, *46*, 372–383.
- [100] Donahue, T. R.; Tran, L. M.; Hill, R.; Li, Y.; Kovochich, A.; Calvopina, J. H.; Patel, S. G.; Wu, N.; Hindoyan, A.; Farrell, J. J.; Li, X.; Dawson, D. W.; Wu, H. Integrative survival-based molecular profiling of human pancreatic cancer. *Clinical Cancer Research* **2012**, *18*, 1352–1363.
- [101] Courtney, K. D.; Corcoran, R. B.; Engelman, J. A. The PI3K pathway as drug target in human cancer. *Journal of Clinical Oncology* **2010**, *28*, 1075–1083.
- [102] Vanhaesebroeck, B.; Leever, S. J.; Panayotou, G.; Waterfield, M. D.; Waterfield, M. D. Phosphoinositide 3-kinases: A conserved family of signal transducers. *Trends in Biochemical Sciences* **1997**, *22*, 267–272.
- [103] O'Donnell, J. S.; Massi, D.; Teng, M. W.; Mandala, M. PI3K-AKT-mTOR inhibition in cancer immunotherapy, redux. *Seminars in Cancer Biology* **2018**, *48*, 91–103.
- [104] Miller, B. W. et al. FDA approval: Idelalisib monotherapy for the treatment of patients with follicular lymphoma and small lymphocytic lymphoma. *Clinical Cancer Research* **2015**, *21*, 1525–1529.
- [105] Hong, D. S.; Bowles, D. W.; Falchook, G. S.; Messersmith, W. A.; George, G. C.; O'Bryant, C. L.; Vo, A. C.; Klucher, K.; Herbst, R. S.; Eckhardt, S. G.; Peterson, S.; Hausman, D. F.; Kurzrock, R.; Jimeno, A. A multicenter phase I trial of PX-866, an oral irreversible phosphatidylinositol 3-kinase inhibitor, in patients with advanced solid tumors. *Clinical Cancer Research* **2012**, *18*, 4173–4182.
- [106] Levine, I. N. *Quantum Chemistry*, 4th ed.; Prentice Hall: Englewood Cliffs, New Jersey, 1991.
- [107] Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Physical Review* **1934**, *46*, 618–622.
- [108] Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 energy evaluation by direct methods. *Chemical Physics Letters* **1988**, *153*, 503–506.
- [109] Krishnan, R.; Pople, J. A. Approximate fourth-order perturbation theory of the electron correlation energy. *International Journal of Quantum Chemistry* **1978**, *14*, 91–100.
- [110] Pople, J. A.; Santry, D. P.; Segal, G. A. Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures. *Journal of Chemical Physics* **1965**, *43*, S129–S135.

- [111] Pople, J. A.; Beveridge, D. L.; Dobosh, P. A. Approximate Self-Consistent Molecular-Orbital Theory. V. Intermediate Neglect of Differential Overlap. *The Journal of Chemical Physics* **1967**, *47*, 2026–2033.
- [112] Pople, J. A.; Beveridge, D. L. *Approximate Molecular Orbital Theory*; McGraw-Hill: New York, 1970.
- [113] Pople, J. A.; Segal, G. A. Approximate Self-Consistent Molecular Orbital Theory. II. Calculations with Complete Neglect of Differential Overlap. *Journal of Chemical Physics* **1965**, *43*, S136–S151.
- [114] Bingham, R. C.; Dewar, M. J.; Lo, D. H. Additions and Corrections: Ground States of Molecules. XXV. MINDO/3. An Improved Version of the MINDO Semiempirical SCF-MO Method. **1975**, *97*, 4787.
- [115] Dewar, M. J.; Thiel, W. Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. *Journal of the American Chemical Society* **1977**, *99*, 4899–4907.
- [116] Dewar, M. J.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. AM1: A New General Purpose Quantum Mechanical Molecular Model. *Journal of the American Chemical Society* **1985**, *107*, 3902–3909.
- [117] Stewart, J. J. Optimization of parameters for semiempirical methods I. Method. *Journal of Computational Chemistry* **1989**, *10*, 209–220.
- [118] Stewart, J. J. MOPAC: A semiempirical molecular orbital program. *Journal of Computer-Aided Molecular Design* **1990**, *4*, 1–103.
- [119] Stewart, J. J. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling* **2007**, *13*, 1173–1213.
- [120] Engel, E.; Dreizler, R. M. *Density Functional Theory*; Theoretical and Mathematical Physics; Springer Berlin Heidelberg: Berlin, Heidelberg, 2011.
- [121] Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Physical Review* **1964**, *136*, B864.
- [122] Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical Review* **1965**, *140*, A1133.
- [123] Leach, A. R. *Molecular Modelling. Principles and Applications (2nd Edition)*; Prentice Hall, 2001; pp 262–269.
- [124] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Physical Review Letters* **1996**, *77*, 3865–3868.
- [125] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.

