

**Virtual Screening and End-point Binding Free
Energy Methods for Developing
Targeted Cancer Therapies**

**Applications on Acetyltransferases and
Protein Kinases as Case Studies**

Dissertation

Zur Erlangung des akademischen Grades

Doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der
Naturwissenschaftlichen Fakultät I
Biowissenschaften
der Martin-Luther-Universität Halle-Wittenberg

von

Herrn Diplom-Pharmazeut Suhaib Shekfeh
geb. am 15 August 1976 in Homs, Syrien

Institut für Pharmazie
Abteilung Pharmazeutische und Medizinische Chemie

Datum der Verteidigung: 26. Februar 2015

Gutachter/in:

1- Prof. Dr. Wolfgang Sippl, MLU Halle-Wittenberg (Germany)

2- Prof. Dr. Gerhard Wolber, Freie Universität Berlin (Germany)

3- Prof. Dr. Maciej Baginski, Gdansk University of Technology (Poland)

This page is intentionally left blank

“I stand at the seashore, alone, and start to think. There are the rushing waves ... mountains of molecules, each stupidly minding its own business ... trillions apart ... yet forming white surf in unison. Ages on ages ... before any eyes could see ... year after year ... thunderously pounding the shore as now. For whom, for what ? ... on a dead planet, with no life to entertain. Never at rest ... tortured by energy ... wasted prodigiously by the sun ... poured into space. A mite makes the sea roar. Deep in the sea, all molecules repeat the patterns of one another till complex new ones are formed. They make others like themselves ... and a new dance starts. Growing in size and complexity ... living things, masses of atoms, DNA, protein ... dancing a pattern ever more intricate. Out of the cradle onto the dry land ... here it is standing ... atoms with consciousness ... matter with curiosity. Stands at the sea ... wonders at wondering ... I ... a universe of atoms ... an atom in the universe. ”

Richard P. Feynman

This page is intentionally left blank

Dedication

This work is dedicated to two special persons from my family.

I will remain forever in their debts.

To my grandfather: *Mohammad Massarani*;
the man who raised me up and was a great example of hard-working, frankness,
open-mindedness, honesty, and sincerity.

and also to my uncle: *Dr. Bassam Massarani*, former professor of solid-state
and quantum physics (University of Damascus);
the man who taught me the love of science, guided me to read great scientific books,
and was many times a source of inspiration.

Both of them left us suffering from pancreatic cancer and GIST.

This page is intentionally left blank

Preface

Approximately twelve years passed since I had the first look at a colorful representation of a pharmacophore model, suggesting a hypothesis to explain the varying binding affinities for a series of enzyme inhibitors. It was 2002, when my interest in the field of drug design started, leading me to a long journey for learning the modern concepts of medicinal chemistry; including Qsar, 3D-Qsar, and pharmacophores. After few years, I had the opportunity to continue working in Germany with more advanced and computationally demanding methods like docking and molecular dynamics. The breakthrough in human knowledge in the biological “Omics” (Genomics and Proteomics) and System biology, combined with the development of computational technologies, gave us a unique opportunity to have a closer look to more realistic model of biological processes and biological phenomenon, including the interactions between the developed small-molecule drugs and their biological partners. These methods will expand our way to understand medicinal chemistry and structure-activity relationship (SAR) studies beyond the traditional methodologies.

Drug discovery process was carried to its best by combining medicinal chemistry, biochemistry, structural and molecular biology, and finally computational physical chemistry with statistical mechanics principles. We have accomplished a great success in fighting major diseases; like cancers and HIV. Despite our expanding knowledge and our triumph over some hard challenges, we still suffer from considerable amount of uncertainties and failures in the process of designing and discovering genuinely novel drugs. Resistance to some therapies can develop in some diseases, and cases of toxicity and side-effects can emerge.

According to some statistics from WHO in 2008, almost 16 million died from communicable, maternal, parental and nutritional conditions, and 36 million from non-communicable diseases. Developing a drug for any disorder or disease can take up to 10 or 20 years with cost of billions of dollars. The process can be accelerated through the developed computational physics-based simulation methods. These methods can help us saving time and money to capture some potential hits, which could be developed to potent inhibitors and drugs, and also for estimating one of the most important physical chemical quantities; ‘Binding Free Energy’. These developed computational simulation methods were able to extend our physical understanding of the universe to the level of biological cellular machineries. The theories and principles of quantum mechanics and statistical mechanics can be applied now to give us an atomistic model of biological processes and the molecular recognition between biological molecules.

The theoretical basics of these methods were established few decades ago by Kirkwood (Applying statistical mechanics on fluid mixtures) and Zwanzig (High-temperature equation of state by a perturbation method. I. Nonpolar gases, *J. Chem. Phys.*, 1954). The force field concept was developed by ideas from Andrews (*Phys. Rev.* 1930, 36, 544.), Hill (*J. Chem. Phys.* 1946, 14, 465), and Westheimer (*J. Chem. Phys.* 1946, 14, 733.). The applications in

the early time were concentrated on simulating small molecules. Further special force fields were later developed by many researchers; including P. Kollman and W. Jorgensen, to simulate the biological macromolecules. The aim of developing the force fields was to combine both experimental work and high-level quantum mechanics calculations in order to provide fast and reliable physical description of big systems comprising thousands of atoms. However, only the increasing power of computational technologies has made applying the atomistic modeling to biomolecular systems feasible. The first molecular dynamics of protein was made in 1977 by McCammon, Gelin, and Karplus. Huge advances in the theory and the development of the right potentials have been made in the nineties of the last century. In 2000, P. Kollman and others started to perform free energy calculations applied on biological macromolecular systems. By combining these methods with our accumulated knowledge about structures of proteins and enzymes and our improved understanding of signaling transduction pathway, we are able to go deeper in understanding the pathological pathways of many challenging diseases.

In spite of all these advances, these developed methods shouldn't be used without careful consideration of the underlying assumptions and simplifications that were used during the development of the used method, force field, or solvent model. This work presents different examples and study cases that show the strength and pitfalls of some computational methods. The presented study cases suggest a combination of these methods to cover their limitation. In the same time, I hope this work will shed a light on some structural and physical factors of designing novel drugs or enzyme inhibitors for developing targeted therapies. The study of enzyme/inhibitor systems using solvent models and physics-based computational methods can provide sometimes an insightful picture of how molecules interact in nature with each others, giving an explanation for challenging observations in SAR studies.

The experimental work in the present dissertation was carried out at the institute of pharmacy (medicinal chemistry department) at Martin-Luther-University Halle-Wittenberg from the middle of 2009 till the end of 2013. PCAF project was a continuation of previous work I started in 2008. The work on kinases started in 2010 till the end of 2012 for the c-Kit project, while most of the work on GSK3/CDK2 kinase inhibitors was done in 2013. The hardest part was performing most of this work, while I was worried about my family living under terrible conditions and civil war in Syria. However, here we are. A lot has happened in those years, and this thesis is sort of a summary of those things: work, struggles, and also the stressful times trying to figure out solutions for challenges.

I hope that this work and these findings will be regarded as a small contribution to the human knowledge.

Suhaib Shekfeh
Halle (Saale), October 30th, 2014

Acknowledgment

I would like to start by thanking my supervisor Prof. Dr. Wolfgang Sippl for giving me the opportunity to work in his group, giving me interesting research topics, and guiding me in the writing process of this dissertation. I would like also to thank our collaborators in university of Freiburg: Prof. Dr. Manfred Jung and his former coworker; Dr. Silviya Furdas, for the successful collaboration in the PCAF project. I owe an acknowledgment to the 'Institute für Dermatopharmazie' for the scholarship during two years of my PhD work, and also to the international office of MLU Halle for different fellowships during my study in Halle.

I feel fortunate to have completed my PhD dissertation in a friendly environment with nice colleagues and friends. Thanks to all the current and former members of the medicinal chemistry group in the institute of pharmacy (Martin-Luther-University Halle-Wittenberg): Tana Uengwetwanit, Inna Slynko, Jelena Melesina, Dat Nguyen, Berin Karaman, and others (Michael Lawson, Ntie Kang Fiedele, Dhilon Patel, Nand Kishor Kuamat). I am also grateful to Dr. Dina Robaa for her help in the proofreading of this dissertation.

I cannot forget the warm welcome from my dear friend Mark Lindner, who was the first colleague I met and the first friend in Germany. I am especially thankful for Urszula Uciechowska for the inspiring ideas and scientific discussions in the beginning of my work. I am grateful for all the support from former members who patiently helped me at the beginning; Rene Meier, Ralf Heinke, German Erenkamp, Kanin Wichapong, and Ragav Kannan.

I feel especially lucky to have worked in the office with Michael Scharfe, Martin Pippl, and Luca Carlino. Luca and Martin are not only great office mates, but also great friends and fun-buddies. We had together memorable moments.

Halle is small city, which can offer a great chance for making international network and meeting people from all around the world. Thanks to the International office of MLU and the amazing efforts of Regine Brandt. I am grateful for having this chance to make a lot of great friends during six years living in Halle.

I really cannot mention the names of all great friends from the scientific community in Halle (Saale); the international group in Leibniz institute of agricultural development (IAMO), Leibniz institute of plant biochemistry (IPB), Helmholtz center of environmental studies (UFZ), the zoologie department of MLU Halle (especially: Jonathan Kidner, Bertrand Fouks, and

ACKNOWLEDGMENT

Alexi Beaurepaire), the physicists and anthropologists in both Max-Planck institutes (especially: Pratyush Das Kanungo (Aka. PDK), Nitin Shingen, Thiago Peixoto, and Mariuz Pazgan).

At times, a PhD can be quite stressful and you need close friends, with whom you can make deep discussions about life, science, .. etc. You need also fun buddies for funny events and amazing times. I feel fortunate that I got chance to meet some great guys in Halle: Joao Afonso Babtista, Denis Montagner, and Christoph Fretter.

Special thanks should indeed go to dear friends who supported me and were like a family for me. I am thankful for the memorable events I had with those great people: Diana Traikova, Ivan Djuric, Alex Popov, Katharina Karsten (Aka. Frau Paprika), Siegfried Vantomme (Aka. Herr Paprika), Markus Franzen, Vasył Kvartiuk, Michał Ochalek, Aga Sendek, Ally Siebenkass, Ilkay Unay, Nicolai Gailhard, Lena Kuhn, Maria Belayva, Domenico Caruso, George Chezia, Kachaber Lominadzi, Mohammad Esmael, Nizami and Guilmira Imamverdiyev, Klodjan Rama, Florian Schierhorn, Brett Hankerson, Denitsa Angelova, and also for my italian-german band: Viola Bruschi, Giulia Furlan, Jörn Klinkenburg, and Frederik Faden. I am thankful for the great hallenesers and german friends: Elke Dobbertin, Uli Heigel, Dajana Tiele, Leandro Gamboa and Sandor Szimeiszter.

Scientific research can be frustrating and I have sometimes wondered if I would be able to pass some stressful times. That would be impossible without some beautiful souls in my life; A thankful hug to Ewa, Lina, Natalia, Lidia, Carla, Kriste, Eleana, Pavlina, and Burcu.

Finally, I would like to express my immeasurable gratefulness to my unconditionally devoted family for their love, trust, and understanding; to my uncle Malaz Massarani, to my beloved sister, and to my angels; my three nieces, whom I incredibly miss, and finally to my devoted mother and grandmother. This thesis is dedicated to them.

Thanks to all the people I acknowledged above,
and to many others who I have not mentioned but they know what they meant to me . . .
Thanks . . .

List of Publications:

- I.** B. Maurer, U. Mathias, S. Shekfeh, P. Papatheodorou, J. Orth, T. Jank, C. Schwan, W. Sippl, K. Aktories, M. Jung. From cosubstrate similarity to inhibitor diversity-inhibitors of ADP-ribosyltransferases from kinase inhibitor screening. *Mol. BioSyst.*, 2011, 7, 799-808.
- II.** S. D. Furdas, S. Shekfeh, E. Bissinger, J. Wagner, C. H. Arrowsmith, M. M. Mangos, V. Valkov, M. Hendzel, M. Jung, W. Sippl. Synthesis and biological testing of novel pyridoisothiazolones as histone acetyltransferase inhibitors. *Bioorganic & Medicinal Chemistry* 19(12), 2011, 3678- 3689.
- III.** S. D. Furdas, S. Shekfeh, S. Kannan, W. Sippl, M. Jung, Rhodanine carboxylic acids as novel inhibitors of Histone acetyltransferases. *Med. Chem. Commun.* 2012, 3, 305-311.

Chapter 2 of this dissertation depends mainly on articles II and III.

Posters:

1. Shekfeh, S.; Hilgeroth, A. and Sippl, W., Computational analysis of conformational changes in GSK3- β /CDK2 kinases for understanding inhibitor selectivity. 3rd International Meeting - Conformational transitions in macromolecular interactions. 4th - 6th November 2013, Halle (Saale), Germany.
2. S. Shekfeh, S.D. Furdas, M. Jung, W. Sippl. Docking study and binary classification model of isothiazolones as irreversible inhibitors of the histone acetyltransferase PCAF. in : (International Workshop – New Approaches in Drug Design & Discovery, Rauschholzhausen, 22.-25.03.2010, Jahrestagung der Deutschen Pharmazeutischen Gesellschaft, Jena, 29.09 – 01.10.2009, 5th Summer School Medicinal Chemistry., 13.-15. 09.2010, Regensburg.)
3. S. D. Furdas, J. M. Wagner, S. Shekfeh, P. Brown, W. Sippl, M. Jung. Histone acetyltransferases inhibition by pyridoisothiazolones. EMBO Conference Series, Chemical Biology 2010 22.-25.09.2010, EMBL Heidelberg.
4. B. Maurer, P. Papatheodorou, C. Schwan, S. Shekfeh, W. Sippl, K. Aktories, M. Jung. Development of an enzymatic assay for ADP-ribosylating enzymes via the chemical quantitation of the co-substrate NAD. 5th Summer School Medicinal Chemistry. 13.-15. 09.2010, Regensburg
5. S. D. Furdas, J. M. Wagner, S. Shekfeh, P. Brown, W. Sippl, M. Jung Pyridoisothiazolones as a novel class of inhibitors of histone acetyltransferase activity: synthesis and biological testing. XXXth European School of Medicinal Chemistry (ESMEC) 04.-09. 07. 2010, Urbino, Italien, and Jahrestagung der Deutschen Pharmazeutischen Gesellschaft, Jena, 29.09 – 01.10.2009.

Contents

1	Introduction: Targeted Cancer Therapy	1
1.1	Mechanisms of Cancer	1
1.2	Targeted Therapy: Definition and Concept	1
1.3	Selective Targeting of Protein Kinases	2
1.4	Epigenetic Deregulation in Cancer	4
1.5	Aim of the work	6
I	Virtual Screening Methods for Novel Epigenetic Modulators	7
2	Novel Inhibitors of Histone Acetyltransferases: Application of Focused Virtual Screening Methods	8
2.1	Histone Acetyltransferases: Biological Role and Classification	8
2.1.1	Structural Overview of Histone Acetyltransferase PCAF	10
2.2	Computational Methods	12
2.2.1	Molecular Docking	12
2.2.2	Similarity Search and Focused Library Design	13
2.2.3	Database Virtual Screening	14
2.2.4	Molecular Dynamics Simulation	15
2.3	Results and Discussion	15
2.3.1	Identification of Isothiazolone-based HAT Inhibitors as Covalent Inhibitors	15
2.3.2	Identification of Rhodanine derivatives as Non-covalent HAT Inhibitors	20
2.3.3	SAR Study and Molecular Interaction Fields	23
2.4	Conclusion	26
II	End-point Free Energy Methods for Selective Kinase Inhibitors	27
3	Structural Aspects of Protein Kinases: The Impact on Designing Selective Kinase Inhibitors	28
3.1	Kinases' Binding Pockets and Catalytic Cleft	28
3.2	Types and Selectivity of Protein Kinase Inhibitors	31

3.3	Binding Mode of Kinase Inhibitors and Overcoming the Mutation-induced Resistance	34
3.4	The Important Role of Water Molecules inside The ATP-binding Pocket	38
3.5	Computational Methods and Structure-based Design of Kinase Inhibitors	40
4	Predicting The Binding Mode for Kinase Inhibitors:	
	1-Aza-9-Oxa-Fluorene Derivatives as GSK3β/CDK2 Inhibitors	41
4.1	Introduction	41
4.1.1	Glycogen Synthase Kinase 3 and CDK2 Kinase	41
4.1.2	GSK3 β /CDK2 Kinase Inhibitors	42
4.1.3	The 1-aza-9-oxafluorene Derivatives	45
4.1.4	Aim of this Study	48
4.2	Computational Methods	48
4.2.1	X-Ray Structures of GSK3 β and CDK2	48
4.2.2	Molecular Docking	49
4.2.2.1	Scoring Functions	49
4.2.3	Molecular Dynamics Simulation and Binding Energy Methods	50
4.2.3.1	Molecular Dynamics Simulation	50
4.2.3.2	MM-PBSA Calculations	51
4.2.3.3	Linear Interaction Energy (LIE)	52
4.2.4	Water Hydration Maps	53
4.3	Results	54
4.3.1	Suggested Binding Modes of 1-aza-9-oxa-fluorenes with GSK3 β	54
4.3.2	Molecular Dynamics Simulations and Trajectory Analysis	57
4.3.3	MM-PBSA Calculations	57
4.3.4	LIE Models for GSK3 β	58
4.3.4.1	LIE models for GSK3 β (Binding mode 1)	58
4.3.4.2	LIE models for GSK3 β (Binding mode 2)	59
4.3.5	LIE Models for CDK2	62
4.3.5.1	LIE models for CDK2 (Binding mode 1)	62
4.3.5.2	LIE models for CDK2 (Binding mode 2)	62
4.3.6	Selection of Most Probable Binding Mode	64
4.4	Discussion	66
4.4.1	Impact of Kinase Conformational Changes	66
4.4.2	Role of Water Molecules and Protein Hydration in the Binding Process	67
4.4.3	Understanding the SAR of 1-aza-9-oxa-fluorene Derivatives	68
4.4.3.1	Derivatives 2c and 2c_2	68
4.4.3.2	Derivatives 2b and 2b_2	69
4.5	Conclusions	72

5 Prediction of Binding Affinities for Kinase Inhibitors: Application on Mutant c-Kit D816V Kinase Inhibitors	73
5.1 Introduction	73
5.1.1 SCF/c-Kit Signaling Pathway: Role and Biological Importance	73
5.1.2 Gain-of-Function Mutations and the Cancer Pathology	74
5.1.3 C-Kit Kinase Inhibitors: Activities with Mutant Forms	76
5.1.4 Aim of this Study	78
5.2 Computational Methods and Application	80
5.2.1 Molecular modeling, Docking and Enrichment Studies	80
5.2.2 Scoring Methods	81
5.2.2.1 X-score	81
5.2.2.2 AMBER GBSA/PBSA scoring after refinement	81
5.2.3 Molecular Dynamics Simulation	82
5.2.4 Free Binding Energies using Linear Interaction Energy and MM-PBSA methods	82
5.2.5 General Structure of the Studied c-Kit inhibitors	83
5.3 Results	84
5.3.1 Docking Studies and Enrichment Assessment	84
5.3.2 Application of LR-MM-PBSA Method	87
5.3.3 Performance of Binding Energy Methods with Anilino-oxazole Derivatives	87
5.3.4 Analysis of the Binding Pocket's Flexibility and the P-loop Fluctuations	88
5.3.5 P-loop/A-loop-distance-dependent LIE Models	89
5.3.6 Extended Validation Using Larger Dataset	92
5.4 Discussion	96
5.4.1 Reliability of Suggested Binding Mode	96
5.4.2 Importance of P-loop Conformation and Fluctuations	96
5.4.3 Role of Water Molecules in the Binding Process	97
5.4.4 Effect of the Inhibitor's Substituents on the Final Stable Conformation	98
5.5 Conclusion	99
6 Summary	101
III Supplementary Materials and References	103
A End-point Free Energy Methods	104
A.1 Implicit-solvent-based Binding Energy methods	104
A.1.1 Pitfalls and critical parameters of PB(GB)SA methods	106
A.2 Linear Interaction Energy (LIE) Method	107
A.2.1 LIE parameters	108
A.2.2 Critical issues in applying LIE	109
A.3 Hybrid Physics-based Methods	110

B	Novel Inhibitors of Histone Acetyltransferases	112
C	Predicting The Binding Mode For Kinase Inhibitors	113
C.1	GSK3 β X-ray Structures	114
C.2	CDK2 X-ray Structure (2WIH)	115
C.3	Docking Scores	116
C.4	MM-PBSA Calculations	119
C.4.1	MM-PBSA Calculations with GSK3 β	119
C.4.2	MM-PBSA Calculations with CDK2	122
C.5	LIE models of GSK3 β binding energy (Binding mode 1)	125
C.6	LIE models of CDK2 binding energy (Binding mode 1)	127
C.7	Prediction of Inactive Compounds	129
D	Prediction of Binding Affinities (c-Kit D816V)	131
D.1	LIE Models for c-Kit D816V Inhibitors	131
D.1.1	LIE model 1-1	132
D.1.2	LIE model 2-1	134
D.1.3	LIE model 3-1	136
D.1.4	LIE model 4-1	138
D.2	PBSA Calculations and LR-MM-PBSA Models	140
D.2.1	LR-MM-PBSA model 1-1	140
D.2.2	LR-MM-PBSA model 2-1	141
D.2.3	LR-MM-PBSA model 3-1	142
D.2.4	LR-MM-PBSA model 4-1	143
	Bibliography	144

This page is intentionally left blank

Chapter 1

Introduction: Targeted Cancer Therapy

1.1 Mechanisms of Cancer

Cancer or malignant neoplasm is a diverse and broad collection of diseases characterized by uncontrolled growth and dividing of some type of cells, leading to a formation of malignant tumors, which invade later other parts of the body. In spite of the complexity and the diversity of cancers' types, causes, and symptoms including the local and systematic symptoms, cancer could be described as defect in regulatory circuits which control the cellular proliferation and homeostasis. Cancer research confirmed that tumorigenesis is a multi-step process driven by **genetic mutations and epigenetic alterations**, which cause the transformation of normal cells to malignant ones with malfunctions in the **regulatory signaling networks** [1, 2]. An overview of major signal transduction pathways reprogrammed in cancer cells is shown in Figure 1.1. Considering the current understanding of cancers, protein kinases have emerged as one of the key regulators of critical signaling pathways, which are possibly involved in the malignant tumor formation [3].

1.2 Targeted Therapy: Definition and Concept

Targeted therapy is the interfering with specific molecular targets needed in special type of cells for carcinogenesis, as these targets play an important regulatory role in cells by being part of a signaling pathway used by cancer cells to grow, divide, or spread throughout the body [4]. The specific and selective interfering/inhibition is usually able to block the growth of cancer cells without affecting other dividing cells (opposite to the traditional cancer chemotherapy). Thus, this selective targeting of tumors is expected to be less harmful for normal tissues. The ideal target to attack for a cancer therapy should exist in cancer tissues more frequently than in normal ones. By adjusting the dose of a selective drug, it might be possible to kill the tumors

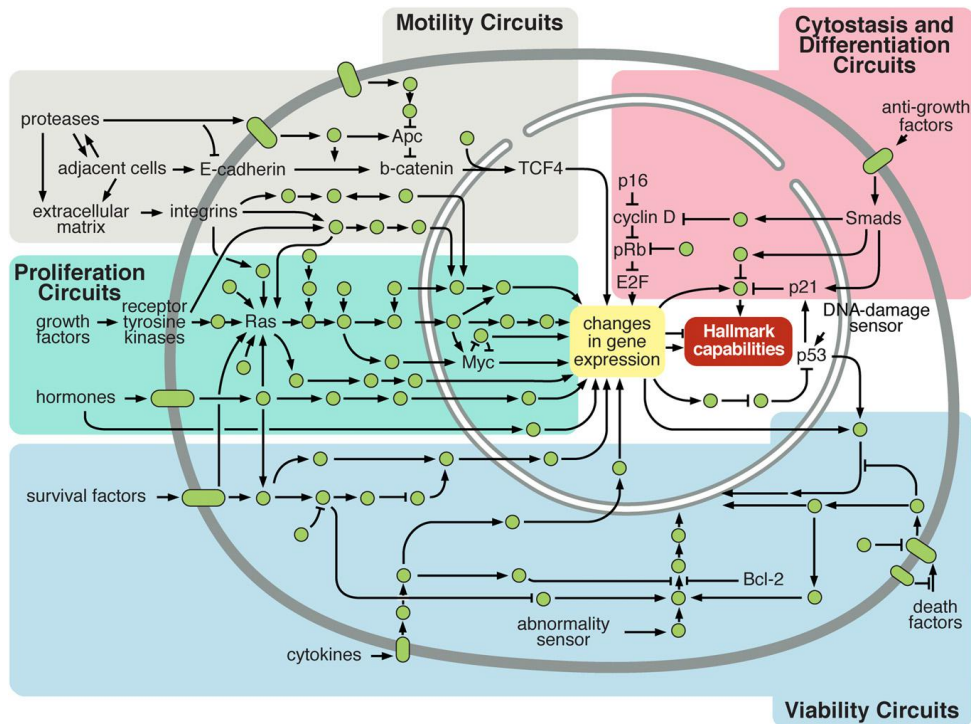


Figure 1.1: An overview of major signal transduction pathways within normal cells, mainly reprogrammed to regulate hallmark capabilities within cancer cells (Adapted from Hanahan and Weinberg [2]).

while normal tissues are unaffected [4, 5, 6, 7].

1.3 Selective Targeting of Protein Kinases

Protein kinases (PK) are essential enzymes for propagation of signal transduction inside the organisms' cells, participating in the regulation of many biological processes, including metabolic processes, transcription, cell cycle progression, apoptosis, and cell differentiation [3, 8]. The propagation of signal transduction cascades is mediated in many biological processes by phosphorylation of amino acid residues of substrate proteins, catalyzed by protein kinases. For this reason, it is not surprising that protein kinases constitute one of the largest enzyme families in the human organism, accounting for ~ 1.7% of the human genome-encoded proteins [8]. By discovering the main role of protein kinases in regulating biological processes and their involvement in many pathological pathways, protein kinases became one of the most important pharmaceutical targets for developing small-molecule inhibitors as new drugs for various severe diseases: cancer, diabetes, inflammation, cardiovascular disorders, and infectious diseases [3].

The analysis of the human genome revealed 518 protein kinase genes, which could be

divided into 478 classical PKs and 40 atypical protein kinases. The collection of 478 protein kinases consists of 388 serine/threonine kinases, 90 tyrosine kinases, and 50 sequences which lack functional catalytic sites. The human protein kinase family is classified into nine groups: Tyrosine kinases group (TK), Tyrosine-kinase like group (TKL), STE kinase group, AGC group, CMGC group, CAMK group, CK1 group, other kinases group, and RGC kinases [8]. In humans, 90 distinct kinases have been classified in the tyrosine kinases group, which can be broadly divided into 58 receptor tyrosine kinases (RTKs) and 32 non-receptor tyrosine kinases (NRTKs) [8]. The structural conservation of the ATP-binding pocket, especially in protein kinases from the same group, makes the development of selective kinase inhibitors a critical issue for developing effective and safe therapies.

The problem of designing specific inhibitors with required selective binding to a given protein target is more complex than just improving the affinity to a single target. In most of the cases, the selectivity problem comes out from an equal affinity of the ligand to homologous proteins, which share a conserved binding pocket. In that case, the binding sites usually exhibit high similarity regarding the shape and the sequence identity, resulting in similar protein/ligand interactions. Physical factors related to the similar ligands; such as the ligand entropy and other ligand-only thermodynamic terms, could be very similar, and thus are playing a small role in explaining selectivity. Other factors related to the protein structure could be more important playing a bigger role in determining the changes of binding affinities and selectivity. These protein-related factors could be small differences in the protein sequence, small differences in the molecular surface shape of the binding pocket, differences in the electrostatic properties generated by some residue changes, receptor's desolvation (hydration sites inside the binding site), or the receptor flexibility; including receptor's reorganization energy (induced fit effects and strain energies) [9]. An important example is the ATP-competitive kinase inhibitors that bind to the conserved and highly similar ATP-binding pocket. It is challenging to develop selective kinase inhibitors, which hit only one kinase without affecting similar kinases from the same group. However, in some cases, it is desirable to get inhibition of multiple kinases, which participate synergistically in one signaling pathway connected to a disease or disorder. Recent pharmaceutical research also shows the importance of estimating the interactions with off-targets like ion channels (including the Kv11.1 potassium ion channel hERG), cytochrome P450s (CYPs), and other proteins that can lead to adverse side effects [9].

1.4 Epigenetic Deregulation in Cancer

Epigenetic mechanisms of regulating gene expression patterns inside the cells have been elucidated during the last decade. Beside the accumulation of genetic mutations and the dysregulation of the cell cycle in cancer cells, epigenetic factors were found to be important in carcinogenesis [10, 11]. Epigenetic mechanisms depend on covalent modifications of the DNA (mainly Cyt-methylation) and the histone proteins for the regulation of gene expression [10, 12, 13]. Histones are the basic core of the nucleosome; which contains the DNA stretches wrapped around the octamer of histone proteins. Post-translational modifications of the histone proteins are covalent addition or removal of specific chemical groups: acetyl groups (acetylation/deacetylation), methyl groups (methylation/demethylation), phosphate groups (phosphorylation/dephosphorylation), or ubiquitin molecules (ubiquitination/deubiquitination). Biochemical researches have reported that histone methylation and acetylation are quite often dysregulated in many types of chronic diseases and cancer tumors [12]. Cancer could possibly be caused by epigenetic dysregulation, like activating tumor-promoting genes and silencing tumor-suppressor genes (e.g. low H3K4 methylation levels in breast cancer patients, high H3K9 acetylation levels in lung cancer patients) [12].

The basic biological tools for these regulatory modifications are different groups of histone-modifying enzymes. For example, the histone acetylation is regulated by opposite functions of histone acetyltransferases (HAT) and histone deacetylases (HDAC) [11, 12]. Several Inhibitors of histone-modifying enzymes have been approved by the FDA for treatment of malignancies, including two DNA-methyltransferase (DNMT) inhibitors (Azacitidine, and Decitabine) and two HDAC inhibitors. The two HDAC inhibitors (HDI) are Vorinostat (Zolinza); developed as a pan-HDAC inhibitor, and Romidepsin (Istodax); developed as a specific class I HDAC inhibitor. The inhibitors of HDACs show an effect on many cellular processes; like differentiation, inhibiting the cell cycle and induction of apoptosis [6]. However, the detailed mechanisms of Histone deacetylases Inhibitors is still unknown and obscure, as most of discovered HDI are non-selective and exhibit toxicity in the clinical tests [6].

The counterparts of HDACs in the process of histone acetylation are the histone acetyltransferases (HATs), which contain a cofactor (Ac-CoA) binding pocket and a substrate binding pocket (for the acetylated lysine). The substrate binding pocket usually appears as a solvent-accessible channel [11, 14]. Most of the reported HAT inhibitors are either natural products or very large peptide inhibitors [11]. Only a few less potent small molecule HAT inhibitors have been reported so far [15]. The therapeutic potential of HAT inhibitors is unclear at the moment, and more potent and selective inhibitors are highly needed to further study the role and involvement of individual HATs in tumorigenesis.

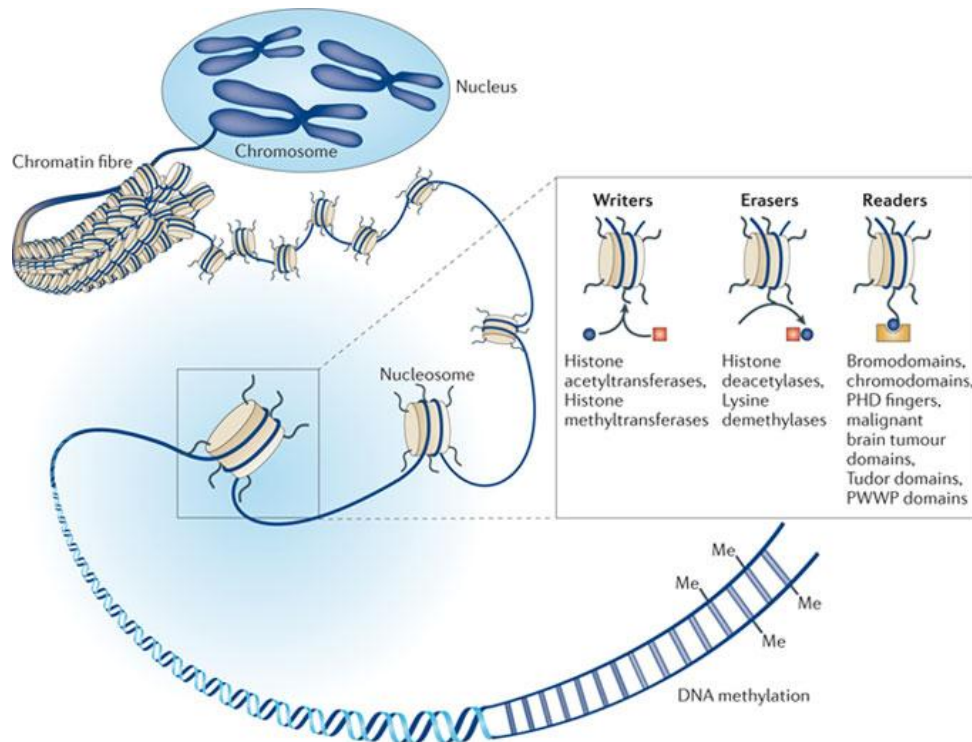


Figure 1.2: DNA is packaged into chromatin by wrapping around histone proteins (two copies each of histones H2A, H2B, H3 and H4) to form the nucleosomes, which are further compacted to form the chromatin. The degree of compactness depends on the types of post-translational modifications on the histones. The combination of modifications on each histone and/or nucleosome establishes an epigenetic code, and mediated by epigenetic enzymes; Proteins that covalently attach acetyl or methyl groups to produce (or 'write') the code (including histone acetyltransferases and histone methyltransferases) and are termed 'writers'. Proteins that recognize and bind to histone modifications are termed 'readers' of the code (including bromodomains, plant homeodomains (PHDs) and members of the royal family of methyl-lysine-binding domains). Enzymes that remove histone marks are termed 'erasers' (including histone deacetylases and lysine demethylases) (Adapted from Arrowsmith *et al.* [12]).

1.5 Aim of the work

Targeted cancer therapy nowadays represents an established paradigm in anti-cancer therapy. The concept of targeted therapy depends on developing antibodies or small-molecules capable of inhibiting specific elements in key signaling pathways or affecting the regulation mechanism of gene expression. **Protein kinases** are versatile and very flexible proteins, which play key role in regulating most, if not all, the critical signaling pathways involved in developing the cancer cells' hallmarks. On other hand, the epigenetic factors and the chemical modifications of the histones, catalyzed by **epigenetic enzymes**, appeared to have critical role in the cell cycle dysregulation, activating tumor-promoting genes, or silencing tumor-suppressor genes.

The current work focuses on the application of computational and structure-based methods for developing novel effective small-molecules, which could be potential inhibitors for targeted cancer therapy. The **first part** focuses on the application of virtual screening techniques to accelerate **the discovery of novel epigenetic modulators**; especially for epigenetic enzymes with little information about their inhibitors. Docking and virtual screening of focused chemical libraries are applied in the case of histone acetyltransferase PCAF to identify and develop novel inhibitors (Chapter 2).

The **second part** will focus on **developing selective kinase inhibitors** and the analysis of the challenges in the field of kinase inhibitor drug design (Chapter 3). A special focus will be given to the challenges associated with the flexibility of kinases and the structural conservation of the ATP-binding pocket by applying computational simulation and binding energy methods. Additionally, we will analyze the role of water molecules located inside the ATP binding pocket for providing water-mediated protein-inhibitor interactions.

To overcome these challenges, computational tools more sophisticated than simple docking and virtual screening are needed. **End-point free energy methods**, such as PBSA/GBSA implicit solvent models and the Linear Interaction Energy (LIE) approach, are applied to address two typical problems in drug design: the **prediction of the actual binding mode** and the **prediction of binding affinities**. In chapter 4, we analyze the performance of implicit solvent models and LIE method for predicting the binding mode for a series of GSK3 β /CDK2 inhibitors. The kinase flexibility and induced-fit effects are studied by molecular dynamics (MD) simulations. Chapter 5 deals with predicting the binding affinities of a large series of inhibitors developed for the constitutively active mutant D816V of c-Kit kinase. This in an example, where a large number of inhibitors, synthesized in a drug discovery project, is studied by a variety of computational methods in order to predict the biological activities of the studied inhibitors and also to lead the design of new small-molecule inhibitors.

Part I

Virtual Screening Methods for Novel Epigenetic Modulators

Chapter 2

Novel Inhibitors of Histone Acetyltransferases: Application of Focused Virtual Screening Methods

2.1 Histone Acetyltransferases: Biological Role and Classification

The genetic material in the nucleus of eukaryotic cells exists in tightly packed form, which functions as a dynamic structure and basic contributor in the regulation of various nuclear processes, including transcription, DNA replication and repair, mitosis, and apoptosis [16]. An important post-translational modification of histones is the acetylation of ϵ -amino groups on conserved lysine residues. Acetylation neutralizes the positively charged lysines and therefore affects interactions of the histones with other proteins and/or with the DNA. Histone acetylation has long been associated with transcriptionally active chromatin and also implicated in histone deposition during DNA replication [17, 18, 19]. The human genome encodes up to 25 proteins that show lysine acetyltransferase activity. At the primary structure level there is little similarity between the different HATs, and even members of the same family usually display considerable sequence diversity. Furthermore, there is no single homolog domain that is conserved in all HATs, although many enzymes contain recognizable Acetyl-CoenzymeA (Ac-CoA) binding motifs and bromodomains [20]. HATs display a conserved core domain, which contains an L-shaped cleft formed by the N- and C-terminal segments of the core domain. This cleft contains the catalytic site, where Ac-CoA binds in the short segment and the macromolecular substrate binds in the long segment. Beyond the core domain, there is little structural similarity between

the different HATs. *In vitro* assays indicated that HATs have different substrate specificities[21].

Important and extensively investigated families of HATs are:

- GNAT family (GCN5-related N-acetyltransferase): includes GCN5, PCAF (p300/CBP-associated factor), other acetyltransferases like serotonin acetyltransferase (AANAT), aminoglycoside N-acetyltransferases (AAC-3, and AAC-6), spermidine/spermine N-acetyltransferase, the elongator subunit Elp3, and Hpa2. HAT1 could be classified to GNAT or as separate family. Similarities are observed at the tertiary structure level for members of GNAT family (Figure 2.1) [14, 22].
- MYST family (named after its founding members, which include MOZ, YBF2/SAS3, SAS2 and TIP60) [11, 12, 23].
- p300/CBP family [24, 25, 26].

Over 40 transcription factors and 30 other nuclear, cytoplasmic, bacterial, and viral proteins have been shown to be acetylated *in vivo* by HATs [20, 27]. For example, p300/CBP proteins are involved in diverse physiological processes, such as proliferation, differentiation and apoptosis [26]. GCN5p is the catalytic subunit of the two multi-protein complexes, ADA and SAGA, involved in remodeling the chromatin structure and acetylation of histone tails at specific lysines [14]. The p300/CREB-binding protein (p300/CBP) works often as transcriptional co-activator, which binds to variety of transcription factors. P300/CBP has intrinsic activity, which plays important role in many biological functions, and could be associated to some tumorigenesis [19, 26]. Most potent and selective p300/CBP inhibitor is Lys_CoA (with $K_i=20$ nM) [28].

Acetylation and deacetylation of histones have emerged as key mechanisms regulating transcriptional activity. Histone acetyltransferases (HATs) catalyze the transfer of acetyl groups to lysine residues in histones, which results in a more open conformation of nucleosomes and increased accessibility of regulatory proteins to DNA [29, 30]. The acetylation status of several non-histone proteins, including p53, ataxia-telangiectasia mutant (ATM), heat shock protein 90, and α -tubulin, is intimately related to their functions [31, 32, 33]. Reversible acetylation of α -tubulin marks stabilized microtubule structures and may contribute to regulating microtubule dynamics. Acetylation of tumor suppressor protein p53 by two HAT subtypes; p300/cAMP-responsive element binding protein (p300/CBP) and p300/CBP-associated factor (PCAF), was linked to its transactivation potential and ability to regulate cell cycle arrest and apoptosis [34, 35].

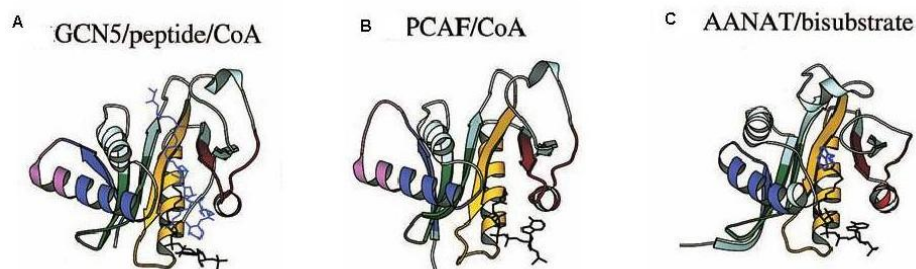


Figure 2.1: Comparison of the three-dimensional structures of GCN5-related N-acetyltransferases: GCN5, PCAF, and AANAT. (A) tGCN5: the ternary complex with CoA and an 11-residue peptide (in blue) is shown. The black line indicates CoA or Ac-CoA. (B) PCAF, complexed with CoA, (c) AANAT: the complex with the bisubstrate analog is shown (indole ring colored blue). The four conserved motifs of the GNAT superfamily C, D, A, and B are shown in purple, green, yellow, and red, respectively (adapted from [14]).

2.1.1 Structural Overview of Histone Acetyltransferase PCAF

In the PCAF crystal structure (1cm0.pdb), CoA is bound in a conformation, forming an extensive set of protein interactions that are mediated predominantly by the pantetheine arm and the pyrophosphate group [36] with motif A-D and motif B (Figure 2.2). All but two groups of the pantetheine arm–pyrophosphate chain make contacts with the protein. Most of the contacts are mediated through either protein backbone hydrogen bonds or protein side chain van der Waals contacts [36]. GNAT conserved residues in PCAF motifs A and B interact extensively with CoA. It could be noticed that residues 580 and 582–587 in the β 4-loop– α 3 region of motif A make direct and water-mediated hydrogen bonds with the pyrophosphate group [37]. T587 also makes a hydrogen bond to the pyrophosphate oxygen. The aliphatic side chain of Q581 and a C–A–V sequence (residues 574–576) at the top of the β 4-strand makes van der Waals contacts with the aliphatic part of the pantetheine arm [36] (see Figure 2.8 for details). In addition, the backbone of C574 and V576 forms hydrogen bonds with the pantetheine arm. Residues in the β 5-loop– α 4 region of GNAT motif B interact by van der Waals contacts with the β -mercaptoethylamine segment of the pantetheine arm and thus play a major role in orienting the reactive sulfhydryl atom for the acetyl transfer [36] (Figure 2.8). Other protein residues, involved in the binding, are A613, Y616 and F617. Also Y616 makes van der Waals contacts with the end of the pantetheine arm near the pyrophosphate group [36] (Figure 2.9 and 2.10).