- [126] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* **1988**, 38, 3098–3100.
- [127] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **1988**, 37, 785–789.
- [128] 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, 194101.
- [129] Slater, J. C. Atomic Shielding Constants. *Physical Review* **1930**, 36, 57.
- [130] Boys, S. F. Electronic wave functions - I. A general method of calculation for the stationary states of any molecular system. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1950**, 200, 542–554.
- [131] Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, 90, 1007–1023.
- [132] Dill, K. A.; Truskett, T. M.; Vlachy, V.; Hribar-Lee, B. Modeling water, the hydrophobic effect, and ion solvation. *Annual Review of Biophysics and Biomolecular Structure* **2005**, 34, 173–199.
- [133] Rinaldi, D.; Rivail, J. L. Polarisabilites moléculaires et effet diélectrique de milieu à l'état liquide. Étude théorique de la molécule d'eau et de ses dimères. *Theoretica Chimica Acta* **1973**, 32, 57–70.
- [134] Rinaldi, D.; Ruiz-Lopez, M. F.; Rivail, J. L. Ab initio SCF calculations on electrostatically solvated molecules using a deformable three axes ellipsoidal cavity. *The Journal of Chemical Physics* **1983**, 78, 834–838.
- [135] Ruiz-López, M. F. The multipole moment expansion solvent continuum model: A brief review. *Challenges and Advances in Computational Chemistry and Physics* **2008**, 6, 23–38.
- [136] Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of AB initio molecular potentials for the prevision of solvent effects. *Chemical Physics* **1981**, 55, 117–129.
- [137] Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chemical Reviews* **1999**, 99, 2161–2200.
- [138] Mark, P.; Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *Journal of Physical Chemistry A* **2001**, 105, 9954–9960.
- [139] De Vivo, M. Bridging quantum mechanics and structure-based drug design. *Frontiers in Bioscience* **2011**, 16, 1619–1633.

- [140] Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713.
- [141] Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *Journal of Chemical Information and Modeling* **2012**, *52*, 3144–3154.
- [142] Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *Journal of the American Chemical Society* **1988**, *110*, 1657–1666.
- [143] Verlet, L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review* **1967**, *159*, 98–103.
- [144] Cerutti, D. S.; Case, D. A. Multi-level ewald: A hybrid multigrid/fast fourier transform approach to the electrostatic particle-mesh problem. *Journal of Chemical Theory and Computation* **2010**, *6*, 443–458.
- [145] Warshel, A.; Lappicirella, A. Calculations of Ground- and Excited-State Potential Surfaces for Conjugated Heteroatomic Molecules. *Journal of the American Chemical Society* **1981**, *103*, 4664–4673.
- [146] Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology* **1976**, *103*, 227–249.
- [147] Levitt, M.; Warshel, A. Computer simulation of protein folding. *Nature* **1975**, *253*, 694–698.
- [148] Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *Journal of Molecular Structure: THEOCHEM* **1999**, *461-462*, 1–21.
- [149] Vreven, T.; Morokuma, K. *Annual Reports in Computational Chemistry*; Elsevier, 2006; Vol. 2; pp 35–51.
- [150] Svensson, M.; Humbel, S.; Froese, R. D.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)₃)₂ + H₂ oxidative addition. *Journal of Physical Chemistry* **1996**, *100*, 19357–19363.
- [151] Senn, H. M.; Thiel, W. *Current Opinion in Chemical Biology*; Elsevier Current Trends, 2007; Vol. 11; pp 182–187.
- [152] Ananikov, V. P.; Musaev, D. G.; Morokuma, K. Real size of ligands, reactants and catalysts: Studies of structure, reactivity and selectivity by ONIOM and other hybrid computational approaches. **2010**, *324*, 104–119.

- [153] Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977**, *23*, 187–199.
- [154] Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **1992**, *13*, 1011–1021.
- [155] Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. Multidimensional free-energy calculations using the weighted histogram analysis method. *Journal of Computational Chemistry* **1995**, *16*, 1339–1350.
- [156] Martins-Costa, M. T.; Ruiz-López, M. F. Highly accurate computation of free energies in complex systems through horsetail QM/MM molecular dynamics combined with free-energy perturbation theory. *Theoretical Chemistry Accounts* **2017**, *136*, 1–7.
- [157] Chipot, C., A., Pohorille In *Free Energy Calculations*; Chipot, C., Pohorille, A., Eds.; Springer Series in CHEMICAL PHYSICS; Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; Vol. 86.
- [158] Retegan, M.; Martins-Costa, M.; Ruiz-López, M. F. Free energy calculations using dual-level Born-Oppenheimer molecular dynamics. *Journal of Chemical Physics* **2010**, *133*, 064103.
- [159] Montecinos, R.; Gazitúa, M.; Santos, J. G. The effect of the electrophilic group on the hierarchy of nucleofuges in the aminolysis reactions of thiol- and dithiocarbonates with secondary alicyclic amines: A kinetic and theoretical study. *New Journal of Chemistry* **2019**, *43*, 6372–6379.
- [160] Montecinos, R.; Aliaga, M. E.; Pavez, P.; Santos, J. G. Experimental and theoretical studies on the nucleofugality ratio in the aminolysis reactions of: O -(4-cyanophenyl) O -(3-nitrophenyl) thionocarbonate with amines in aqueous ethanol. *New Journal of Chemistry* **2017**, *41*, 9954–9962.
- [161] 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.; Son, D. J. Gaussian 09, Revision B.01. 2010.
- [162] Jin, L.; Wu, Y.; Xue, Y.; Guo, Y.; Xie, D. Q.; Yan, G. S. Theoretical studies on the aminolysis of phenyl formate. Mechanism and solvent effect. *Acta Chimica Sinica* **2006**, *64*, 873–878.
- [163] Fukui, K. The Path of Chemical Reactions - The IRC Approach. *Accounts of Chemical Research* **1981**, *14*, 363–368.
- [164] Derbel, N.; Clarot, I.; Mourer, M.; Regnouf-De-Vains, J. B.; Ruiz-López, M. F. Intramolecular interactions versus hydration effects on p-guanidinoethyl-phenol structure and pK_a values. *Journal of Physical Chemistry A* **2012**, *116*, 9404–9411.