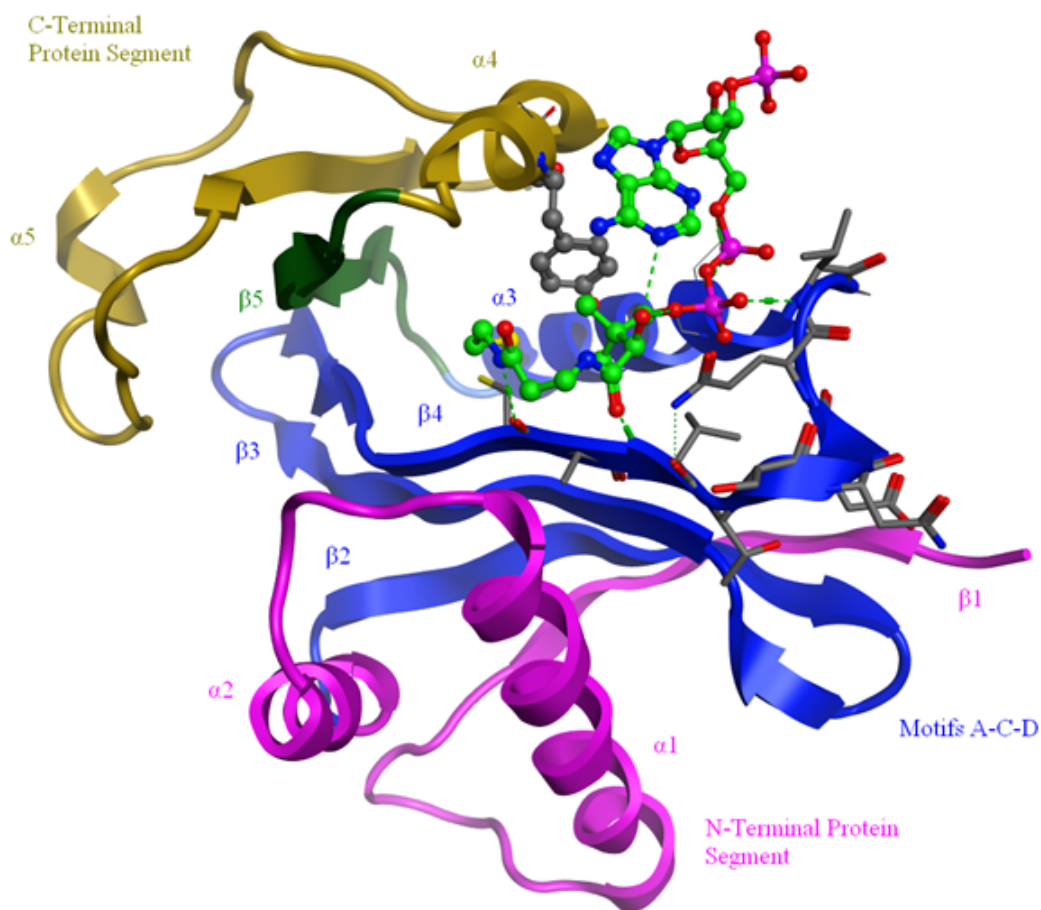


Figure 2.2: Structure of the PCAF-CoA complex representing the general secondary structure of GNAT family acetyltransferases and the location of the acetyl-CoA binding site. The four domains of the protein are color-coded. Motifs A-D and motif B (based on structural conservation) are colored blue and green, respectively. The N- and C-terminal protein segments flanking the core are colored magenta and gold, respectively. CoA is colored light green. Y616 is shown in capped sticks colored grey.

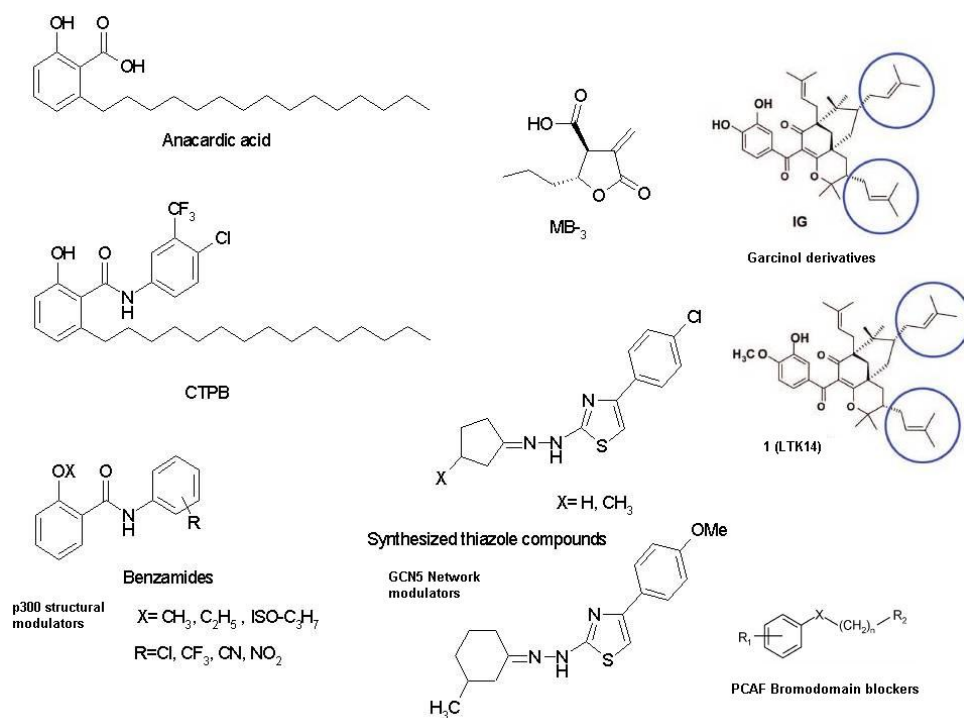


Figure 2.3: Some known HAT inhibitors from the literature. For more details, see [15].

2.2 Computational Methods

2.2.1 Molecular Docking

The program GOLD 4.0 [38] (Cambridge Crystallographic Data Centre) was used for docking, while the calculation of all molecular descriptors and the analysis of the docking results were carried out by MOE2008.10 (Chemical Computing Group). The crystal structure of human PCAF in complex with Acetyl-CoA (PDB id: 1cm0, chain B) resolved at 2.30 Å was taken from the Protein Data Bank. The cofactor and the water molecules were removed, hydrogen atoms were added and the protein was minimized using the AMBER force field and the conjugate gradient minimization (MOE 2008.10) until the gradient of 0.1 kcal/mol was reached. Docking of the ligands was carried out using the GOLD 4.0 program with default settings. A sphere of 20 Å around the oxygen atom of Y616 was defined for ligand docking. For covalent docking, a sphere of 14 Å around the S-atom of C574 was defined for ligand covalent docking.

To test the applicability of the docking tool, a control docking was carried out with the cofactor acetyl-CoA. Using GOLD score as scoring function, an RMSD value of 1.45 was derived for the top-ranked conformation of Acetyl-CoA (data not shown). For all compounds under study, the GOLD score as well as Chemscore (applied in GOLD) were calculated and analyzed. To support the obtained docking poses, we calculated the molecular interaction fields for

the PCAF binding pocket using MOE2008.10. The favorable interaction field of a carboxylate probe (Fig. 2.10) as well as of a hydrophobic methyl (C3) probe was calculated. The results agree well with the location of the carboxylic head group as well as the hydrophobic parts of the active inhibitors (Figure 2.10).

2.2.2 Similarity Search and Focused Library Design

Molecular Chemical fingerprints could be used to search in large compound databases for structurally related molecules to a given search query. Several fingerprint systems were implemented in Chemical Computing's Molecular Operating Environment (MOE). Moreover each fingerprint system would support a number of similarity metrics and use different representation. The following fingerprints systems are applied in MOE:

1. MACCS Structural Keys (feature list version). Each feature indicates the presence of one of the 166 public MDL MACCS structural keys computed from the molecular graph. The fingerprint is represented as a sparse list of keys present in the molecule.
2. Bit MACCS: MACCS Structural Keys (bit packed version). Each feature indicates the presence of one of the 166 public MDL MACCS structural keys calculated from the molecular graph. The fingerprint is a dense bit vector of feature bits 6 words long.
3. Protein Ligand Interactions Fingerprints: Each feature represents a protein-ligand interaction type, e.g. hydrogen bond or ionic interaction.
4. PiDAPH3: 3-point pharmacophore based fingerprint calculated from a 3D conformation. Each atom is given one of 8 atom types computed from 3 atomic properties: "in pi system", "is donor", "is acceptor". Anions and cations are not represented. Then, all triplets of atoms are coded as features using the three inter-atomic distances and three atom types of each triangle.
5. piDAPH4: 4-point pharmacophore based fingerprint calculated from a 3D conformation. Each atom is given one of 8 atom types computed from 3 atomic properties: "in pi system", "is donor", "is acceptor". Anions and cations are not represented. Then, all quadruplets of atoms are coded as features using the six inter-atomic distances, four atom types and chirality of each quadruplet.
6. GpiDAPH3: 3-point pharmacophore based fingerprint calculated from the 2D molecular graph. Each atom is given one of 8 atom types computed from 3 atomic properties: "in

pi system", "is donor", "is acceptor". Anions and cations are not represented. Then, all triplets of atoms are coded as features using the three graph distances and three atom types of each triangle.

Tanimoto similarity search could be later accomplished using MOE. Tanimoto similarity module calculates the similarity values for each target molecule with respect to one or more reference molecules using molecular fingerprints systems. The Tanimoto similarity search is defined by the expression: $\text{Similarity} = N_{ab} / (N_a + N_b + N_{ab})$ where : N_{ab} is the number of fingerprint bits presented in both reference and target molecule, N_a is the number of fingerprint bits presented only in the Reference molecule, N_b is the number of fingerprint bits presented only in the Target molecule. Tanimoto similarity index ranges from zero (no common bits) to one (exact same bits).

2.2.3 Database Virtual Screening

The 260,000 3D structures of the National Cancer Institute (NCI) database, generated with the program CORINA, were obtained from the NCI homepage and were imported as Mol2 files into the MOE program. The 3D structures of the Ambinter, ChemDiverse, Chemical Block and Enamine were retrieved from the ZINC database [39, 40, 41]. Using an isothiazolone/isothiazolidinone ring as search query, we identified 51 isothiazolone compounds from the NCI database that were subsequently docked into the PCAF protein structure using the GOLD program as described above. Thirty-two (32) molecules were successfully docked into the CoA binding pocket showing a distance $<5 \text{ \AA}$ between the S–N bond and the thiol group of Cys574 (distance S to S) to facilitate the nucleophilic attack from the cysteine residue on the isothiazolone ring. From the 32 compounds selected, 15 compounds could be obtained from the NCI. By using the same isothiazolone search query, six further compounds were identified and purchased from Ambinter, ChemDiverse, Chemical Block, and Enamine.

To find novel non-covalent PCAF inhibitors, MACCS Structural Keys (feature list version) and GpiDAPH3 fingerprints were used to filter compounds from 21 commercial chemical databases (Abc-eurochem, Ambinter, Asinex, Chembridge, ChemDiv, Com-Genex, Enamine, Ibscreen, Interchem, Keyorganics, Life Chemicals, Maybridge, Nanosyn, NCI, Otava, PeakDale, PHARMEKS, PUBCHEM, Ryan-Scientific, Sigma-Aldrich, Spec, UkrOrgSynth). MACCS fingerprints generated from published rhodanine-indolinone AANAT inhibitors were used as a search query to identify chemically similar compounds by applying a tanimoto coefficient of (0.85). The similarity based screening identified 6423 compounds by taking into consideration the Lipinski rule of five [42].

2.2.4 Molecular Dynamics Simulation

Molecular dynamics was carried out using AMBER 9.0 Molecular Dynamics package. Protein and ligand molecule were parameterized using AMBERff03 [43] and general AMBER force field (GAFF) [44] respectively. The initial structures of the PCAF–inhibitor complexes were taken from the GOLD docking study. Complex systems were neutralized with 8 Cl⁻ counter-ions by using the xleap module of AMBER 9.0. The structures were solvated in an octahedral box with TIP3P [45] water molecules leaving at least 10 Å between the solute atoms and the borders of the box. The fully solvated and neutralized systems were subjected to energy minimization with the sander module of the AMBER 9.0 package. Following minimization the systems were gradually heated from 50 to 300 K with positional restraints (force constant: 50 kcal/mol/Å) on protein-ligand complex over a period of 0.25 ns allowing water molecules and ions to move freely. A 9 Å cutoff for the short-range non-bonded interactions was used in combination with the Particle Mesh Ewald (PME) option [46] using a grid spacing of ~0.9 Å to account for long-range electrostatic interactions. The Settle algorithm [47] was used to constrain bond vibrations involving hydrogen atoms. During additional 0.25 ns the positional restraints were gradually reduced to allow finally unrestrained MD simulation of all atoms over a subsequent equilibration time of 1 nanosecond (ns). Further 10 ns free MD simulation was carried using these equilibrated structures as a start structures. VMD [48] was used for visualization of trajectories and preparation of figures.

2.3 Results and Discussion

In order to identify new inhibitors of HATs, we applied a strategy that depends on creating a focused virtual database based on chemical similarity with known inhibitors of a homologous target, and subsequently filtered according to Lipinski rule of five. These focused virtual databases were used to conduct a virtual screening after optimizing the molecular docking protocol regarding the conditions, constraints, and scoring functions.

2.3.1 Identification of Isothiazolone-based HAT Inhibitors as Covalent Inhibitors

Aryl and alkyl N-substituted isothiazolone compounds (e.g., CCT077791) have been shown to inhibit irreversibly the acetylation of histones H3 and H4 by PCAF and p300. Stimson *et al.* showed that a series of isothiazolones, identified from high-throughput screening, inhibits HAT catalytic activity [49]. These isothiazolones are also cell permeable and can reduce global acetylation, as well as acetylation of specific histones (H3 and H4) and non-histone proteins, such as alpha-tubulin. HAT inhibition by isothiazolones is abolished in the presence of thiol-reducing agents; such as dithiothreitol (DTT) or glutathione. Furthermore, HAT activity was not restored

in experiments involving the incubation of PCAF with the two isothiazolones CCT077791 and CCT077792, followed by dialysis for 24 hours. The SAR study of this series of compounds proved that their activity is related to the nature and electron withdrawing/pushing properties of the substitutes. According to the supposed mechanism of disulfide bridge formation of the isothiazolone compounds to produce irreversible inhibition of HAT [49], it is thought that isothiazolones could form a covalent bond with the side chain of C574 in PCAF (Figure 2.5a and 2.5b). As this residue is taking place in the acetyl transfer reaction and it contributes in adjusting the binding of the cofactor, such covalent binding will produce a permanent loss of HAT catalytic activity. Using the MOE rotamer explorer, we calculated five conformations for the SH-side chain of the C574 in PCAF. By choosing the preferred orientation, we could identify a favorable docking solution for the isothiazolone compound **NCI694616** that was identified as potent inhibitor in the *in vitro* testing (Table 2.1). The calculated binding mode showed that the nitro group is located in the cofactor binding pocket, while the bromo-phenyl is interacting with the residues of the substrate histone binding pocket. Also the other isothiazolones could be docked to PCAF and showed the same orientation (Figure 2.5b).

We selected first the National Cancer Institute (NCI)'s 3D database for virtual screening, since it contains structurally diverse synthetic compounds collected from many laboratories around the world as well as natural products. The recently identified isothiazolone inhibitor **NCI694616**, which acts as an irreversible inhibitor by covalently binding to the residue C574 at the PCAF active site, was used to derive a search query for the *in-silico* screening. The crystal structure of PCAF complexed with the cofactor Acetyl-CoA (PDB id: 1cm0) was used for docking studies. We identified 51 compounds with an isothiazolone or isothiazolidinone substructure that were subsequently docked into the PCAF substrate binding site to test whether they are able to bind at the catalytic site. Thirty-two compounds were identified to contain a reactive S–N bond, which was located (by docking) in close proximity to the active site C574 in the model (S–S distance below 4.5 Å). Among these, 15 compounds (Table 2.1) could be obtained from the NCI. Using the same search query six other pyridoisothiazolones were identified and purchased from commercial suppliers (Ambinter, ChemDiverse, Enamine).

The 21 selected compounds were tested for *in vitro* inhibition of histone acetyltransferase activity using recombinant human PCAF by S. Furdas (University of Freiburg). Biotinylated oligopeptide sequences from both histones H3_(aa1–21) and H4_(aa2–24) were used as the substrates. The conversion was detected with a primary antibody against acetylated histone H3 and H4 and quantification was achieved by using a secondary Europium-labeled antibody with a final measurement of time resolved fluorescence. This assay had been used before for the determination of cellular hypoacetylation caused by isothiazolones. The published isothiazolone **v** (Figure 2.4) was synthesized and used as reference inhibitor [50].

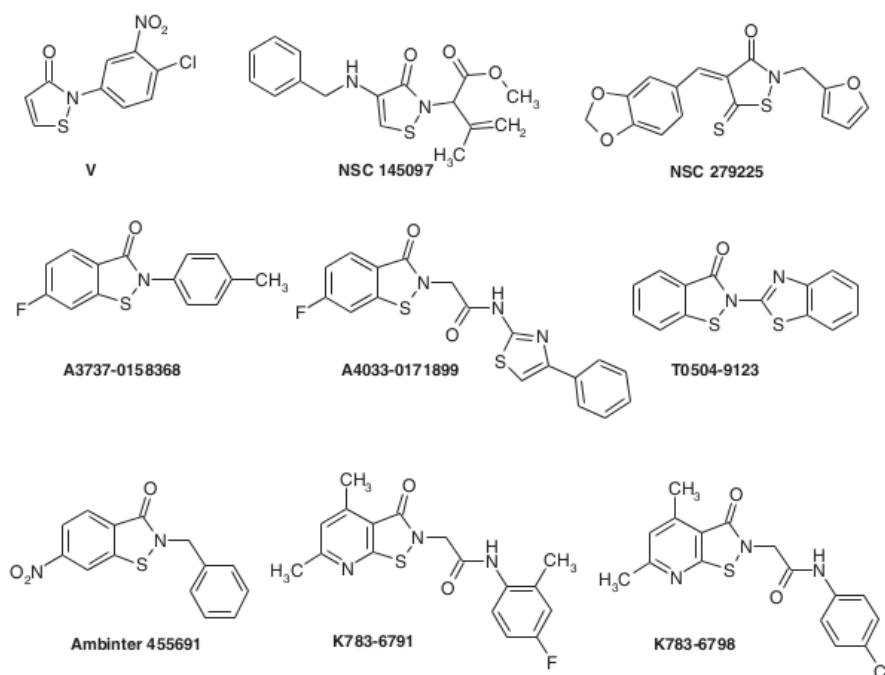
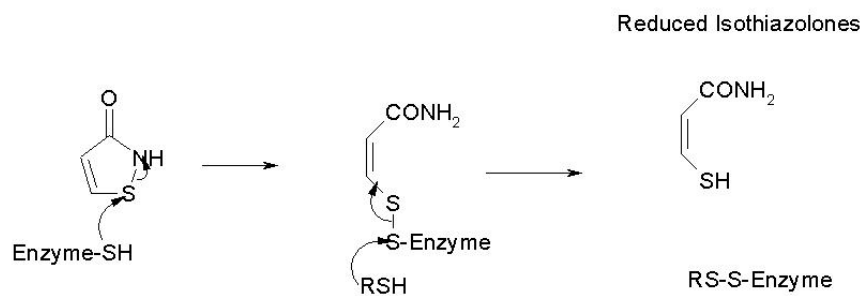
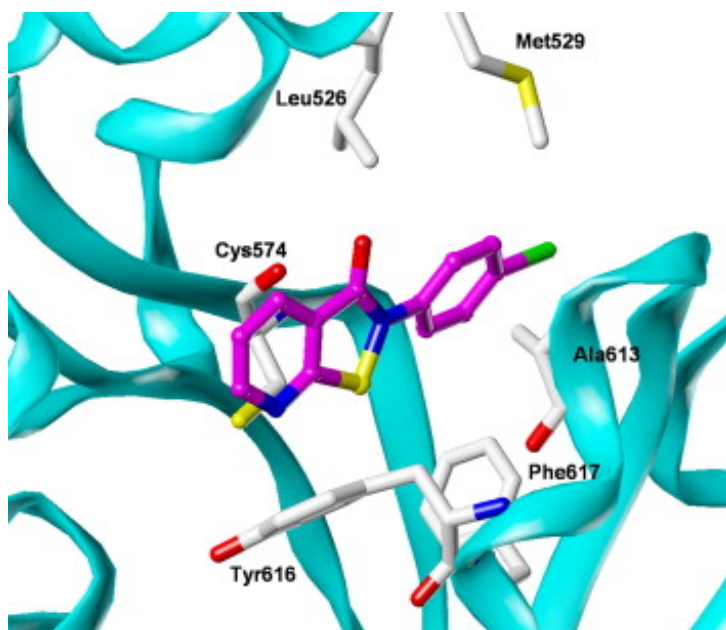


Figure 2.4: Chemical structures of some selected isothiazolones.

The most active pyridoisothiazolone showed PCAF inhibition in the submicromolar range. In most cases, a higher potency was observed on the H4 substrate. Thus, the binding of the histone substrate seems to influence the structure or reactivity (or both) of the acetyltransferase with regard to inhibitor responsiveness. The amino-substituted isothiazolone **NSC145097** (Figure 2.4) as well as the isothiazolone-thione **NSC279225** (Figure 2.4) did not show significant HAT inhibition. Isothiazolones substituted with an aminocarbonylmethyl linker (**A4033-0171899**, **K783-6791**, and **K783-6798**) showed only moderate inhibition on PCAF inhibition. The highest potency was observed for the pyridoisothiazolones **NSC694614** and **NSC694622**. Further compounds **4a–g** and **8a–h** (Table 2.1) were also synthesized by S. Furdas (University of Freiburg) [50]. The inhibitory activity of **NSC694622** against HAT was significantly reduced in the presence of 1 mmol/L DTT; confirming the role of thiol group and the reactivity of N-S bond in the inhibitory activity of the pyridoisothiazolones toward PCAF (Figure 2.5a). Additionally, introducing a methylene group between the aromatic substituent and the isothiazole core (e.g., compound **8h**) results in increasing the distance C574/S–N bond and is therefore unfavorable.



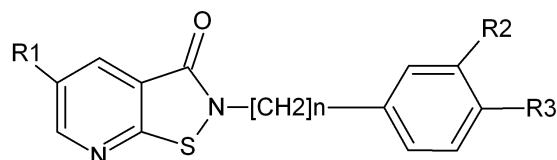
(a) Proposed mechanism of the covalent binding of isothiazolones to thiol groups.



(b) GOLD docking solution for pyridoisothiazolone **4d** (colored magenta) at the PCAF binding site. The sulphur of the isothiazolone inhibitor is in close proximity to the active site residue C574, thus enabling the formation of a covalent bond between the enzyme and the isothiazolone.

Figure 2.5: The proposed mechanism of the isothiazolones' covalent binding and docking solution for pyridoisothiazolone

Table 2.1: Inhibition of PCAF histone acetyltransferase activity for histone substrates H3 (amino acid residues 1–21) and H4 (amino acid residues 2–24): IC₅₀ value [μ M] \pm standard error [μ M] or enzyme inhibition [%] at the specified concentration. (n.t.: not tested; n.i.: no enzyme inhibition (<5%) at the specified assay concentration.)



Cmp.	n	R1	R2	R3	PCAF, H3	PCAF, H4
NSC694614	0	CH3	H	Br	2.99 \pm 0.27	n.t.
NSC694615	0	H	H	Br	4.43 \pm 0.20	n.t.
NSC694616	0	NO ₂	H	Br	4.91 \pm 0.39	0.86 \pm 0.08
NSC694617	0	H	H	NO ₂	4.10 \pm 0.60	2.99 \pm 0.32
NSC694618	0	NO ₂	H	NO ₂	11.8 \pm 1.83	15.8 \pm 5.76
NSC694619	0	CH3	H	OC ₆ H ₅	12.9 \pm 0.95	n.t.
NSC694620	0	NO ₂	H	OC ₆ H ₅	48% @ 10 μ M	n.t.
NSC694621	0	H	H	OCH ₃	5.71 \pm 0.30	n.t.
NSC694622	0	NO ₂	H	OCH ₃	3.42 \pm 0.35	1.83 \pm 0.47
NSC694623	0	H	H	C ₄ H ₉	15.9 \pm 2.20	n.t.
NSC698599	0	CH ₃	H	NO ₂	15.3 \pm 6.21	2.72 \pm 0.60
NSC698600	0	CH ₃	H	OCH ₃	6.51 \pm 0.49	n.t.
NSC700864	0	Phe	H	C ₄ H ₉	49% @ 25 μ M	n.t.
4a	0	H	H	H	7.85 \pm 0.46	88% @ 50 μ M
4b	0	H	H	Cl	3.53 \pm 0.07	87% @ 50 μ M
4c	0	H	Cl	Cl	4.57 \pm 0.23	91% @ 50 μ M
4d	0	H	H	F	1.64 \pm 0.11	0.72 \pm 0.05
4e	0	H	F	F	5.90 \pm 0.85	n.t.
4f	0	H	H	CH ₃	4.63 \pm 0.24	75% @ 50 μ M
4g	0	H	OCH ₃	H	5.03 \pm 0.14	n.t.
8a	1	H	H	H	22.96 \pm 2.06	n.t.
8b	1	H	H	Cl	17.5 \pm 1.27	n.t.
8c	1	H	Cl	Cl	6.94 \pm 0.57	n.t.
8d	1	H	H	F	86.3 \pm 3.40	n.t.
8e	1	H	F	F	101 \pm 4.92	n.t.
8f	1	H	H	CH ₃	27.9 \pm 4.50	n.t.
8g	1	H	OCH ₃	H	22.2 \pm 1.97	n.t.
8h	1	H	H	CF ₃	130 \pm 8.49	25.5 \pm 2.85
V		See figure 2.4			4.80 \pm 0.20	n.t.
NSC145097		See figure 2.4			48% @ 300 μ M	20% @ 200 μ M
NSC279225		See figure 2.4			n.i. @ 25 μ M	n.t.
A455691		See figure 2.4			7.66 \pm 1.29	51.3 \pm 8.83
A3737-0158368		See figure 2.4			3.69 \pm 0.17	n.t.
A4033-0171899		See figure 2.4			21% @ 50 μ M	n.t.
T0504-9123		See figure 2.4			46% @ 50 μ M	n.t.
K783-6791		See figure 2.4			48% @ 50 μ M	n.t.
K783-6798		See figure 2.4			38% @ 50 μ M	n.t.

2.3.2 Identification of Rhodanine derivatives as Non-covalent HAT Inhibitors

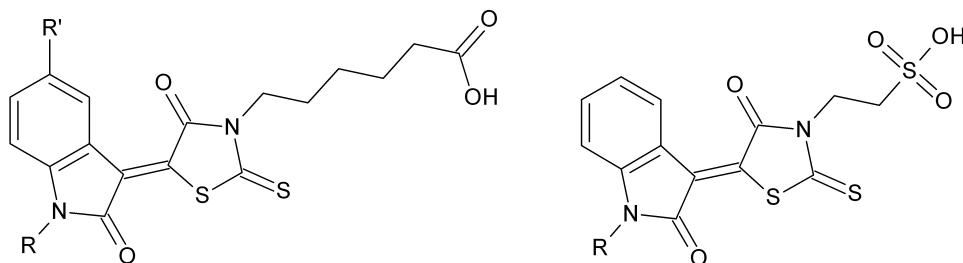


Figure 2.6: AANAT inhibitors from the literature [51], used as search query for generating focused library for virtual screening.

As a different starting point, we decided to use the structure of rhodanine-indolinone derivatives (shown in Figure 2.6), which were previously identified as inhibitors of the homologue serotonin N-acetyltransferase AANAT [51], another member of the GNAT family of acetyltransferases [14]. The aim was to find non-covalent small-molecule inhibitors of PCAF, as most of the known PCAF inhibitors are covalent inhibitors, peptide inhibitors, or very large natural products.

We used a synergistic approach that combines the benefits of structure-based virtual screening and experimental testing using validated PCAF inhibition assay to subsequently screen a limited number of the top-ranked compounds. We carried out a multi-step virtual screening experiment starting with a similarity based screening followed by docking. MACCS fingerprints derived from published rhodanine-indolinone AANAT inhibitors [51] (Figure 2.6) were used for a search query in 21 commercial databases. A total of 6423 compounds were identified from the *in silico* screening and subsequently docked into the PCAF binding pocket in order to test whether they are able to bind at the CoA binding site. The crystal structure of PCAF complexed with the cofactor acetyl-CoA (PDB id: 1cm0) was used for docking studies using GOLD4.1 and GOLD score as scoring functions. A rescoring with Chemscore, followed by visual inspection, was carried out for the 100 top ranked solutions and 11 compounds were further considered for biological testing (Figure 2.7). The purchased compound **PHAR037680** was detected as most promising lead inhibiting PCAF with $IC_{50} = 97.7 \mu\text{M}$ (Table 2.2). Additional compounds with the same core rhodanine-indolinone-scaffold (**12a-g**, **13**, **14**, and **15**) were synthesized by S. Furdas (University of Freiburg) (Table 2.2) [52].

The predicted binding mode for the rhodanine-indolinone-carboxylate analogs suggests that the inhibitor's acidic group, aliphatic linker, as well the rhodanine ring interact with the cofactor binding pocket, whereas the rest of the inhibitor interacts with the substrate binding pocket (Figure 2.9). After performing MD simulation study, the derived RMSD plot (calculated for

heavy atoms of PCAF protein and ligand **12e**, respectively) shows that the complex, as well as the interaction of the inhibitor with the protein, is stable over the simulation time of 10 ns (See Appendix B, Figure B.1). Similar results were obtained for compounds **14** and **15**, which are analogs of inhibitor **PHAR037680**, supporting the selected docking binding mode.

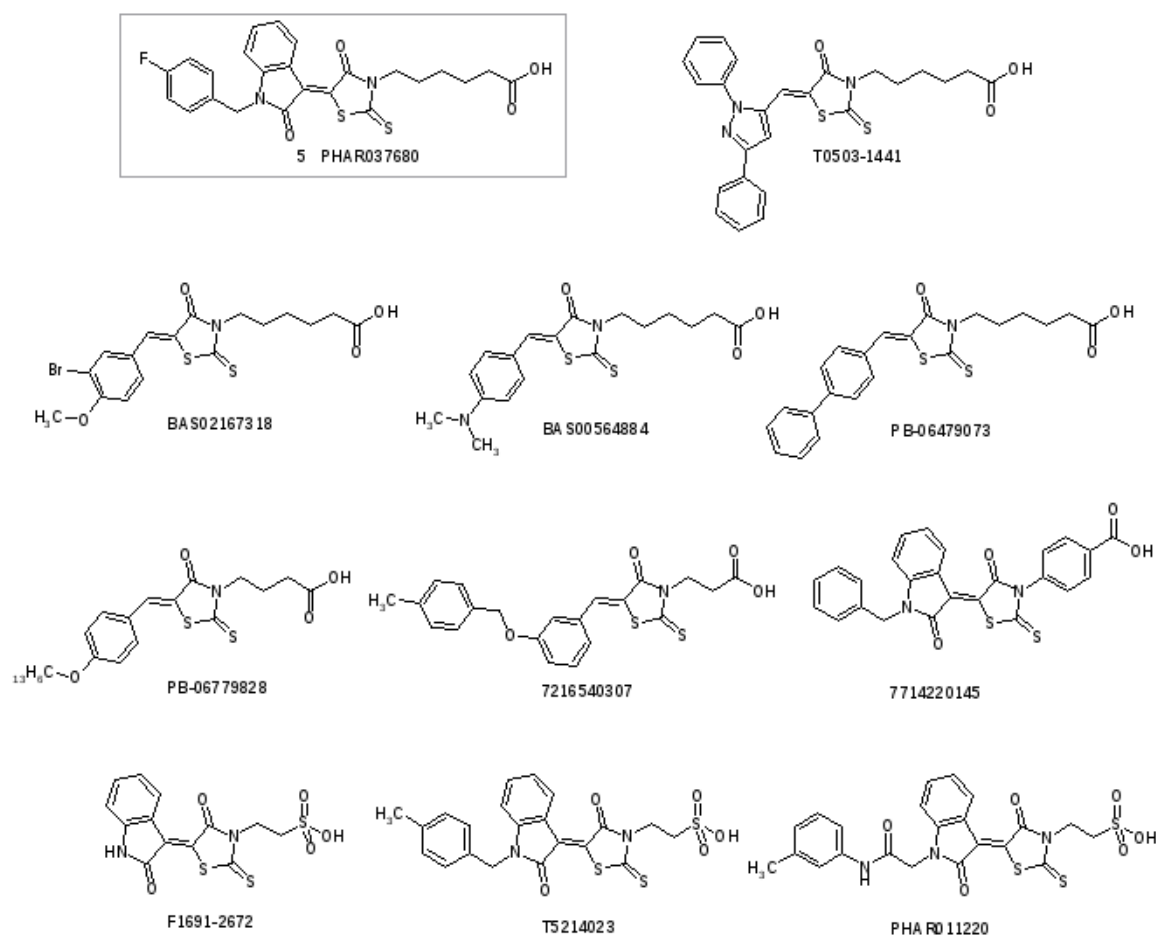
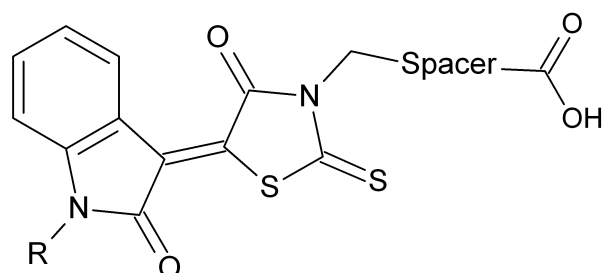


Figure 2.7: Selected compounds from the virtual screening with rhodanine-indolinone scaffold. Detailed information concerning the suppliers and docking scores can be found in Appendix B.

Table 2.2: Inhibition of PCAF histone acetyltransferase activity for histone substrate H3 (amino acid residues 1-21); IC₅₀ value [μM] \pm standard error [μM] or enzyme inhibition [%] at the specified concentration for synthesized compounds according to the scaffold.



Cmp.	Spacer	R	IC ₅₀ or inhibition (%) (PCAF, H3 _{aa1-21})
12a	C ₄ H ₈ (<i>butylene</i>)	H	9% @ 50 μM
12b	C ₄ H ₈ (<i>butylene</i>)	CH ₃ (<i>methyl</i>)	5% @ 50 μM
12c	C ₄ H ₈ (<i>butylene</i>)	C ₂ H ₅ (<i>ethyl</i>)	6% @ 50 μM
12d	C ₄ H ₈ (<i>butylene</i>)	(C ₆ H ₅)CH ₂ (<i>benzyl</i>)	13% @ 50 μM
12e	C ₄ H ₈ (<i>butylene</i>)	4-CH ₃ (C ₆ H ₄)CH ₂ (<i>4-methylbenzyl</i>)	67.2 \pm 2.3 μM
12f	C ₄ H ₈ (<i>butylene</i>)	4-OCH ₃ (C ₆ H ₄)CH ₂ (<i>4-methoxybenzyl</i>)	28.8 \pm 2.1 μM
12g	C ₄ H ₈ (<i>butylene</i>)	4-NO ₂ (C ₆ H ₄)CH ₂ (<i>4-nitrobenzyl</i>)	78.2 \pm 3.6 μM
13	C ₃ H ₆ (<i>propylene</i>)	4-CH ₃ (C ₆ H ₄)CH ₂ (<i>4-methylbenzyl</i>)	8% @ 50 μM
14	C ₆ H ₄ (<i>phenylene</i>)	4-CH ₃ (C ₆ H ₄)CH ₂ (<i>4-methylbenzyl</i>)	63.7 \pm 1.6 μM
15	C ₆ H ₁₀ (<i>trans-cyclohexylene</i>)	4-CH ₃ (C ₆ H ₄)CH ₂ (<i>4-methylbenzyl</i>)	41.8 \pm 4.2 μM
PHAR037680		See figure 2.7	97.7 \pm 10.4 μM
T0505-1441		See figure 2.7	12% @ 50 μM
BAS02167318		See figure 2.7	8% @ 50 μM
BAS0056484		See figure 2.7	n.i. @ 50 μM
PB-06479073		See figure 2.7	n.i. @ 50 μM
PB-06779828		See figure 2.7	8% @ 50 μM
7216540307		See figure 2.7	5% @ 50 μM
7714220145		See figure 2.7	14% @ 50 μM

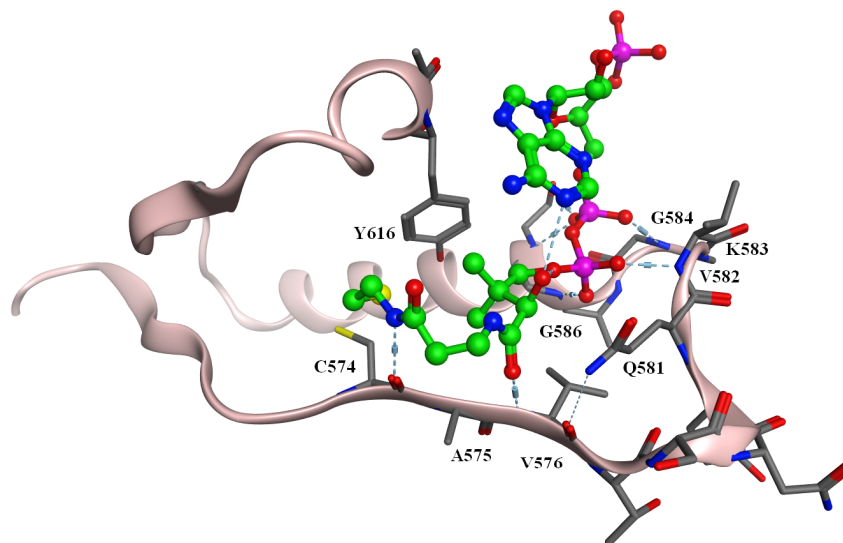


Figure 2.8: Interaction of CoA in the PCAF crystal structure (PDB id: 1cm0)

2.3.3 SAR Study and Molecular Interaction Fields

SAR study has been performed by synthetic modification of the active scaffold 'rhodanine-indolinone-carboxylate' after purchasing the compound **PHAR037680** (Figure 2.7). This SAR is in well agreement with the proposed binding mode. A negatively charged group (carboxylate) is necessary for the activity. Substituting the charged carboxylate with a neutral methoxylate would abolish the binding affinity. The highest binding affinity was for the analogs **12e**, **12f**, **12g**, **14**, and **15** (Figure 2.9b). That indicates that the binding affinity would be optimal with a specific spacer length (butylene or phenylene), which corresponds together with rhodanine ring to a distance of 10 Å, which is similar to the distance between the V562 and the central residue C574. Shortening the alkyl spacer by only one methylene group (in compound **13**) led to a strong decrease in activity but rigidification with a 1,4-phenylene (compound **14**) or a trans-cyclohexylene (compound **15**) spacer was possible and able to keep the inhibitory activity. On the other side of the scaffold, the Indolinone ring should be substituted at the aromatic nitrogen by a benzyl group, which orients its aromatic ring inside the substrate binding pocket (Figure 2.10).

Molecular interactions fields further support the predicted binding mode. They indicated the strongest affinity for a carboxylic group in the region around the residues V582 and G586 where the pyrophosphate part of CoA is interacting (2.10a). On the other hand, the highest affinity for a hydrophobic probe exists in the neighborhood of the reaction center where the cofactor's sulfur is located (Figure 2.10b).