- [165] Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual density functional theory. *Chemical Reviews* **2003**, *103*, 1793–1873.
- [166] Bruice, T. C.; Mayahi, M. F. The Influence of the Leaving Tendency of the Phenoxy Group on the Ammonolysis and Hydrolysis of Substituted Phenyl Acetates. *Journal of the American Chemical Society* **1960**, *82*, 3067–3071.
- [167] Satterthwait, A. C.; Jencks, W. P. The Mechanism of the Aminolysis of Acetate Esters. *Journal of the American Chemical Society* **1974**, *96*, 7018–7031.
- [168] Serjeant, E.; Dempsey, B. *International Union of Pure and Applied Chemistry (IUPAC)*; Pergamon Press: New York, 1979; Vol. 23; p 26.
- [169] Slater, B.; McCormack, A.; Avdeef, A.; Comer, J. E. Ph-metric log P. 4. Comparison of partition coefficients determined by HPLC and potentiometric methods to literature values. *Journal of Pharmaceutical Sciences* **1994**, *83*, 1280–1283.
- [170] D.A. Case, R.M. Betz, W. Botello-Smith, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Y.; York, P. K. M.; Kollman, P. A. Amber 16. 2016.
- [171] Zhu, J.; Ke, K.; Xu, L.; Jin, J. Discovery of a novel phosphoinositide 3-kinase gamma (PI3K γ) inhibitor against hematologic malignancies and theoretical studies on its PI3K γ -specific binding mechanisms. *RSC Advances* **2019**, *9*, 20207–20215.
- [172] Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; De Beer, T. A.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research* **2018**,
- [173] Wang, J.; 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.
- [174] Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.; Domański, J.; Dotson, D.; Buchoux, S.; Kenney, I.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference. 2016.
- [175] Olsson, M. H.; SØndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical p K a predictions. *Journal of Chemical Theory and Computation* **2011**, *7*, 525–537.
- [176] André, I.; Linse, S.; Mulder, F. A. Residue-specific pKa determination of lysine and arginine side chains by indirect ¹⁵N and ¹³C NMR spectroscopy: Application to apo calmodulin. *Journal of the American Chemical Society* **2007**, *129*, 15805–15813.

- [177] Chung, L. W.; Sameera, W. M.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H. B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chemical Reviews* **2015**, *115*, 5678–5796.
- [178] Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theoretical Chemistry Accounts* **2008**, *120*, 215–241.
- [179] Stewart, J. J. Application of the PM6 method to modeling proteins. *Journal of Molecular Modeling* **2009**, *15*, 765–805.
- [180] Akyüz, M. A.; Erdem, S. S. Computational modeling of the direct hydride transfer mechanism for the MAO catalyzed oxidation of phenethylamine and benzylamine: ONIOM (QM/QM) calculations. *Journal of Neural Transmission* **2013**, *120*, 937–945.
- [181] Cakir, K.; Erdem, S. S.; Atalay, V. E. ONIOM calculations on serotonin degradation by monoamine oxidase B: Insight into the oxidation mechanism and covalent reversible inhibition. *Organic and Biomolecular Chemistry* **2016**, *14*, 9239–9252.
- [182] Akyüz, M. A.; Erdem, S. S.; Edmondson, D. E. The aromatic cage in the active site of monoamine oxidase B: Effect on the structural and electronic properties of bound benzylamine and p-nitrobenzylamine. **2007**, *114*, 693–698.
- [183] Kellie, J. L.; Wilson, K. A.; Wetmore, S. D. Standard role for a conserved aspartate or more direct involvement in deglycosylation? an ONIOM and MD investigation of adenine-DNA glycosylase. *Biochemistry* **2013**, *52*, 8753–8765.
- [184] Deng, L.; Hu, C.; Qin, X.; Li, L.; Zhang, Y.; Li, P.; Chen, X. The remote arginine promoting the dehydrogenation of glucose in glucose oxidase via a proton-coupled double-electron transfer mechanism. *Journal of Catalysis* **2018**, *367*, 150–158.
- [185] Himo, F. Recent Trends in Quantum Chemical Modeling of Enzymatic Reactions. *Journal of the American Chemical Society* **2017**, *139*, 6780–6786.
- [186] Harris, T. K.; Turner, G. J. Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. *IUBMB Life* **2002**, *53*, 85–98.
- [187] Cuesta, A.; Taunton, J. Lysine-targeted inhibitors and chemoproteomic probes. *Annual Review of Biochemistry* **2019**, *88*, 365–381.
- [188] Carroll, T. X.; Thomas, T. D.; Bergersen, H.; Børve, K. J.; Sæthre, L. J. Fluorine as a π donor. Carbon 1s photoelectron spectroscopy and proton affinities of fluorobenzenes. *Journal of Organic Chemistry* **2006**, *71*, 1961–1968.
- [189] Rosenthal, J.; Schuster, D. I. The anomalous reactivity of fluorobenzene in electrophilic aromatic substitution and related phenomena. *Journal of Chemical Education* **2003**, *80*, 679–690.

- [190] Cortopassi, W. A.; Kumar, K.; Duarte, F.; Pimentel, A. S.; Paton, R. S. Mechanisms of histone lysine-modifying enzymes: A computational perspective on the role of the protein environment. *Journal of Molecular Graphics and Modelling* **2016**, *67*, 69–84.
- [191] Guo, H. B.; Guo, H. Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 8797–8802.
- [192] Karsten, W. E.; Liu, D.; Rao, G. S.; Harris, B. G.; Cook, P. F. A catalytic triad is responsible for acid-base chemistry in the *Ascaris suum* NAD-malic enzyme. *Biochemistry* **2005**, *44*, 3626–3635.
- [193] Saha, S. K.; Maniscalco, S. J.; Fisher, H. F. Mechanistic interpretation of tryptophan fluorescence quenching in the time courses of glutamate dehydrogenase catalyzed reactions. *Biochemistry* **1996**, *35*, 16483–16488.
- [194] Sass, H. J.; Büldt, G.; Gessenich, R.; Hehn, D.; Neff, D.; Schlesinger, R.; Berendzen, J.; Ormos, P. Structural alterations for proton translocation in the M state of wild-type bacteriorhodopsin. *Nature* **2000**, *406*, 649–653.
- [195] Rammelsberg, R.; Huhn, G.; Lübken, M.; Gerwert, K. Bacteriorhodopsin's intramolecular proton-release pathway consists of a hydrogen-bonded network. *Biochemistry* **1998**, *37*, 5001–5009.
- [196] Kipp, D. R.; Quinn, C. M.; Fortin, P. D. Enzyme-dependent lysine deprotonation in EZH2 catalysis. *Biochemistry* **2013**, *52*, 6866–6878.
- [197] Zhang, X.; Bruice, T. C. Product specificity and mechanism of protein lysine methyltransferases: Insights from the histone lysine methyltransferase SET8. *Biochemistry* **2008**, *47*, 6671–6677.
- [198] Hermann, J. C.; Hensen, C.; Ridder, L.; Mulholland, A. J.; Höltje, H. D. Mechanisms of antibiotic resistance: QM/MM modeling of the acylation reaction of a class A β -lactamase with benzylpenicillin. *Journal of the American Chemical Society* **2005**, *127*, 4454–4465.
- [199] Ribeiro, A. J.; Santos-Martins, D.; Russo, N.; Ramos, M. J.; Fernandes, P. A. Enzymatic Flexibility and Reaction Rate: A QM/MM Study of HIV-1 Protease. *ACS Catalysis* **2015**, *5*, 5617–5626.
- [200] Calixto, A. R.; Ramos, M. J.; Fernandes, P. A. Conformational diversity induces nanosecond-timescale chemical disorder in the HIV-1 protease reaction pathway. *Chemical Science* **2019**, *10*, 7212–7221.
- [201] Awoonor-Williams, E.; Walsh, A. G.; Rowley, C. N. Modeling covalent-modifier drugs. **2017**, *1865*, 1664–1675.