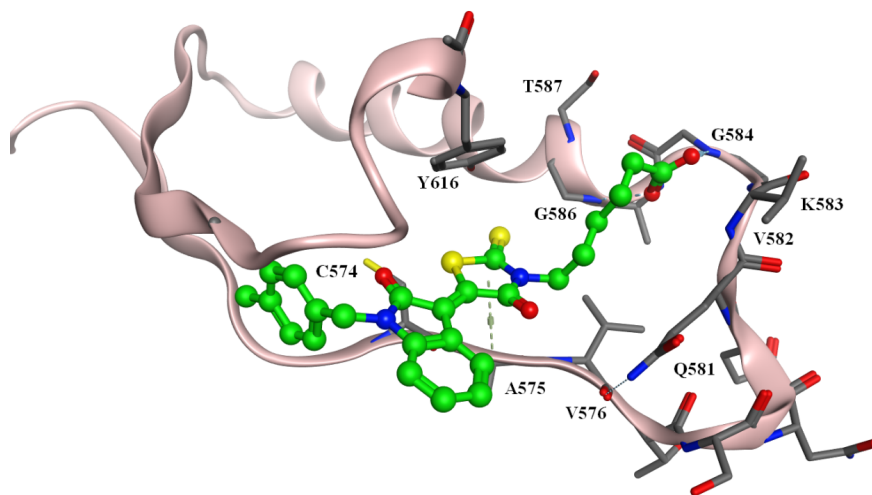
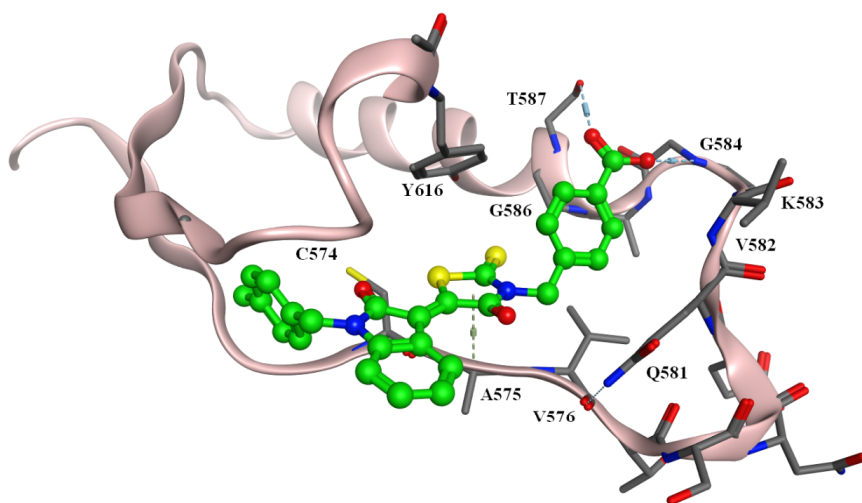
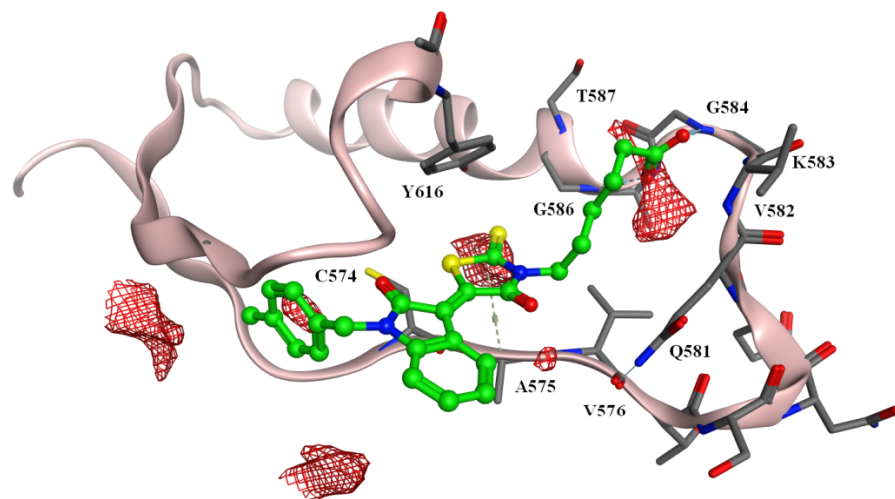
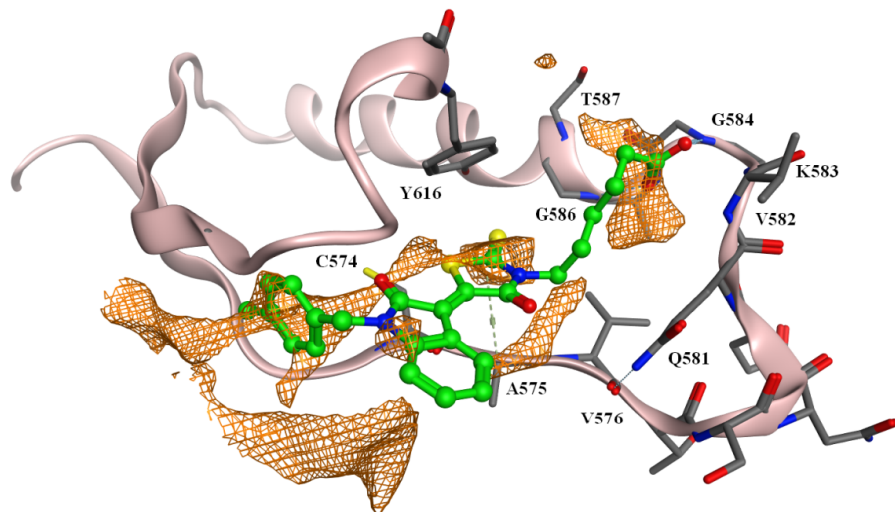
(a) Docking pose calculated for compound **12e**(b) Docking pose calculated for compound **14**.

Figure 2.9: Comparison of the docking pose for inhibitor **12e** in the binding pocket of PCAF crystal structure (PDB id: 1cm0), in comparison with the docked inhibitor **14**. Docking solutions propose that: the carboxylic group is interacting with V582, the rhodanine's sulfur is positioned beside the C574 and the benzyl group on the indolinone is positioned deeply inside the substrate binding pocket (compare with Fig. 2.8). SAR studies showed that a specific length for the linker is necessary to maintain the inhibitory activity, supporting the proposed binding mode.



(a) the predicted binding mode of the active inhibitor **12e** (colored green) is shown. Molecular interaction fields were calculated for the binding pocket using a) a carboxylate probe (contour level -4.5 kcal/mol, colored red).



(b) the predicted binding mode of the active inhibitor **12e** (colored green) is shown. Molecular interaction fields were calculated for the binding pocket using a hydrophobic methyl (C3) probe (contour level -2.5 kcal/mol, colored orange).

Figure 2.10: Docking pose calculated for compound **12e** with HAT PCAF (PDB id: 1cm0) with the Molecular interaction fields calculated for the PCAF binding pocket.

2.4 Conclusion

In order to identify new inhibitors of HATs, we applied a strategy that combined computational screening methods with a robust biochemical assay. Virtual database filtering was used to conduct virtual screening with the National Cancer Institute (NCI) compound collection and several commercial compound libraries. As search query we used the isothiazolone and isothiazolidinone ring system, which was derived from published isothiazolone HAT inhibitors. Isothiazolones were presented as new lead structures for HAT inhibition in the literature but suffer from general high reactivity. A covalent attachment to thiol groups does not necessarily rule out clinical application as can be shown, for example, for the antiulcer drug omeprazole. In this study, **pyridoisothiazolones** were identified via *in-silico* virtual screening and *in-vitro* enzyme testing as new covalent PCAF inhibitors.

To expand the chemical knowledge of HAT inhibitors, we tried to find further non-covalent inhibitors of Histone acetyltransferase PCAF by choosing a different starting structures, which are the rhodanine–indolinone derivatives, identified as inhibitors of the homologue serotonin N-acetyltransferase AANAT from the GNAT family of acetyltransferases. The result of the virtual screening was promising leading to the discovery of **rhodanine carboxylic acids** as HAT non-covalent inhibitors.

Rhodanine carboxylic acids are new but unselective HAT inhibitors with activity in the two-digit micromolar range. While the initial SAR study has better determined the pharmacophore for PCAF inhibition, the beneficial effects of certain structural elements (e.g. 4-methoxybenzyl group, and trans-cyclohexylene as spacer) point out to directions for further optimization. The new inhibitors are broadband HAT inhibitors while a pyridoisothiazolone reference inhibitor was shown to possess increased activity on CBP and to some extent on PCAF but is less potent on Gcn5.

Part II

End-point Free Energy Methods for Selective Kinase Inhibitors

Chapter 3

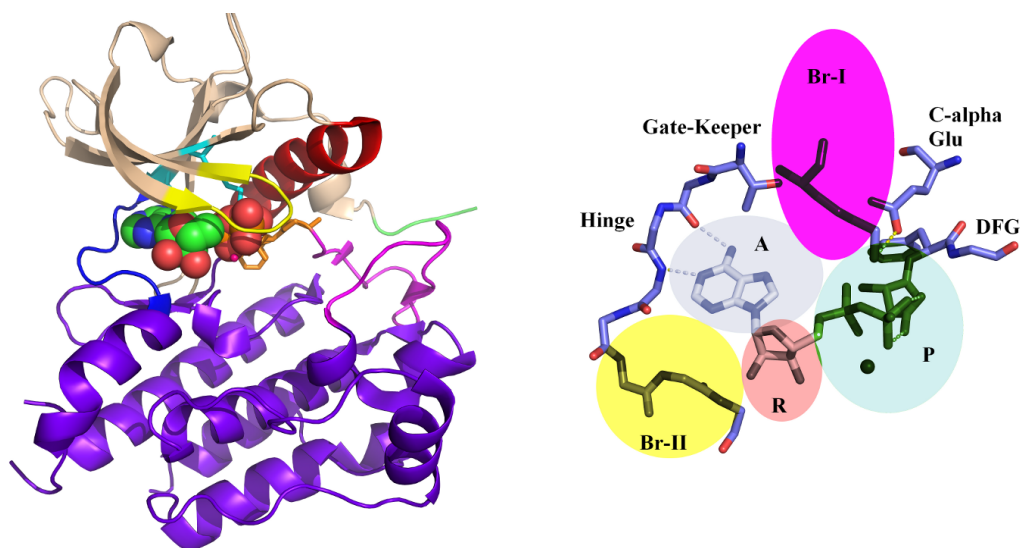
Structural Aspects of Protein Kinases: The Impact on Designing Selective Kinase Inhibitors

3.1 Kinases' Binding Pockets and Catalytic Cleft

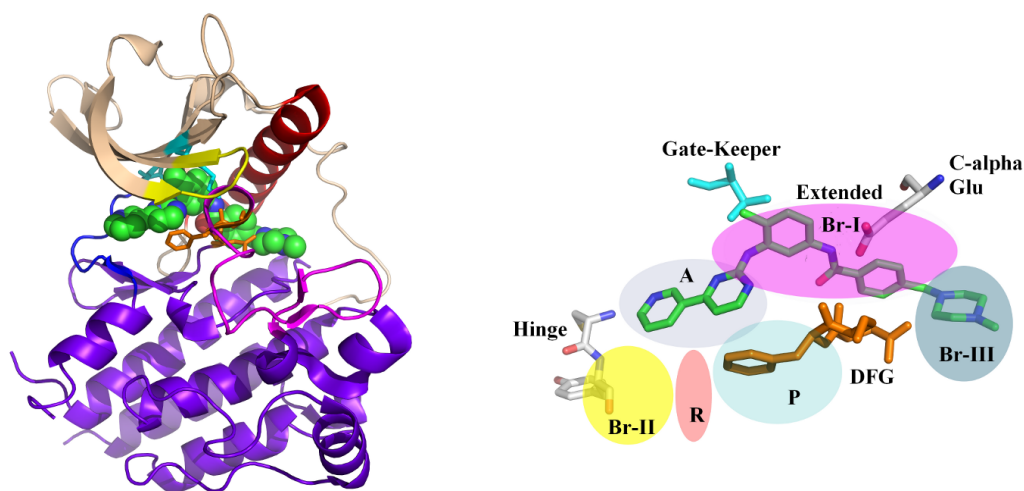
The catalytic domain of all kinases is located between two lobes: the N-terminal lobe and the C-terminal lobe; the N-lobe is formed of one alpha helix (called αC) and five beta strands forming the ceiling of the catalytic cleft, while the C-lobe is mainly composed of alpha helices. The two lobes are connected by a linker that includes hinge region and convex-shaped motif (usually 6-8 residues) (Figure 3.1a) [53, 54]. In the N-lobe, there is the flexible Glycine-rich **P-loop** (also called G-loop), which is usually composed of the sequence [Gly-X-Gly-X-X-Gly]. The P-loop adopts different conformations depending on the state of the kinase and on the nature of the bound ligand. The first Gly position in that motif could be glycine, alanine, or serine, while the second (Gly) position is always glycine in all kinase families.

The catalytic cleft could be divided into the front cleft, which is the ATP-binding pocket, and back cleft, which contains important residues for kinase regulation. Between these two parts, there is the activation segment (**A-segment**), which is responsible for changing the conformation of the kinase. Both A-segment (in the C-lobe) and αC helix (in the N-lobe) are essential for regulating the kinase activity. The A-segment contains the DFG motif, the activation loop (**A-loop**), and other secondary structural elements. In a fully active state, the A-segment adopts an open conformation so that the A-loop takes a position away from the catalytic center, providing suitable pocket for substrate binding (compare Figures 3.1a and 3.1b) [53, 55].

The borders between the two clefts are usually composed of two beta sheets, which contain



(a) **(left)** Ribbon diagram of the x-ray crystal structure of c-Kit kinase (PDB id 1PKG): in the middle between the two lobes, there is ADP (carbons in green). **(right)** Overview of the kinase ATP-binding pocket (active/"DFG-in" conformation), highlighting regions that are relevant for small molecule inhibitor binding: **A**, adenine binding site/linker region; **R**, ribose binding region; **P**, phosphate binding/catalytic aspartate/salt bridge region; **Br_I**, back hydrophobic Binding region; **Br_II**, front specificity binding region; also used for solubilizing groups in type-I kinase inhibitors.



(b) **(left)** Ribbon diagram of the x-ray crystal structure of c-Kit kinase in complex with Imatinib (PDB id: 1T46): In the middle between the two lobes, there is the type-II inhibitor Imatinib (carbons in green). **(right)** Overview of the imatinib binding with (inactive/"DFG-out" conformation), highlighting regions that are relevant for small molecule inhibitor binding: **A**, adenine binding site/linker region; **R**, ribose binding region; **P**, phosphate binding region occupied by DFG motif; **Br_I**, extended back hydrophobic binding region after changing the orientation of the DFG motif; **Br_II**, front specificity binding region. **Br_III**, additional binding region in the allosteric site, used for solubilizing groups in type-II kinase inhibitors.

Figure 3.1: Ribbon diagrams of x-ray crystal structures of c-Kit kinase; with ADP (**top**), and with Imatinib (**bottom**), combined with overview of the binding regions in the catalytic cleft. The N-lobe is shown in wheat and the C-lobe in blue violet, they are linked by the hinge region in blue, the α C-helix in red, the P-loop in yellow, the A-loop in magenta, the DFG motif in orange, gatekeeper and catalytic lysine residues in light blue.

two important residues: a quite flexible and functionally important **conserved lysine** residue and a residue called the gate-keeper (**GK**). The gate-keeper residue is less conserved and could be a small residue like threonine or alanine, or a bulky one like phenyl-alanine (e.g. in CDK2), leucine (e.g. in GSK3 β), or methionine. The importance of the gate-keeper emerges from controlling the access to the back selectivity cleft; whose size mainly depends on the size of gate-keeper (Figure 3.2) [54, 56]. The conserved lysine usually makes salt bridge with the catalytic glutamate residue of the α C helix, and additionally helps anchoring the alpha- and beta-phosphate groups of ATP.

The front cleft consists of the ATP-binding pocket and relatively small non-ATP regions. The ATP-binding pocket provides the aromatic ‘adenine’ region for the adenine, beside of additional regions for the ribose and the phosphate, donated respectively **A**, **R**, and **P**. The hydrophobic adenine binding region (**A**) is surrounded on one side by the gate keeper residue and the DFG motif, while on the other side adjoins the hinge region, which provides the traditional two hydrogen bonds with the ATP’s adenine. The ribose region (**R**) consists of hydrophobic residues shared with the adenine pocket and other residues from the P-loop, very close to the hydrophilic and solvent-exposed entrance of the binding pocket. The phosphate region (**P**) is located between the P-loop and the catalytic loop (activation loop), and it is highly flexible, hydrophilic, and solvent-exposed especially in the active state. As the ribose pocket is not very conserved among kinases, it could be used for gaining selectivity and optimizing the binding affinity. Additionally, there is a ‘front selectivity’ binding region (**Br-II**) or the solubilizing region; which is a small hydrophobic region between the hinge region and the ribose region, usually not occupied by ATP, often serves as an entrance for ligand binding, and could be used to gain selectivity because of its diversity in sequence and conformation (Figure 3.1a) [53, 54, 56].

In the back pocket, three or four sub-pockets, which are hydrophobic and make no ATP-contacts, can be recognized. The hydrophobic ‘back selectivity’ binding region (**Br-I**) is adjacent to the adenine pocket in the active conformation, and its size is dependent on the gate-keeper residue. Therefore, Br-I is quite often used for getting better selectivity for kinases with small-size gate keeper (Figure 3.2). The second hydrophobic binding region (labelled **Br-III**) locates behind the DFG motif on the opposite side of the ATP binding pocket in the DFG-in conformation. Therefore, this pocket is not accessible for ligand binding in the DFG-in conformation, but it is connected to the ATP binding pocket after the change of the DFG orientation and the conversion from DFG-in to the inactive DFG-out conformation (Figure 3.1b) [53, 57, 58].

3.2 Types and Selectivity of Protein Kinase Inhibitors

The major problem of finding selective small-molecule kinase inhibitors is the high conservation of the ATP-binding catalytic domain. Typically, protein kinases could be found in two functional states: active and inactive. The position of the conserved motif DFG (or rarely D-[LWY]-G) in the kinases' activation loop is used to distinguish between these two functional states. The most reported state of crystallized kinases is the active state, which is called the DFG-in conformation and characterized by the DFG's aspartate orientation inside the ATP catalytic domain. The majority of the kinases' crystal structures in the Protein Data Bank represent this conformation of kinase, which is bound to protein substrate and the ATP molecule with the help of the chelating Mg^{+2} ion [56, 57].

In the decade of nineties, the main efforts in the academic and industrial research for developing kinase inhibitors concentrated on mimicking the ATP binding to the kinase catalytic domain. These efforts resulted in a good number of diverse scaffolds, which competitively inhibit the ATP binding in a low concentration [59, 60, 61]. This first generation of kinase inhibitors, which targets the kinases in the **DFG-in active conformation**, has been called **type I kinase inhibitors**. The common feature of type I kinase inhibitors is their mimicking of ATP binding by conserved hydrogen bonds to the hinge region and less frequently hydrogen bonds with DFG motif or the flexible conserved lysine residue. In most cases, type I kinase inhibitors have major problem of cross-reactivity and lack of selectivity, as they inhibit broad range of kinases [57, 62, 63, 64]. A better selectivity profile could be obtained, when the kinase inhibitor is designed to use the back selectivity binding region (Br-I) for making more protein-inhibitor interactions; then they are called **type II/2 kinase inhibitors** (Figure 3.2).

Later, a new conformation of some kinases, characterized by a closed conformation of the activation loop, was identified and called **the inactive DFG-out conformation**. The closed conformation of the activation loop prevents binding of both the cofactor 'ATP' and the protein substrate. In the inactive state, the DFG motif is recognized by special positioning of its residues, resulting in a 'DFG Asp-out' conformation. In this last conformation, the aspartate residue points out of the ATP-binding pocket, while the phenylalanine residue occupies the place of the phosphate region inside the ATP-binding pocket (Figure 3.1b). The last conformational change provides additional hydrophobic pocket, only available in the inactive conformation, and could be used for obtaining higher selectivity and discovering wider chemical space for inhibiting the protein kinases [65].

The first small molecule inhibitor, which targeted this inactive DFG-out conformation, was Imatinib (GleevecTM). Imatinib is considered the lead of the second generation of kinase inhibitors, called **type II kinase inhibitors**, showing good selectivity profile by inhibiting only three kinases (Abl, c-Kit, and PDGFR). Imatinib was approved by the FDA as first kinase in-

hibitor for the treatment of chronic myeloid leukemia (CML). Since then, Imatinib has been widely used in the targeted therapy of CML (especially Philadelphia-chromosome-positive chronic myeloid leukemia), GIST, and mastocytosis, as it is potent specific inhibitor of the signaling kinases related to these tumors (Bcr-ABL, Kit, and PDGFR) [65, 66].

Till 2009, only eight kinase inhibitors have been approved by the FDA for cancer treatment; four of them belong to type I1/2 kinase inhibitor class (Erlotinib, Gefitinib, Dasatinib, and Lapatinib), while the other four belong to type II class (including: Imatinib, Sorafinib, Nilotinib, and Sunitinib). More type II multiple kinases inhibitors were later approved. Axitinib, Masitinib, Pazopanib, and Toceranib were approved by the FDA as c-Kit/VEGFR inhibitors for usage for different kinds of tumors. Other approved inhibitors for VEGFR kinase family include: Cediranib, Regorafenib, Semaxanib (SU5416, another Indolinone derivative) and Vandetanib (VEGFR/RET/EGFR inhibitor) [67, 68]. Bosutinib (approved in 2012 as Bcr-Abl/Src kinases inhibitor; including all Src family members: Src, Lyn and Hck) [69], Lestaurtinib (Staurosporin-related, Flt3/JAK2 inhibitor), and also Ruxolitinib (JAK1/2 kinase inhibitor) [70] have been recently approved.

The first inactive DFG-out conformation was reported for insulin receptor kinase (IRK) and later for Abl, p38 α MAPK, b-RAF, FLT3, Kit, LCK, HCK, KDR, TIE-2, Aurora-A, FMS/CSFR, MET, Src, Pyk2, and CDK6. In the case of type-II inhibition, it was firstly proposed that the inactive DFG-out conformation is specific to only few members of kinases, which are characterized by the existence of a small gate-keeper residue. However, inactive DFG-out conformation was later recognized for TIE and MET kinases in spite of a medium size gate-keeper. One of the difficulties is that the majority of reported kinase crystal structures represent the DFG-in conformation (representing 70% of the mammalian kinome crystal structures) rather than DFG-out conformation (only 3%), while 22% represents an intermediate conformation [55].

What is important to notice is the fact, that there is no exclusive linkage between the inhibitor type and the kinase conformation that it binds to. X-ray crystallography showed the type-II kinase inhibitor imatinib with un-phosphorylated/DFG-in active confirmation of SYK kinase in a cis-binding mode making two hydrogen bonds with the hinge region (PDB id: 1xbb) [71]. However, Imatinib obtains its selectivity from its higher binding affinity to DFG-out conformation, as its inhibitory activity against phosphorylated active kinases are always weaker (K_i for Imatinib inhibition of phosphorylated Abl is 0.9 μ M, whereas against unphosphorylated Abl it is in the low nanomolar range; 0.014 μ M) [71]. **Type-I1/2 kinase Inhibitors** not only bind to catalytically active “DFG-in” kinases (dasatinib with ABL kinase; PDB id: 1m52), but it could also bind to catalytically inactive “DFG-in/alpha-C-helix-out” conformation (Lapatinip with EGFR kinase, PDB id 1xkk), and could even bind to the “DFG-out” conformation if it makes hydrogen bonds with the hinge region and doesn't occupy the sugar/phosphate region

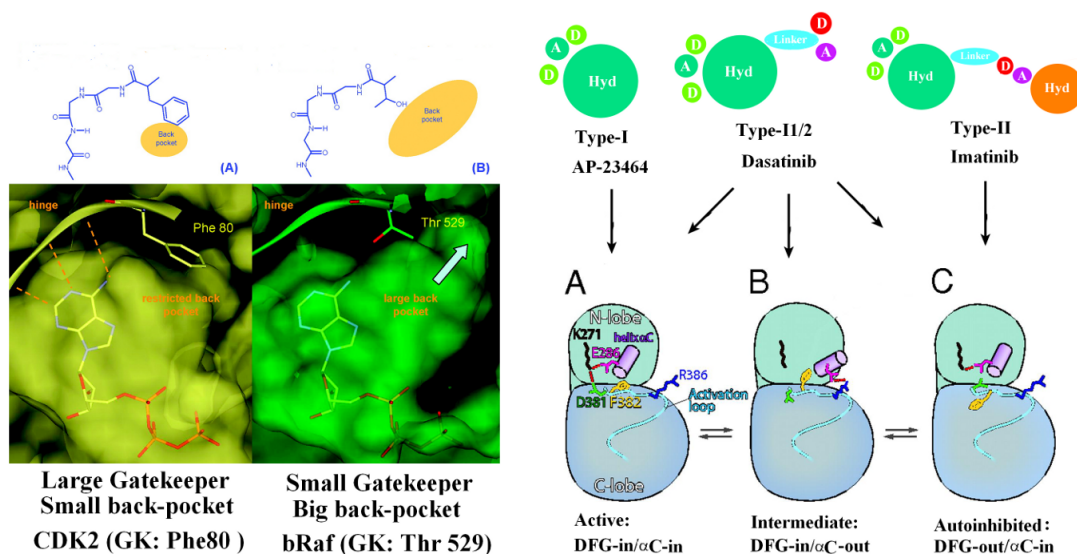


Figure 3.2: **(left)** Effect of gate-keeper on the size of back selectivity pocket **(right)** Type I kinase inhibitor pharmacophore fits only the DFG “in” conformation, while Type II kinase inhibitor pharmacophore fits only the DFG “out” conformation. Type II/2 pharmacophore represents the potential interactions at the hinge and in the back cavity Br-I, which all conformations; active and inactive and intermediate, can provide. (adapted from [54])

Table 3.1: Comparison of general properties of types: I, II, and III kinase inhibitors.

KI	Type I and II/2	Type II	Type III
Activation state	Active or inactive	inactive	Active or inactive
Require DFG-out conf.?	No	Yes	No
Phosphorylation state- sensitive	Usually not	Usually sensitive	Usually not
Binding site	ATP (+Br-I)	ATP + allosteric (Ext. Br-I)	Allosteric, far from ATP
Hinge binding	Yes	Possible, Not required	No
ATP-competition	Yes	Yes, indirectly	No
Apply to every kinase	Yes	Depends on the stability of DFG-out conf.	No, very few kinases
Selectivity	Usually low, better selectivity for II/2	High selectivity, Allosteric site provide a key for improvement.	High selectivity
Developed resistance	Mutations inside ATP-pocket only	More mutations: ATP-pocket, P-loop, A-loop, and Juxtamembrane	Not reported

(the location of DFG's phenylalanine in the case of the DFG-out conformation). An example is **SB203580**; a p38 MAP kinase inhibitor that binds equally to both DFG-in and DFG-out conformations [72, 73, 74]. This new type of kinase inhibitors, which bind to the two conformations, was called type I1/2 kinase inhibitor (Figure 3.2).

Type I1/2 KI emerged as hybrid combination of the two types, making the same traditional hydrogen bonds with the hinge region and extending into the back selectivity region **Br-I** (in the active conformation DFG-in); establishing interactions with some residues, usually involved in type-II inhibitors binding. The size and the shape of that back selectivity hydrophobic pocket depend mainly on the size of the gate-keeper residue. By targeting the kinase with small gate-keeper, it is possible to use the large back selectivity cavity of the ATP-binding pocket to get more interactions, allowing for more novelty in the chemical space and better affinity and selectivity. Another type of kinase inhibitors is **type III kinase inhibitors**, which are actually allosteric inhibitors that don't target the ATP-binding pocket, but other allosteric sites far from ATP-pocket. Therefore; they are non-ATP-competitive and usually they show high selectivity, because the allosteric sites are highly specific and found only in some kinases (Table 3.1).

3.3 Binding Mode of Kinase Inhibitors and Overcoming the Mutation-induced Resistance

It is well reported that many tumor types are actually associated with persistent activation of kinases, either by over-expression of the kinases themselves or their ligands, by autocrine loops, or by activating gain-of-function mutations. The most occurring molecular event is actually the activating gain-of-function mutations in the kinase domain, which could be divided to the two types: deletions/mutations in the juxtamembrane domain, and mutations residing close to the activation loop. The difference between these two types of the mutations is mainly the different sensitivity of the mutant kinases to the inhibition by different kinase inhibitors. Ongoing research revealed that both kinds of gain-of-function mutations, which reside in the kinase domain, are able to cause a complex effect on the kinase conformations, reactions, and its affinity to the natural ligand/cofactor or other discovered inhibitors. One of the most important effects of kinase domain's mutations is changing the dynamic balance between the different kinase conformations (Active vs. Inactive); leading to the dominance of the active conformation (Figure 3.3) [58, 75, 76, 77].

Another important conformational change occurs in the phosphate-binding loop mutations (P-loop's mutants) [58]. The P-loop conformation also plays a significant role in determining the kinase binding pocket's shape and its dynamics [78, 79]. Mutations of P-loop and the activation loop (A-loop) often destabilize the inactive conformation (DFG-out) in favor of the active conformation (DFG-in).

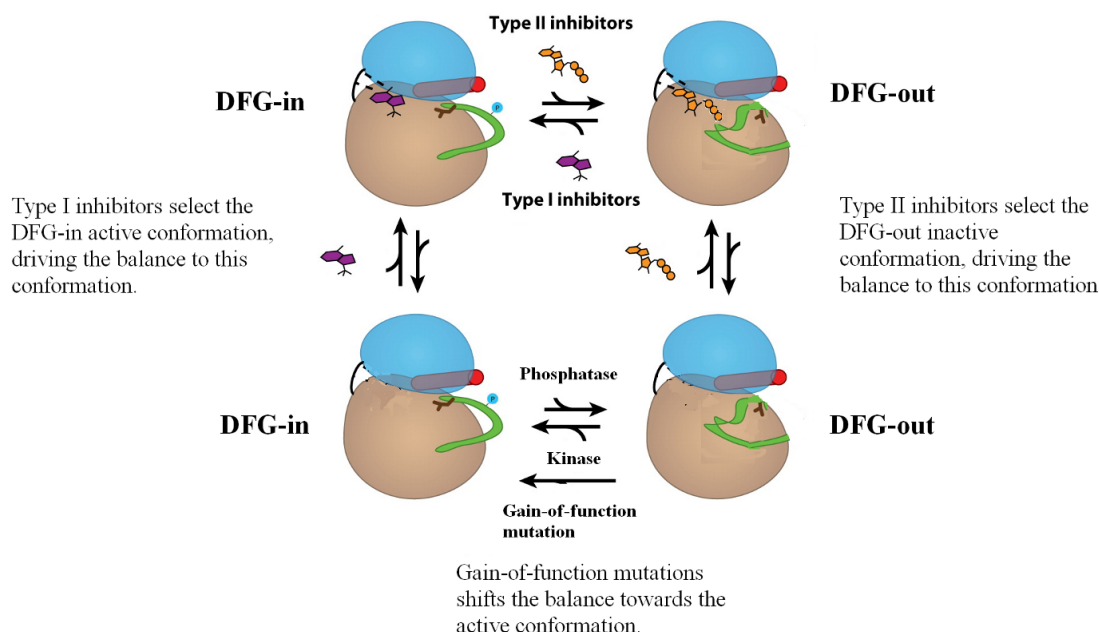


Figure 3.3: Schematic representation of the dynamic equilibrium between active, inactive, apo and type-I and type-II inhibitor-bound kinase conformations. The activating mutations usually shift the dynamic balance in the direction of the active conformation.

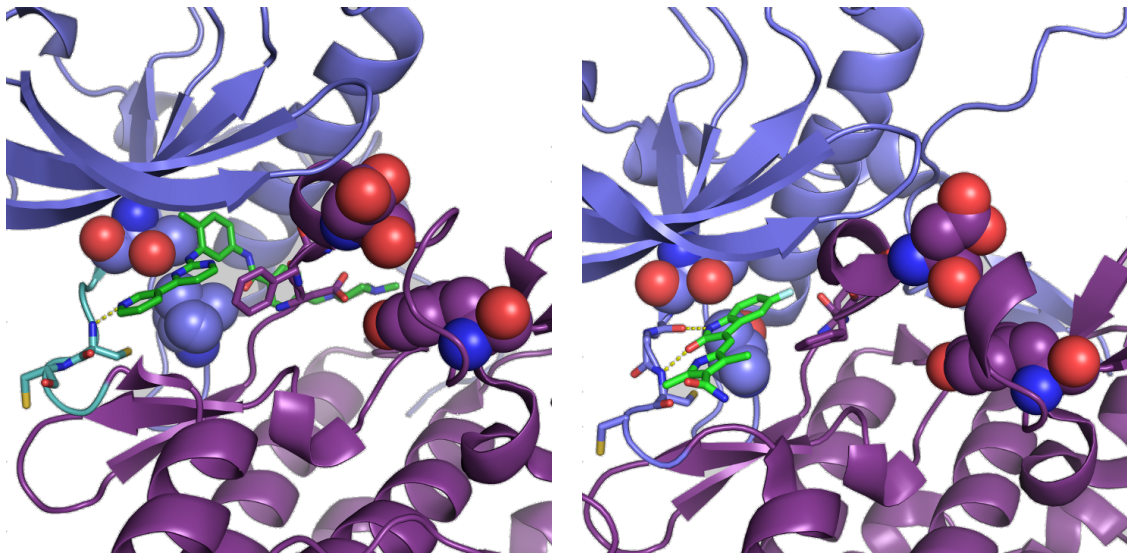
Some other types of mutations have been reported to cause resistance against some developed kinase inhibitor (e.g. imatinib), as a result of losing of a hydrogen bond or steric clash effect. The size of the “gatekeeper residue” appeared to be quite important in determining the size of the kinase active site, and subsequently its mutation is a main cause for developing a resistance. An example of the gatekeeper mutation is the mutation T670I/E and T315I/F/D/N in c-Kit and Abl kinase respectively [76, 78].

Type-II kinase inhibitors have usually better selectivity profile obtained by their preferential binding to the DFG-out conformation, and consequently their potent inhibition activity is restricted to only few kinases [65, 66, 80]. However, they are also affected by many resistance-conferring mutations; including the gate-keeper mutation, P-loop mutations, and gain-of-function mutations. Developing resistance to the therapy by **Imatinib** was reported in some cases and studies, and was attributed to some mechanisms; such as Imatinib’s extracellular sequestration by P-glycoprotein active transport, the compensation for the Imatinib-inhibited signaling pathway by another kinase pathway (Src-kinases-regulated signaling pathway), or possibly by some point mutations, which affect the Inhibitor-kinase binding **directly** or **indirectly**; by changing the protein/inhibitor interactions inside the binding pocket or by changing the kinase’s conformational balance (Figure 3.4a).

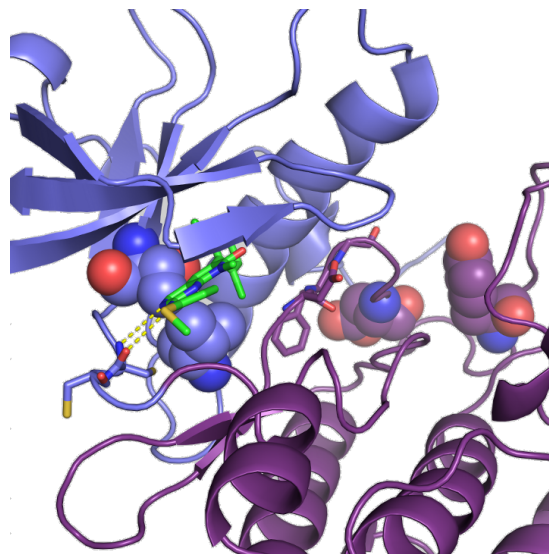
It is important to differentiate between two kinds of resistance to therapy by Imatinib; an ‘early’, ‘innate’, ‘primary’ resistance and a ‘delayed’, ‘acquired’, ‘secondary’ resistance. The primary resistance to the therapy with kinase inhibitor appears in 3-6 months from the start of the therapy, while the secondary resistance develops after the response to the therapy [76, 78, 81, 82]. Most of the early resistance cases of GIST therapy were reported with mutations in exon 9, while secondary resistance could be associated with mutations in exons 13 and 17, which could affect the binding pocket and the activation loop [81, 83, 84, 85].

Sunitinib binds to the DFG-out conformation of the kinases, making two hydrogen bonds with the hinge region; the NH and O of the dihydrooxaindole ring forms the donor-acceptor feature of the kinase inhibitor, while the phenyl ring of the DFG’s phenylalanine residue (F811 in c-Kit structure) contributes in making pi-pi hydrophobic interactions with the inhibitor. This interaction with the phenylalanine residue of DFG motif is not possible in the other DFG-in conformation, because the phenylalanine orients away outside of the binding pocket in this last conformation. Therefore, Sunitinib has low activity against the mutant forms of kinases with activating mutations; mainly D816 and Y823 in the case of c-Kit kinase [75] (Figure 3.4b). On the other hand, **Dasatinib** (as representative of type II/2 kinase inhibitor) occupies only the ATP-binding pocket, making double hydrogen bonds with the hinge residues and extending inside the back selectivity pocket. According to this binding mode, dasatinib binds close to the gatekeeper residue and it is affected by the gatekeeper’s mutation, but its binding affinity would not be affected by the activating mutations; such as the activation loop’s mutations or the juxtamembrane mutations, because it binds to both conformations [86, 87, 88] (Figure 3.4c). The previous three examples of kinase inhibitors suggest that there is a trade off between improved selectivity and overcoming mutation-based resistance.

Beside the problem of resistance-conferring mutations, the importance of the kinase inhibitor’s selectivity should be kept in mind. In most cases, the broad range of kinase inhibition (like the case of Dasatinib and Sunitinib) could be a reason of ‘off-target side-effects’, especially the off-target cardiotoxicity. The main important mechanisms of this cardiac toxicity could be listed as: Mitochondrial abnormalities observed in cardiomyocytes, inhibition of AMP-activated Protein Kinase (a known off-target), PDGFR inhibition (especially PDGFR β), inhibition of kinases that affect the vasculature (especially the VEGFR family), inhibition of cardiac ion channel inhibition (especially the hERG potassium channel), and also the interaction with adenosine receptors [89, 90, 91].



(a) **Imatinib** bound to the cKit DFG-out conformation (PDB id: 1T46). Imatinib is affected by two types of mutations; the binding pocket's mutations and activation loop's mutations. (b) **Sunitinib** bound to c-Kit DFG-out conformation (PDB id: 3F0G), but without using the allosteric pocket. It is affected by the activation loop's mutations, but not by binding pocket's mutations.



(c) **Dasatinib** docked to DFG-in cKit structure (PDB id: 1PKG) by aligning the last to the mutant c-ABL structure with Dasatinib (PDB id: 1M52); Dasatinib is not affected by activation loop's mutations, but is affected by binding-pocket's conferring mutations

Figure 3.4: Mutational hotspots in the c-Kit kinase: mutations conferring the ATP-binding pocket (T670 and V654; shown in blue) and the activation loop's mutations (D816 and Y823; shown in violet).

3.4 The Important Role of Water Molecules inside The ATP-binding Pocket

The binding process between a small-molecule ligand and a protein usually incorporates displacement of the water molecules, which fill the binding pocket of the protein satisfying the hydrogen donors/acceptors of the protein structure. This process of water displacement could have two energetic advantages: enthalpically favorable when a polar/charged group or hydrogen donor/acceptor of the ligand displace the water making alternative hydrogen bonds, and entropic gain as the released water molecule get more freedom in the bulk solvent rather being restricted inside the protein pocket [92, 93, 94]. Therefore, displacing water molecules from the binding pocket by a ligand is mostly considered an energetically favorable process, leading to entropically-driven higher binding affinity. However, many studies have showed the importance of stable structural water inside the binding pockets. These stable structural water molecules should maintain 3 to 4 hydrogen bonds in order to keep the same number of hydrogen bonds, which a water molecule can usually make in the bulk phase [92, 93]. That is usually considered as compensation between the enthalpy gain and the entropy penalty.

In a case of a strongly bound water molecule inside a decoy kinase and a weakly bound water molecule inside a targeted kinase in identical locations, removing the water by a designed inhibitor would give it higher affinity with the targeted kinase than with the decoy kinase [95, 96, 97, 98, 99, 100, 101]. However, it is not an easy task to determine the thermodynamic stability of water molecules inside the protein binding pockets, as it is important to give special care for the enthalpy-entropy compensation [98, 99, 100, 102, 103]. Water molecules could also play bridging roles in the protein-ligand interactions, and these water bridges could also be thermodynamically favorable or unfavorable depending on the targeted system [92, 104, 105, 106]. Rigorous free energy methods (e.g. FEP and TI) have been used to estimate the thermodynamic contribution of the structural water molecules to the binding process [107, 108, 109, 110], in spite of their high sensitivity to the details of computation [103]. The performance of the implicit solvent binding energy methods (PBSA/GBSA) have been investigated and reviewed in numerous papers, showing different degrees of success and failure depending on the studied enzyme/inhibitor systems [111, 112, 113, 114, 115]. One of the problems of implicit solvent models is their incapability to account for the role of water molecules inside solvent-exposed binding pockets like the ATP-pocket in the kinases.

In the absence of a crystal structure of the kinase-inhibitor complex, one of the critical issues for performing precise free energy calculations is the right prediction of the water content and location sites inside the binding pocket, especially if these hydration sites locate deep inside the binding pocket and participate in mediating the protein-ligand interactions [116, 117, 118].

Some computational algorithms have been suggested to predict the location of the potential hydration sites inside the binding pocket and then to determine their occupancies and thermodynamic properties [98, 116, 119, 120, 121].

A water placement algorithm called JAWS has been developed by Michel and Jorgensen *et al.* for predicting the hydration sites in the protein-ligand complexes [98]. This method has been applied on a series of 17 inhibitors of p38 MAP kinase, previously reported by Pearlman and Charifson [122]. A previous application of MM-PBSA method was performed by D. Pearlman showing that this series is a challenging case study for many scoring functions [114]. The application of a very rigorous free energy method, namely Monte-Carlo/free energy perturbation (MC/FEP) simulations, was applied to this series of p38 kinase inhibitors, and was only able to correctly predict the affinities after using JAWS-derived water distributions [99].

Another approach is the inhomogeneous solvation theory, which was proposed by Lazaridis [123, 124], and was applied to different systems of protein/ligand complexes. Later, the inhomogeneous solvation theory was used in developing a method called WaterMap (Schrodinger, Inc.), applied in different examples to explain the selectivity of some inhibitors of FXa, HIV-1 protease, PDZ domains, and different kinases [125, 126, 127, 128, 129]. This approach performs both prediction of the water locations (hydration sites) and their thermodynamic properties (enthalpy and entropy). In a study comparing the hydration sites thermodynamics between SRC and GSK3 kinases, it was shown that the water molecules located close to the hinge are similar in the energetic aspects, while there are more differences in the location and thermodynamic properties of the water molecules in the deep selectivity pocket [128]. While the WaterMap method correctly describes the hydrophobic interactions and the hydrophobic effect resulting from displacing the solvent by the ligand, WaterMap doesn't account for binding energy terms, such as protein-ligand electrostatic interactions, or ligand-based terms like intramolecular strain and desolvation penalty. Some publications suggest combining the WaterMap method with a continuum electrostatics method like GBSA [130, 131]. Other theories for the solvation have been developed depending on integral equation method; such as 3D-RISM [132], which was later applied in AMBER package.

A. Fernandes has introduced the concept of 'dehydron', defined as "Water-exposed intramolecular hydrogen bonds in native folds of the proteins" [133, 134]. These dehydrons represent wrapping defects in the protein, which favor the removing of the surrounding water to enhance the electrostatic interaction in that area. He later used the dehydron theory to reengineer the kinase inhibitor imatinib by adding a methyl group on the pyridine ring. It was hypothesized, that the small methyl group could efficiently dehydrate the hinge residues C673 and G676 in c-Kit, making the modified inhibitor (called WBZ_4) more selective toward c-Kit and less active for Abl kinase, which is responsible for the cardiotoxicity of the inhibitor [135]. This observation gave new insight about the importance of hydration sites in the binding

pocket for determining the binding affinities and explaining the selectivity between the different kinases.

3.5 Computational Methods and Structure-based Design of Kinase Inhibitors

Selective inhibition of kinases is now well established as a new paradigm in targeted therapy of cancer. The selectivity of the developed kinase inhibitors is a crucial feature, as the inhibition should be restricted to one or two kinases in the targeted signaling pathway for the considered kind of tumor. Another challenge in developing selective kinase inhibitors is the capability of the tumor cells to develop acquired resistance during the therapy.

After the emergence of the highly-selective type-II kinase inhibitors, the major problem of identifying this type of kinase inhibitors is the application of traditional assay methods and high-throughput screening, as the traditional assay methods mainly depend on the activated phosphorylated kinases. To overcome this problem, new inhibition assays should be developed depending on phosphorylation-state-independent binding assays [74], competition binding to immobilized probes, or temperature-dependent unfolding of the protein [136, 137]. However, the high expense of these kinds of bioassays would make the high-throughput screening method very expensive and impractical. The computational structure-based methods and QSAR-guided modifications have proved to be helpful in the development of type-II kinase inhibitors [55]. Computational methods and binding energy approaches are helpful to get insight into the inhibitor-kinase binding process, to predict the conformational changes, and to generate predictive models for ranking and estimating the binding affinity of novel compounds. Two challenges still exist when we are dealing with flexible targets like kinases:

A) Can we predict the ‘actual binding mode’ of novel series of kinase inhibitors, without having an x-ray crystal structure of the kinase/inhibitor complex?

B) Can we estimate the binding affinities or obtain the right ranking for novel kinase inhibitors using the computational free binding energy methods?

In the next chapters, we will analyze the performance of two end-point free binding energy methods applied to two cases of kinase inhibitors, for predicting the binding mode in the first case study, and for predicting and ranking the inhibitors according to their estimated binding affinities in the second case study.

Chapter 4

Predicting The Binding Mode for Kinase Inhibitors: 1-Aza-9-Oxa-Fluorene Derivatives as GSK3 β /CDK2 Inhibitors

4.1 Introduction

4.1.1 Glycogen Synthase Kinase 3 and CDK2 Kinase

Glycogen synthase kinase-3 β (abbreviated as GSK3 β) or tau-phosphorylating kinase is one of the serine/threonine kinases, which was first discovered for its role in glycogen synthesis. Later, GSK3 β was also identified to play an important role in the neurons of the brain [138, 139, 140, 141]. GSK3 β phosphorylates tau, which is a microtubule-associated protein, leading to the disassociation of this protein from the microtubules [139, 142, 143]. The detached tau protein is the essential component that leads by aggregation to the formation of neuro-fibrillary tangles (NFT), which is one of the remarkable pathological hallmarks in the brains of Alzheimer disorder (AD) patients. Increased levels of activated GSK3 β were observed in the brains of Alzheimer patients [143, 144, 145, 146]. GSK3 β was also linked to the aggregation of PolyQ protein, which is considered one of the hallmarks of another neurodegenerative disease, namely Huntington's disease [139, 147].

Additionally, GSK3 β inhibits the insulin signaling by phosphorylation of the insulin receptor substrate proteins IRS1 and IRS2, while insulin stimulation, in turn, inhibits GSK3 β by activating Akt kinase (PKB), which phosphorylates GSK3 β at a conserved N-terminal serine [146]. The involvement of GSK3 β in glycogen metabolism and insulin signaling could result

in an increased resistance to insulin and developing of diabetes type II [145, 146]. In a variety of recent studies, GSK3 β over-expression and activation were noticed and associated with cancer progression [145, 148]. GSK3 β is also involved in stem-cell renewal, cell-division cycle, differentiation, apoptosis, circadian rhythm, transcription, and insulin action [145, 149, 150].

Cyclin-dependent kinases (CDKs) are one of the key regulators of the cell cycle; therefore they became important pharmaceutical targets for the discovery of anti-proliferative drug candidates [151, 152, 153]. Their activity is dependent on the binding to other partner proteins called cyclins. CDK2 is one of these kinases, whose complex with cyclin E forms a restriction point at S-phase checkpoint, while its complex with cyclin A is necessary for the completion of phase S, controlling G1- to S-phase checkpoint [153, 154].

Both GSK3 and CDKs kinases belong to the CMGC class of serine/threonine protein kinases sharing high homology and sequence similarity, especially within the ATP binding pocket. This homology would explain why many CDK2 inhibitors are also potent inhibitors of GSK3 β kinase [149, 155, 156]. One challenge of the kinase research is optimizing the selectivity for one of these two kinases (CDK2 or GSK3 β) over the other. One concern is that the involvement of GSK3 β in Wnt and hedgehog signaling pathways would make the correct assessment of CDK2 inhibition effects quite difficult [148]. Another reason for developing novel selective inhibitors is that GSK3 β is considered a potential target for treating Alzheimer's disease (AD) [147, 157].

4.1.2 GSK3 β /CDK2 Kinase Inhibitors

Some of the well-known pan-kinase inhibitors were co-crystallized with GSK3 β kinase. The non-hydrolysable ATP-analog (AMP-PNP) is a famous pan-kinase inhibitor, which binds to many kinases making similar interactions with the ATP-binding pocket residues. The structure of GSK3 β was resolved with AMP-PNP showing a binding mode, where the adenine ring makes hydrogen bonds with the hinge region residues, D133 and V135, and hydrophobic interactions with I62, V70, A83, V110, L132, Y134 and L188. Other important functional residues, such as the conserved lysine K85 and the α C helix's glutamate E97, also interact with other parts of the inhibitors. The glycine-rich P-loop interacts with the phosphate groups through bridging water molecules protecting them from the bulk water. Another example of pan-kinase inhibitor is the natural product Staurosporine, which is a potent ATP-competitive kinase inhibitor for the vast majority of the kinases. Staurosporine mimics the binding of ATP by two conserved hydrogen bonds with the hinge residues D133 and V135 [158].

Among the selective inhibitors of both GSK3 β and CDK2 kinases, alsterpaullone (9-nitro-paullone) is the most potent GSK3 β inhibitor of the paullones series (GSK3 β IC₅₀ value of 4 nM) and also potent inhibitor of CDK1/2/5. The interactions between the alsterpaullone

and the kinases are composed of two direct hydrogen bonds with the hinge residue V135 and a water-mediated interaction with D133. Two indirubin derivatives; indirubin-3'-monoxime (IC₅₀ against GSK3 β 22 nM) and 5-iodoindirubin-3'-monoxime (IC₅₀ against GSK3 β 9 nM), have been recognized as the most potent GSK3 β inhibitors among the series of indirubins (IC₅₀ values against GSK3 β in the 5–50 nM range and against CDKs in the 50–100 nM range)[158] (Figure 4.1).

The ATP binding pockets of these two kinases were compared with the intention of highlighting potential regions for gaining selectivity for GSK3 β versus CDK2 (Table 4.1). Some of the significant differences between GSK3 β /CDK2 sequences that could be used for developing selective inhibitors are the gatekeepers' change L132 (GSK3 β)/F80 (CDK2), the residue opposite to gatekeepers M101 (GSK3 β)/L55 (CDK2), and inside the binding pocket C199 (GSK3 β)/A144 (CDK2). These differences change the entrance size to the back selectivity pocket (Br-I), and also the inside surface of the ATP-binding pocket. The inner part of ATP pocket becomes more flat in the case of GSK3 β and more hilly (more hills and valleys) in the case of CDK2. Another structural difference in the sequence and structure at the entrance of the ATP pocket is the sequence P136-E137-T138-Y140-R141(in GSK3 β)/H84-Q85-D86-K88-K89 (in CDK2), which is determinant for defining the boundary region of the ATP pocket (Table 4.1 and Figure 4.3). A salt bridge (E137-R141), which is noticed in GSK3 β , is not observed in CDK2. In CDK2, K89 is pointing inwards to the ATP pocket interacting with residue I10 from the P-loop. These last differences contribute to the formation of a different opening in the ATP-pocket's entrance at the front selectivity region (Br-II) (See Figure 4.3) [159].

The first selective GSK3 β inhibitor was reported in 2003 by J. Avila *et al.* and named AR-A014418 [159]. The GSK3 β -selective inhibitor AR-A014418 is a thiazole derivative, which is an ATP-competitive kinase inhibitor showing specific inhibition against GSK3 β in a panel of 26 kinases. AR-A014418 was co-crystallized with GSK3 β showing three hydrogen bonds with the hinge residue V135. The nitro group attached to the thiazole is able to make polar interaction with the known salt bridge Lys85-Glu97, mediated by water molecules. The aromatic thiazole ring makes hydrophobic interactions with the hydrophobic residue L188, while the attached phenyl ring can make hydrophobic interactions with I62 from the glycine-rich P-loop. The differences in the entrance region of the ATP pocket could be used for explaining the selectivity of AR-A014418. The side chain of K89 in CDK2 would prevent the inhibitor's interaction with residue I10, which corresponds to I62 in GSK3 β . The residue change T138 (GSK3 β)/D86 (CDK2) also prevents the good fitting of the phenyl ring in that area [159, 160] (Figure 4.2).

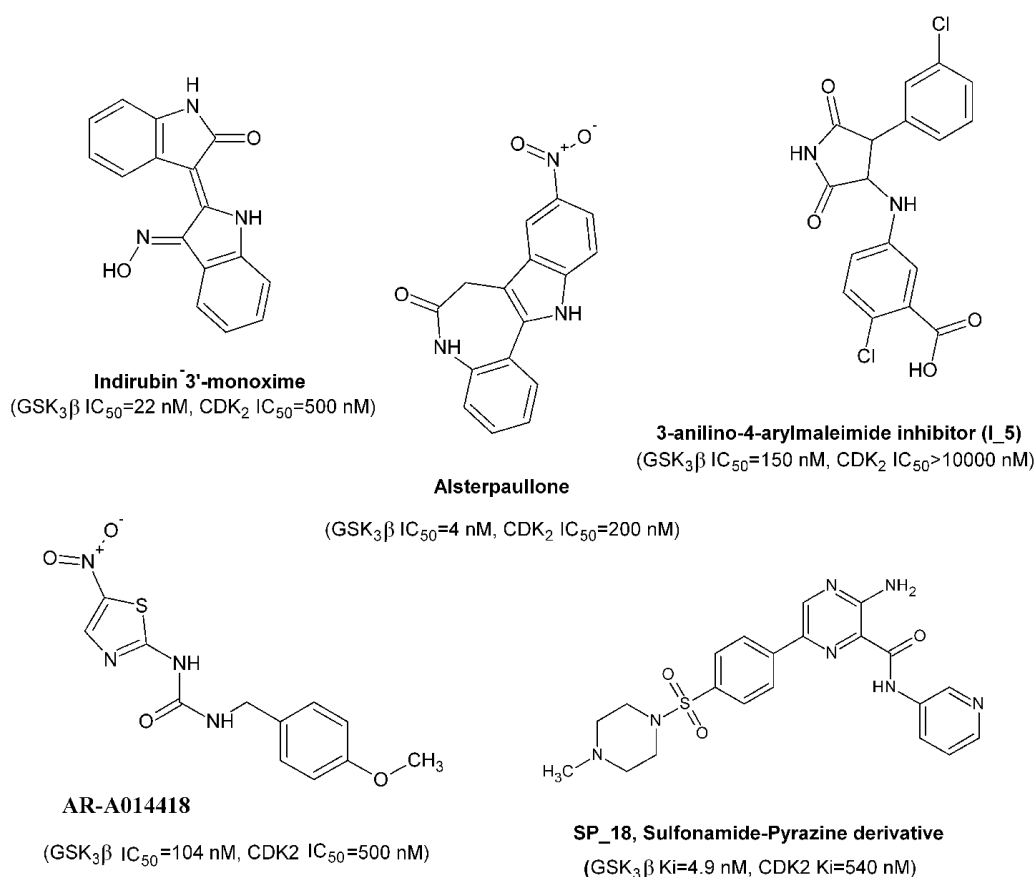


Figure 4.1: Chemical structures of some reported GSK3β/CDK2 kinase inhibitors.

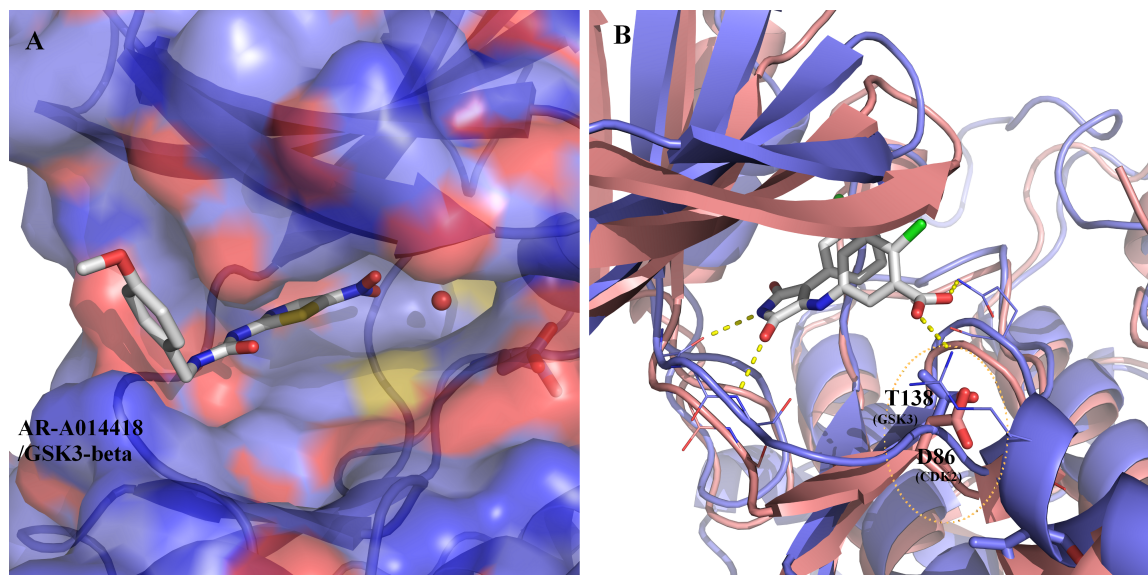


Figure 4.2: A) Binding of **AR-A014418** to GSK3β. its selectivity for GSK3β inhibition emerges mainly from its fitting to the molecular surface of GSK3β kinase. B) Inhibitor **I-5** bound in GSK3β active site (colored in blue), aligned with CDK2 active site (colored in pink). The residue change T138(GSK3)/D86(CDK2) explains the selective inhibition of GSK3β by **I-5**; as its carboxylate is incompatible to bind close to CDK2 aspartate D86.

A series of sulphonamide-pyrazine derivatives was also reported as potent GSK3 β inhibitors with 100-fold selectivity versus CDK2 for some derivatives (**SP-18**) [160]. GSK3 β and CDK2 were both co-crystallized with a piperazine-sulphonamide derivative **23** (GSK3 β K_i = 20 nM, CDK2 K_i = 10 nM); (PDB id: 4acc and 4acm respectively) (Figure 4.3). The comparison of the two crystal structures shows that the non-planar part of the sulfonamide-pyrazines is located at the hilly region of CDK2's ATP pocket. The selectivity of these piperazine-sulphonamide derivatives is explained by the shape complementary; the less hilly and flat ATP site's surface of GSK3 β is more able to leverage the non-planar scaffold of sulfonamide-pyrazine derivatives [160].

Another inhibitor; called I-5 (a 3-anilino-4-arylmaleimides derivative), was found to be a selective inhibitor for GSK3 β , FGFR-1, and VEGFR-3, while it doesn't inhibit CDK2 [158]. Beside the usual hydrogen bonds with the hinge region, the carboxylate group makes salt bridge with the residue R141 in GSK3 β and another water-mediated hydrogen bond with residue T138. T138 in GSK3 corresponds to N571 in FGFR-1 and D86 in CDK2. That could explain the unfavorable binding of I-5 with CDK2, as the similar negatively charged groups of D86 and the inhibitor's carboxylate repel each other (Figure 4.2).

4.1.3 The 1-aza-9-oxafluorene Derivatives

Novel 1-aza-9-oxafluorene derivatives have been synthesized at the Institute of Pharmacy (Martin-Luther-University Halle) in the group of PD Dr. A. Hilgeroth. Different substituents at the position **3**, **5**, **6**, and **7** produce inhibitors with variable potency of inhibiting GSK3 β /CDK2 kinases [161, 162, 163]. The main substituents are substituent R1 on location **6**, and R2 on location **3**, location **4** is always substituted by a phenyl (Table 4.2).

Table 4.1: Comparison between ATP Pocket Residues of CDK2 and GSK3, non-conserved residues are marked in bold

GSK3	K60	I62	G63	N64	G65	S66	F67	G68	V69	V70
CDK2	E8	I10	G11	E12	G13	T14	Y15	G16	V17	V18
GSK3	Q72	L81	A83	K85	E97	M101	V110	L112	L130	L132
CDK2	K20	V29	A31	K33	E51	L55	V64	L66	L78	F80
GSK3	D133	Y134	V135	P136	E137	T138	Y140	R141	K183	P184
CDK2	E81	F82	L83	H84	Q85	D86	K88	K89	K129	P130
GSK3	Q185	N186	L187	L188	C199	D200	G202			
CDK2	Q131	N132	L133	L134	A144	D145	G147			

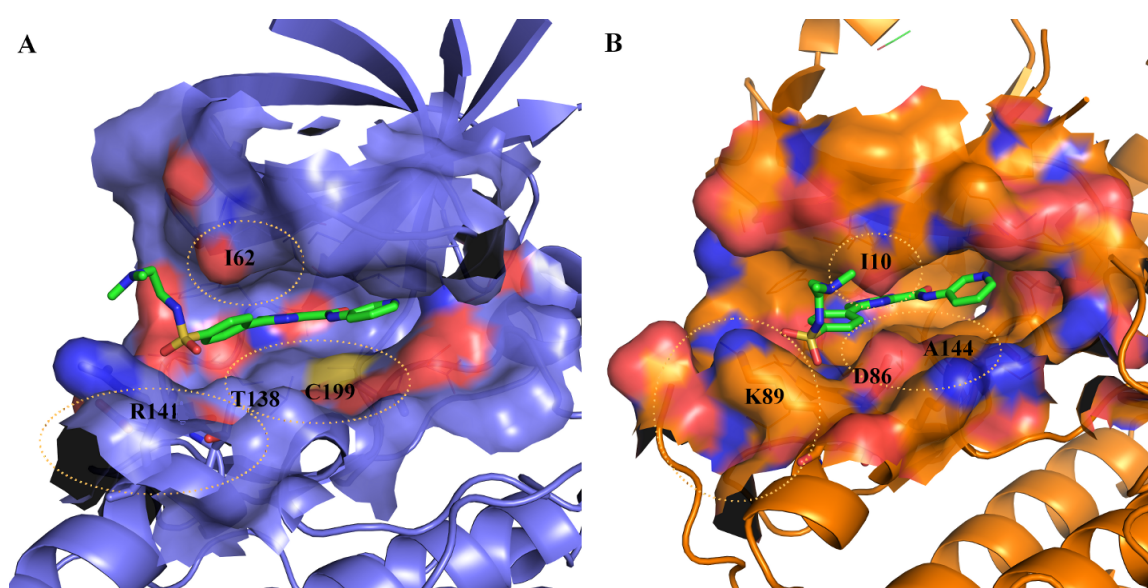
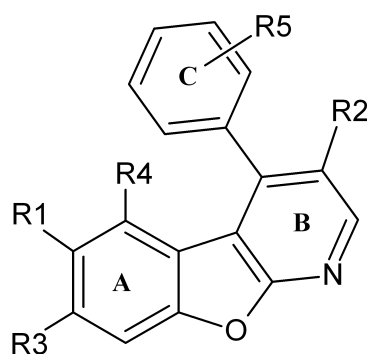


Figure 4.3: X-ray crystal structure of compound **SP-23** in the GSK3 β ATP site (A, 4acc.pdb, blue surface) and with CDK2 ATP site (B, 4acm.pdb, orange surface). Several residue changes have been marked to show their effect on the molecular surface; C199 (GSK3 β)/A144 (CDK2), T138 (GSK3 β)/D86 (CDK2), and R141 (GSK3 β)/K89 (CDK2).

Table 4.2: General Structures and binding affinities of the 1-aza-9-oxa-Fluorene derivatives.



Compound code	R1	R2	R3	R4	R5	GSK3- β Ki (μ M)	CDK2/Cyclin E Ki (μ M)
2a	OH	OMe	H	H	H	14.8	24
2a_2	OMe	OMe	H	H	H	n.a (≥ 1000)	147
2a_3	OH	OMe	H	H	4-Me	9.9 \pm 0.8	50
2a_4	OH	OMe	H	H	2-OMe	n.a (≥ 1000)	267
2a_5	OH	OMe	H	H	4-OMe	16.3 \pm 3.1	241
2b	OH	OBn	H	H	H	5.8 \pm 1.2	6.4 \pm 0.2
2b_2	OMe	OBn	H	H	H	0.02	n.a (≥ 1000)
2c	OH	OH	H	H	H	0.02	0.6
2c_2	OMe	OH	H	H	H	n.a (≥ 1000)	n.a (≥ 1000)
2d_1	OH	-CONH2	H	H	H	4.1 \pm 0.4	n.a (≥ 1000)
2d_2	OH	-CON(CH3)2	H	H	H	1.5 \pm 0.2	n.a (≥ 1000)
2d_3	OH	-CONH(CH3)	H	H	H	9	-
2d_5	OH	-CONH(C3H7)	H	H	H	4.4	-
2d_6	OH	CONH(C2H4)NH2	H	H	H	4.8	-
3a_2	OH	OBn	Me	H	H	5.8	-
3a_3	OH	OBn	OMe	OMe	H	5.4	-

4.1.4 Aim of this Study

Protein flexibility has a significant effect on docking experiments and the prediction of protein/ligand binding modes. In the case of flexible targets like kinases, it is quite common to notice that crystal structures of the same kinase complexed with different ligands show induced fit changes. In a virtual screening experiment, cross-docking using different crystal structures might be helpful for generating two or more clusters of docking solutions as possible binding modes. Often, fast conventional scoring functions, which are used in docking programs, are unable to choose the ‘actual’ or ‘correct’ binding mode. Here, we will compare the power and performance of two kinds of end-point physics-based scoring methods (MM-PBSA and Linear Interaction Energy) for choosing the most probable binding mode for a series of novel 1-aza-9-oxafluorene derivatives with two kinases; GSK3 β and CDK2. A detailed analysis of detected possible binding modes will be performed using MD simulations. Later, an analysis of the trajectories and the induced fit effects will be conducted, together with applying different binding free energy methods.

4.2 Computational Methods

4.2.1 X-Ray Structures of GSK3 β and CDK2

More than 25 x-ray crystal structures are available for the GSK3 β with varieties of substrates and inhibitors. Analyzing these crystal structures shows quite low RMSD when comparing the backbone of these different kinase crystal structures. Still, we can notice a large flexibility with the orientations of the side chains for some critical residues in the binding pocket: like K85, R141, and L132 (the Gatekeeper). Moreover we can notice differences in the conformation of the glycine-rich P-loop depending on the size and the binding mode of the bound inhibitor. Table C.1 in Appendix C presents a list of the available GSK3 β structures with the name of the bound ligand and the noticed distance between residue F67 (from the P-loop) and residue D186 (from the DFG motif), while Table C.2 presents an RMSD matrix comparing 10 of GSK3 β structures after performing superposition of the 3D-structures.

The published structure (1j1b.pdb) owns the best crystal resolution at 1.80 Å, representing the complex of GSK3 β kinase with the inhibitor AMP-PNP, while the other structure (1q3d.pdb) represents the complex of GSK3 β kinase with the natural inhibitor Staurosporine (crystal resolution at 2.20 Å). The RMSD between the two structures is 0.69 Å. The structure (1q3d.pdb) was chosen because it owns a special orientation of R141’s side chain outwards the binding pocket as a result of binding with large inhibitor (Staurosporine). The orientation of the residue R141

has an impact on the resulting docking solutions, as it will be explained in the results (4.3.1). The structure of CDK2 (2wih.pdb) represents a complex of CDK2 kinase with pyrazolo[4,3-h]quinazoline derivative (resolution at 2.50 Å). It is mainly chosen because it is a complex of CDK2 with a tricyclic-core inhibitor [164] (See Figure C.1 in Appendix C). The crystal structures of both kinases; GSK3β (PDB id: 1q3d and 1j1b) and CDK2 (PDB id: 2wih), were prepared using the software MOE 2011 by adding the hydrogen atoms using protonate3d protocol. For the further docking studies, the bound ligand and crystal waters were removed.

4.2.2 Molecular Docking

The 1-aza-9-oxafluorene compounds were docked into the crystal structure of GSK3β (PDB id: 1q3d and 1j1b) using GOLD 5.0, considering the Nitrogen NZ of the residue K85 as center of the binding pocket, while the radius of the binding pocket was set to 15 Å. For CDK2, the crystal structure (PDB id: 2wih) was taken for the docking using GOLD 5.0 considering the Nitrogen NZ of the residue K33 as the center of the binding pocket with radius of 15 Å.

In the case of GSK3β kinase, the used constraints were one protein hydrogen bond to one of the residues: the hinge residue V135 (in the hinge region), residue D200 (the aspartate of the DFG motif), or the flexible lysine K85. The used constraints for docking with CDK2 were one protein hydrogen bond to one of the residues: L83 (in the hinge region), D145 (the aspartate of the DFG motif), or the flexible lysine K33. All the docking solutions were later collected and clustered by calculating the rmsd from the top-scored solution combined with visual inspection, in order to obtain possible binding modes.

4.2.2.1 Scoring Functions

GOLD and Glide scores were used as scoring functions, applied on all obtained docking solutions; GOLD score is a force-field-like function with four (or five) terms:

$$\text{GOLD score} = S_{\text{hb ext}} + S_{\text{vdW ext}} + S_{\text{int tor}} + S_{\text{vdW int}} + (S_{\text{hb int}})$$

where $S_{\text{hb ext}}$ is the protein–ligand hydrogen-bond score, $S_{\text{vdW ext}}$ is the protein-ligand van der Waals score, $S_{\text{int tor}}$ is the ligand torsional strain energy score (internal torsion), and $S_{\text{vdW int}}$ is the contribution due to intra-molecular strain in the ligand [165]. $S_{\text{hb int}}$ is the contribution to the fitness due to intra-molecular hydrogen bonds in the ligand (switched off usually in the settings of GOLD, by default). GOLD score is taken as the negative of the sum of the component energy terms, so that larger fitness scores are better.

The Standard version of Glide score is developed starting from the empirical fitness function ChemScore (by Eldridge *et al.* [166]); which estimates the binding energy as follows:

$$\Delta G_{\text{bind}} = \Delta G_0 + \Delta G_{\text{hbond}} S_{\text{hbond}} + \Delta G_{\text{metal}} S_{\text{metal}} + \Delta G_{\text{lipo}} S_{\text{lipo}} + \Delta G_{\text{rot}} H_{\text{rot}}$$

S_{hbond} , S_{metal} , S_{lipo} are scores for hydrogen-bonding, acceptor-metal, and lipophilic interactions, respectively. H_{rot} is a score representing the loss of conformational entropy of the ligand upon binding to the protein. Glide score was developed using the same electrostatic, van-der-Waals, and lipophilic–lipophilic interaction terms (of ChemScore), while the hydrogen-bonding term of ChemScore is separated in Glide into differently weighted components (the donor and acceptor are both neutral, one is neutral and the other is charged, or both are charged) [167]. Glide score is trained to give a predicted estimation of the binding affinities.

4.2.3 Molecular Dynamics Simulation and Binding Energy Methods

The docking complexes of GSK3 β or CDK2 with the docked ligands were processed using AMBER11 and AMBER tools 11 packages. In the case of the GSK3 β (PDB id: 1q3d and 1j1b), 7 chloride ions were added to neutralize the protein charges, then the protein complex was centered inside a cubic solvent box of TIP3P water molecule's model with a buffer zone of 10 Å.

4.2.3.1 Molecular Dynamics Simulation

AMBER package version 11 was mainly used for performing the molecular dynamic simulations of the proposed binding modes with the protein kinases' crystal structures. *Antechamber* was used to parameterize the docked ligands and assign the atom types to the generalized AMBER force field (GAFF) [44], with partial charges calculated according to the semi-empirical method AM1-BCC, while the kinase structures were processed using the software *tleap* to be parameterized using the force field AMBERff99SB [168, 169] and generating the final parameters of docking complexes.

Three minimization steps were carried out: the first step is 1000 minimization cycles using conjugate gradient with restrains of (50 kcal.mol⁻¹.Å⁻¹) applied on the whole complex, the second minimization step is also 1000 cycles of conjugate gradient with restrains of (50 kcal.mol⁻¹.Å⁻¹) applied on only the protein while leaving the ligand free, the third step is 1000 of cycles of conjugate gradient without any restrains. The minimized system was passed to a heating phase for 50 ps, during which the temperature of the system was raised from 0 K to 300 K; using Langevin temperature equilibration scheme and restrains of (10 kcal.mol⁻¹. Å⁻¹) applied on the whole complex (the protein and the ligand). After heating the system in constant volume, water density relaxation was done by using constant pressure 1.0 bar under constant pressure periodic boundary (isotropic position scaling method with relaxation time 2 ps). The

pressure relaxation phase lasted another 50 ps and also using restrains of (10 kcal.mol⁻¹.Å⁻¹) applied on the whole complex. Equilibration simulation was performed later for 150 ps: in the first 50 ps with a restrain force of (10 kcal.mol⁻¹.Å⁻¹) applied on the whole complex, the second 50 ps the same restrain force was applied only on the protein, and then for the last 50 ps the simulation was done without restrains. SHAKE algorithm was employed to remove the bond stretching freedom. After performing the equilibration phase, free molecular dynamics was performed for 10 nanoseconds with (2 fs) time step and with (9 Å) cut-off for the non-bonded interactions. Particle Mesh Ewald was employed to compute the electrostatic interactions during the simulation.

4.2.3.2 MM-PBSA Calculations

The free energy is estimated by summation of the molecular mechanics energy (ΔE_{MM}) of the complex estimated by the force field as combination of the van der Waals Interactions (ΔE_{vdW}) and electrostatic interaction (ΔE_{ele}), the electrostatic solvation penalty ($\Delta G_{ele-sol}$) estimated by Poisson-Boltzman model (ΔG_{PB}) or Generalized Born model (ΔG_{GB}), and the non-polar solvation term (ΔG_{SA}), which is dependent on the solvent-accessible surface area (SASA). An additional solute entropy term ($-T\Delta S_{solute}$) should be added to get an estimation of the binding free energy, if the ligands belong to different series. For the complex (protein/ligand), we can estimate the binding energy as the difference between the free energy of the complex and the summation of the free energies of the free protein and the free ligand [170, 171, 172, 173] (For detailed description of this method, see Appendix A, A.1).

$$\Delta G_{bind} = G_{complex} - (G_{protein} + G_{ligand}) = \Delta E_{vdW} + \Delta E_{ele} + \Delta G_{PB} + \Delta G_{SA} - T.\Delta S_{solute}$$

The MM-PBSA (molecular mechanics Poisson–Boltzmann surface area) calculations were performed on 25 snapshots obtained from the last nanosecond of an MD trajectory. The van der Waals (ΔE_{vdW}) and electrostatic (ΔE_{ele}) interaction between ligand and protein in gas phase are calculated with an infinite cutoff using the SANDER module. The electrostatic free energy of solvation ($\Delta G_{ele-sol}$) is calculated with numerical solvation of the Poisson-Boltzmann (PB) equation as implemented in the python script MMPBSA.py in AMBER11. Default parameters for the PB solver; such as a grid spacing at 0.5 Å, dielectric constants of 1.0 for solute and 80.0 for implicit PB solvent (others values like 2 and 4 has also been used as dielectric constants of the solute), solvent probe radius at 1.4 Å (mbondi radii), and ionic strength at 0 M concentration, are used. The non-electrostatic free energy of solvation (G_{SA}) was calculated as linear function of the solvent accessible surface area (SASA); $\Delta G_{SA} = \gamma.SASA + b$ where γ and b were set at default values for the applied implicit solvent model. For MM-PBSA calculations: the nonpolar

contribution was determined on the basis of solvent-accessible surface area (SASA) using the LCPO method: $\Delta G_{SA} = 0.0072 \times \Delta SASA$ (Amber 10 user's manual). The Entropy term will not be considered, as we are dealing with a congeneric series of ligands.

For LR-MM-PBSA (LIECE) models, the energy components of PBSA/GBSA calculations, depending on one snapshot or multiple MD snapshots, were used to be fitted to the experimental binding energy ΔG_{bind} ($RT \ln K_i$), according to the following equations:

1- three-parameter model with decomposed electrostatics: $(\Delta G_{bind}) = \alpha \Delta E_{vdW} + \beta_1 \Delta E_{coul} + \beta_2 \Delta G_{solv}$

2- four-parameter model with decomposed electrostatics: $(\Delta G_{bind}) = \alpha \Delta E_{vdW} + \beta_1 \Delta E_{coul} + \beta_2 \Delta G_{solv} + \gamma \Delta G_{SA}$

4.2.3.3 Linear Interaction Energy (LIE)

According to the LIE approach, the free binding energy could be approximated by this equation:

$$\begin{aligned} \Delta G_{bind} &= \Delta G_{polar} + \Delta G_{nonpolar} = \beta \Delta E_{ele} + \alpha \Delta E_{vdW} \\ &= \beta (\langle V_{lig-surr}^{el} \rangle_{bound} - \langle V_{lig-surr}^{el} \rangle_{free}) \\ &+ \alpha (\langle V_{lig-surr}^{vdW} \rangle_{bound} - \langle V_{lig-surr}^{vdW} \rangle_{free}) + \gamma \end{aligned} \quad (\text{equation LIE})$$

Where $\langle \rangle$ denotes MD or MC averages of the non-bonded van der Waals (vdW) and electrostatic (el) interactions between the ligand and its surrounding environment (lig-surr), i.e. either the solvated receptor binding site (bound state) or just solvent (free state). The Δ denotes the change in these averages when transferring the ligand from the receptor binding site (bound state) to the solution (free state). The parameters of this equation are the weight coefficients α and β for the non-polar and polar binding energy contributions respectively. An additional constant γ is possibly needed to enhance the description of the non-polar contribution [174, 175, 176] (For detailed description of this method, see Appendix A, A.2).

After performing MD simulation for sufficient time to reach stable inhibitor/protein complex, the protein and the ligand from the last snapshot were parameterized to be simulated in Gromacs4.5. Parameterization of the protein was done according to AMBERff99SB force field, while the parameters of the small-molecule ligand was obtained using AmberTools11's antechamber, in order to calculate the ligand partial charges according to AM1-BCC method using gaff atom-types. ACPYPE program was used to convert antechamber parameters to Gromacs topology and coordinates' files. These complexes were solvated in 15 Å octahedral box of

water solvent in TIP3P model and then simulated for 500 ps. Two simulations were carried out: one of the ligand in water (“free simulation”) and another of the ligand bound to the solvated protein (“complex simulation”).

Steepest descent method was used for energy minimization with 500 steps as maximum number of iterations. Position-restrained equilibrium was performed later for 20 picoseconds using Berendsen thermostat with time step 2 fs (10000 steps), in which the protein atoms were restrained by a force constant of $1000 \text{ kJ.mol}^{-1}.\text{nm}^{-2}$ to their initial position. During the simulation, The LINCS algorithm was used to constrain bonds, allowing for an integration step of 2 fs. The long electrostatic interactions were calculated at every step with the LRF (local reaction field) method with a short-range electrostatic interaction cut off of 1 nm. The V-rescale thermostat was used to keep simulation temperature constant by coupling ($\tau = 0.1 \text{ ps}$ as relaxation time) the protein, ligand, and solvent separately to a temperature bath of 300 K. The pressure was maintained isotropic using Berendsen barostat with coupling ($\tau_p = 0.5 \text{ ps}$ as relaxation time).

The gromacs’ analysis programs (`g_dist` and `g_energy`) were used to control the changes of the distance between the P-loop and A-loop during the simulation, `g_energy` was used for the calculation of different ligand-environment energies in the complex bound form: the van-der-Waals interactions between ligand-protein, the van-der-Waals interactions between ligand-solvent, electrostatic interactions ligand-protein, electrostatic interactions ligand-solvent

($\Delta\langle V^{\text{vdw}}_{\text{lig-prot}} \rangle_{\text{bound}}$, $\Delta\langle V^{\text{vdw}}_{\text{lig-solv}} \rangle_{\text{bound}}$, $\Delta\langle V^{\text{el}}_{\text{lig-prot}} \rangle_{\text{bound}}$, $\Delta\langle V^{\text{el}}_{\text{lig-solv}} \rangle_{\text{bound}}$ respectively).

The ligand alone was also solvated in a octahedral box of water TIP3P model, and simulated for 500 ps, again to use `g_energy` for calculating different ligand-environment energies in the free form: vdW interaction energy between ligand-solvent and electrostatic interaction energy between ligand-solvent are abbreviated as ($\Delta\langle V^{\text{vdw}}_{\text{lig-solv}} \rangle_{\text{free}}$, $\Delta\langle V^{\text{el}}_{\text{lig-solv}} \rangle_{\text{free}}$), respectively.

4.2.4 Water Hydration Maps

Solvent distributions grid were calculated using `ptraj` by binning atom positions from RMS coordinate fit frames over all protein heavy atoms at 0.8-ps intervals into $(0.5\text{-}\text{\AA})^3$ grids over 10-ns trajectories. A value was attributed to each grid element to represent the number of times the coordinates of water oxygen atoms were within the $(0.5\text{-}\text{\AA})^3$ grid element. These grids can then be contoured using the density delegate of UCSF Chimera. In the graphics of the water hydration presented (Figure 4.9), the contouring of the water density was performed at 175 hits per $(0.5\text{-}\text{\AA})^3$, which represents 175 visits (three times expected bulk water density) to each $(0.5\text{-}\text{\AA})^3$ grid element over the 10 ns trajectory (12500 frames). For 12500 frames, the expected

number of waters per grid element, assuming bulk water density, is 52.25 (For 1000 MD frames, the expected visits for water to a grid element is 4.18).

4.3 Results

4.3.1 Suggested Binding Modes of 1-aza-9-oxa-fluorenes with GSK3 β

Docking of the available compounds with two of the available crystal structures of GSK3 β (PDB id 1q3d, and 1j1b) suggests mainly two different binding modes, which were obtained by collecting and clustering all the docking solutions. The different potential binding modes are as follows:

Binding mode 1: the tricyclic 1-aza-9-oxa-fluorene binds adjacent to the hinge region where the nitrogen can make a hydrogen bond with the hinge residue V135. Ring B and its substituent (R2) are positioned inside the binding pocket to be close from the gate keeper residue L132. In this case, the hydroxyl or methoxy group (R1) on ring A is directed towards the solvent boundary close to the residue R141 (Figure 4.4).

Binding mode 2: this binding mode is distinct from the first suggested binding mode, as the nitrogen and the oxygen atoms are located in the direction of the activation loop. They could make hydrogen bonds either with the flexible lysine residue K85 or the activation loop's aspartate D200. The hydroxyl or methoxy group (R1) on ring A makes a hydrogen bond with the hinge residue V135, while the substituent of ring B (R2) are directed to the solvent boundary in the phosphate-binding region (Figure 4.5).

The first binding mode was mainly derived while docking the 1-aza-9-oxafluorene compounds to the crystal structures of 1q3d, as the orientation of the residue R141 outwards from the binding pocket allows the rigid tricycles to bind adjacent to the hinge residues. In other crystal structures, such as the case of the crystal structure 1j1b, the side chain of R141 side-chain takes another orientation inside the binding pocket, preventing close binding to the hinge region residues (Compare the orientation of R141 in both figures 4.4 and 4.5).

In the case of docking with CDK2 structure, both GOLD and Glide scores gave preference for binding mode 1. Both scores of 1-aza-9-oxa-fluorene in binding mode 1 are higher than the scores in binding mode 2, as the (R2) substituent on location 3 can make higher van-der-Waals interactions with the aromatic gate-keeper of CDK2 (F80). In the case of CDK2, the gate-keeper; F80, and the opposite residue; L55, offer wide opening for the back selectivity pocket; providing enough space for a bulky substituent in that area. On the other hand, binding mode 2 in docking solutions with GSK3 β structure obtained higher scores than binding mode 1 for most of 1-aza-9-oxa-fluorene derivatives. The opening of the back selectivity pocket in GSK3 β structure between the gate-keeper; L132, and the opposite residue; M101, is relatively

narrow; resulting in steric clashes with bulky R2 substituents. These steric clashes could explain why binding mode 1 is less favorable in the case of GSK3 β . However, for both binding modes neither of the docking scores can generate significant correlation with the experimental binding affinities (Tables C.3, C.4 and Figure C.2 in Appendix C).

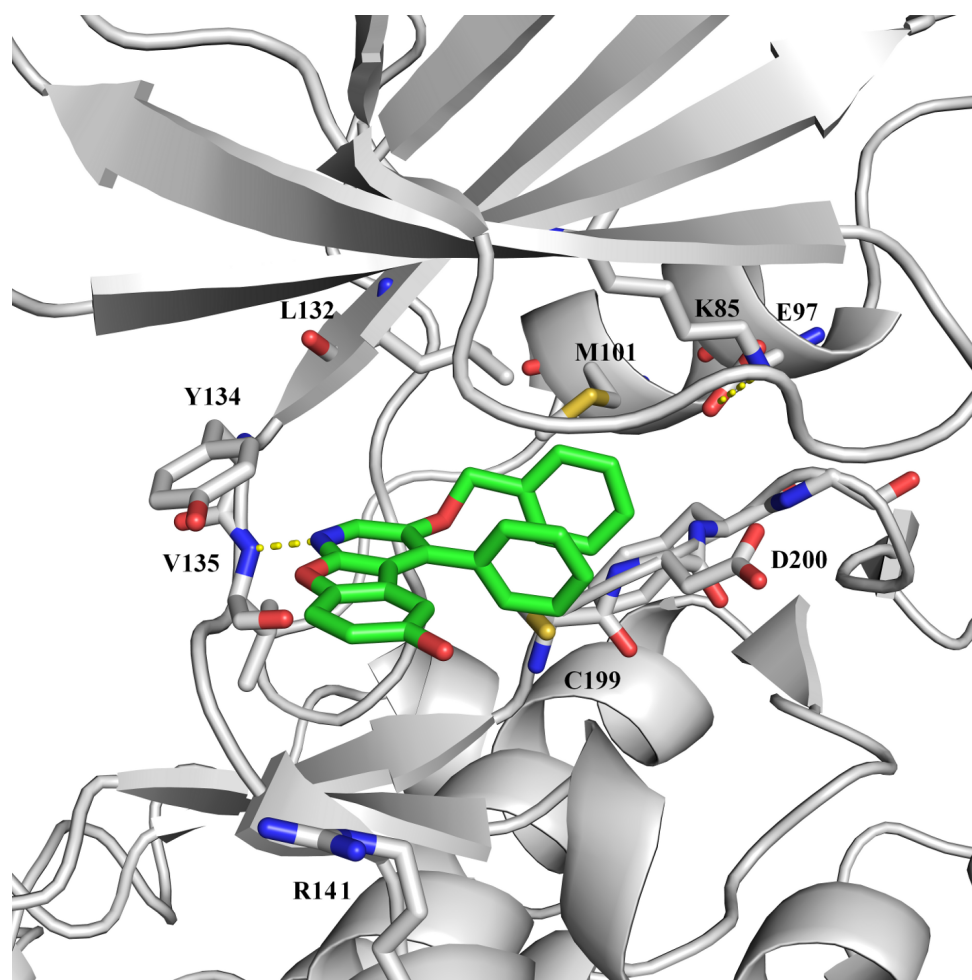


Figure 4.4: **Binding mode 1:** Docking solution of compound **2b** with GSK3 β (PDB id: 1q3d).