Appendix

Publication List

- (i) Theoretical Investigation of the Biogenetic Pathway for Formation of Antibacterial Indole Alkaloids from *Voacanga Africana*.
Soysal, E.N.; Fındık, V.; Dedeoglu, B.; Aviyente, V.; Tantillo, D.J., *ACS Omega*, 7, 35, 31591–31596, 2022.
- (ii) Mechanistic Investigation of Lysine-Targeted Covalent Inhibition of PI3K via ONIOM QM:QM Computations.
Fındık, V.; Varınca Gerçik, B.; Sinek, Ö.; Sağ Erdem, S.; Ruiz-Lopez, M.F., *Journal of Chemical Information and Modeling*, 2022. Doi: doi.org/10.1021/acs.jcim.2c00569
- (iii) Influence of ionic liquids on the electronic environment of atomically dispersed Ir on (MgO) (100).
Akgul, D., Öztulum-Kurtoğlu, S.F., Zhao, Y., Fındık, V., Monari, A., Uzun, A., Aviyente, V., *Physical Chemistry Chemical Physics*, 24, 11305-11314, 2022.
- (iv) Mechanistic insights into lysine-targeting covalent inhibition through a theoretical study of ester aminolysis.
Fındık, V., Ruiz-Lopez, M.F., Erdem, S.S., *Organic Biomolecular Chemistry*, 19, 9996-10004, 2021.
- (v) Origins of the photoinitiation capacity of aromatic thiols as photoinitiators: a computational study
Fındık, V., Fındık, B.K., Aviyente, V., Monari, A., *Physical Chemistry Chemical Physics*, 23, 24377-24385, 2021,
- (vi) Insight into the thiol-yne kinetics via a computational approach.
Fındık, V., Varınca, B.T., Degirmenci, I., Erdem, S.S., *Journal of Physical Chemistry A*, 125, 17, 3556-3568, 2021.
- (vii) Solvent effects on thiol-ene kinetics and reactivity of carbon and sulfur radicals.
Munar, İ., Fındık, V., Degirmenci, I., Aviyente, V., *Journal of Physical Chemistry A*, 124, 13, 2580-2590, 2020.
- (viii) Theoretical investigation of thiol-ene click reactions: A DFT perspective.
Fındık, V., Degirmenci, I., Çatak, S., Aviyente, V., *European Polymer Journal*, 110, 211-220, 2019.

Résumé

Les inhibiteurs covalents ciblés (TCI) sont très prometteurs pour la recherche de nouveaux médicaments. Ils offrent un certain nombre d'avantages par rapport aux inhibiteurs réversibles traditionnels, comme un temps de séjour prolongé, une puissance accrue et la possibilité d'apporter des modifications pour une conception efficace. Les inhibiteurs de kinases sont les exemples les plus courants d'ITC. Les enzymes phosphoinositide 3-kinase (PI3K) sont des cibles médicamenteuses importantes en oncologie car elles sont impliquées dans la voie de signalisation de nombreuses fonctions cellulaires telles que le contrôle de la croissance, le métabolisme et l'initiation de la traduction. Les résidus lysine (Lys) ont suscité un intérêt croissant comme alternative pour l'inhibition covalente ciblée. Récemment, les premiers inhibiteurs sélectifs et irréversibles avec des groupes esters comme tête électrophile ciblant le résidu Lys779 et inactivant de manière covalente l'enzyme PI3K δ ont été rapportés. L'objectif principal de cette thèse est d'élucider le mécanisme de l'inhibition covalente de PI3K δ par ces inhibiteurs ester afin d'aider à la conception future de nouveaux inhibiteurs avec des activités supérieures.

Avant les études mécanistiques sur l'enzyme, nous avons d'abord effectué des calculs *ab initio* et DFT sur la réaction modèle entre la méthylamine et les acétates de méthyle, phényle et *p*-NO₂ phényle en solution aqueuse. Les mêmes systèmes modèles ont ensuite été étudiés par l'approche de dynamique moléculaire QM/MM "à double niveau". Pour l'option "bas niveau", les simulations QM/MM ont été conduites au niveau PM3/TIP3P, et elles ont permis d'obtenir l'échantillonnage du système dans le cadre de la technique « umbrella sampling ». Les structures obtenues ont ensuite été utilisées pour obtenir des corrections perturbatives à l'énergie libre avec une région QM de "haut niveau" décrite par la méthode M06-2X/6-311+G(d,p). Les résultats montrent que la première étape implique la formation d'un intermédiaire tétraédrique zwitterionique. Ensuite, pour des esters suffisamment électrophiles, tels que le dérivé *p*-NO₂, la réaction se déroule par dissociation du zwitterion sous forme de paire d'ions, suivie d'un transfert de protons conduisant à la formation des produits attendus.