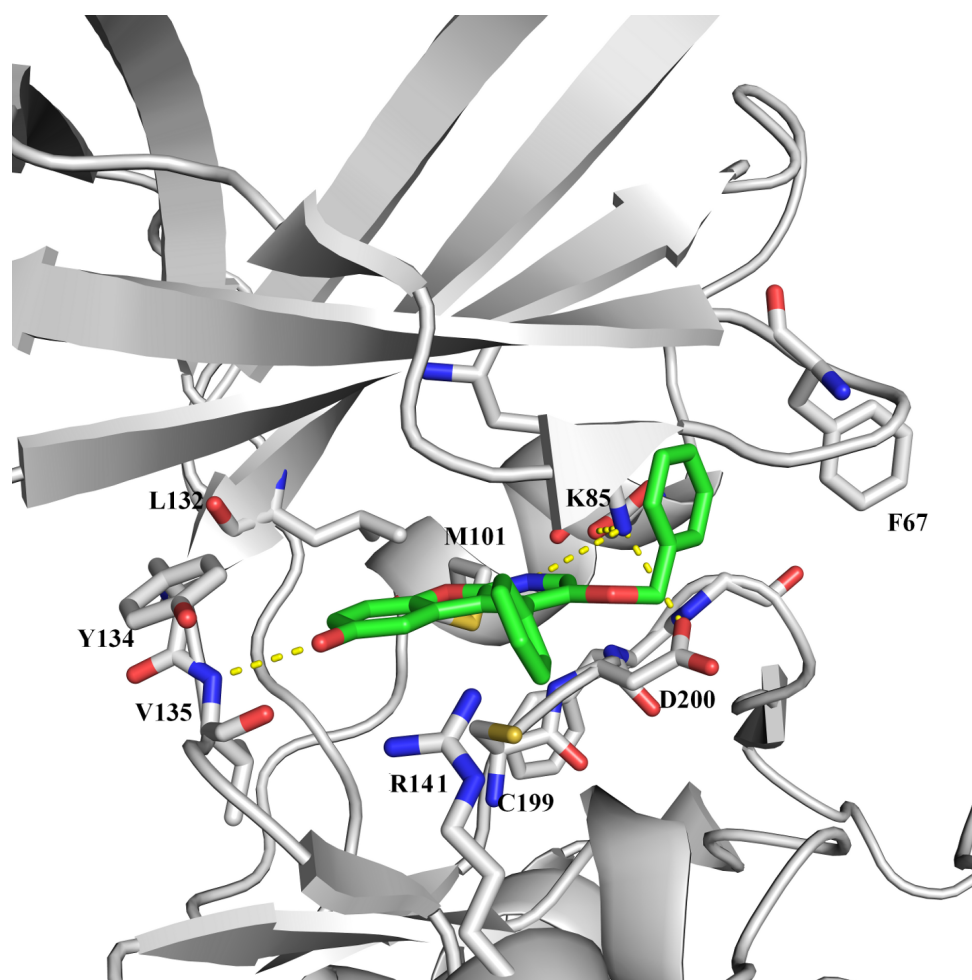


Figure 4.5: **Binding mode 2:** Docking solution of compound **2b** at GSK3-β (PDB id: 1j1b).

4.3.2 Molecular Dynamics Simulations and Trajectory Analysis

To have further insights into the binding of the 1-aza-9-oxafluorenes trying to determine the correct binding mode, and to explain the selectivity profile of these kinase inhibitors, binding free energy methods were applied on both suggested binding modes, including the implicit solvent methods (PBSA and GBSA) and the Linear Interaction Energy method (LIE) using explicit solvent. The correct binding mode is supposed to provide the right ranking of the ligands' affinities using one of the binding energy methods.

The protein conformation of the kinase has possibly an impact on the estimated binding energy value. Therefore, monitoring the conformation of the glycine-rich P-loop was performed in combination with the two applied methods. The distance between the P-loop and A-loop, defined as the distance between CA of F67 and CA of D200 (in the case of GSK3 β), shows significant differences in the P-loop movement depending on the nature of R1 and R2 substituents of the inhibitors. Monitoring the P-loop fluctuations during the course of MD simulation runs with different binding modes can show different induced-fit effects depending on the binding mode of the inhibitor. For derivative **2b** in binding mode 1 with GSK3 β , the steric clashes between the bulky benzyl and GSK3 gate-keeper residue (L132) would require bigger induced-fit effects; compared to the induced-fit effects in binding mode 2. The average P-loop/A-loop distance for simulation with **2b** in binding mode 1 will reach 13.5 Å, while it is only 12.3 Å in binding mode 2. The important impact of the glycine-rich P-loop on the protein kinase conformational energy was addressed previously in some research studies [115, 177]. To consider the impact of the P-loop conformation and its distance from the activation loop, the binding energy calculations have been performed on different kinase conformations, recognized depending on the P-loop/A-loop average distance as a descriptor for the conformational change.

4.3.3 MM-PBSA Calculations

Both methods, which depend on implicit solvent model; i.e. MM-PBSA and LR-MM-PBSA, did not result in a correct ranking of the studied inhibitors or a significant correlation with the experimental binding affinities. Both methods were applied on the total dataset of 1-aza-9-oxa-fluorene derivatives, and also on subsets, chosen on the basis of the P-loop/A-loop distance (average value) observed during the MD simulation. However, for all studied subsets both implicit solvent models were not able to generate acceptable models for predicting the binding affinity. Neither for binding mode 1 nor for binding mode 2 was there a significant correlation. (The results are summarized in the Appendix C; Tables C.5, C.6, and C.7 for GSK3, Tables C.8 and C.9 for CDK2, and Figure C.3).

4.3.4 LIE Models for GSK3 β

To select the right binding mode, which is able to explain the variation of the binding affinities of the 1-aza-9-oxa-fluorene series with the two kinases, Linear Interaction Energy (LIE) method was applied in a trial to obtain predictive models of the binding free energy ΔG_{bind} with the two studied kinases.

We converted the experimental values of the binding assays (K_i or IC_{50}) to binding free energy values using the equation:

$$\Delta G_{\text{bind}} = RT \ln (K_i)$$

which could also be written as:

$\Delta G_{\text{bind}} = RT \ln (IC_{50}) + c$; where c is a constant ($c = -RT \ln [1 + S/K_m]$), which doesn't change as long as we measure IC_{50} using identical assays methods [178].

All LIE calculations were performed, starting from snapshots of the inhibitors' series with GSK3 β structures (3dqw.pdb was used for binding mode 1 and 1j1b.pdb for binding mode 2). The snapshot was taken from the last nanosecond and selected to have a distance, between the atom CA of activation loop residue D200 and the atom CA of the P-loop's residue F67, equal to the average distance between these two exact atoms during 10 ns MD simulation.

To consider the effects of the induced fit, recognized by the change of P-loop/A-loop distance, we tried to classify the dataset into subsets depending on the average distance observed during the MD simulations. Multiple linear regression was applied on to inhibitor subsets showing a similar A-loop/P-loop distance, i.e. keeping A-loop/P-loop average distances within a small range (1-1.5 Å). The generated LIE models were assessed by statistical parameters: the coefficient of determination (r^2) and root mean squared error (rmse). A special care was given to the coefficients of van-der-Waals energy and electrostatic energy terms; not to violate the basic principles of molecular recognition and binding. As the physical understanding of molecular recognition process applies that higher van-der-Waals ligand-protein interactions results in higher binding affinities, a positive coefficient of van-der-Waals energy term is required when LIE models are generated using the free binding energy (ΔG_{bind}) [175, 174]. Different combination of the empirical values for the electrostatic and van-der-Waals terms coefficient were applied in LIE equation to get the best correlation with the experimental data ($\beta = 0.5, 0.4$ or 0.3 , and $\alpha = 0.6, 0.3, 0.18$, or 0.13) [174, 179] (See Appendix A, A.2).

4.3.4.1 LIE models for GSK3 β (Binding mode 1)

In the case of GSK3 β kinase, applying multiple linear regression on the whole dataset of 1-aza-9-oxa-fluorene compounds in the binding mode 1 using LIE electrostatic and hydrophobic

energy terms gave a very poor correlation with $r^2 = 0.08$ and root mean square error (RMSE= 1.40). Applying linear regression on individual subsets, considering the different P-loop/A-loop distance, didn't improve the correlation. All the developed models in this binding mode own negative coefficient for one of the energy terms at least. The results are summarized in the Appendix C; Tables C.10, C.11, and C.12).

4.3.4.2 LIE models for GSK3 β (Binding mode 2)

In the case of GSK3 β kinase inhibitors in binding mode 2, applying multiple linear regression on the whole dataset of 1-aza-9-oxa-fluorene compounds using the LIE electrostatic and hydrophobic energy terms gave a model with poor correlation ($r^2 = 0.14$ and $rmse=1.35$) (Table 4.3). However, applying multiple linear regression on individual subsets, clustered depending on the observed P-loop/A-loop distance, improved the correlation significantly.

Applying the multiple linear regression on the first subset (larger average P-loop/A-loop distances in the range 12.1-13.15 Å) including compounds (**3a_3**, **2d_1**, **2b_2**, and **2b**) generated two significant models with good correlation and low RMSE values (Table 4.4). The second subset of compounds (smaller average P-loop/A-loop distances in the range 10.37 – 11.95 Å), including compounds (**2d_5**, **2a_5**, **2d_6**, **2a_3**, **2a**, **3a_2**, **2d_3**, and **2c**) generated three statistically significant models (Table 4.5).

The best model for predicting the binding energies of the first subset is:

$$\Delta G_{\text{bind}} (\text{pred.}) = 0.80(\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle) + 0.16 (\Delta \langle V^{\text{vdW}}_{\text{lig-surr}} \rangle) - 12.5 (r^2=0.97, \text{RMSE}=0.41)$$

The second significant model predictable model for the same dataset:

$$\Delta G_{\text{bind}} (\text{pred.}) = 0.63 (\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle) + 0.32 (\Delta \langle V^{\text{vdW}}_{\text{lig-surr}} \rangle) - 8.22 (r^2=0.91, \text{RMSE}=0.73)$$

For the second dataset, a significant model with low RMSE was obtained:

$$\Delta G_{\text{bind}} (\text{pred.}) = 0.86 (\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle) + 0.23 (\Delta \langle V^{\text{vdW}}_{\text{lig-surr}} \rangle) - 7.89 (r^2=0.78, \text{RMSE}=0.65)$$

Another model, which owns closer values to the standard coefficients of LIE approach's coefficients, could also be developed:

$$\Delta G_{\text{bind}} (\text{pred.}) = 0.48 (\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle) + 0.13 (\Delta \langle V^{\text{vdW}}_{\text{lig-surr}} \rangle) - 7.40 (r^2=0.78, \text{RMSE}=0.93)$$

By applying linear regression using partial least squares on the second subset, a model with higher correlation could be developed, but with low value for the coefficient of van-der-Waals term (α). Cross-validation shows that this model is over-fitted: ($\beta = 0.93$, $\alpha = 0.01$, $\gamma = -11.71$, $r^2 = 0.93$, $\text{RMSE} = 0.42$, $q^2 = 0.05$), See Table 4.5.

Table 4.3: LIE model obtained for the whole dataset of 1-aza-9-oxafluorene derivatives using their binding affinities against GSK3 β , starting from docking solutions in binding mode 2. Compounds are ranked according to the descending order of P-loop/A-loop distance.

$\Delta G_{\text{bind}} (\text{pred.}) = \beta(\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle) + \alpha(\Delta\langle V^{\text{vdW}}_{\text{lig-surr}} \rangle) + \gamma$						
Cmp code	Ki	ΔG_{bind}			P-loop/A-loop	$\beta = 0.21 \alpha = 0.13 \gamma = -6.95$
BM2	GSK3 β	(obsrv.)	$\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta\langle V^{\text{vdW}}_{\text{lig-surr}} \rangle$	Average dist (10 ns)	$r^2 = 0.14, \text{RMSE} = 1.35$
3a_3	5.4	-7.28	9.95	-18.25	13.5 \pm 0.7	-7.27
2d_1	4.1	-7.44	9.07	-13.12	13.3 \pm 0.8	-6.78
2b_2	0.02	-10.6	6.29	-18.81	13.15 \pm 0.7	-8.11
2b	5.8	-7.23	10.64	-18.28	12.3 \pm 0.8	-7.13
2d_5	4.4	-7.40	4.71	-17.37	11.9 \pm 0.8	-8.25
2a_5	16.3	-6.61	4.68	-14.60	11.8 \pm 1.2	-7.89
2d_6	4.8	-7.34	4.89	-14.45	11.6 \pm 0.7	-7.83
2a_3	9.9	-6.91	4.93	-14.92	11.48 \pm 1.1	-7.88
2a	14.8	-6.67	5.76	-13.28	11.43 \pm 1.4	-7.49
3a_2	5.8	-7.23	4.8	-18.2	11.35 \pm 0.7	-8.34
2d_3	9	-6.90	5.09	-13.06	10.99 \pm 0.8	-7.61
2c	0.02	-10.6	1.31	-12.88	10.37 \pm 0.7	-8.37
2d_2	1.5	-8.04	7.18	-13.01	8.2 \pm 0.6	-7.16

Table 4.4: LIE models obtained for subset 1 (average P-loop/A-loop distance between 12.1 to 13.15 Å) using their binding affinities against GSK3 β , starting from docking solutions in binding mode 2. Compounds are ranked according to the descending order of P-loop/A-loop distance.

Comp code	$\Delta G_{\text{bind}} (\text{pred.}) = \beta(\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle) + \alpha (\Delta\langle V^{\text{vdW}}_{\text{lig-surr}} \rangle) + \gamma$					
GSK3 β , BM2	Avg. Dist	$\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta\langle V^{\text{vdW}}_{\text{lig-surr}} \rangle$	exp.	$\beta=0.80$ $\alpha=0.16$ $\gamma=-12.5$	$\beta=0.63$ $\alpha=0.32$ $\gamma=-8.22$
3a_3	13.5 \pm 0.7	9.95	-18.25	-7.28	-7.51	-7.86
2d_1	13.3 \pm 0.8	9.07	-13.12	-7.44	-7.39	-6.76
2b_2	13.15 \pm 0.7	6.29	-18.81	-10.6	-10.55	-10.35
2b	12.3 \pm 0.8	10.64	-18.28	-7.23	-6.96	-7.44
Model r^2 (RMSE)					0.97 (0.41)	0.91 (0.73)

Table 4.5: LIE models obtained for subset 2 (average P-loop/A-loop distance between 10.37 to 12.02 Å) using their binding affinities against GSK3 β , starting from docking solutions in binding mode 2. Compounds are ranked according to the descending order of P-loop/A-loop distance.

Comp code	$\Delta G_{\text{bind}} (\text{pred.}) = \beta(\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle) + \alpha (\Delta\langle V^{\text{vdW}}_{\text{lig-surr}} \rangle) + \gamma$						
GSK3 β , BM2	Avg. Dist	$\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta\langle V^{\text{vdW}}_{\text{lig-surr}} \rangle$	exp.	$\beta=0.93$ $\alpha=0.01$ $\gamma=-11.71$	$\beta=0.86$ $\alpha=0.23$ $\gamma=-7.89$	$\beta=0.48$ $\alpha=0.13$ $\gamma=-7.40$
2d_5	11.9 \pm 0.8	4.71	-17.37	-7.40	-7.50	-7.83	-7.53
2a_5	11.8 \pm 1.2	4.68	-14.60	-6.61	-7.50	-7.22	-7.05
2d_6	11.6 \pm 0.7	4.89	-14.45	-7.34	-7.30	-7.00	-7.04
2a_3	11.48 \pm 1.1	4.93	-14.92	-6.91	-7.27	-7.08	-6.97
2a	11.43 \pm 1.4	5.76	-13.28	-6.67	-6.48	-5.99	-6.45
3a_2	11.35 \pm 0.7	4.8	-18.2	-7.23	-7.4	-7.94	-7.8
2d_3	10.99 \pm 0.8	5.09	-13.06	-6.90	-7.1	-6.51	-6.75
2c	10.37 \pm 0.7	1.31	-12.88	-10.6	-10.48	-9.72	-8.55
Model r^2 (RMSE)					0.94 (0.42)	0.78 (0.65)	0.78 (0.93)

4.3.5 LIE Models for CDK2

4.3.5.1 LIE models for CDK2 (Binding mode 1)

In the case of binding mode 1 with CDK2 kinase, applying multiple linear regression on the whole dataset of 1-aza-9-oxa-fluorene compounds using LIE energy terms as descriptors gave no significant model ($r^2 = 0.42$, RMSE=0.95, with negative coefficient for the van-der-Waals energy change). Considering different subsets; clustered according to the observed A-loop/P-loop distance, did not improve the LIE models. In all cases, the generated models had problems in the physical sense, as they obtained negative coefficient for the van-der-Waals energy change (The results are summarized in the Appendix C; Tables C.13 and C.14).

4.3.5.2 LIE models for CDK2 (Binding mode 2)

In the case of binding mode 2 with CDK2 kinase, performing multiple linear regression on the LIE electrostatic and hydrophobic energy differences gave a good correlation ($r^2 = 0.56$, rmse=1.05); Table 4.6.

We could improve this model by applying the linear regression on a subset of compounds in a small range of P-loop/A-loop distance. Applying the multiple linear regression on the LIE electrostatic and van-der-Waals energies terms for a subset with average P-loop/A-loop distance in the range (8.1-9.39 Å); including (**2a**, **2a_3**, **2a_4**, and **2a_5**), generated three statistically significant models with good correlation coefficient (Table 4.7).

The best model for predicting the binding energies of that subset is:

$$\Delta G_{\text{bind}}(\text{pred.}) = 0.40(\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle) + 0.21(\Delta \langle V^{\text{vdW}}_{\text{lig-surr}} \rangle) - 4.11 \quad (r^2=0.88, \text{RMSE}=0.42)$$

Another significant model could be generated with higher correlation, but also with higher RMSE:

$$\Delta G_{\text{bind}}(\text{pred.}) = 0.24(\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle) + 0.09(\Delta \langle V^{\text{vdW}}_{\text{lig-surr}} \rangle) - 4.8 \quad (r^2=0.91, \text{RMSE}=0.73)$$

Table 4.6: LIE model obtained for the whole dataset of 1-aza-9-oxafluorene derivatives using their binding affinities against CDK2, starting from docking solutions in binding mode 2. Compounds are ranked according to the descending order of P-loop/A-loop distance.

$\Delta G_{\text{bind}} (\text{pred.}) = \beta(\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle) + \alpha (\Delta \langle V^{\text{vdW}}_{\text{lig-surr}} \rangle) + \gamma$							
Cmp code	Ki	ΔG_{bind}	P-loop/A-loop		$\beta = 0.46$	$\alpha = 0.06$	$\gamma = -6.9$
BM2	CDK2	(obsrv.)	$\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta \langle V^{\text{vdW}}_{\text{lig-surr}} \rangle$	Average dist (10 ns)	$r^2 = 0.56$, RMSE=1.05	
2b	6.4	-7.17	4.16	-14.52	11.11±1.6	-5.92	
2a	24	-6.38	0.55	-12.03	9.93±1.4	-7.45	
2a_3	50	-5.94	3.40	-13.2	9.43±1.2	-6.19	
2a_4	267	-4.93	4.68	-13.5	9.56±1.2	-5.61	
2a_5	241	-4.99	6.83	-16.94	8.1±0.85	-4.83	
2c	0.6	-8.50	0.33	-11.04	7.51±0.70	-7.49	
2a_2	147	-5.30	4.78	-15.88	7.6±0.60	-5.72	

Table 4.7: LIE models obtained for compounds' subset (average P-loop/A-loop distance between 8.1 to 9.39 Å) using their binding affinities against CDK2, starting from docking solutions in binding mode 2. Compounds are ranked according to the descending order of P-loop/A-loop distance.

$\Delta G_{\text{bind}} (\text{pred.}) = \beta(\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle) + \alpha (\Delta \langle V^{\text{vdW}}_{\text{lig-surr}} \rangle) + \gamma$							
Comp code	Avg. Dist	$\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta \langle V^{\text{vdW}}_{\text{lig-surr}} \rangle$	exp.	$\beta = 0.40$	$\beta = 0.24$	$\beta = 0.5$
CDK2, BM2					$\alpha = 0.21$	$\alpha = 0.09$	$\alpha = 0.16$
					$\gamma = -4.11$	$\gamma = -4.8$	$\gamma = -4.8$
2a	9.93±1.4	0.55	-12.03	-6.38	-6.50	-5.70	-6.45
2a_3	9.43±1.2	3.40	-13.2	-5.94	-5.59	-5.17	-5.21
2a_4	9.56±1.2	4.68	-13.5	-4.93	-5.14	-4.89	-4.62
2a_5	8.1±0.9	6.83	-16.94	-4.99	-5.01	-4.68	-4.09
Model r^2 (RMSE)					0.88 (0.42)	0.91 (0.73)	0.86 (0.85)

4.3.6 Selection of Most Probable Binding Mode

LIE calculations show that **binding mode 2** is more likely to be the actual binding mode of the 1-aza-9-oxa-fluorene derivatives. **Binding mode 2** is able to generate significant LIE models, describing the variance of binding affinities of the 1-aza-9-oxa-fluorene derivatives with both kinases GSK3 β and CDK2, and correlating the binding affinity with the energy calculation of these ligands. This idea of selecting the '*correct*' binding mode depending on the strength of LIE models and their predictive potential was previously suggested and applied by Nervall and Aqvist *et al.* [178]. Considering the high flexibility of kinases and the significant induced fit effects observed for different inhibitors, a special care should be given to the conformational changes. LIE models, which are statistically significant and physically acceptable, were generated using subsets of compounds, clustered according to their P-loop/A-loop average distance observed in the MD simulations.

The applied method, which is named "**P-loop/A-loop-distance-dependent LIE approach**", successfully generated predictive models after dividing the complete dataset of 1-aza-9-oxa-fluorene derivatives into two subsets, depending on the observed P-loop/A-loop average distance, calculated over an MD simulation for 10 nanoseconds. The LIE models of **binding mode 2** can predict the free binding energy (ΔG_{bind}) to GSK3 β kinase with significant correlation; giving r^2 equal to 0.97 and 0.78 for the two subsets. In the case of binding mode 2 with CDK2 kinase, it was possible to obtain another LIE models, predicting the binding energy (ΔG_{bind}) with r^2 of 0.56 for the whole dataset of inhibitors, and 0.88 for one subset with restricted range of P-loop/A-loop distance. However, the high correlation obtained by these models shouldn't be given a high importance, as the number of the included data points is small. The strength of LIE models, generated by binding mode 2, is that the statistical parameters are much higher than the statistical parameters of binding mode 1's models, and that the energies' coefficients are physically acceptable and improved after restricting the range of induced fit effects. Binding mode 2 could also rationally explain the variation of the binding affinity in accordance with the chemical modifications and the change in the protein/ligand interactions, together with the desolvation energies and induced-fit effects. On the other hand, binding mode 1 could only generate weak models in both cases: with GSK3 β and CDK2 kinase. Moreover, all models, generated by binding mode 1, have a problem in the physical sense; as the coefficients of van-der-Waals energy term are negative in spite of restricting the range of induced-fit effects.

Simple docking scores (GOLD score and Glide score) gave contradicting results for the favorable binding mode in the case of two kinases: GSK3 β and CDK2. While binding mode 1 was mostly favored over binding mode 2 in the case of CDK2, the second was favored by docking scores for GSK3 β . However, the LIE method doesn't show such a conflict and support the same binding mode for both kinases. The discussion will analyze the fitting of the binding mode

2 with the structures of the two studied kinases, trying to understand the physical and structural basis of the 1-aza-9-oxa-fluorene derivatives' binding affinities with these two kinases. For example, converting the hydroxyl group (R1) to methoxy (in the pairs of compounds; **2a** and **2a_2**, also **2c** and **2c_2**) results in a drastic drop in the inhibition, which could only be explained by binding mode 2; as a result of losing one hydrogen bond in the hinge region (Section 4.4.3.1 will discuss this drop in detail in the case of derivatives **2c** and **2c_2**). On the other hand, derivative **2b_2** shows the exact expected drop of inhibition in the case of CDK2, while surprisingly it shows improved inhibition in the case of GSK3 β . However, this surprising change could be understood, if we combine the desolvation energy in binding mode 2 with the larger induced-fit effect, as a result of higher mobility of GSK3 β 's P-loop (Section 4.4.3.2 will discuss in detail the case of derivatives **2b** and **2b_2**).

The predictivity and ranking power of LIE models in binding mode 2 were tested with the inactive compounds ($K_i \geq 1000$). The inactive compounds were excluded from the linear regression calculations of LIE models, as their inhibition constants were not precisely determined. Therefore, the prediction of these compounds could be considered as a kind of external validation. Only the LIE models of binding mode 2 could predict the highly active compounds, and also predict the binding energies of inactive compounds lower than the active compounds. The statistically weak models of binding mode 1 have failed, as expected, to give any reasonable ranking of 1-aza-9-oxa-fluorene compounds (Figure 4.6 shows the difference of LIE models' predictions in both cases: binding mode 1 and binding mode 2; for both kinases).

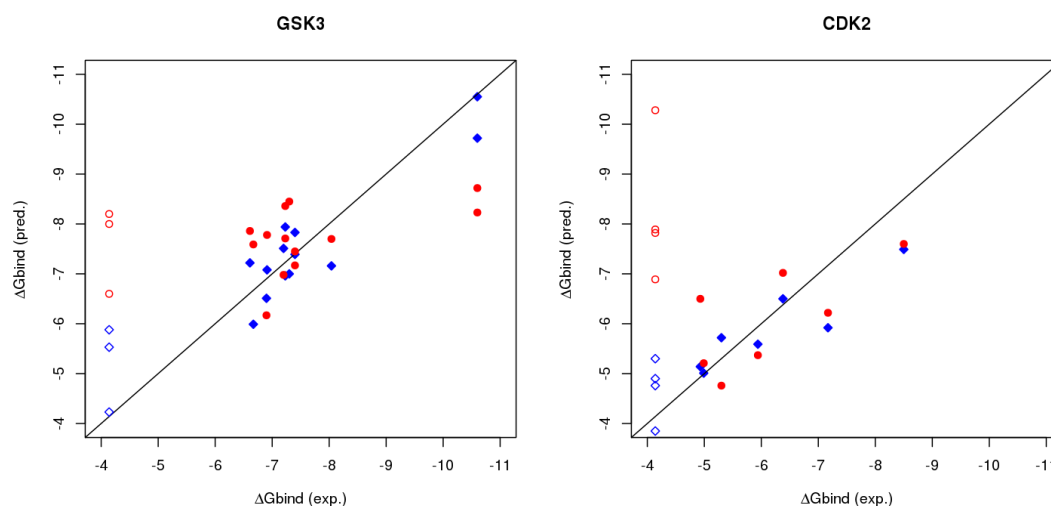


Figure 4.6: Scatter plots of predicted binding free energy estimates from LIE models vs. experimental binding free energies; for GSK3 (left) and CDK2 (right). Red points (circles) represent the predictions according to LIE models derived from **binding mode 1**, while blue points (diamonds) represent the predictions according to LIE models derived from **binding mode 2**. Empty circles represent the inactive compounds plotted at ($\Delta G_{\text{bind}}(\text{exp.}) = -4.14$).

4.4 Discussion

4.4.1 Impact of Kinase Conformational Changes

In a computational study to predict the binding affinity of a series of PIM-1 kinase inhibitors (imidazo[1,2-b]pyridazine derivatives) [177], a PMF simulation was performed to estimate the free energy of glycine-rich P-loop's displacement. The calculation showed a linear relationship between the free energy of P-loop's displacement and the displacement distance estimated as the distance between the two C α atoms of two residues: the ASP residue of the activation loop (A-loop) and the PHE residue of the glycine-rich P-loop. This study highlights the importance of the energy penalty for displacing the P-loop during every binding process with protein kinase in order to accommodate the ligand inside the binding pocket [177]. A computational estimation of the P-loop's displacement penalty for about 5 Å was about (13 kcal.mol⁻¹) for PIM-1 kinase[177]. Another study investigated the capability of MM-PBSA method to describe the selectivity of six kinase inhibitors against a panel of six kinases, and gave also an attention to the fluctuations of the binding pocket's residues. This last study reported that the most fluctuating part of the kinase binding pocket is actually the glycine-rich P-loop, and that the natural inhibitor tends to simulate with its corresponding kinase structure with the least rms fluctuations for P-loop residues [115].

In our case study of the 1-aza-9-oxa-fluorene derivatives with GSK3 β /CDK2, the smallest derivative **2c** obtains its high affinity to GSK3 β kinase from its low desolvation penalty and its small required displacement of the P-loop. Compound **2c** has a remarkable low desolvation penalty, associated with its ability to use two water molecules enhancing the hydrogen bonds' network with the hinge residues and the activation loop's residues. The small size of derivative **2c** also requires very small displacement of the P-loop to be accommodated inside the binding pocket. On the other hand, derivatives with bulkier substituents, such as **2b** and **2b_2**, would require higher displacement of the P-loop to be accommodated inside the binding pocket.

Despite the fact that the P-loop displacement costs energy, this displacement could help in: optimizing the enthalpic interactions, changing the desolvation penalty of the ligand binding, or changing the hydration of the kinase's binding pocket itself. Hydrophobic van-der-Waals interactions and also the solvation/desolvation effects could be optimized by the displacement of the P-loop, such as the pi-pi interaction between the ligand and the aromatic residue of the P-loop (F67 in GSK3 β and Y15 in CDK2). This optimized pi-pi interaction could play an important role in improving the binding affinity and compensating for the displacement penalty.

The derivative **2b_2** (K_i = 20 nM with GSK3 β) can exactly compensate for the energy cost of the P-loop's displacement, obtaining a binding affinity similar to derivative **2c**. In the case of **2b_2** with GSK3 β , the displacement of the P-loop will increase the van-der-Waals

interactions through optimizing the stacking between the benzylate group (R2 in **2b_2**) and the aromatic P-loop residue, while keeping the desolvation penalty of **2b_2** low by maintaining solvent interactions with the hinge region (See section 4.4.3.2).

4.4.2 Role of Water Molecules and Protein Hydration in the Binding Process

In the kinase structures, it is almost confirmed that the hinge region residues should be satisfied either by hydrogen bonding with a ligand/substrate or by water molecules. Some studies refer to a fact that satisfying a hydrogen donor like the backbone NH of a hinge residue is more important and has bigger effect on the binding energy than satisfying hydrogen acceptors like the backbone carbonyl groups [180]. Almost all kinase inhibitors satisfy the condition of making one hydrogen bond at least with the hinge region NH with a possibility to leave the adjacent carbonyl desolvated. However, providing extra hydrogen bonds with the two adjacent carbonyls in the hinge region is one way to increase the binding affinity of the kinase inhibitors [180]. If we consider the equation $\Delta G_{\text{bind}} = RT \cdot \ln(K_i)$, it could be concluded that one order of magnitude in the equilibrium constant K_i (or IC_{50}) costs around $1.4 \text{ kcal}\cdot\text{mol}^{-1}$ at 300 K. That means that the difference between a kinase inhibitor in the nano-molar range like (**2b_2** with GSK3 β , $K_i = 20 \text{ nM}$; $\Delta G_{\text{bind}} = -10.6$) to an inhibitor in micro-molar range like (**2b** with GSK3 β , $K_i = 5.8 \mu\text{M}$, $\Delta G_{\text{bind}} = -7.23$) is less than $4 \text{ kcal}\cdot\text{mol}^{-1}$ at temperature 300 K. Any physical factor, which subtly interplays in the binding process, would be expected to play a significant role in the change of the binding affinity, considering such small range of energy differences. These factors include the ligand desolvation, protein hydration effects, the structural changes and induced fit effects. The role of the water molecules in mediating the protein-ligand interactions and its involvement in the entropic effects would also explain many cases of the affinities' changes.

The role of the solvent's molecules in the electrostatic ligand-protein interaction couldn't be easily accounted for by using the methods of implicit solvent (MM-PBSA/GBSA). Pearlman has pointed out to the issue of the MM-PBSA method's poor performance for the scoring of a congeneric series of p38 MAP kinase inhibitors [114]. An initial placement of structural water inside the binding pocket using an algorithm called JAWS was necessary to get the right prediction of the binding affinities for this series [99]. The implicit solvent models (MM-PBSA) are not able to capture such factors of molecular-length-scale solvation physics, which is better described by the explicit solvent models [125]. The study of Page and Bates, which also tested MM-PBSA for describing the selectivity of kinase inhibitors, also showed a poor performance to rank a collection of six kinase inhibitors depending on their binding affinities to six different kinases [115].

4.4.3 Understanding the SAR of 1-aza-9-oxa-fluorene Derivatives

The comparison between the two ATP binding pockets of both GSK3 β and CDK2 kinases reveals some minor differences in the amino acid sequence, which in turn result in differences in the topological shape of the molecular surface. One of these differences is the gatekeeper residue (L132/F80 in GSK3 β /CDK2) and the opposite residue (M101/L55 in GSK3 β /CDK2), which results in a different size of the back selectivity pocket's entrance. The second difference is the change of a sequence (P136-E137-T138) in GSK3 β to (H84-Q85-D86) in CDK2. The impact of this change is significant on the shape of molecular surface around the binding pocket of GSK3 β /CDK2 in the adjacent area to the hinge region (Br-II), inducing differences in the contact opening between the hinge region and the bulk solvent. The third difference is the residue change C199 (GSK3 β)/A144 (CDK2). These three differences result in a substantial difference in the overall shape of the binding site's molecular surface, making the molecular surface of CDK2 curvier. Other residue changes in the P-loop of the two kinases result in different conformation of the aromatic residue of P-loop and different mobility of the P-loop (Table 4.1).

4.4.3.1 Derivatives **2c** and **2c_2**

The interactions of the most active derivative of the 1-aza-9-oxa-fluorene series (**2c**), when it is bound according to binding mode 2, include strong van-der-Waals interactions and also extensive network of hydrogen bonds with important kinase residues (the hinge residues and the activation loop's aspartate); see Figure 4.7. These interactions provide the essential foundation of the strong binding affinity of the small inhibitor **2c**. The desolvation penalty of compound **2c** is notably small, according to LIE calculation, when binding mode 2 is considered. On the other hand, PBSA calculations cannot capture the important role of water molecules in mediating many of the protein-ligand interactions, and consequently couldn't correctly predict the strong binding of derivative **2c**.

Derivative **2c** makes direct hydrogen bond between its hydroxyl group (R1) and the backbone NH of the hinge residue V135, while it uses a water molecule to mediate additional hydrogen bond with the backbone carbonyl of the same hinge residue. The methylation of hydroxyl (R1) would obviously be a reason for losing one of the hydrogen bonds between this hydroxyl and the hinge residue, making **2c_2** a weaker inhibitor. The second hydroxyl (R2) makes important hydrogen bonds with the activation loop residue D200 and with residue N186, mediated by one of the bulk water molecules. The tricyclic core of 1-aza-9-oxa-fluorene makes strong van-der-Waals interactions with V70, L188 and C199, and a hydrogen bond between the fluorene's nitrogen and the flexible lysine K85.

After MD simulation of **2c** with CDK2, similar interactions between the inhibitor and the kinase residues could be noticed; similar hydrogen bonds network between the hydroxyl group (R1) and the hinge residue L83 directly with the backbone NH and indirectly with the backbone carbonyl mediated by water molecule, a direct hydrogen bond between the 1-aza-9-oxa-fluorene's nitrogen and the flexible lysine K33. The other hydroxyl group (R2), close to the activation loop, makes a network of hydrogen bonds with the DFG aspartate D145 and another residue N132 with assistance from a bulk water molecule. The derivative **2c** also makes van-der-Waals interactions with the CDK2 residues V18 and L134. The residue A144 (in CDK2) cannot make equal van-der-Waals interactions comparing to the corresponding C199 (in GSK3 β). Rather, the residue Y15 (the aromatic residue of CDK2's P-loop) is usually pointing inside the ATP pocket aligning to the inhibitor's aromatic tricycles.

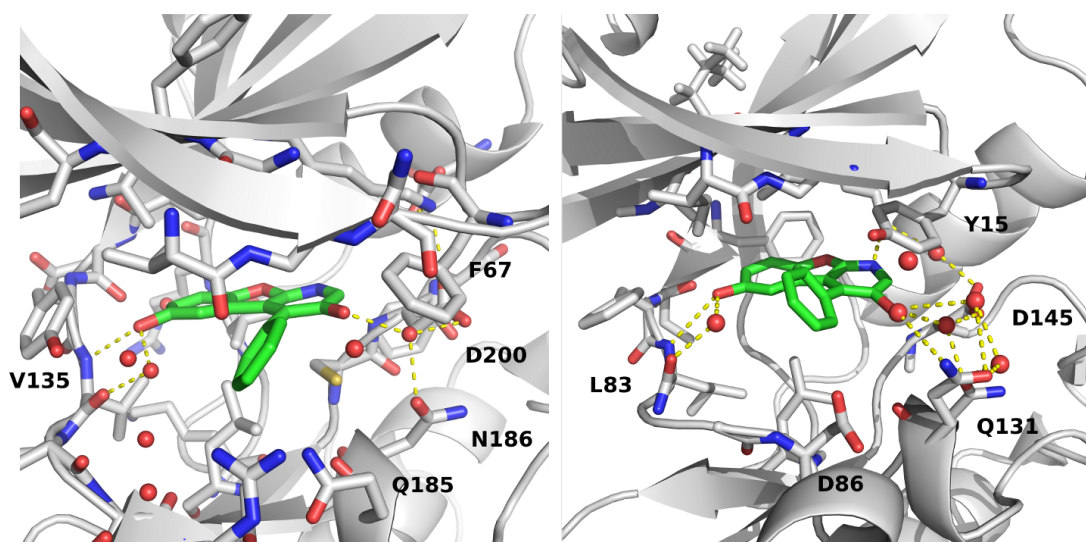


Figure 4.7: 10 ns MD simulation of compound **2c** complexed with GSK3 β structure (PDB id: 1j1b) (left) and CDK2 (PDB id: 2wih) (right) shows a stable hydrogen bonds network, established between the ligand and protein with the help of water molecules.

4.4.3.2 Derivatives **2b** and **2b_2**

Similarly to the case of derivative **2c**; derivative **2b** shows similar hydrogen bonds network with both kinases GSK3 β and CDK2 using two hydrogen bonds between the hinge region and the hydroxyl group (R1) (Figure 4.8). The hydrogen bonds' network is composed of one direct hydrogen bond with the amide NH of the hinge residue, while donating another hydrogen bond to the hinge residue carbonyl mediated by a water molecule. In the case of the **2b_2** derivative, the methoxy (R1) has no hydrogen participating in this hydrogen bond, resulting in weaker hydrogen bonds network comparing to compound **2b**. What makes the difference in the case of derivative **2b_2** is the difference of P-loop's mobility between GSK3 β and CDK2 kinases.

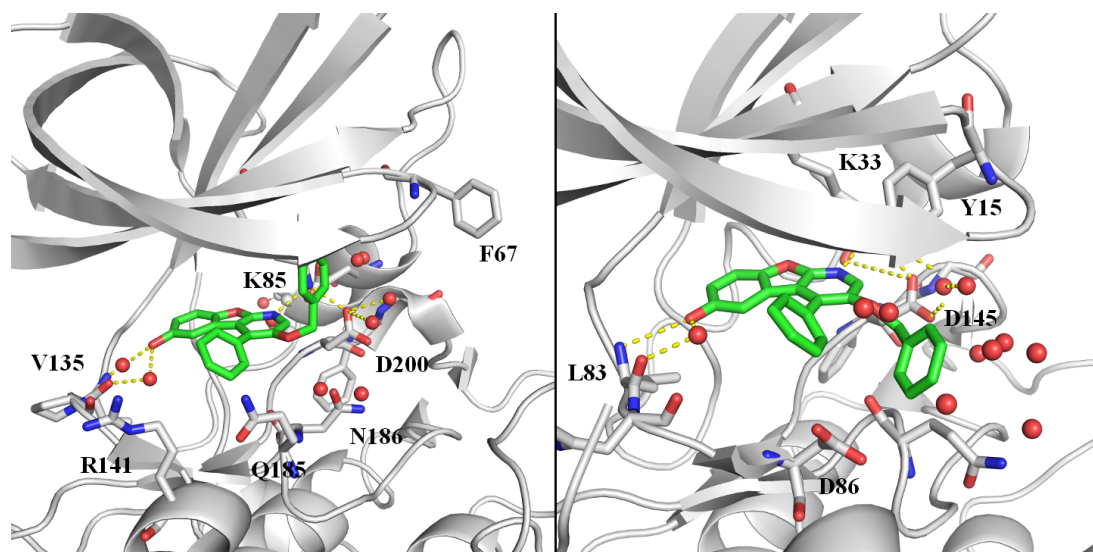


Figure 4.8: 10 ns MD simulation of compound **2b** complex with GSK3 β structure (PDB id: 1j1b, left) and CDK2 (PDB id: 2wih, right) shows a stable hydrogen bonds network, established between the ligand and protein with the help of water molecules.

In the case of **2b** and **2b_2** complex with GSK3 β , MD simulation shows up-movement of the P-loop, leading to increased P-loop/A-loop average distance, which reaches 13.15 Å for **2b_2**. Consequently, this increase in the P-loop/A-loop average distance makes the GSK3 β 's binding pocket more exposed to the bulk water. The calculation of the hydration map (water's occupancy map) around the complex of **2b_2**/GSK3 β refers again to the existence of a favorite hydration site beside the hinge region. The up-movement of the P-loop keeps the region beside the hinge more hydrated, making the impact of losing one hydrogen bond much less. According to LIE calculations, the desolvation penalty of compound **2b_2** is lower than the desolvation penalty of compound **2b**, explaining the increasing activity of compound **2b_2** with GSK3 β ($K_i = 20$ nM) comparing to compound **2b** ($K_i = 5.8$ μ M with GSK3 β).

On the other hand, when simulating **2b_2** with CDK2, the P-loop is aligning down inside the binding pocket, enhancing the pi-pi staking between the residue Y15 and the inhibitor, which in turn reduces the exposure of the binding pocket to bulk water and subsequently decrease the hinge region's hydration (Figure 4.9). The weaker hydrogen network in the hinge region and the less exposure to the bulk solvent might explain the reduction of **2b_2** derivative's binding affinity to CDK2 ($K_i \geq 1000$ μ M), comparing to more favorable binding affinity of derivative **2b** to CDK2 ($K_i = 6.4$ μ M).

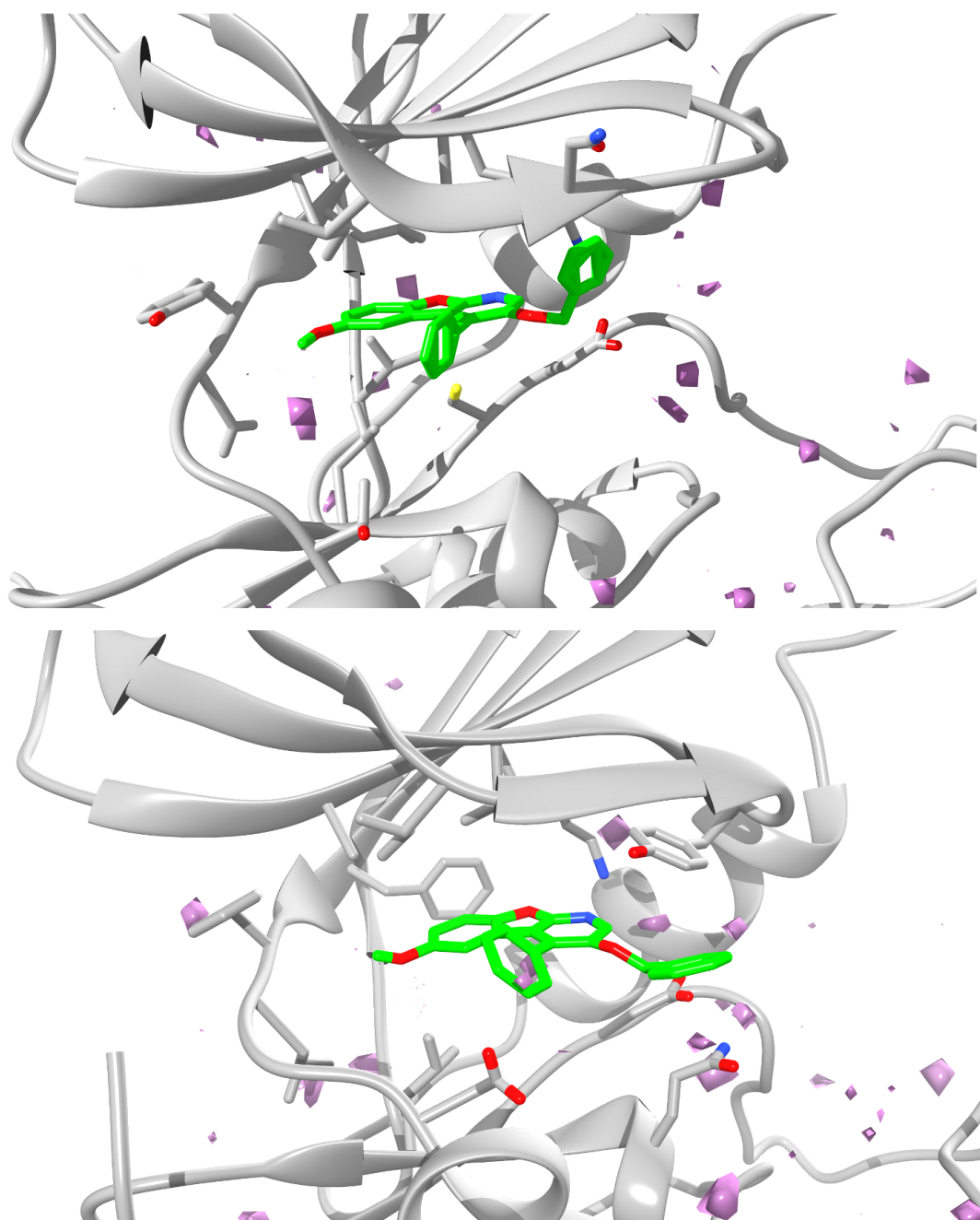


Figure 4.9: Hydration map for compound **2b_2** after 10 ns MD simulation with GSK3 β structure (PDB id: 1j1b, top) and CDK2 (PDB id: 2wih, bottom) with the highly occupied hydration sites. The hydration map (colored magenta) shows that the methoxy cannot keep a hydrogen bond network in the case of **2b_2**/CDK2, while the thermodynamically favorable hydration site, beside the hinge, will be kept in the case of **2b_2**/GSK3 β . The P-loop in the case GSK3 β moves up optimizing the pi-pi stacking with the inhibitor and consequently making the binding pocket more hydrated. The figure shows the average structures from the 10 nanoseconds trajectory for every compound presented along with contoured water oxygen atom density. The water oxygen density were calculated at 0.8-ps intervals over 10-ns trajectories into $(0.5\text{-}\text{\AA})^3$ grid elements over a 50 \AA^3 cubed grid centered at the ATP-pocket. The contours represent 175 hits per $(0.5\text{-}\text{\AA})^3$ grid element.

4.5 Conclusions

We have conducted a structure-based study together with MD simulations and physics-based binding energy methods in order to predict the binding mode and to understand the structural basis of 1-aza-9-oxa-fluorene derivatives' binding to two homologous kinases: GSK3 β and CDK2. LIE approach appeared to have advantage over implicit solvent methods (MM-PBSA), obtained by better accounting for the important water molecules' role in mediating the protein-ligand interactions. However, generating acceptable LIE model required careful consideration of the different induced-fit effects. The changes of P-loop/A-loop average distance appeared to be suitable parameter to estimate the induced-fit effects. Statistically significant and physically acceptable LIE models could be obtained, when the average P-loop/A-loop distances were considered for the compounds included in the model.

LIE calculations favor **binding mode 2** for both kinases GSK3 β and CDK2. In this binding mode, the substituent (R1) on the location **6** is located in the direction of hinge region, while the substituent on the location **3** (R2) is located in the phosphate binding pocket. Additionally, this binding mode allows a hydrogen bond between the pyridine's nitrogen and the flexible lysine. The substituent R1 (hydroxyl or methoxy) participates in making hydrogen bond network with the hinge residue, while different groups as substituent (R2) in the phosphate-binding region would make different contacts with the A-loop residues (especially the DFG's aspartate) and with the P-loop's aromatic residue; F67 (GSK3 β)/Y15 (CDK2). The different interactions with A-loop and P-loop result in different conformation changes, which can be estimated by measuring the distance between the P-loop's aromatic residue and the aspartate of DFG motif. Such conformational change significantly affects other properties, which play an important role in the protein-ligand binding; including the protein-ligand interactions, the protein's conformational energy, the protein hydration, and also the ligand desolvation penalty.

Chapter 5

Prediction of Binding Affinities for Kinase Inhibitors: Application on Mutant c-Kit D816V Kinase Inhibitors

5.1 Introduction

5.1.1 SCF/c-Kit Signaling Pathway: Role and Biological Importance

The Stem cell factor receptor (c-Kit or named CD117) is a member of the platelet-derived growth factor receptor (PDGFR) family from the group of receptor tyrosine kinases (RTK). It is known that c-Kit plays critical roles in regulating numerous aspects of cellular processes; like cell growth and survival, differentiation of germ cells and melanocytes, and maturation. It is mainly expressed in the hematopoietic system, in the gastrointestinal system, in germ cells, and in melanocytes. SCF/c-Kit plays a significant role in activating different signal transduction pathways: mainly PI3K-Akt signaling pathway, SRC-family kinases pathway, JAK-STAT signaling pathway, and the MAPK pathway, also known as RAS/RAF-MEK-ERK pathway [181]. The critical role of SCF/c-Kit signaling is in hematopoietic cells such as stem and progenitor cells, but its importance decreases after the differentiation of these cells. The only type of cells, which stays dependent on SCF/c-Kit pathway for growth and survival, are mast cells [181].

The c-Kit receptor tyrosine kinase consists of a ligand-binding extracellular domain, a transmembrane region, a cytoplasmic domain which contains the regulatory juxtamembrane region, and the tyrosine kinase domain. The extracellular domain is composed mainly of five immunoglobulin-like domains, which play a role in the kinase activation. Some mutations of

the c-Kit gene could lead to a loss-of-function, which could result in defects in hematopoiesis, melanogenesis, and gametogenesis (Figure 5.1) [182, 183].

The c-Kit ligands (e.g. the stem cell factor SCF) bind to the extracellular immunoglobulin-like domains of the c-Kit leading to a receptor homodimerization and activation of its intrinsic kinase activity. The autophosphorylation of specific tyrosine residues in the c-Kit kinase (such as Y568 and Y570 of the juxtamembrane domain) activates the kinase function leading to phosphorylation of some intracellular substrates triggering some downstream signaling pathways, like RAS-ERK pathway and PI3-Kinase pathway [181, 184].

Auto-regulation of the kinase activity is controlled mainly by a conformational change of the activation loop (A-loop). In the auto-inhibitory state, A-loop adopts an inactive conformation, interrupting the access of the cofactor (ATP) and the protein substrate to the kinase catalytic site. Upon the phosphorylation of some conserved tyrosine residues of the juxtamembrane domain, the A-loop adopts an active conformation, which facilitates the approach of the chelating complex (Mg^{+2}/ATP , magnesium ion with adenosine triphosphate) and substrates to their binding pockets in the kinase domain. It is proposed that Juxtamembrane (JM) region plays an autoinhibitory role, as JM domain interacts with the N-terminal lobe of the monomeric kinase domain causing the autoinhibition, until the ligand-induced dimerization leads to phosphorylation of JM residues and consequently to c-Kit activation [185].

The activation of protein kinases may gain an independence from external stimuli by different mechanisms. One of these mechanisms is specific mutation of different codons inside the encoding genes of the proto-oncogene kinases. For c-Kit, one of the most frequent mutations is the mutation of codon 816 of c-Kit kinase gene in exon 17, which results in substitution of residue D816 with valine, tyrosine, phenylalanine, or asparagine [186]. The mutation D816V has been often reported in most patients of known mastocytosis [187]. Another mechanism is a deletion in the juxtamembrane domain, which affects the regulatory role of the JM domain, which in turn results in a constitutive activation of c-Kit without binding to any stimuli or ligand.

5.1.2 Gain-of-Function Mutations and the Cancer Pathology

Many types of tumors have been associated with over-expression or activation mutation of c-Kit. It is estimated that 50-80% of the gastrointestinal stromal tumors (GISTs) have activation mutation in the expressed c-Kit, while 28-88% of the small cell lung cancer (SCLC) cell line is characterized by an over-expression of c-Kit. Activation (gain-of-function) mutations; like D816V, have been reported in many types of tumors: the mastocytosis, acute myeloid leukemia (AML), germ cell tumors, and gastrointestinal stromal tumors (GISTs) [181, 184]. Systematic

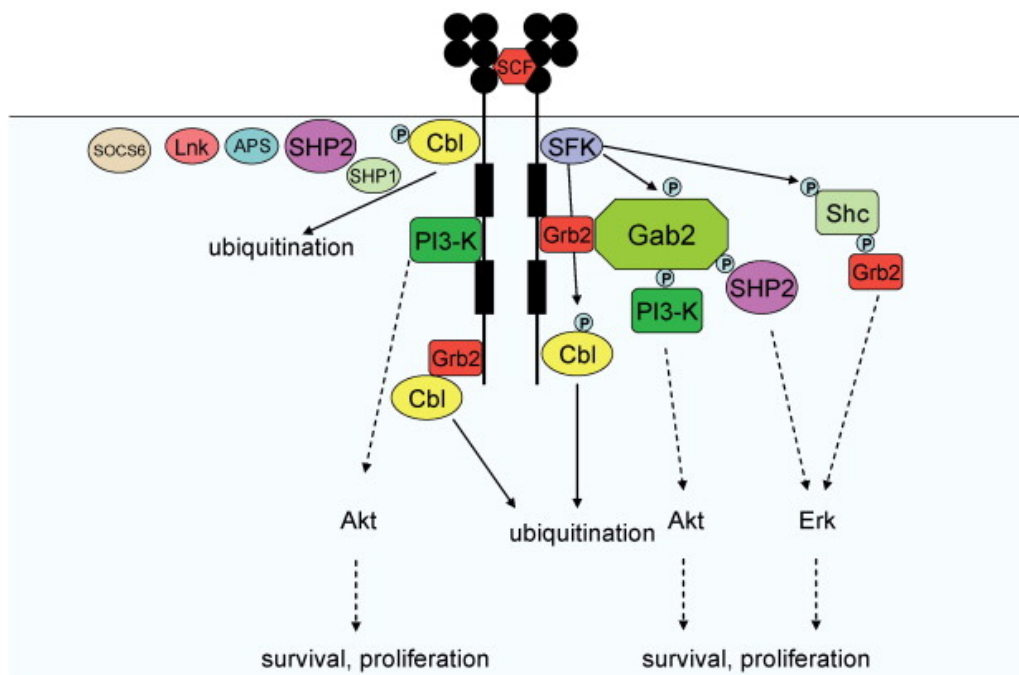


Figure 5.1: c-Kit signaling pathway, SCF indicates stem cell factor; STAT3, signal transducer and activator of transcription 3; K, tyrosine kinase domain. The MAPK pathway is denoted as RAS/RAF- MEK-ERK and the PI3K pathway as PI3K-AKT-mTOR (Adapted from Rönstrand *et al.* [183]).

mastocytosis (SM) is highly associated with mutations in exon 17 (codon 816); the mutation D816V is found in almost 90% of SM patients, beside other activating point mutations such as D816Y, D816F, and D816H (See Table 5.1).

Table 5.1: The common c-Kit Mutations in different Human Tumors

TUMOR TYPE	C-Kit MUTATION
Mastocytosis	D816V, D816Y, D820G, V560G
GIST	V559A, V559D, W557R, dup 502-503, various Δ 551-576
AML	Δ 418, Δ 419, Δ 418-419, D816V, D816Y
Sinonasal NK/T cell lymphoma	V825A, D816N
Germ cell tumor	D816H

Results from Rönstrand *et al.* also indicated that the signal transduction pathways, activated by the mutant c-Kit D816V, are different from the ones activated by the wild-type c-Kit [181, 184]. The mutation of the “gatekeeper residue”, which affects the size of the kinase active site, was also detected. An example of the gatekeeper mutation is the mutation T670I/E and T315I/F/D/N in c-Kit and Abl kinases, respectively.

The GIST tumor is mostly characterized by constitutively activated c-Kit, as a result of losing the autoinhibitory function. In the case of GIST, the activation of c-Kit is a result of

deletion or mutation in the juxtamembrane domain [188, 189]. In other types of tumors, like acute myeloid leukemia (AML) and systemic mast cell disease (SMCD), the activation of c-Kit is attributed to mutations in the proximity of the activation loop; mainly D816V and D816Y. While the first mutant activated form of c-Kit (the JM mutant form) usually shows good sensitivity to the famous kinase inhibitor **Imatinib** (GleevecTM), the second kind of mutations (Activation loop mutations) D6816V/Y and Y823D shows more resistance to the therapy by Imatinib [65, 83, 84, 190, 191, 192]. Indolinone derivatives, e.g. SU6577, SU11652, SU11654, SU11655, SU6668, and SU11248 (**Sunitinib**, SutentTM), have shown good activity in inhibiting some mutant forms of c-Kit; like T670I (gatekeeper mutation) [75, 193].

5.1.3 C-Kit Kinase Inhibitors: Activities with Mutant Forms

After the emergence of imatinib, it became common to categorize the known kinase inhibitors into two wide types:

- **known Type I c-Kit kinase inhibitors** include: (Figure 5.2a)

1. **PKC412** is a developed Staurosporine-derived tyrosine kinase inhibitor, which is found to be active inhibitor against PKC, KDR (VEGFR2), PDGFR α , Flt3, and c-Kit. PKC412 shows high activity against activated c-Kit with activation loop mutation (IC_{50} against mutant form D816V IC_{50} =33-95 nM, against the wild type c-Kit IC_{50} =138 nM, while against other mutations IC_{50} =146-365 nM) [184, 194, 195].
2. **Nocodazole** was originally identified as anti-microtubule agent, and then it was used in clinical cancer therapy. Nocodazole shows high affinity to some kinases like Abl, c-Kit, BRAF, and MEK. A special feature of Nocodazole is its activity against the gatekeeper mutant Abl T315A and the c-Kit T670I [196].
3. **Dasatinib** is more potent against the WT c-Kit and other Imatinib-resistant mutant types, including activation loop's mutations D816H/V, while it is 1000-fold weaker against the gatekeeper mutant c-Kit T670I [78, 86, 87, 88].

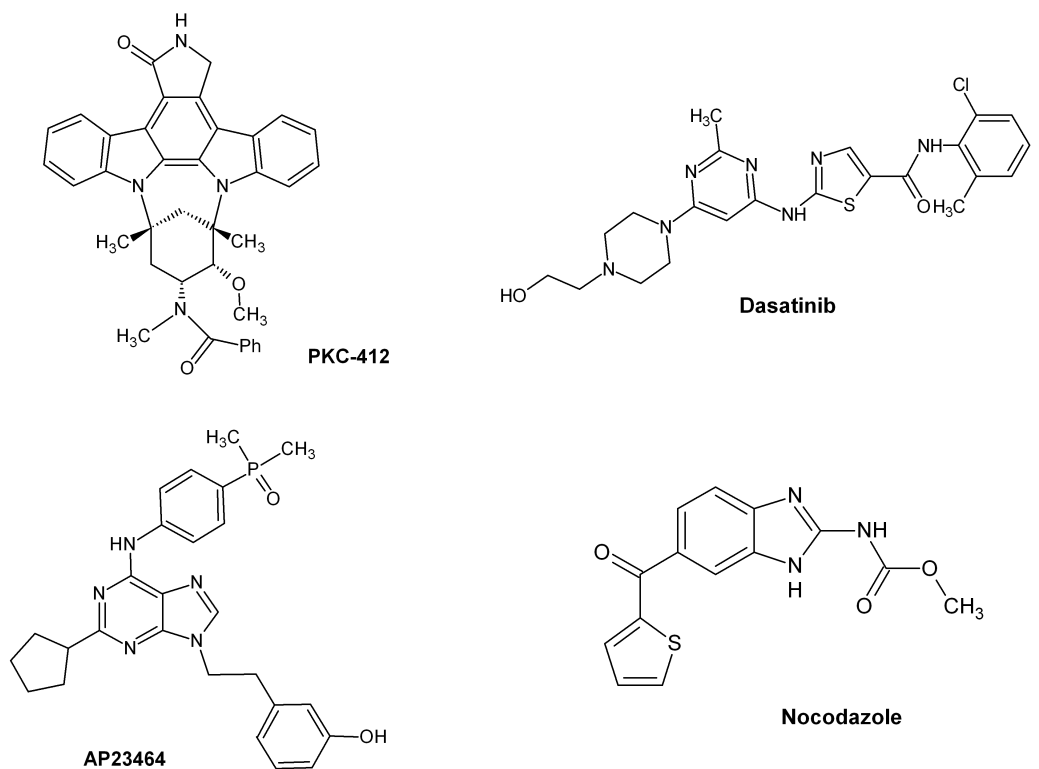
- **Type II c-Kit kinase inhibitors** include: (Figure 5.2b).

1. **Imatinib**: is the first developed type II kinase inhibitor (also known as GleevecTM (USA), GlivecTM (Europe) or STI571). It received FDA approval for treatment of CML and GIST, after it had shown good selectivity profile as potent inhibitor of the kinases BCR-Abl, c-Kit, and PDGFR. However, CML was found resistant to therapy with Imatinib, when the kinase BCR-Abl is mutated on some points in the P-loop (Y253F and E255K), mutation of the gatekeeper residue M315I, and the activation loop's mutation H396R [77]. GIST has also showed resistance to imatinib therapy in the cases of c-Kit mutations, mainly the gatekeeper mutation T670I, another mutation in the ATP binding pocket V654A, and mutation points in the Activation loop (D816V/H and Y823D) [65, 83, 84, 191, 197, 198, 199, 200].
2. **Sunitinib or SU11248** (SutentTM) was approved for treatment of advanced GIST in some Imatinib-resistant cases, as it is effective against some imatinib-resistant Kit mutant like the ATP-binding pocket's mutants V654A and the gatekeeper mutant T670I. However, Sunitinib couldn't inhibit the activation loop's mutants like D816H/V [75, 81, 82, 193, 201, 202, 203].
3. **Sorafinib** is a c-Kit inhibitor with good activity against all Imatinib-resistant mutant c-Kit, except the mutations that involve the codon 816 (D816V/H) [204]. Sorafinib has relatively better binding affinity to DFG-out conformation than imatinib, obtained possibly through double hydrogen bonds to bidentate carboxylate of the conserved glutamate in α C helix, and different binding mode which positions its aromatic rings away from the gatekeeper [205].
4. **Motesanib** was developed as type II kinase inhibitor, which binds in a binding mode similar to Imatinib's binding to DFG-out conformation (as the crystal structure of Motesanib with VEGFR2 has showed). Motesanib has good inhibitory potency against WT c-Kit and some Imatinib-resistant mutant c-Kit, such as the double mutant types (V560D/V654A and V560D/T760I), the juxtamembrane domain and extracellular domain c-Kit mutants (V560 D, Δ 552-559, and AYins503-504), and also one of the A-loop mutant (Y823D). Motesanib fails to inhibit the other A-loop mutant (D816V). That binding affinity profile of Motesanib refers that Y823D has less effect in the shifting of the dynamic conformational equilibrium towards the active conformation, in contrast to the case of D816V [206].
5. **Masitinib** is a novel inhibitor for c-Kit and PDGFR α/β with IC₅₀ of 200 nM and 540 nM/800 nM, with weak inhibition to Abl and c-Fms. It shows the highest activity in inhibiting only Kit and PDGF kinases, beside good inhibition of Lyn kinase (IC₅₀ = 510 \pm 130 nM compared to 2200 \pm 100 nM for imatinib), and to lesser extent FGFR3. Additionally, it is relatively weak against Abl kinase (Abl IC₅₀/Kit IC₅₀ = 6.0 for Masitinib

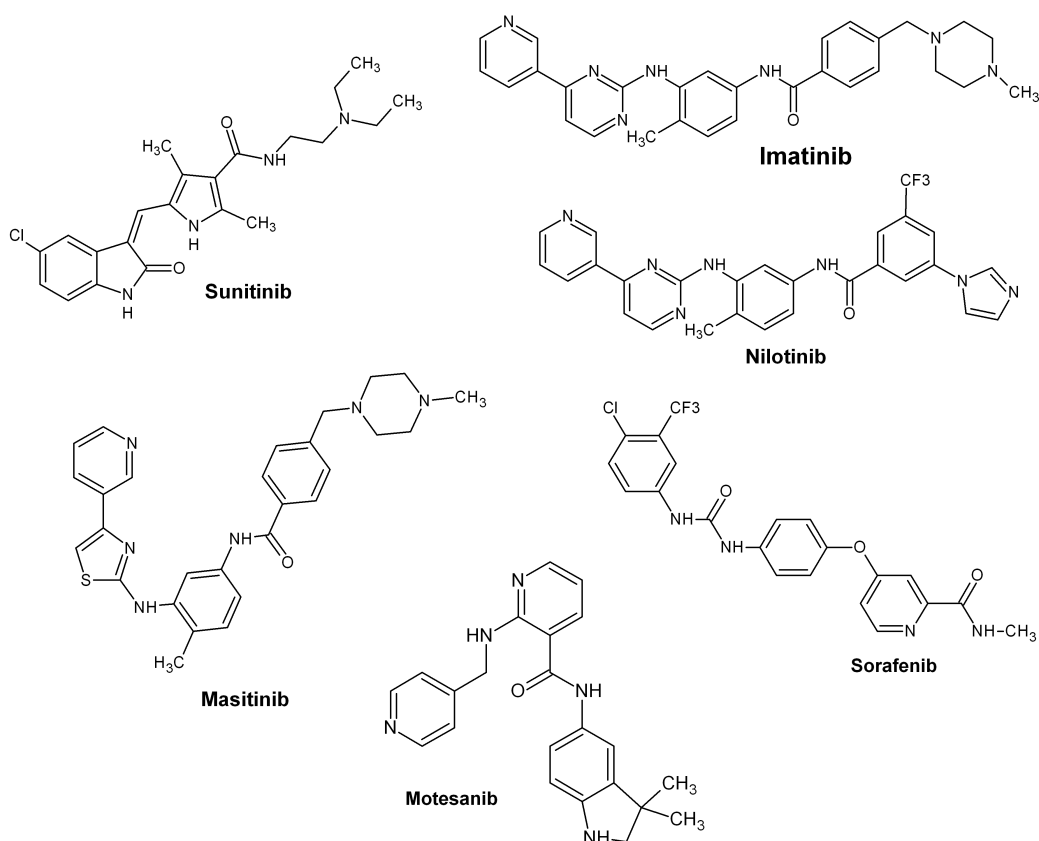
versus 0.6 for Imatinib), and also inactive against VEGFR [207]. It is at the moment in clinical phase 3.

5.1.4 Aim of this Study

An in-house database of 1215 synthesized compounds; comprising two congeneric series of anilino-oxazoles and anilino-thiazoles, is used in the current study to test the performance and accuracy of a variety of computational methods. The studied inhibitors were tested *in vitro* on the wild-type and D816V mutant of c-Kit (c-Kit D816V). To estimate the binding affinities of the inhibitors, MD simulations are carried out followed by extensive analysis of different scoring methods and physics-based binding free energy calculations. Since the whole dataset is too large for several computationally demanding binding free energy methods, several subsets are compiled. Different computational methods for predicting the binding mode and the activities of the inhibitors against c-Kit D816V are tested in order to find out the best model and the optimal method. Also, a detailed analysis of the factors, which affect the binding of this type of inhibitors to the active conformation of c-Kit, is performed by means of different computational methods.



(a) Chemical structures of the known type I c-Kit kinase Inhibitors



(b) Chemical structures of known type II c-Kit kinase Inhibitors

Figure 5.2: Chemical structures of known c-Kit kinase Inhibitors

5.2 Computational Methods and Application

5.2.1 Molecular modeling, Docking and Enrichment Studies

The crystal structure of c-Kit in its active conformation (PDB id: 1pkg) was used during this study to represent the mutant D816V c-Kit structure. The mutation D816V is located far away from the binding pocket and is only affecting the equilibrium between active and inactive conformations. Thus, the mutant c-Kit D186V is constitutively active and it is supposed to exist mainly in the active conformation.

To perform an enrichment study for optimizing the docking procedure with the mutant c-Kit D816V, a reduced dataset of 329 compounds; which contains only the very potent inhibitors and very weak inhibitors, was used in order to test the ability of docking protocol to discriminate between these two classes of activities. The dataset includes 122 highly active compounds (IC_{50} between 5 nM and 100 nM) and 207 weakly actives/inactives (IC_{50} above 10 μ M). The compounds of the dataset were docked to the prepared structure of c-Kit (PDB id: 1pkg) using GOLD4.1 with Goldscore as fitness function and different hydrogen bond constraints. The used constraints were hydrogen bonds to different residues at the binding pocket: two hydrogen bonds with the backbone NH and CO of the hinge residue C673, the backbone NH of D810 (the aspartate in the DFG sequence) or side chain of K623, carboxylic oxygen or backbone NH of D677. Enrichment ROC curves were generated for the above described dataset of 329 inhibitors.

Enrichment indexes were calculated: The first is the sensitivity (Se, true positive rate, = Number of Selected Actives/ Number of total actives), which is the ratio of the number of active molecules found by the virtual screening method to the number of all active database compounds. The second index of enrichment is the specificity (Sp, false positive rate, = Number of discarded inactives/Number of total inactives), which represents the ratio of the number of inactive compounds that were not selected by the virtual screening methods to the total number of inactives in the whole database. One of the most used methods currently to describe the enrichment is the receiver of operator curve (ROC), which describes the selectivity (Se) as a function of (1-Sp), which is the ratio of selected inactives (Number of selected inactives/Number of total inactives). As Sp is the ratio of discarded inactives to the total inactives, then 1-Sp is the ratio of the selected inactives, or in another words the selected decoys. The ROC curve is plotted by considering the different scores of actives as thresholds. For every threshold, the number of decoys and number of actives within this cut-off is counted. Then, the ROC curve is obtained as map of the distribution of actives and decoys according to their scores. By this method, we avoid the selection of arbitrary threshold by considering all Se and Sp pairs for each score threshold. The quality of the enrichment can be measured by the Area under the curve

(AUC); higher values represent more discriminative models.

5.2.2 Scoring Methods

5.2.2.1 X-score

X-score is an empirical scoring function developed in University of Michigan by S. Wang *et al.*, estimating the binding energy as follows:

$$\Delta G_{\text{bind}} = \Delta G_{\text{vdW}} + \Delta G_{\text{H-bond}} + \Delta G_{\text{deformation}} + \Delta G_{\text{hydrophobic}} + \Delta G_0$$

X-score is designed to be a consensus score by taking the average of three empirical scoring functions; HSScore, HMScore, and HPScore. The main difference between the three scoring functions is the implemented algorithm to account for the hydrophobic effect. HSScore uses a hydrophobic surface algorithm, HMScore uses hydrophobic matching algorithm, while HPScore uses hydrophobic pair contact algorithm [208].

5.2.2.2 AMBER GBSA/PBSA scoring after refinement

The docking complexes (c-Kit with the docking solutions) were prepared for calculating the interaction energy using the *tleap* module in AMBER tools 10. The parameters of the general Amber force field (GAFF) were used for the ligands with the semi-empirical method AM1-BCC for ligand's charges while the AMBER99SB force field atom-types and charges were used for the protein kinase. All inhibitor/protein complexes were minimized by the conjugate gradient algorithm to a root mean square of the energy gradient of $0.01 \text{ kcal.mol}^{-1}.\text{\AA}^{-1}$, and then simulated for 250 picoseconds with a GB implicit solvent model (A modified GB model developed by A. Onufriev, D. Bashford and D. A. Case; abbreviated as GBOBC) with modified bondi. A Perl script for PBSA was used for calculating the PBSA/GBSA on the final minimized complexes. In the PBSA/GBSA calculations, The van der Waals (E_{vdW}) and electrostatic (E_{ele}) interaction between ligand and protein in gas phase were calculated with an infinite cutoff using the SANDER module in AMBER10. The electrostatic free energy of solvation ($G_{\text{ele-sol}}$) was calculated with numerical solvation of the Poisson-Boltzmann (PB) equation as implemented in the mm-pbsa.pl module in AMBER10. Default parameters for the PB solver such as a grid spacing at 0.5 \AA , dielectric constants of 1.0 for solute and 80.0 for implicit PB solvent, solvent probe radius at 1.4 \AA , and ionic strength at 0 M concentration were used. The non-electrostatic free energy of solvation (G_{SA}) was calculated as linear function of the solvent accessible surface area (SASA); $\Delta G_{\text{SA}} = \gamma \cdot \text{SASA} + b$ where γ and b were set at the default values ($\gamma = 0.00542 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$ and $b = 0.92 \text{ kcal.mol}^{-1}$).

5.2.3 Molecular Dynamics Simulation

The crystal structure (1pkg.pdb) was used with the top ranked docking solution, after mutating the residue D816 to valine using MOE. Two chloride ions were added to neutralize the protein charges, and then the protein complex was centered inside cubic solvent box of TIP3P water molecule's model with buffer zone of 10 Å. The parameterization of the system was performed using *antechamber* and *tleap*; of Amber tools 11; exactly as explained in chapter 4 (4.2.3.1). After performing the minimization and the equilibration phase, also as described in chapter 4 (4.2.3.1), free molecular dynamics was performed for 6 nanoseconds with (2 fs) time step and with (9 Å) cut-off for the non-bonded interactions.

5.2.4 Free Binding Energies using Linear Interaction Energy and MM-PBSA methods

The MM-PBSA (molecular mechanics Poisson–Boltzmann surface area) methodology was applied on 25 snapshots obtained from the last nanosecond of the 6 nanoseconds MD trajectory, starting from c-Kit/docking complexes of the 200 top-scored anilino-oxazole compounds. The calculations were performed as described in chapter 4 for the CDK2/GSK3 β inhibitors (4.2.3.2).

To perform LIE calculations, A snapshot was taken from the last nanosecond of the MD simulation and selected to have a P-loop/A-loop distance equal to the average distance. The chosen snapshot was parameterized to be simulated in Gromacs4.5 suit. The parameterization of the small-molecule ligand was performed using Amber Tools10's *antechamber* to calculate the ligand partial charges according to AM1-BCC method and using gaff atom-types. The ACPYPE program was used to convert *antechamber*'s parameters to Gromacs topology and coordinates' files, while the protein kinase was parametrized according to AMBER99SB force field. The complex was solvated in octahedral box of water solvent in TIP3P model and then simulated for 500 ps. Later, the Gromacs' tools were used; *g_dist* was used to control the P-loop/A-loop changes during the simulation and *g_energy* for the calculation of different ligand-environment energies in the complex bound form. The details of calculations were mentioned in chapter 4 (4.2.3.3).

For LR-MM-PBSA (also named LIECE) models, the energy components were calculated from PBSA/GBSA using a single snapshot or using multiple snapshots from MD. The terms were used for fitting with the experimental binding affinities as pIC_{50} ($-\log IC_{50}$) or experimental binding energy ΔG_{bind} ($RT \ln IC_{50}$) according to the following equations:

1- three-parameter model with decomposed electrostatics: $(pIC_{50}) = \alpha \Delta E_{vdW} + \beta_1 \Delta E_{coul} + \beta_2 \Delta G_{solv}$

2- four-parameter model with decomposed electrostatics: $(pIC_{50}) = \alpha \Delta E_{vdW} + \beta_1 \Delta E_{coul} + \beta_2 \Delta G_{solv} + \gamma \Delta G_{SA}$

5.2.5 General Structure of the Studied c-Kit inhibitors

The whole dataset that was studied in the current work includes 1215 compounds, belonging to two series of anilino-oxazoles and anilino-thiazoles (Figure 5.3). A series of anilino-oxazoles compounds was previously reported as inhibitors of structurally related kinases: VEGFR1/2/3, SRC and PDGFR1 β [209]. Also, X-ray structures of two anilino-oxazole derivatives in a complex with VEGFR2 (PDB id: 1y6a, and 1y6b) have been reported, and were used in the current study to test the accuracy of the chosen docking procedure. In case of anilino-thiazoles, one compound (Masitinib, Figure 5.2b) has entered clinical phase 3 for the treatment of GIST.

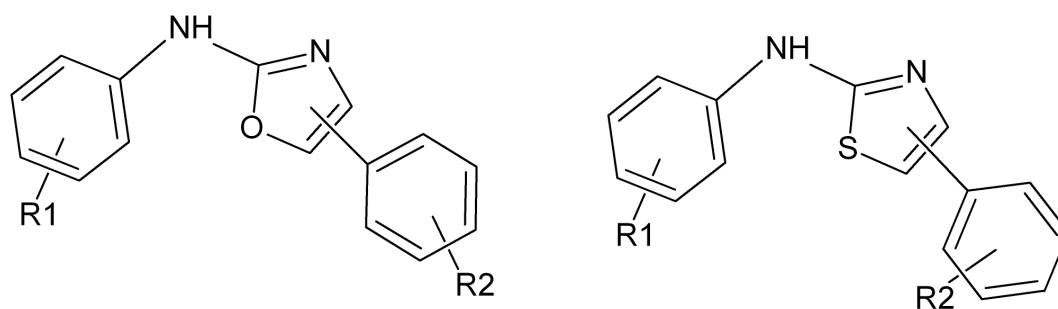


Figure 5.3: General structure of anilino-oxazoles and anilino-thiazoles derivatives in in-house database of c-Kit inhibitors, tested against the mutant c-Kit D816V.

5.3 Results

5.3.1 Docking Studies and Enrichment Assessment

Different docking setups including different sets of constraints were considered for the docking runs with the crystal structure of c-Kit in the active (DFG-in) conformation and the inhibitor dataset. To select the preferred binding mode, we generated ROC enrichment curves for every docking run using the above described dataset of 329 inhibitors; including 122 highly active compounds (IC_{50} between 5 nM and 100 nM) and 207 weakly actives/inactives (IC_{50} above 10 μ M).

The used docking constraints were derived from the c-Kit/ATP complex (Figure 5.4). The best results of the docking runs were obtained with collection of constraints with the GOLD program: Two protein hydrogen bonds to the hinge region residue: C673 (backbone NH and backbone CO), hydrogen bond to the backbone NH of D810 (one residue of DFG sequence) or the flexible lysine K623, and hydrogen bond to the carboxylic oxygen of D677 (Figure 5.6a). The best enrichment was obtained with the last mentioned settings with area under enrichment ROC curve (AUC=0.86) (Figure 5.5). A rescoring of all docking solutions was performed using X-score and PBSA/GBSA calculations. However, none of these rescoring methods gave any increase in the enrichment (i.e. AUC) or improvement in the discrimination between actives and inactives, compared to the simple Goldscore.

The optimized docking conditions were applied to the complete data set of 1215 compounds (anilino-oxazole and anilino-thiazole derivatives) in order to include compounds from all activity ranges: highly active, moderate active (886 compounds with IC_{50} between 100 nM and 10000 nM), and weakly active. The correlation between different scores and experimental activities for the complete dataset of 1215 compounds gave an r^2 between 0.19 for X-score to 0.27 for the MM-GBSA score (see Table 5.2a).

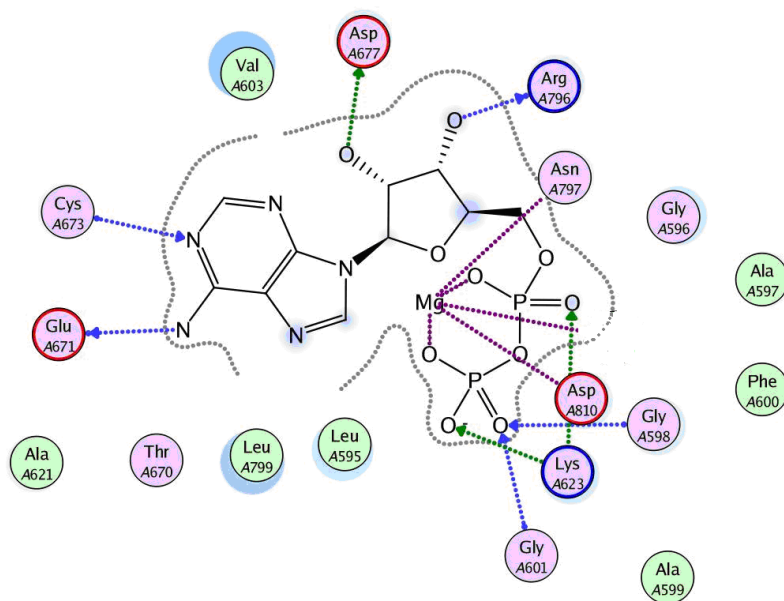


Figure 5.4: Schematic representation of the interaction of ATP at the binding pocket of c-Kit (PDB id: 1pkg). Interacting residues: C673, E671 (in the hinge region), D810 (from DFG motif), G601, the flexible lysine K623, D677 and R796 (from the ribose/binding region).

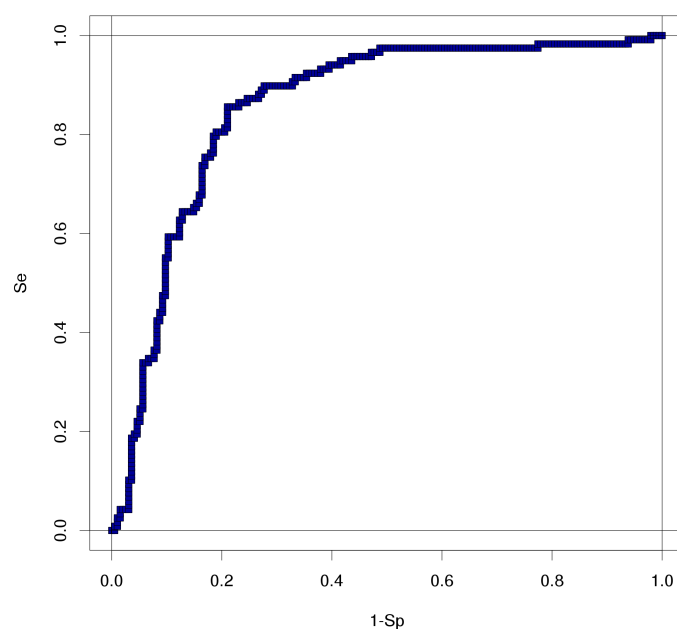
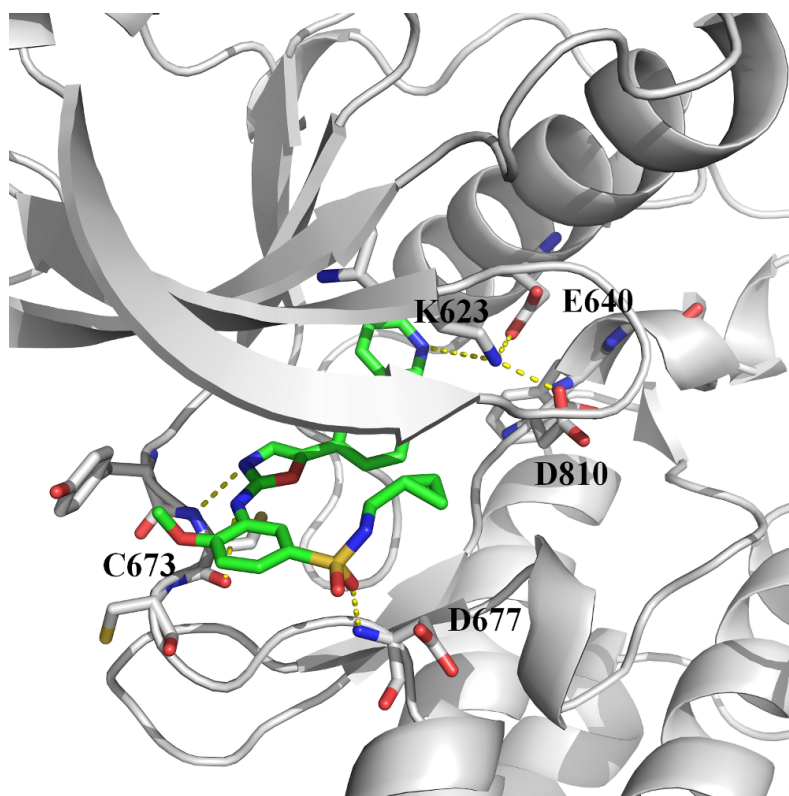
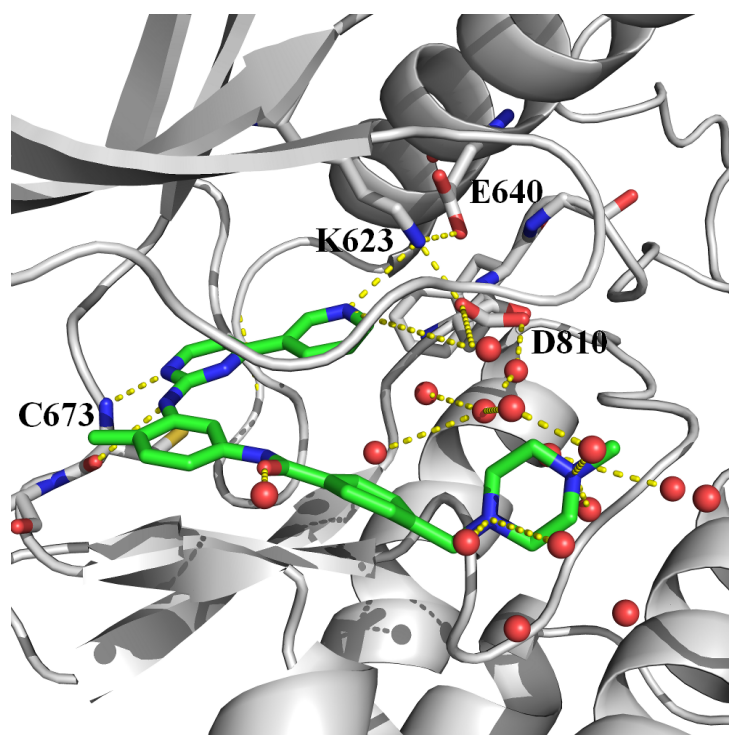


Figure 5.5: Enrichment ROC curve (GOLD score) for 122 actives and 207 inactive c-Kit D816V inhibitors, giving the best discrimination between active inhibitors (IC_{50} between 5 nM and 100 nM) and inactives (IC_{50} above 10 μ M) with Area under curve (AUC=0.86).