Nous avons utilisé des outils théoriques similaires pour étudier les mécanismes d'inhibition dans le cas de l'enzyme. Tout d'abord, un modèle de site actif de l'enzyme a été construit par des simulations classiques de dynamique moléculaire. Ensuite, l'approche ONIOM QM:QM au niveau M06-2X/6-31+G(d,p):PM6 a été appliquée pour obtenir les mécanismes de réaction envisageables dans ce site actif. Ces calculs nous ont permis d'affiner les mécanismes de réaction dans l'environnement enzymatique qui confirment globalement les étapes obtenues à partir du petit système modèle. Nous avons finalement utilisé ces informations pour aborder une étude QM/MM dynamique sur l'enzyme en utilisant le même protocole "double niveau" établi pour le petit système modèle, ce qui nous a permis d'obtenir le profil d'énergie libre du mécanisme d'inhibition de PI3K δ pour le dérivé *p*-NO₂ de l'inhibiteur ester. La barrière calculée est en bon accord avec les données cinétiques expérimentales disponibles, ce qui valide l'approche théorique proposée et les mécanismes obtenus.

Grâce à l'élucidation du mécanisme d'inhibition de composés précédemment testés expérimentalement, notre étude ouvre la voie à la découverte de nouveaux inhibiteurs à activité améliorée à l'aide des outils de la chimie théorique.

Mots-clés: Dynamique Moléculaire; ONIOM; QM/MM; Dual-Level; DFT; PI3K δ ; Inhibition Covalente; Simulation Atomistique, Lysine

Abstract

Targeted Covalent Inhibitors (TCIs) hold great promise for search of new drugs. They offer a number of potential advantages over traditional reversible inhibitors, such as extended residence time, increased potency, and the ability to make modifications for effective design. Kinase inhibitors are the most common examples of TCIs. Phosphoinositide 3-kinase (PI3K) enzymes are important drug targets in oncology as they are involved in the signaling pathway for many cellular functions such as growth control, metabolism and translation initiation. Lysine (Lys) residues have gained increasing interest as an alternative for targeted covalent inhibition. Recently, the first selective and irreversible inhibitors with ester groups as electrophilic head targeting the Lys779 residue and covalently inactivating the PI3K δ enzyme were reported. The main objective of this thesis is to elucidate the mechanism of the covalent inhibition of PI3K δ by these ester inhibitors in order to assist future design of new inhibitors with superior activities.

Prior to the mechanistic studies on the enzyme, initially, we performed ab initio and DFT calculations on the model reaction between methylamine and methyl, phenyl and *p*-NO₂ phenyl acetates in aqueous solution. The same model systems were then studied by the "dual-level" QM/MM molecular dynamics approach. For the "low-level" option, PM3/TIP3P umbrella sampling QM/MM simulations were applied for the sampling. The obtained structures were then used to obtain perturbative corrections to the free energy with a "high-level" QM region at the M06-2X/6-311+G(d,p) level. The results show that the first step involves the formation of the zwitterionic tetrahedral intermediate. Then, for sufficiently electrophilic esters, such as the *p*-NO₂ derivative, the reaction proceeds by dissociation of the zwitterion as an ion pair, followed by proton transfer leading to the formation of the expected products.

We, then, employed similar computational tools to shed light on the mechanistic aspects of the enzyme. First, an active site model of the enzyme was built through classical molecular dynamics simulations. Then, ONIOM QM:QM approach at the M06-2X/6-31+G(d,p):PM6 level was applied to get possible reaction mechanisms in this active site. These calculations guided us to refine the reaction mechanisms in enzyme environment which globally confirm the steps obtained from the small model system. We finally used this information to approach a dynamic QM/MM study on the enzyme using the same "dual-level" protocol established for the small model system, which allowed us to obtain the free energy profile of the inhibition mechanism of PI3K δ for *p*-NO₂ derivative of the ester inhibitor. The calculated barrier is in good agreement with the available experimental kinetic data, which validates the proposed theoretical approach and the obtained mechanisms.

Through the elucidation of the inhibition mechanism of previously experimentally tested compounds, our study paves the way for the discovery of new inhibitors with improved activity with the help of theoretical chemistry tools.

Keywords: Molecular Dynamics; ONIOM; QM/MM; Dual-Level; DFT; PI3K δ ; Covalent Inhibition, Atomistic Simulation, Lysine