(a) Docking solution of an anilino-oxazole derivative **46** (from Harris *et al.* [209]) with *c*-Kit kinase (active conformation, PDB id: 1pkg). The same inhibitor was co-crystallized with VEGFR2 (PDB id: 1y6b) showing the same binding mode.



(b) Binding mode of Imatinib docked with *c*-Kit in the active conformation (PDB id: 1pkg) after 6 ns MD. Interacting water molecules forming a hydrogen bond network are shown as red balls.

Figure 5.6: Suggested Binding mode for anilino-oxazole derivative and Imatinib by docking to active conformation of *c*-Kit kinase.

5.3.2 Application of LR-MM-PBSA Method

First, different docking and scoring settings were tried for the complete data set of 1215 anilino-oxazole and anilino-thiazole derivatives, tested *in vitro* against c-Kit D816V. Neither for GOLD nor for other docking programs (results not shown) was there a significant correlation between docking scores and *in vitro* activities (given as pIC₅₀ values) (see Table 5.2a). Rescoring of docking solutions by MM-PBSA or MM-GBSA didn't improve the statistical results. We tested either a single protein-inhibitor complex that was minimized with MM-PBSA or MM-GBSA (1 snapshot) or multiple snapshots derived from MD simulation. MD simulation followed by MM-PBSA/GBSA calculations were done for a smaller data set of 110 top-scored compounds in order to test more sophisticated MM-PBSA binding free energy using multiple snapshots (Table 5.2b). In addition, we used the individual terms from MM-PBSA/GBSA calculations trying to establish linear regression based models.

1. Using one snapshot for MM-PB(GB)SA calculations: the following model was derived

$$\text{pIC}_{50} = 4.68 - 0.15 (\Delta E_{\text{coul}}) - 0.079 (\Delta G_{\text{gb}}) + 0.037 (\Delta E_{\text{vdW}}) - 0.21 (\Delta G_{\text{SA}})$$

$$r^2 = 0.35, \text{rmse} = 0.68, q^2(\text{loo}) = 0.34, \text{rmse}_{\text{loo}} = 0.76 \text{ (number of compounds 1215).}$$

2. Using multiple snapshots MM-PB(GB)SA calculations from MD simulation: the following model was derived

$$\text{pIC}_{50} = 6.16 - 0.28 (\Delta E_{\text{coul}}) - 0.162 (\Delta G_{\text{gb}}) + 0.00675 (\Delta E_{\text{vdW}}) - 0.003 (\Delta G_{\text{SA}})$$

$$r^2 = 0.20, \text{rmse} = 0.85, q^2(\text{loo}) = 0.094, \text{rmse}_{\text{loo}} = 0.922 \text{ (number of compounds 110).}$$

The results show that the consideration of more snapshots derived from the MD simulations did not improve the model accuracy.

5.3.3 Performance of Binding Energy Methods with Anilino-oxazole Derivatives

In order to test further computational methods, we subsequently used a smaller data set of compounds representing the whole 1215 inhibitor data set. 40 compounds were selected out of the 200 top-scored anilino-oxazole derivatives. The selected 40 compounds distribute over a range of IC₅₀ values between the nanomolar to the micromolar range, including moderately active inhibitors (IC₅₀ = 100-10000 nM). All compounds possess the same core (as shown for the anilino-oxazoles in Figure. 5.3); whereas the R groups vary among the different inhibitors. No anilino-thiazole was included in the smaller data set. The 40 selected compounds were docked and then used for MD simulation using the c-Kit D816V structure and PBSA/GBSA as well as LIE calculations as described earlier.

Table 5.2: Performance of scoring methods and implicit solvent's calculations for different subsets of c-Kit inhibitors.

(a) Correlation between different scoring values with the IC₅₀ values derived from testing against mutant c-Kit D816V.

Scoring method:	Num. cmp.	r ²	RMSE	q ² (LOO)	RMSE _{LOO}
Correlation between pIC ₅₀ and					
GOLD score	1215	0.26	0.74	0.24	0.77
X-score	1215	0.19	0.92	0.15	0.98
GBSA (GBTOT)	1215	0.27	0.66	0.25	0.64
PBSA (PBTOT)	1215	0.21	0.83	0.16	0.89

(b) Linear regression based models for 1215/110 c-Kit D816V inhibitors using energy components of MM-PB/GB-SA calculations.

Linear Model pIC ₅₀ with	Num. cmps	Model coefficients					Statistical values			
		Ele	vdW	GB/PB	Surf.	Const.	r ²	RMSE	q ²	RMSE ₁₀₀
GBSA (1 snap)	1215	0.15	0.037	0.079	0.21	4.58	0.35	0.74	0.34	0.77
PBSA (1 snap)	1215	0.20	0.045	0.036	0.34	5.84	0.27	0.84	0.25	0.95
GBSA (25 MD snapshots)	110	0.28	0.0067	0.16	0.003	6.15	0.20	0.85	0.094	0.92
PBSA (25 MD snapshots)	110	0.19	0.0084	0.24	0.055	7.24	0.15	0.89	0.085	0.99

Multiple linear regression models were generated for the 40 compounds using the *in vitro* activities and the individual energy terms from LIE or PBSA calculations. Neither the PBSA binding free energy (ΔG including the entropy calculation) nor the regression method with individual terms resulted in a significant correlation with the experimental data:

LR-MM-PBSA model:

$$\Delta G_{\text{bind}}(\text{pred.}) = -0.0065 (\Delta E_{\text{ele}}) + 0.01 (\Delta E_{\text{vdW}}) - 0.03 (\Delta G_{\text{pb}}) - 6.44$$

$$(r^2=0.096, \text{rmse}=1.15, q^2=0.013, n=40)$$

LIE model:

$$\Delta G_{\text{bind}}(\text{pred.}) = 0.082 (\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle) + 0.15 (\Delta \langle V_{\text{lig-surr}}^{\text{vdW}} \rangle) - 5.8$$

$$(r^2=0.277, \text{rmse}=1.01, q^2=0.23, n=40)$$

5.3.4 Analysis of the Binding Pocket's Flexibility and the P-loop Fluctuations

After performing 6 ns of MD simulation for all inhibitor/c-Kit complexes of the previously mentioned 40 selected compounds, a careful analysis of MD trajectories showed that different substituents at R1 result in different induced fit effects. The induced-fit effects were character-

ized by comparing the RMSD and the distance between CA of F600 (from the P-loop) and D810 (from the DFG motif in the A-loop). For most of the studied compounds, the P-loop/A-loop distance takes a value in the range 8.4 to 11 Å. In some rare cases, the conformational change was more dramatic and the distance between P-loop and A-loop reached 12-14 Å. Such a change of the kinase conformation affects not only the protein structure, but also the protein hydration. A comparison of similar anilino-oxazole compounds during the MD simulation showed that small differences at R1 or R2 (Figure 5.3) resulted in different movements of the P-loop (see Figure 5.7).

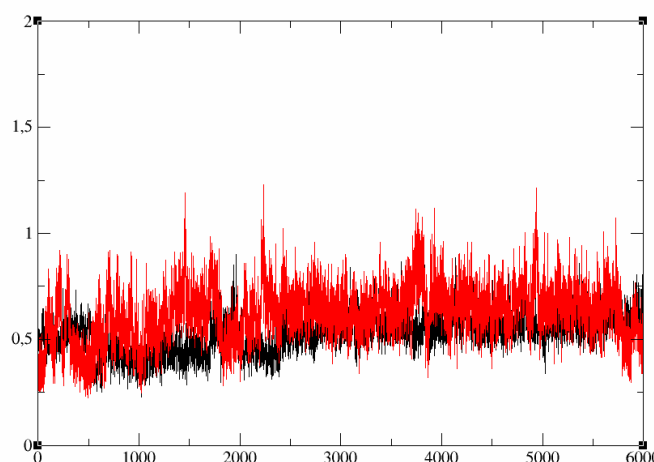


Figure 5.7: RMSD plot for the residues 596-603 (P-loop) of the c-Kit kinase (PDB id: 1pkg) during MD simulation with two anilino-oxazole derivatives, showing the RMSD of P-loop's fluctuation. The active anilino-oxazole derivative (in Black) shows less fluctuation comparing to weak anilino-oxazole derivative (in red) as a result of different interactions with the P-loop using their different substituents (R1 and R2).

To consider the different induced-fit effects and the changes of P-loop conformation, beside of the contribution of the bulk water molecules in the protein-ligand interactions, the methodology that has been applied to GSK3- β and CDK2 inhibitors (Chapter 4) was also used for studying the c-Kit D816V inhibitors. This methodology depends on developing different Linear Interaction Energy (LIE) models by applying the linear regression only on subset of compounds with similar P-loop/A-loop average distance.

5.3.5 P-loop/A-loop-distance-dependent LIE Models

To test the P-loop/A-loop LIE method, 40 further (anilino-oxazole) inhibitors were taken from the 200 top-ranked complexes (GOLD score) and then the total 80 compounds were di-

vided into training and validation (test) sets. The considered 80 compounds were grouped into four training and four validation sets depending on the average P-loop/A-loop distance observed during the 6 ns MD simulation. The following subsets were generated: compounds showing a P-loop/A-loop distance in the range (8.4 – 8.7 Å) were included in the first subset, (8.75 – 9 Å) in the second subset, (9.1 – 9.7 Å) in the third subset, and (10 - 10.5 Å) in the fourth subset. The *in vitro* activities for all training and validation set compounds span about 3 orders of magnitude. The four training sets include 37 compounds from the previously simulated compounds, while the four validations sets include 43 compounds from the previous mentioned compounds (See the detailed calculations in Appendix D, D.1). By applying linear regression using the derived LIE energy terms, statistically sound models were obtained. Models were generated for the training sets and were used to predict the activities of validation set compounds. The statistical parameters of these generated LIE models are summarized in Table 5.3a.

To perform fair comparison between LIE approach and LR-MM-PBSA method, LR-MM-PBSA models were generated for the individual subsets showing different P-loop/A-loop average distances (same subsets used in LIE method). In general, all models generated by linear regression using the PBSA energy terms resulted in only weak correlation. The statistical parameters of these generated LR-MM-PBSA models are summarized in Table 5.3b (See the detailed calculations in Appendix D, D.2).

LIE Model 1-1: generated from the compounds with distance between P_loop and A_loop = 8.4 - 8.7:

$$\Delta G_{\text{bind}}(\text{pred.}) = 0.10 (\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle) + 0.16 (\Delta \langle V_{\text{lig-surr}}^{\text{vdW}} \rangle) - 5.01$$

($r^2=0.66$, $\text{rmse}=0.81$, adjusted $r^2=0.49$, $q2=0.40$, $n=7$)

LIE Model 2-1: generated from the compounds with distance between P_loop and A_loop = 8.75 - 9.0:

$$\Delta G_{\text{bind}}(\text{pred.}) = 0.16 (\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle) + 0.45 (\Delta \langle V_{\text{lig-surr}}^{\text{vdW}} \rangle) + 1.56$$

($r^2=0.74$, $\text{rmse}=0.77$, adjusted $r^2=0.67$, $q2=0.55$, $n=11$)

LIE Model 3-1: generated from the compounds with distance between P_loop and A_loop = 9.4 - 9.9:

$$\Delta G_{\text{bind}}(\text{pred.}) = 0.047 (\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle) + 0.31 (\Delta \langle V_{\text{lig-surr}}^{\text{vdW}} \rangle) - 0.84$$

($r^2=0.628$, $\text{rmse}=0.70$, adjusted $r^2=0.52$, $q2=0.48$, $n=10$)

LIE Model 4-1: generated from the compounds with distance between P_loop and A_loop = 10 - 10.5:

$$\Delta G_{\text{bind}}(\text{pred.}) = 0.24 (\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle) + 0.21 (\Delta \langle V_{\text{lig-surr}}^{\text{vdW}} \rangle) - 7.5$$

($r^2=0.776$, $\text{rmse}=0.61$, adjusted $r^2=0.70$, $q2=0.65$, $n=9$)

Table 5.3: Summary of generated models derived from the different training sets selected depending on the final P-loop/A-loop average distance: depending on LIE calculations (top) and depending on MM-PBSA calculations (bottom). The four training set are donated as TS1, TS2, TS3, and TS4, while the validation sets are donated VS1, VS2, VS3, and VS4. Models generated using TS1, TS2, TS3, or TS4 are validated against VS1, VS2, VS3, or VS4 respectively, and vice versa.

(a) Summary of generated LIE models derived from the different training sets selected depending on the final P-loop/A-loop average distance. q^2 is the leave-one-out cross-validation correlation coefficient; while pred. r^2 is the correlation of predicted affinities against experimental ones for the model's external 'validation set'.

(LIE model) : $\Delta G_{\text{bind}}(\text{pred.}) = \beta(\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle) + \alpha(\Delta \langle V^{\text{vdW}}_{\text{lig-surr}} \rangle) + \gamma$										
Model #	P/A distance	Dataset	num cmp (n)	r^2	q^2	pred r^2 (valid.)	SE	γ	β	α
1-1	8.4 - 8.7	TS1	7	0.66	0.40	0.66	0.81	-5.01	0.10	0.16
1-2	8.4 - 8.7	VS1	10	0.80	0.67	0.16	0.35	-4.54	0.06	0.16
2-1	8.75 - 9	TS2	11	0.74	0.55	0.64	0.77	1.56	0.15	0.45
2-2	8.75 - 9	VS2	12	0.73	0.52	0.64	0.77	1.59	0.14	0.42
3-1	9.4 - 9.9	TS3	10	0.62	0.48	0.69	0.7	-0.84	0.047	0.31
3-2	9.4 - 9.9	VS3	12	0.76	0.59	0.56	0.52	-1.85	-0.002	0.25
4-1	10 - 10.5	TS4	9	0.77	0.65	0.45	0.61	-7.5	0.24	0.21
4-2	10 - 10.5	VS4	9	0.68	0.50	0.76	0.67	-9.11	0.24	0.125

(b) Summary of generated LR-MM-PBSA models derived from the different training sets selected depending on the final P-loop/A-loop average distance. r^2_{PBSA} is the correlation coefficient between PBSA_{tot} and the binding affinities as pIC_{50} ; $r^2_{\text{LR-PBSA}}$ is the model's correlation coefficient, while pred. r^2 is the correlation of predicted affinities against experimental ones for the model's external 'validation set'.

(LR-MM-PBSA model) : $\Delta G_{\text{bind}}(\text{pred.}) = \beta \Delta E_{\text{coul}} + \alpha \Delta E_{\text{vdW}} + \beta_2 \Delta G_{\text{pb}} + \gamma$											
Model #	P/A distance	Dataset	num cmp (n)	r^2_{PBSA}	$r^2_{\text{LR-PBSA}}$	pred r^2 (valid.)	SE LR-PBSA	β	α	β_2	γ
1-1	8.4 - 8.7	TS1	7	0.38	0.47	0.27	0.81	0.45	0.23	0.37	-1.26
1-2	8.4 - 8.7	VS1	10	0.35	0.49	0.32	0.61	0.08	0.11	0.05	-0.44
2-1	8.75 - 9	TS2	11	0.31	0.5	0.082	1.14	0.01	0.08	-0.02	-2.32
2-2	8.75 - 9	VS2	12	0.10	0.3	0.08	1.24	0.009	0.11	0.02	-2.81
3-1	9.4 - 9.9	TS3	10	0.01	0.23	0.24	1.09	0.01	0.11	0.02	-2.91
3-2	9.4 - 9.9	VS3	12	0.38	0.6	0.021	0.73	0.09	0.1	0.046	-2.38
4-1	10 - 10.5	TS4	9	0.00	0.43	0.20	0.77	-0.02	0.09	-0.03	13.5
4-2	10 - 10.5	VS4	9	0.42	0.46	0.0048	0.67	-0.08	-0.07	-0.06	3.34

5.3.6 Extended Validation Using Larger Dataset

For further validation, we decided to apply the four P-loop/A-loop-distance-dependent LIE models on all 200 top-scored compounds from the GOLD docking run. The activities of these 200 compounds spans over almost 4 orders of magnitude (pIC₅₀ from 8.3 to 4.6, $\Delta G_{\text{bind}} = -11.6$ to -7.2), including the previously considered 80 compounds from the training and validation sets. The 37 compounds of the four training sets were excluded, while the remaining 163 compounds were considered as external test set (including 43 compounds used in the previous validation set). The predictions of binding affinities were calculated using the generated LIE model: 1-1, 2-1, 3-1, and 4-1 (Table 5.3a).

All 200 top-scored compounds were processed according to the protocol mentioned in the methods (6 ns MD simulation and measuring the average distance between P-loop and A-loop during the simulation, and then calculating LIE energy terms for the prediction).

To validate the LIE models, the coefficient of determination and root mean squared error are calculated for the external validation sets according to the following equations:

$$r^2 = 1 - \text{SS}_{\text{res}}/\text{SS}_{\text{tot}} = 1 - \sum (y_i - y_{\text{pred}})^2 / \sum (y_i - y_{\text{mean}})^2$$

$$\text{RMSE} = \sqrt{(\sum (y_i - y_{\text{pred}})^2 / n)}$$

As y_i the experimental value, n is the total number of observations, y_{pred} is the predicted value for this observation according to the tested model, y_{mean} is the mean value of the actual values for all observations (i.e. $y_{\text{mean}} = \frac{\sum (y_i)}{n}$).

To have fair comparison between the different scoring methods, a further parameter was calculated. The prediction index (PI) suggested by D. Pearlman [114, 122] is helpful as a standard index, which estimates how well the prediction and the ranking of these different scores are. It could be considered as general index to compare different scoring methods together for their predictivity and ranking powers. The prediction index is defined, assuming that E is the experimental binding affinity of compounds and P is the predicted binding affinity, as follows:

$$\text{PI} = \sum_{j>i} \sum_i W_{ij} C_{ij} / \sum_{j>i} \sum_i W_{ij}$$

$$\text{With } W_{ij} = |E(j) - E(i)|$$

Either $C_{ij} = +1$ if $[E(j) - E(i)]/[P(j) - P(i)] > 0$ (correctly predicted; both differences should have similar sign),

$$\text{Or } C_{ij} = -1 \text{ if } [E(j) - E(i)]/[P(j) - P(i)] < 0 \text{ (wrongly predicted),}$$

$$\text{Or } C_{ij} = 0 \text{ if } [P(j) - P(i)] = 0 \text{ (completely random prediction).}$$

This prediction index ranges from -1 to $+1$ depending on of the model's prediction or the score's prediction considering two aspects: the right ranking of the pairs of all the compounds,

and the precise tracking of the order of the experimental values: +1 arises from perfect prediction, -1 arises from predictions that are always wrong, and 0 arises from a predictions that are completely random. Pearlman has used opposite conditions when he compared negative score or estimation of ΔG_{bind} against positive experimental values pK_i or pIC_{50} [114, 122]. In our study, we have compared always values with similar sign (Goldscore against pIC_{50} , PBSA against ΔG_{bind}).

To determine the wrongly predicted compounds, a two-sided 90% confidence interval is located between two lines (the fitness line $\pm 1.65 \cdot \text{RMSE}$). Thirty-seven observations appeared as prediction outliers depending on the following criteria ($|E(i)-P(i)| > 1.65 \cdot \text{RMSE} = 1.45$). After removing these 37 wrongly predicted compounds, the prediction (r^2) increases significantly to 0.56 (Figure 5.8a).

In summary, out of 163 compounds validation set, 123 compounds were predicted with ($r^2=0.56$, $\text{RMSE}=0.48$) and prediction index ($\text{PI}=0.7$). Table 5.4 compares the correctness and ranking power of three scoring methods: P-loop/A-loop-distance-dependent LIE models, PBSA score (PBSA_{tot}) as estimation of the binding energy (performed on 25 snapshots taken from 6ns MD simulation), as well as GOLD score.

The performance of the three methods was compared considering four different sets of compounds (Table 5.4):

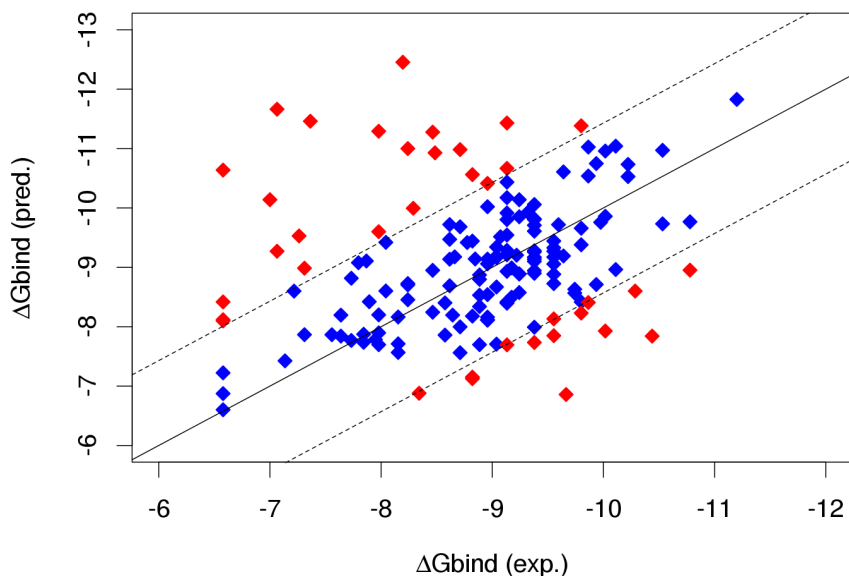
1. All the 37 compounds, which are included in the four training sets (TS1, TS2, TS3, and TS4).
2. All the 43 compounds, which are included in the four validation sets (VS1, VS2, VS3, and VS4).
3. 163 compounds, chosen from the top high-scored 200 compounds after excluding the compounds of training sets. Compounds in the four validation sets (VS1, VS2, VS3, and VS4) are included (Figure 5.8a).
4. 126 compounds set, after excluding the 37 outliers from the previous set identified as explained above (Figure 5.8b).

A simple comparison between the different indexes in (Table 5.4) shows that the application of the P-loop/A-loop-distance-dependent LIE models improves the prediction and the ranking of kinase inhibitors. The ranking for both PBSA binding energy and GOLD score was almost comparable and close to random, despite the longer time required by the MD simulation and MM-PBSA binding energy calculations.

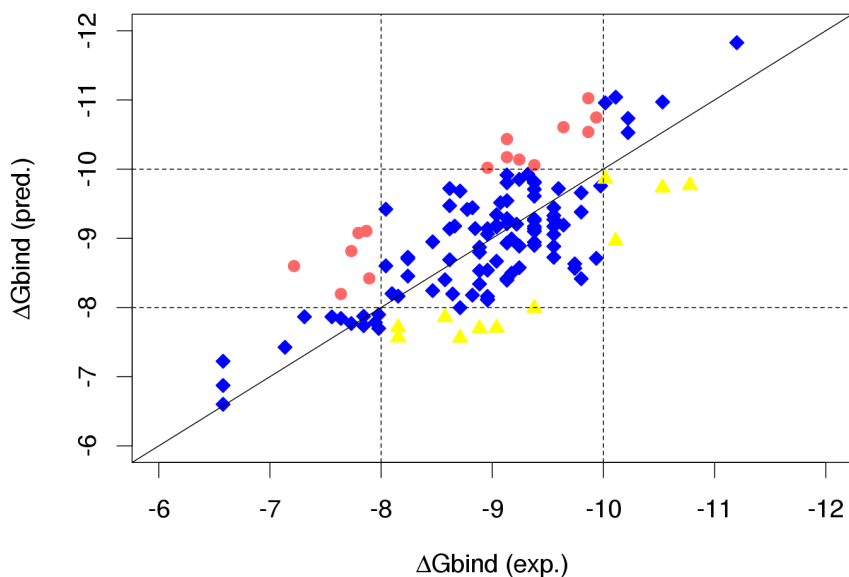
A structural analysis was made for the compounds, which appeared as prediction outliers according to the LIE models' prediction. The outlier compounds have special differences from the chemical structures of the compounds in the training sets. The most important differences could be classified in two categories: either having an aromatic ring fused directly with the aromatic ring A or having long alkyl chain with multiple hydrogen donating/accepting groups as R2 substituent, which significantly affects the conformation of P-loop.

Table 5.4: Comparison of the statistical parameters: (**predictive r^2 , RMSE and prediction index PI**) for three scoring and binding energy methods.

Cmp. set	Num. Cmp.	LIE (4 models) ($\Delta G_{\text{bind}}(\text{pred.})$ vs. $\Delta G_{\text{bind}}(\text{observ.})$)			MM-PBSA (PBSA _{tot} vs. $\Delta G_{\text{bind}}(\text{observ.})$)			GOLD Score (Goldscore vs. pIC ₅₀)		
		r^2	RMSE	PI	r^2	RMSE	PI	r^2	RMSE	PI
1	37	0.65	0.41	0.85	0.0009	6.19	0.04	0.08	5.04	0.2
2	43	0.59	0.45	0.78	0.08	8.7	0.21	0.07	7.11	0.15
3	163	0.07	0.95	0.33	0.085	8.5	0.08	0.009	3.9	0.06
4	126	0.56	0.48	0.7	0.095	8.23	0.10	0.046	3.4	0.15



(a) Prediction plot for 163 compounds depending on the P-loop/A-loop-distance LIE models. Red points refers to 37 points considered as prediction outliers (Wrongly-predicted compounds) outside two-sided 90% confidence interval (Fitness line $\pm 1.65 \cdot \text{RMSE}$). Fitness line was forced to pass from origin to compare between prediction and experimental values.



(b) Prediction plot for 126 compounds depending on four P-loop/A-loop-distance-dependent LIE models, after removing 37 prediction outliers (Prediction $r^2=0.56$, $\text{RMSE}=0.48$). If we divide the activity range into categories: highly active compounds (ΔG_{bind} between -12 and -10), moderately active (ΔG_{bind} between -10 and -8), and weakly actives (ΔG_{bind} between -8 and -6), 15 compounds are the predicted to have higher activity than actual (pink circles), and 11 compounds are predicted to have lower activity than the actual (yellow triangles). Fitness line was forced to pass from origin to compare between prediction and experimental values.

Figure 5.8: Prediction plots depending on four P-loop/A-loop-distance-dependent LIE models.

5.4 Discussion

5.4.1 Reliability of Suggested Binding Mode

The GOLD docking and enrichment assessment showed a preferred binding mode, which was able to differentiate between potent inhibitors (defined as with $IC_{50} = 1-100$ nM) and weak inhibitors (IC_{50} above $10 \mu\text{M}$) (Figure 5.5). The proposed binding mode of the most potent inhibitor is highly similar to the binding mode of the known inhibitor; Dasatinib, which makes two hydrogen bonds with the hinge region (Figure 5.6a). The same binding mode was also noticed in two related crystal structures of other two anilino-oxazoles derivatives co-crystallized with VEGFR2 kinase (PDB ID: 1y6a, and 1y6b) [209].

It is also note-worthy to mention that Imatinib binds in a similar “closed conformation” to un-phosphorylated/DFG-in active SYK kinase (PDB id: 1xbb) [71]. That confirms that Imatinib is able to bind to DFG-in conformation (Figure 5.6b), in spite that the high activity and selectivity appear only when it is bound to the inactive DFG-out conformation in elongated binding mode[198] (e.g. Imatinib’s structure with c-Kit in PDB id: 1t64). The crystal structure of SYK kinase with Imatinib gives more support to the ‘conformational selection’ mechanism in the binding of type-II kinase inhibitors, as other computational studies concluded [210].

5.4.2 Importance of P-loop Conformation and Fluctuations

The glycine-rich P-loop has recently been shown as an important factor for adjusting the kinase conformation and the accommodation of the kinase inhibitors inside the ATP binding pocket. Thus, the P-loop could also be considered as a determinant factor of the kinase inhibitor’s selectivity. A good example is the different binding affinity of Imatinib to some homologous kinases; like Abl kinase and c-SRC kinase [211]. For optimal binding of Imatinib to the DFG-out conformation, the P-loop appears to play a critical role by forming a kinked pocket, which perfectly interacts with Imatinib in Abl kinase [212]. While the P-loop’s residue Y253 in Abl kinase makes a favorite contribution to the binding in the case of Imatinib/Abl kinase, the corresponding aromatic residue in c-SRC kinase F278 has a negligible contribution in the Imatinib/c-SRC kinase interactions[211]. Some significant alterations in the P-loop conformation could also be noticed in some kinases’ crystal structures like ABL, ACK1, Aurora, c-Met, FGFR, and p38 kinases. These P-loop’s alterations are driven by special interactions with selective ligands/inhibitors. It is suggested that kinked conformations of P-loop could improve the selectivity of these co-crystallized inhibitors [79]. Some developed type I Abl-kinase inhibitors with high selectivity profile were complexed with Abl kinase showing a special P-loop conformation rather than the usual extended conformation [79].

In a research study published by C. Page *et al.* [115], different protocols of MM-PBSA

calculations failed to explain the selectivity of the kinase inhibitor SB203580 ([3H-imidazol-4-yl]pyridine derivative), which is active inhibitor against SABK2a/p38 and LCK kinases and weak inhibitor for GSK3 β and PKA kinases. The different potencies of this last inhibitor against different kinases could not be described by traditional MM-PBSA calculations. However, a striking observation was that the main fluctuations of the binding site occur in the glycine-rich P-loop: showing relatively small RMSF for P-loop residues in the case of SABK2a/p38 and LCK, and larger RMSF for GSK3 β and PKA. The last difference of P-loop fluctuations could be explained by the different contacts made by the kinase inhibitor with the aromatic residue Y34 of the P-loop in p38, but not with its equivalent phenylalanine in GSK3 β [115]. One of the remarkable results was that the natural ligand/inhibitor shows smaller fluctuations of the binding site (especially P-loop) when simulated with the native kinase [115]. We observed a similar effect in our MD simulations of the of anilino-oxazole compounds: c-Kit-inhibitor complex of potent inhibitors always gave smaller atomic fluctuation of the P-loop residues compared to the complexes of weak/inactive inhibitors (Figure 5.7).

5.4.3 Role of Water Molecules in the Binding Process

Many protein/ligand binding process might require reorganization of the water network around both the protein and the ligand. Water molecules can play an important role in mediating protein/ligand interactions by forming hydrogen bonds' bridges. The ATP binding pocket of the protein kinases is featured by high exposure to the bulk water, which could be the main reason of the important water contribution in the molecular recognition of protein kinases and their ligands/inhibitors. Considering this important role of the water molecules, we can describe the protein/ligand complex formation as a result of different molecular interactions; including hydrogen bonding, van-der-Waals interactions, and the enthalpy of the hydration. Enthalpy-entropy compensation could also play an important role in defining this important role of the water molecules [92].

The analysis of the MD trajectories of inhibitor/c-Kit complexes showed important contributions of water molecules' network in mediating the ligand/kinase interactions in 'the phosphate-binding region'; located between the DFG motif, the glycine-rich P-loop, and different substituents of the inhibitors. The MD results also showed that different substituents R1 and R2 would have an impact on the shape of the water molecules network, which interacts with both the inhibitor and the protein kinase.

A general problem for simulations in continuum solvent models (such as PB or GB) is the simplification that the system is divided into several regions having different dielectric constants (internal medium - binding pocket - with a dielectric constant between 1 and 4 depending on the system, and an external medium - the bulk solvent - with a dielectric constant of 80). The boundaries between these two mediums depend on calculating the molecular surface of the so-

lute (the enzyme or the protein), which in turn, depends on the used set of atomic van-der-Waals radii. In case of the studied inhibitors with c-Kit active conformation, the prediction of binding energies appeared to be a challenging task for implicit solvent methods. Two features of the studied inhibitors' binding to c-Kit D816V could explain this difficulty. The first feature is that a considerable part of these inhibitors; the substituents R1 and R2, can be solvent-exposed to some extent and not totally desolvated. The second feature is the participation of the water molecules as mediating bridges in the interactions between the protein kinase and the inhibitor (Figure 5.6b). In addition, the substituents change the water network in a different way depending on the chemical properties. The implicit solvent model is not able to describe these effects properly without an explicit inclusion of some water molecules [106]. However, this inclusion of explicit water molecules is problematic unless these water molecules are stable 'structural water molecules' with fixed position contributing similarly to all inhibitors of a congeneric series of compounds.

5.4.4 Effect of the Inhibitor's Substituents on the Final Stable Conformation

Linear interaction energy method can be considered as physics-based approach for determining the relative protein-ligand binding energy. For a congeneric series of compounds and a relatively rigid binding site, it is expected to have a comparable entropy term and equivalent 'induced fit' effects, upon the binding of a ligand to the receptor. In the case of c-Kit kinase, the high flexibility of the ATP binding pocket and its high exposure to the water could result in some complications for predicting the binding affinity. For the studied inhibitors, diverse substituents interacting in the selectivity region (Br-I) and the phosphate region (P) can be regarded as source of differences in the P-loop's fluctuations. The different interactions and the P-loop's movement result in consequently changes in the binding pocket's conformational energy and the protein's exposure to the solvent. Among the 37 prediction outliers, 18 compounds were detected to have long flexible chains with multiple hydrogen donating/accepting groups, which could alter the P-loop conformation significantly. The big margin of error in the prediction of these compounds can be attributed to larger changes in the conformation of the P-loop.

5.5 Conclusion

We have extensively analyzed the performance of different binding energy and scoring methods in order to find the optimal method for predicting the binding affinities for a series of c-Kit inhibitors. By Analyzing the MD simulations, a significant movement of loop regions flanking the ATP-binding pocket was observed, making the estimation of the binding affinities problematic. In the case of the active conformation of c-Kit, the P-loop conformation and positioning appeared again as determinant factor for the ATP binding pocket's shape and its conformational energy. The changes of the P-loop would be quite important to be considered in the case of every bound inhibitor, together with other flexible parts of the ATP pocket, in order to estimate the induced-fit effects. Two case studies have shown that the distance between two kinase loops; P-loop and A-loop, is a good parameter to estimate the induced-fit effects, and to classify a congeneric series of kinase inhibitors after a sufficient time of MD simulation. Another important factor is the role of water molecules in mediating the protein/ligand interactions. Our developed methodology, which depends on generating **P-loop/A-loop-distance-dependent LIE models**, was able to tackle these two problems of induced fit effects and the role of the water molecules as mediating bridges between the kinase and the inhibitors. 'P-loop/A-loop-distance-dependent LIE method' was also able to provide us with predictive models, which can explain the varying binding affinity. The usage of this method might help to design more selective kinase inhibitors in the future.

This page is intentionally left blank

Chapter 6

Summary

Targeted kinase inhibition emerged as a promising strategy for targeted therapy of different kinds of diseases, related to cell regulation and proliferation; as many of these cellular processes and signaling pathways are controlled mainly by protein kinases. On the other hand, the epigenetic mechanisms of regulating cellular gene expression were also found to be important in many types of cancer and diseases. In the first part of this work, we successfully applied a virtual screening strategy, based on screening a focused chemical library on one of the epigenetic enzymes; namely Histone Acetyltransferase (PCAF). Our work was successful in capturing new scaffolds for inhibiting PCAF; **pyridoisothiazolones** as new covalent PCAF inhibitors, and **rhodanine-carboxylic acids** as non-covalent PCAF inhibitors.

The second part of our work focused on developing selective kinase inhibitors. Kinase inhibitors should be designed to inhibit specific kinases in some cellular signaling pathway (on-targets) and prevent other homologous kinases (off-targets), whose inhibition could result in side effects or toxicity. With targeting the ATP-binding pocket of kinases, a problem of selectivity emerges. Beside the difficult issue of selectivity, developing resistance to the therapy by kinase inhibitors could appear, adding more difficulties to the obstacles in front of developing and designing selective inhibitors. One of the common mechanisms of resistance to kinase inhibitors is point mutations in different parts of the kinase domains. Targeting the DFG-out inactive conformation by type II kinase inhibitors faces the problem of developing resistance to its binding; either directly by mutations in the binding pocket, or indirectly by changing the balance between the kinase conformations (e.g. the mutations in the juxtamembrane domain or in the A-loop). The mutation D816V of c-Kit kinase is an example of gain-of-function mutations, which result in a resistance against type-II kinase inhibitors. One of the attractive strategies for overcoming the double challenge is by developing a hybrid type-I/type-II kinase inhibitor, which has the ability to inhibit both DFG-in and DFG-out conformations.

However, optimizing the selectivity and affinity of developed kinase inhibitors requires efficient methods for:

- 1) Predicting the binding mode of novel series of kinase inhibitors, without having an x-ray crystal structure of the kinase/inhibitor complex.
- 2) Predictive scoring to estimate the binding affinities for novel kinase inhibitors.

End-point free energy methods, such as Linear Interaction Energy and implicit solvents' methods, are computational methods based on the physical laws of molecular recognition and still less computationally demanding than the more rigorous physics-based methods (e.g. FEP and PMF calculations). In the second part of this work, we extensively analyzed the performance of end-point physics-based binding energy approaches and other scoring methods to find the optimal methodology for **predicting the binding mode and binding affinities for kinase inhibitors**. To handle the flexibility of protein kinases, Molecular dynamics simulations were carried out to estimate the degree of induced-fit effects upon the binding of inhibitors to their targets, while LIE and MM-PBSA methods were tested on different series of kinase inhibitors. On the other hand, the choice of the applied solvent model, whether explicit or implicit, should be made depending on the case study in order to obtain the best description of the studied system. In the case of kinase inhibitors' series studied in this work, the importance of water-mediated interactions gave a clear advantage to the explicit solvent model.

The high flexibility of the ATP binding pocket and its high exposure to water can be reasons of some complications in predicting the binding affinities of some kinase inhibitors series. As many researches pointed out to the importance of kinases' P-loop flexibility, we tried to examine its impact on the performance of binding energy methods. The conformation of P-loop appeared as determinant factor for kinase conformational energy and the binding pocket's hydration. Depending on the ligand/protein interactions in the phosphate-binding region, the conformation of P-loop and its distance from A-loop could vary considerably. Therefore, the distance between P-loop and A-loop can serve as a good descriptor for the induced-fit effects upon the inhibitor/kinase binding. A developed method, which is called **P-loop/A-loop-distance-dependent LIE models**, was able to tackle the problems of induced fit effects and the role of water molecules as mediating bridges between the protein kinase and its ligands. The last method was also successful in providing predictive models for the varying binding affinities of a series of c-Kit D816V kinase inhibitors. This work sheds a light on important features of kinase structures and important factors that affect the binding affinities of kinase inhibitors. We show that the right combination of computational tools could tackle the problems and challenges of predicting the most probable binding mode (the case study of 1-aza-9-oxa-fluorene derivatives with two homologous kinases: GSK3 β and CDK2) and also for obtaining predictive models for estimating the binding affinities (the case study of c-Kit inhibitors).

Part III

Supplementary Materials and References

Appendix A

End-point Free Energy Methods

The physics-based Free Energy methods could be categorized in two families: The first category is the rigorous, but computationally expensive, '**pathway MD methods**' [213, 171, 170, 214], including the alchemical double decoupling method, potential of mean force (PMF) method, replica-exchange-based free energy methods, or meta-dynamics-based methods, while the second category is the '**Endpoint free energy methods**' [215, 216, 217, 218], including the Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) model and the Linear Interaction Energy (LIE) method.

The rigorous physics-based methods or the **free energy perturbation (FEP) methods** depend on slow transformations between multiple states of the considered system [107, 219, 220]. In order to get convergent results, extensive conformational sampling is needed and a big number of interactions are to be calculated and considered. These heavy requirements and big number of computations make these methods very slow and computationally-demanding. Moreover, these alchemical transformations can pose serious problems, if the considered ligands are substantially different. **Endpoint free energy methods** are less accurate, but more computationally efficient because they evaluate only initial and final states of the system. This work concentrates on these last methods (End-point free energy methods); trying to compare their performance and application on different study cases related to the selectivity of kinase inhibitors.

A.1 Implicit-solvent-based Binding Energy methods

The **MM-PB(GB)SA** method depends on performing Monte Carlo (MC) or molecular dynamics (MD) simulation with explicit solvent, then estimating the enthalpic energy differences between the bound and unbound solute states using the Poisson-Boltzmann model or Generalized-Born model of implicit solvent. MM-PBSA was applied successfully by P. Kollman on a series of 12 TIBO-like HIV-1 RT inhibitors and biotin analogues with good agreement between the

results and experimental affinities [170, 221, 222, 223]. MM-GBSA was also successfully applied on diverse sets of inhibitors, including CDK2, Factor Xa, thrombin and HIV-RT inhibitors, obtaining a good correlation between its results and the experimental $\log(\text{IC}_{50})$ [111]. Similar success was noticed for GBSA method in predicting the relative binding affinities of a series of kinase inhibitors [112]. As MM-PB(GB)SA is computationally more efficient than rigorous binding free energy methods (e.g. FEP and TI), many research has been conducted to compare its performance to other computational methods, regarding predicting the free binding energy and its ability to recognize the right binding pose and correctly ranking the ligands [112, 113, 224, 225, 226, 227, 228, 229].

Two approaches are suggested for extracting the snapshots, which should represent the conformational space of the system's components; namely the single-trajectory approach and three-trajectory approach [172, 173]. In the single-trajectory approach, multiple snapshots of the system components: complex, receptor, and ligand are extracted from one trajectory generated by MC or MD simulation in explicit solvent. In the second approach, the snapshots are extracted from three trajectory generated from separated MD simulations of the complex, free receptor, and free ligand in a box of explicit solvent. The single trajectory approach is currently widely used in the research, because it gives less noisy results and helps for the cancellation of intramolecular interactions' effects, making the convergence of the results faster [230, 231, 232], although it neglects the conformational flexibility of unbound components. On the other hand, the three-trajectory approach shows big energy fluctuations; and needs more extensive sampling and longer simulation times [230]. The free energy is estimated by summation of the molecular mechanics energy (E_{MM}) of the complex estimated by the force field, the electrostatic solvation penalty estimated by Boltzman-Poisson model (G_{PB}) or Generalized Born model (G_{GB}), and the non-polar solvation term (G_{SA}), which is dependent on the solvent-accessible surface area (SASA). This estimation of the binding enthalpy could be added to an additional solute entropy term ($-\text{TS}_{\text{solute}}$) to give an estimation for the binding free energy.

$$G = E_{\text{MM}} + G_{\text{PB}} + G_{\text{SA}} - \text{TS}_{\text{solute}} = E_{\text{el}} + E_{\text{vdW}} + G_{\text{PB}} + G_{\text{SA}} - \text{TS}_{\text{solute}} \quad (\text{equation PBSA-1})$$

For the complex Protein-Ligand, we can estimate the binding energy as the difference between the free energy of the complex and the summation of the free energies of the free protein and the free ligand.

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}}) \quad (\text{equation PBSA-2})$$

A.1.1 Pitfalls and critical parameters of PB(GB)SA methods

The main shortcoming of implicit solvent methods is in the accounting for structural water molecules. Therefore, errors and pitfalls could arise, when interfacial water molecules exist inside the binding pocket, and participate in mediating interactions between the ligand and the protein [96, 113, 173]. Another pitfall arises, when there is considerable ligand-reorganization free energy [113, 232]. Another problem of MM-PBSA/GBSA method is its high sensitivity of some calculation parameters; PBSA could be very sensitive to the solute dielectric constant and the characteristics of the interface [113], while GBSA results is dependent on the used GB model and the used van-der-Waals radii [173, 233]. Moreover, the estimation of entropy is still challenging and represents the main source of uncertainty for determining the absolute free binding energy. The vibrational entropy is estimated usually using normal mode analysis or quasi-harmonic analysis [234, 235, 236, 237, 238]. However, both methods show pitfalls and problems in estimating the entropic contribution [222, 223, 113, 231, 234, 236, 237, 238, 239]. It is helpful to assume that entropic contributions are equal and canceled, when similar ligands of similar size with similar ligand-protein interactions are studied. This last assumption helps to make the PBSA calculations a better estimation for relative binding affinities, but not for the absolute ones [236].

The effect of the dielectric constant on the description of the electrostatic shielding was addressed in different studies [240, 241]. A model of variable dielectric constant based on residue types was developed for better description of protein-ligand electrostatics in MM-GBSA scoring [240, 241]. The protein desolvation term is also not optimal in the GBSA calculations and replacing it with a value extracted from explicit solvent method like WaterMap could improve the results [241]. Another uncertainty in the MM-PBSA calculation is the estimation of solvation free energy for extremely polar or charged ligand [113]. Additionally, the solvation estimation of the solvent-exposed residues or ligands could be source of another uncertainty [221, 113, 231, 233, 239, 242, 243]. This uncertainty becomes especially problematic, when we have part of the bound ligand exposed to the bulk solvent more than other buried parts, and consequently creating an inhomogeneous interior of the solute. This inhomogeneous screening of the electrostatic interactions couldn't be described using a single dielectric constant for the solute [233, 239]. This problem could be addressed by selecting a suitable value for the dielectric constant, which could be 1, 2, or 4 [221, 113, 231, 233, 239, 243]. Another issue is that the effect and role of structural water molecules could not be described efficiently by the implicit solvent models. In some cases, these stable structural water molecules should be included explicitly in the MM-PBSA/GBSA calculations [113, 244]. In some publications, it is suggested to include the whole first solvation shell in the calculation [245, 246]. However, that will add more complications to the procedure and some critical issues should be considered,

regarding the number of the included water molecules and to which system part (free complex or free ligand), the water molecule are associated. In other studies, it was suggested to combine the PBSA estimation for the receptor solvation with linear response approximation (LRA) in explicit solvent for the ligand solvation [247].

A.2 Linear Interaction Energy (LIE) Method

The linear interaction energy (LIE) is another end-point physics-based binding energy method, developed by J. Aqvist and tested for the first time on a set of endothiapepsin inhibitors [175]. The method relies also on molecular dynamics or Monte Carlo simulations for generating ensemble averages of two reference states of the system: the free ligand in solution, the bound complex receptor/ligand in solution [174]. LIE method depends, in principle, on linear response approximation to estimate the electrostatic contribution to the binding energy [248]. Linear response approximation was successful in its application on the ion solvation energetics, describing it accurately in previous researches [175, 249]. The non-polar part of the binding energy is supposed to be derived by scaling the Lennard-Jones energies using an empirically derived coefficient. That last assumption depends on the linear relationship between the solvation free energies for non-polar compounds and the solute size; estimated as solvent-accessible surface area (SASA) [174]. According to the LIE approach, the free binding energy could be approximated by this equation:

$$\begin{aligned} \Delta G_{bind} &= \Delta G_{polar} + \Delta G_{nonpolar} = \beta \Delta E_{ele} + \alpha \Delta E_{vdw} \\ &= \beta (\langle V_{lig-surr}^{el} \rangle_{bound} - \langle V_{lig-surr}^{el} \rangle_{free}) \\ &+ \alpha (\langle V_{lig-surr}^{vdw} \rangle_{bound} - \langle V_{lig-surr}^{vdw} \rangle_{free}) + \gamma \end{aligned} \tag{equation LIE-1}$$

Where $\langle \rangle$ denotes MD or MC averages of the non-bonded van der Waals (vdW) and electrostatic (el) interactions between the ligand and its surrounding environment (l-s), i.e. either the solvated receptor binding site (bound state) or just solvent (free state). The Δ denotes the change in these averages when transferring the ligand from solution (free state) to the receptor binding site (bound state). The parameters of this equation are the weight coefficients α and β for the non-polar and polar binding energy contributions respectively. An additional constant γ is possibly needed to enhance the description of the non-polar contribution [109, 250, 251]. The constant γ could be added to get reasonable binding energy [252], which is considered similar to the addition of SASA term introduced by Jorgensen *et al.*

Another general formula has been written in some publications [179] as

$$\begin{aligned} \Delta G_{bind} = & \beta_{prot} \langle V_{lig-surr}^{el} \rangle_{prot} - \beta_{wat} \langle V_{lig-surr}^{el} \rangle_{wat} \\ & + \alpha_{prot} \langle V_{lig-surr}^{vdw} \rangle_{prot} - \alpha_{wat} \langle V_{lig-surr}^{vdw} \rangle_{wat} + \gamma \\ & + \delta_1 (\langle V_{lig-lig}^{el} \rangle_{prot} - \langle V_{lig-lig}^{el} \rangle_{wat}) + \delta_2 (\langle V_{lig-lig}^{vdw} \rangle_{prot} - \langle V_{lig-lig}^{vdw} \rangle_{wat}) \end{aligned}$$

(equation LIE-2)

Where $\langle V_{lig-lig}^{el} \rangle$ and $\langle V_{lig-lig}^{vdw} \rangle$ are the electrostatic and van der Waals intramolecular ligand interaction energies, also referred to as strain energy, have been used in some publications for correcting the difference in ligand shape in the free and protein-bound simulations [179]. It is possible to use different terms from the last equation to obtain the suitable LIE model equation depending on the study case.

A.2.1 LIE parameters

The “linear response approximation” is well applicable for ionic solutes according to the calculation of Aqvist and Hansson. In some cases, significant deviations have been noticed, such as for neutral dipolar solutes containing hydroxyl groups, and solutes which can interact with the solvent by specific hydrogen bonding [176]. The derivation of the linear interaction energy method results in a generalized form, which allows different values for the electrostatic coefficient β (equal to 0.5, 0.33, or 0.29) for different sets of ligands [176]. It has been proven that the value of β is dependent on the charge and polar properties of the ligands [250, 179, 176]. Moreover, Application of LIE method to a set of charged neuraminidase inhibitor complexes obtained very low value of the electrostatic coefficient β [253]. This low value of β was explained by Aqvist, as a result of big number of charged residues in the binding pocket and the big number of polar groups of the inhibitors [174]. Similar low (β) value was obtained in developing LIE model for a series of MMP3 ligands which bind to zinc ion inside the binding pocket [179].

The van-der-Waals coefficient α appears to take also different values depending on the studied ligand-protein system. While the first publications of J. Aqvist suggested a transferable parameter of the van-der-Waals interaction as 0.16 or 0.18, other research found α with other values, like 0.32 or 0.34 for a set of FKBP12 complexes by Lamb and Jorgensen [254], 0.47 or 0.23 in the work of Jones-Hertzog and Jorgensen on thrombin inhibitors [255]. In a study of Paulsen and Ornstein on cytochrome P450/camphor analogue complexes, α obtained a value of (1.04) [256]. In a study of Wang et al, an estimation of α was suggested on the basis of desolvation surface areas, explaining the difference between the studies on trypsin and HIV protease (that gave α a value of 0.16 or 0.18) and the studies on thrombin, avidin and cytochrome P450cam complexes that required higher values of α [257]. The lack of polarization

effect in most of the current force-fields would lead probably to a higher value of α [174].

A.2.2 Critical issues in applying LIE

The LIE method neglects an explicit consideration of the internal energy's change and configurational entropy of both ligand and protein. Therefore, its success in many cases could be attributed to similarity of the compared ligands, and the cancellation between solute entropy drop and solvation entropy rise as part of Entropy-Enthalpy compensation [258, 259, 260]. The LIE method resembles the one-trajectory MM-PBSA method in the point of neglecting energy's change and configurational entropy. However, LIE still owns an advantage of exploring the configurational space of the ligand in its free state in solution. Moreover, Aqvist argued that intra-molecular terms like intra-molecular relaxation/strain, entropy, receptor desolvation, etc., are not neglected but embedded in the linear response approximation or in the LIE parameterization [175, 174].

Dealing with charged groups in the LIE methods could be somehow critical or problematic to get accurate description of the electrostatic contribution to the free energy. A critical point is how to get equal net charge of the surrounding medium around the ligand within its interaction range (e.g. a possible cutoff) in both states; the bound and free states, if the protein has a non-zero charge within the interaction sphere, or ions must be added in the free (solvent) simulation [250, 261]. Cutoffs of the electrostatic interactions could also lead to an artificial over-polarization of the surrounding of a charge. Therefore the solvation energy of the charge is usually overestimated in water more than in the protein, resulting in an artificial "anti-binding" contribution [262]. Performing the calculations in a big simulation system with higher cutoff could solve these problems, and the Local Reaction Field (LRF) method for treating the electrostatic during the simulation is also recommended by J. Aqvist, in order to avoid using cutoffs [261, 174].

One of the important notes, when the LIE calculations are performed, is that turning off the net charges on distant ionized groups would need a small electrostatic correction term for these neutralized charges should be added to the calculated binding free energy, according to equation:

$$\Delta\Delta G_{\text{el corr}} = 1/4 \pi \cdot \epsilon_0 \sum \frac{q_p \cdot q_l}{\epsilon \cdot r_{p-l}}$$

Summation is performed for all $p \in$ neutralized ionic residues, $l \in$ ligand atoms. Where q_p is the formal charge of the residue that has been neutralized, q_l is the partial charge of the ligand atom, ϵ is the dielectric constant which was set to 80, and r_{p-l} is the distance between the ligand atom and a central atom of the charged group of the residue. However, when selecting the neutralized residues far from the binding pocket, i. e. $r_{p-l} > 15 \text{ \AA}$, then $\Delta\Delta G_{\text{el corr}} < 6.63 \cdot 10^{-5} \text{ kcal.mol}^{-1}$.

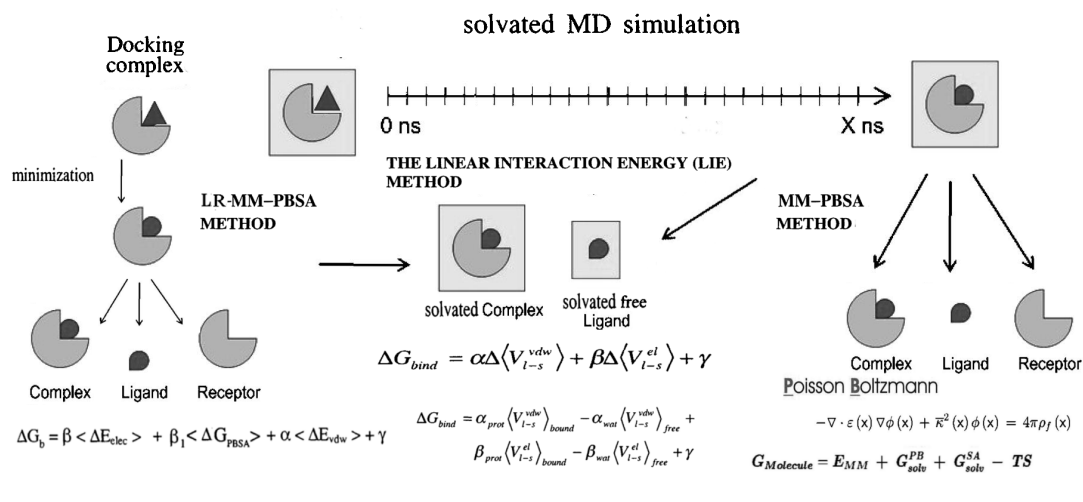


Figure A.1: General scheme represents the general procedure for performing three end-point physics-based binding energy methods

A.3 Hybrid Physics-based Methods

The performance of both explicit solvent models and implicit solvent models in LIE method was compared by J. Carlsson *et al.*, showing good agreement between the explicit solvent and PB model for a small set of malarial aspartic protease's inhibitors (plasmepsin II inhibitors). The last study asserts the importance of choosing the right van-der-Waals radii and the GB model to get comparable results [263]. Additionally, Another comparison study was made between multiple varieties of end-point binding energy methods, combined with implicit solvent models [215], including hybrid methods combining PB(GB)SA calculations with LRA and LIE. This research work showed conflicting results for the MM-PBSA and MM-GBSA methods when treating charged ligands. The precision and accuracy of the methods prediction are dependent on the receptor/ligand system [215, 264]. Specific models have been derived theoretically from linear response theory depending on the fact, that receptor and the solvent respond differently, when the solvent is treated implicitly [264, 265].

LR-MM-PBSA method emerged later as a combination of the linear response approximation (LRA) with the MM-PBSA method for estimating the binding affinity [266, 267, 268]. It depends mainly on combining scaled MM-PBSA energy terms with LR optimization of coefficients against known activity. This method was applied by P. Kolb *et al.* [269], and was called Linear Interaction Energy with Continuum electrostatics (LIECE). The models are developed depending on different equations for estimating the experimental free energies of binding (ΔG) = $RT \ln(IC_{50})$. The used equations give one parameter, two-parameters or three-parameter model, as follows:

1. one-parameter model:

$$(\Delta G) = \alpha \Delta E_{\text{vdw}} \quad (\text{LR-MM-PBSA-1 equation})$$

2. a two-parameter model with continuum electrostatics:

$$(\Delta G) = \alpha \Delta E_{\text{vdw}} + \beta \Delta G_{\text{ele}} \quad (\text{LR-MM-PBSA-2 equation})$$

3. three-parameter model with decomposed electrostatics:

$$(\Delta G) = \alpha \Delta E_{\text{vdw}} + \beta_1 \Delta E_{\text{ele}} + \beta_2 \Delta G_{\text{solv}} \quad (\text{LR-MM-PBSA-3 equation})$$

In the publication of P. Kolb *et al.* [269], 50 oxindole-based compounds from a series of LCK/CDK2 kinase inhibitors, 23 O6-substituted guanine derivatives as CDK2 kinase inhibitors, and 41 dihydroquinazolinone derivatives as p38 MAP kinase inhibitors were docked manually to native X-ray structures. Structure (1ke5.pdb) was used as crystal structure of CDK2 complexed with Oxindole-Based compound, while structure (1m7q.pdb) was used as crystal structure of p38 MAP kinase complexed with dihydroquinazolinone compound. Also in the study of K. Wichapong *et al.* [270], 174 pyrrolocarbazole-dione derivatives (pyrrolo[3,4-c]carbazole-1,3(2H,6H)-dione) and 48 pyridopyrimidinones derivatives (2-anilio-6-phenylpyrido[2,3-d]pyrimidin-7(8H)-ones) were docked to a crystal structure of WEE-1 kinase (1x8b.pdb), which is complexed with PD0407824 (a potent Wee-1 kinase inhibitor and pyrrolocarbazoledione derivative). It is expected that the success of the LIECE transferable models depends mainly on using a crystal structure of the same kinase complexed with a ligand from the same series, and showing the same contacts with the hinge region and the activation segment.

Appendix B

Novel Inhibitors of Histone Acetyltransferases

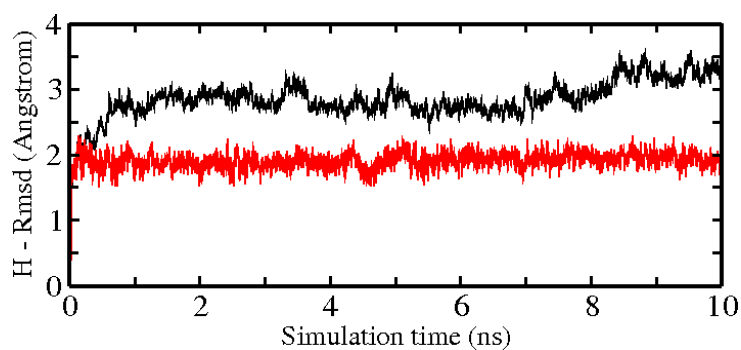


Figure B.1: Heavy atom root mean square deviation (H-Rmsd) of PCAF protein (black) and ligand (12e, red) conformations sampled during 10 ns MD simulation with respect to the initial structure versus simulation time.

Table B.1: Virtual Screening Hits with Docking Scores

Cmp. code	GOLD score	IC ₅₀ (PCAF, H3 _{aa1-21})
PHAR037680	55.86	97.7 ± 10.4 μM
T0505-1441	61.34	12% @ 50 μM
BAS02167318	51.84	8% @ 50 μM
BAS0056484	56.34	n.i. @ 50 μM
PB-06479073	63.98	n.i. @ 50 μM
PB-06779828	51.83	8% @ 50 μM
7216540307	58.49	5% @ 50 μM
7714220145	64.85	14% @ 50 μM
F1691-2672	58.37	5% @ 50 μM
T5214023	64.28	n.i. @ 50 μM
PHAR011220	61.36	15% @ 50 μM

Appendix C

Predicting The Binding Mode For Kinase Inhibitors

C.1 GSK3 β X-ray Structures

Table C.1: Available GSK3 β X-ray Structures in the Protein Data Bank

PDB id	Resolution	GSK3 complex with	P-loop/A-loop dist. (Å)	R141 orientation
1J1B	1.80 Å	AMP-PNP	10.93	inwards
1Q5K	1.94 Å	Aminothiazole (AR)	9.34	outwards
4AFJ	1.98 Å	5-aryl-4-carboxamide-1,3-oxazoles	9.89	outwards
1J1C	2.10 Å	ADP	11.36	inwards
1Q41	2.10 Å	Indirubin-3'-monooxime	9.53	outwards
1Q3D	2.20 Å	Staurosporine	11.25	outwards
3DU8	2.20 Å	Pyrrolopyridinone	9.81	outwards
4ACC	2.21 Å	sulphonamide-pyrazine	9.81	inwards
1R0E	2.25 Å	3-indolyl-arylmaleimide	7.22	outwards
1Q3W	2.30 Å	Alsterpauellone	10.71	outwards
3I4B	2.30 Å	Pyrimidylpyrrole	12.90	outwards
2JLD	2.35 Å	Ruthenium complex	7.54	outwards
3ZRK	2.37 Å	2-(4-pyridyl)Thieno-Pyridinones	10.12	outwards
1O9U	2.40 Å	ATP (Axin peptide)	11.21	inwards
1PYX	2.40 Å	AMP-PNP	12.13	inwards
3F7Z	2.40 Å	1,3,4-oxadiazole	11.31	outwards
3GB2	2.40 Å	1,3,4-oxadiazole	10.83	outwards
3ZRL	2.48 Å	2-(4-pyridyl)Thieno-Pyridinones	10.15	inwards
3ZRM	2.49 Å	2-(4-pyridyl)Thieno-Pyridinones	10.42	inwards
4ACD	2.60 Å	sulphonamide-pyrazine	9.83	inwards
4ACG	2.60 Å	sulphonamide-pyrazine	8.84	outwards
4ACH	2.60 Å	sulphonamide-pyrazine	8.94	outwards
1GNG	2.60 Å	- (Fratide peptide)	11.97	inwards
3F88	2.60 Å	1,3,4-oxadiazole	7.92	outwards
1I09	2.70 Å	-	11.63	outwards
1Q4L	2.77 Å	Anilino-maleimide (I-5)	9.77	inwards
1H8F	2.80 Å	-	11.33	inwards
1UV5	2.80 Å	Indirubin-3'-monooxime	9.08	outwards
2OW3	2.80 Å	Bis-indolyl-maleimide	7.45	outwards
3L1S	2.90 Å	3-Aryl-1H-pyrazole-5-one	9.50	outwards
2O5K	3.20 Å	benzoimidazole	8.59	inwards

Table C.2: RMSD (root mean squared deviation) matrix between some studied GSK3 β X-ray structures after superposition.

RMSD	1j1b	1q5k	1j1c	1q41	1q3d	3du8	1r0e	1q3w	3i4b	2jld
1j1b	0.0	0.35	0.15	0.57	0.69	0.31	0.68	0.61	0.50	0.51
1q5k	0.35	0.0	0.46	0.40	0.78	0.40	0.71	0.56	0.46	0.47
1j1c	0.15	0.46	0.0	0.63	0.64	0.31	0.67	0.60	0.53	0.54
1q41	0.57	0.40	0.63	0.0	0.84	0.60	0.86	0.50	0.45	0.50
1q3d	0.69	0.78	0.64	0.84	0.0	0.65	0.67	0.58	0.73	0.66
3du8	0.31	0.40	0.31	0.60	0.65	0.0	0.60	0.56	0.54	0.45
1r0e	0.68	0.71	0.67	0.86	0.67	0.60	0.0	0.74	0.80	0.56
1q3w	0.61	0.56	0.60	0.50	0.58	0.56	0.74	0.0	0.49	0.51
3i4b	0.50	0.46	0.53	0.45	0.73	0.54	0.80	0.49	0.0	0.46
2jld	0.51	0.47	0.54	0.50	0.66	0.45	0.56	0.51	0.46	0.0

C.2 CDK2 X-ray Structure (2WIH)

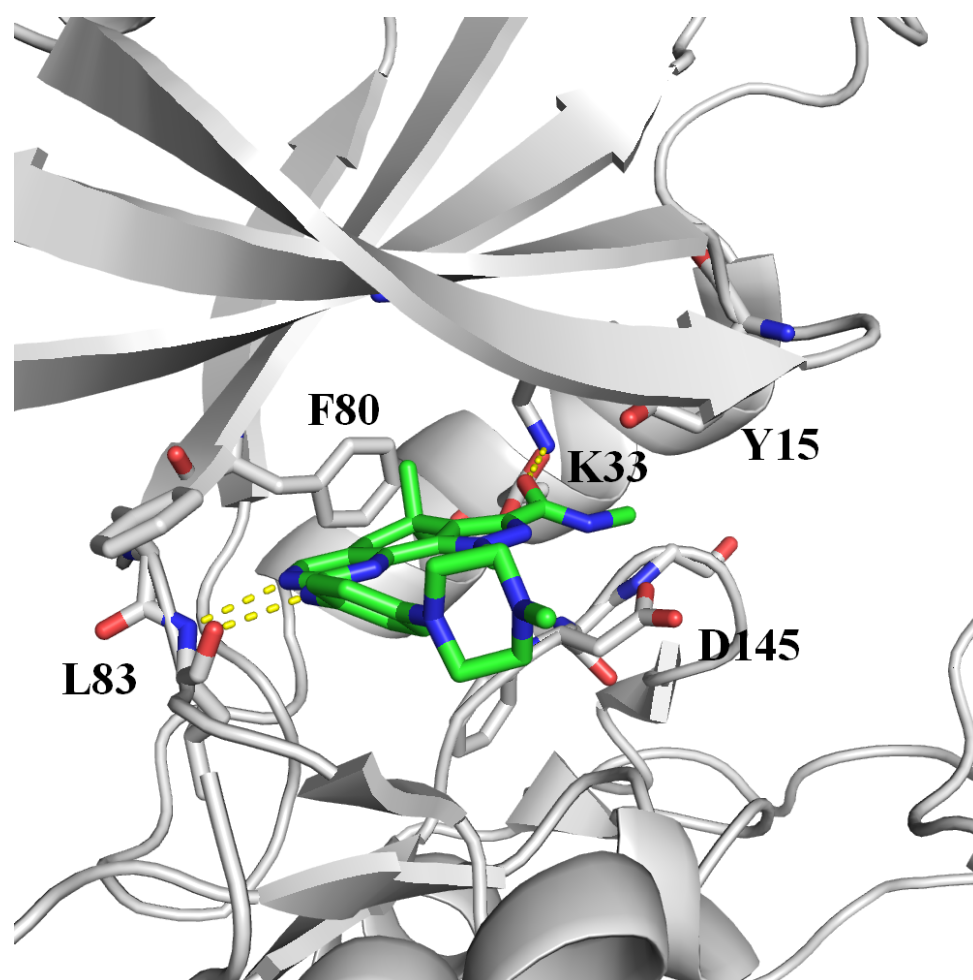


Figure C.1: Crystal structure of compound 28 (PHA-848125) in complex with CDK2/cyclinA (PDB code 2WIH).

C.3 Docking Scores

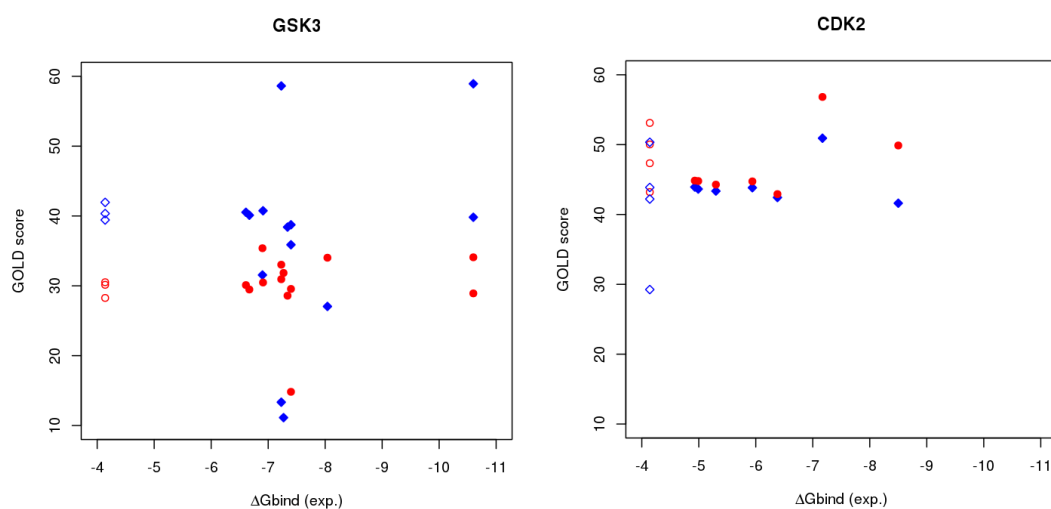
Table C.3: Docking score of top-scored solutions with GSK3 β structures in binding mode 1 and binding mode 2; (3dqw.pdb and 1j1b.pdb have been used for bm1 and bm2, respectively).

<i>Cmp code</i>	Ki	ΔG_{bind}	Binding mode 1		Binding mode 2	
	(GSK3- β)	(obsrv.)	Gold score	Glide Sp	Gold score	Glide Sp
2b	5.8	-7.23	33.03	-6.08	58.62	-6.69
2b_2	0.02	-10.6	34.09	-5.94	58.93	-6.54
2c	0.02	-10.6	28.92	-6.18	39.82	-6.67
2c_2	(≥ 1000)	-4.14	28.27	-6.96	39.43	-6.56
2d_1	4.1	-7.4	14.83	-6.24	38.74	-7.31
2d_2	1.5	-8.04	34.03	-7.18	27.05	-7.5
2d_3	9	-6.9	35.39	-4.48	31.55	-6.62
2d_5	4.4	-7.4	29.56	-4.52	35.88	-6.64
2d_6	4.8	-7.34	28.59	-3.89	38.41	-6.47
2a	14.8	-6.67	29.48	-4.95	40.11	-7.11
2a_2	(≥ 1000)	-4.14	30.15	-4.94	40.35	-6.94
2a_3	9.9	-6.91	30.48	-4.98	40.75	-7.23
2a_4	(≥ 1000)	-4.14	30.52	-4.78	41.95	-7.36
2a_5	16.3	-6.61	30.11	-4.65	40.53	-7.19
3a_3	5.4	-7.27	31.85	-4.74	11.13	-5.15
3a_2	5.8	-7.23	30.95	-4.56	13.34	-5.25

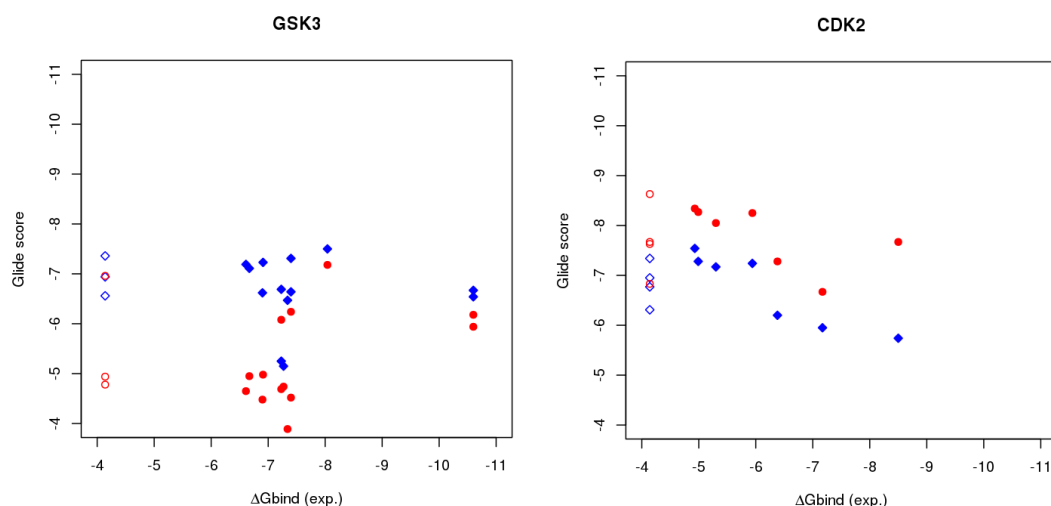
APPENDIX C: PREDICTING THE BINDING MODE FOR KINASE INHIBITORS

Table C.4: Docking score of top-scored solutions with CDK2 structure (2w1h.pdb) in binding mode 1 and binding mode 2.

<i>Cmp code</i>	Ki	ΔG_{bind}	Binding mode 1		Binding mode 2	
	(μM)	(obsrv.)	Gold score	Glide SP	Gold score	Glide SP
2b	6.4	-7.17	56.83	-6.67	50.92	-5.95
2b_2	(≥ 1000)	-4.14	53.11	-6.83	50.34	-6.31
2a	24	-6.38	42.92	-7.28	42.43	-6.20
2a_2	147	-5.3	44.28	-8.05	43.35	-7.17
2a_3	50	-5.94	44.74	-8.25	43.83	-7.24
2a_4	267	-4.93	44.85	-8.34	43.93	-7.54
2a_5	241	-4.99	44.79	-8.27	43.64	-7.28
2c	0.6	-8.5	49.88	-7.67	41.61	-5.74
2c_2	(≥ 1000)	-4.14	50.03	-7.67	42.21	-6.77
2d_1	(≥ 1000)	-4.14	43.21	-8.63	43.87	-7.34
2d_2	(≥ 1000)	-4.14	47.34	-7.63	29.25	-6.95



(a) Scatter plots of GOLD score vs. experimental binding free energies. Red points (circles) represent the GOLD scores for **binding mode 1**, while blue points (diamonds) represent GOLD scores for **binding mode 2**. Non-filled points represent the inactive compounds plotted at ($\Delta G_{\text{bind}} = -4.14$).



(b) Scatter plots of Glide score vs. experimental binding free energies. Red points (circles) represent the Glide scores for **binding mode 1**, while blue points (diamonds) represent Glide scores for **binding mode 2**. Non-filled points represent the inactive compounds plotted at ($\Delta G_{\text{bind}} = -4.14$).

Figure C.2: Scatter plots of docking scores (in the two suggested binding modes) vs. experimental binding free energies

C.4 MM-PBSA Calculations

C.4.1 MM-PBSA Calculations with GSK3 β

Table C.5: MM-PBSA and LR-MM-PBSA applied on Inhibitors/GSK3 β complexes in both binding modes

(a) MM-PBSA methods (GSK3- β , binding mode 1)								
LR-MM-PBSA : $\Delta G_{\text{bind}} = \alpha \Delta E_{\text{vdw}} + \beta 1 \Delta E_{\text{coul}} + \beta 2 \Delta G_{\text{solv}} + \gamma$								
P/A avg. dist. (comp. num.)	$\beta 1$	$\beta 2$	α	γ	r^2	rmse	q^2 (rmse_{loo})	r^2 ($\text{PBSA}_{\text{total}}$)
10.92-14.5 Å (13)	0.19	0.10	0.10	-5.09	0.12	1.44	0.03 (2.9)	0.22
13.5-14.5 Å (7)	-0.30	0.16	0.24	-4.33	0.47	1.27	0.05 (2.03)	0.67
11.25-12.45 Å (6)	0.65	0.16	-0.5	-24.06	0.98	0.28	0.006 (12.22)	0.0004

(b) MM-PBSA methods (GSK3- β , binding mode 2)								
LR-MM-PBSA : $\Delta G_{\text{bind}} = \alpha \Delta E_{\text{vdw}} + \beta 1 \Delta E_{\text{coul}} + \beta 2 \Delta G_{\text{solv}} + \gamma$								
P/A avg. dist. (comp. num.)	$\beta 1$	$\beta 2$	α	γ	r^2	rmse	q^2 (rmse_{loo})	r^2 ($\text{PBSA}_{\text{total}}$)
10.37-13.15 Å (13)	0.01	-0.01	0.02	-6.49	0.011	1.53	0.005 (2.76)	0.038
12.2-13.15 Å (4)	-	-	-	-	-	-	-	0.92
10.37-12.0 Å (7)	-0.25	-0.22	-0.34	-16.82	0.25	1.48	0.0004 (2.87)	0.25

APPENDIX C: PREDICTING THE BINDING MODE FOR KINASE INHIBITORS

Table C.6: LR-MM-PBSA models obtained for the whole dataset of 1-aza-9-oxafluorene derivatives using energy components of MM-PBSA calculations with their binding affinities against GSK3 β , considering docking solutions in binding mode 1. The compounds are ranked according to the descending order of P-loop/A-loop distance.

LR-MM-PBSA : $\Delta G_{\text{bind}} = \alpha \Delta E_{\text{vdw}} + \beta_1 \Delta E_{\text{coul}} + \beta_2 \Delta G_{\text{solv}} + \gamma$							
Cmp code	Ki	ΔG_{bind} (obsrv.)	ΔE_{coul}	ΔG_{solv} (PB)	ΔE_{vdw}	P-loop/A-loop Average dist (10 ns)	$\beta = 0.19$ $\beta_1=0.10$ $\alpha= 0.10$ $\gamma= -5.09$ $r^2= 0.12, \text{rmse}=1.44$
2d_5	4.4	-7.40	-6.34	23.78	-36.23	14.5 \pm 0.9	-7.5266
2b_2	0.02	-10.6	-6.38	24.76	-43.56	14.2 \pm 0.7	-8.1702
2d_2	1.5	-8.04	-6.08	24.9	-37.83	14.09 \pm 0.6	-7.5245
3a_3	5.4	-7.28	-6.36	25.26	-35.57	13.89 \pm 0.8	-7.3147
3a_2	5.8	-7.23	-6.54	24.87	-34.38	13.75 \pm 0.5	-7.2690
2d_6	4.8	-7.34	-5.34	23.54	-35.64	13.77 \pm 1.7	-7.3017
2b	5.8	-7.23	-6.65	25.25	-42.65	13.5 \pm 0.5	-8.0808
2a_5	16.3	-6.61	-8.77	25.58	-36.76	12.53 \pm 1.3	-7.8598
2a_3	9.9	-6.91	-5.7	20.73	-34.71	12.05 \pm 1.4	-7.5607
2d_3	9	-6.9	-11.45	25.45	-39.90	11.76 \pm 1.7	-8.6972
2a	14.8	-6.67	-6.72	27.03	-33.65	11.75 \pm 1.5	-7.0118
2c	0.02	-10.6	-12.36	26.96	-33.47	11.36 \pm 0.8	-8.0728
2d_1	4.1	-7.44	-13.49	36.58	-38.4	11.25 \pm 0.4	-7.8101
	Ki	$\Delta G_{\text{bind}} \geq \Delta E_{\text{coul}}$		ΔG_{solv} (PB)	ΔE_{vdw}	Average dist.	Pred. ΔG_{bind}
2a_4	(≥ 1000)	-4.14	-7.32	26.39	-34.89	12.55 \pm 1.7	-7.3308
2c_2	(≥ 1000)	-4.14	-7.42	26.37	-34.76	12.45 \pm 1.6	-7.3388
2a_2	(≥ 1000)	-4.14	-11.54	30.42	-34.36	10.92 \pm 0.9	-7.6766

APPENDIX C: PREDICTING THE BINDING MODE FOR KINASE INHIBITORS

Table C.7: LR-MM-PBSA models generated to predict the free binding energy depending on MM-PBSA calculations with Inhibitor/GSK3 β complex in binding mode 2. The calculation are performed including all the compounds. The compounds are ranked according to the descending order of P-loop/A-loop distance.

LR-MM-PBSA : $\Delta G_{\text{bind}} = \alpha \Delta E_{\text{vdw}} + \beta_1 \Delta E_{\text{coul}} + \beta_2 \Delta G_{\text{solv}} + \gamma$							
Cmp code	Ki	ΔG_{bind} (obsrv.)	ΔE_{coul}	ΔG_{solv} (PB)	ΔE_{vdw}	P-loop/A-loop Average dist (10 ns)	$\beta = 0.01$ $\beta_1 = -0.01$ $\alpha = 0.02$ $\gamma = -6.49$ $r^2 = 0.011$, rmse=1.53
3a_3	5.4	-7.28	-10.27	33.37	-44.58	13.5 \pm 0.7	-7.7829
2d_1	4.1	-7.40	-5.35	24.04	-39.57	13.29 \pm 0.4	-7.5420
2b_2	0.02	-10.6	-15.09	34.83	-43.80	13.15 \pm 0.7	-7.83278
2b	5.8	-7.23	-14.15	41.32	-46.92	12.2 \pm 0.8	-7.9488
2d_5	4.4	-7.44	-17.27	38.04	-41.95	11.9 \pm 0.8	-7.8537
2a_5	16.3	-6.61	-12.51	28.58	-37.54	11.8 \pm 1.2	-7.6244
2d_6	4.8	-7.34	-22.37	41.16	-35.45	11.6 \pm 0.7	-7.81655
2a_3	9.9	-6.91	-10.37	26.02	-37.71	11.48 \pm 1.1	-7.5791
2a	14.8	-6.67	-19.13	33.25	-37.07	11.43 \pm 1.4	-7.7318
3a_2	5.8	-7.23	-11.45	35.26	-39.86	11.35 \pm 0.7	-7.7261
2d_3	9	-6.90	-11.36	21.05	-32.41	10.99 \pm 0.8	-7.4385
2c	0.02	-10.6	-13.7	30.1	-34.7	10.37 \pm 0.7	-7.5991
2d_2	1.5	-8.04	-11.28	34.61	-40.23	8.2 \pm 0.6	-7.7245
	Ki	$\Delta G_{\text{bind}} \geq \Delta E_{\text{coul}}$		ΔG_{solv} (PB)	ΔE_{vdw}	Average dist.	Pred. ΔG_{bind}
2a_2	(≥ 1000)	-4.14	-12.96	26.27	-37.11	12.2 \pm 0.9	-7.6245
2a_4	(≥ 1000)	-4.14	-13.67	29.47	-36.94	11.95 \pm 1.4	-7.6602
2c_2	(≥ 1000)	-4.14	-9.35	22.37	-33.54	11.9 \pm 1.6	-7.478

C.4.2 MM-PBSA Calculations with CDK2

Table C.8: LR-MM-PBSA models generated to predict the free binding energy depending on MM-PBSA calculations with Inhibitor/CDK2 complex in binding mode 1. Calculations are performed including all the compounds. The compounds are ranked according to the descending order of P-loop/A-loop distance.

LR-MM-PBSA : $\Delta G_{\text{bind}} = \alpha \Delta E_{\text{vdw}} + \beta_1 \Delta E_{\text{coul}} + \beta_2 \Delta G_{\text{solv}} + \gamma$							
Cmp code	Ki (μM)	ΔG_{bind}				P-loop/A-loop	$\beta = -0.19$ $\beta_1 = -0.7$ $\alpha = -0.34$ $\gamma = +7.9$
(BM1)	(CDK2)	(obsrv.)	ΔE_{coul}	ΔG_{solv} (PB)	ΔE_{vdw}	Average dist (10 ns)	$r^2 = 0.67$, $\text{rmse} = 0.75$ $q^2 = 0.06$ (5.19)
2c	0.6	-8.5	-25.9	44.82	-32.18	11.22±0.69	-7.7012
2a_2	147	-5.3	-35.22	46.23	-33.21	11.23±0.78	-6.5628
2a_3	50	-5.94	-35.45	45.84	-34.37	11.41±0.94	-5.8459
2a_4	267	-4.93	-35.64	45.35	-34.87	11.52±0.60	-5.2924
2a_5	241	-4.99	-35.77	44.93	-34.85	11.25±0.83	-4.9780
2a	24	-6.38	-38.36	47.97	-38.36	10.0±0.66	-5.4252
2b	6.4	-7.17	-15.72	49.65	-48.66	9.78±1.3	-7.4044
	K _i	$\Delta G_{\text{bind}} \geq \Delta E_{\text{coul}}$	ΔG_{solv} (PB)	ΔE_{vdw}	Average dist.	Pred. ΔG_{bind}	
2d_2	(≥1000)	-4.14	-3.63	26.72	-38.69	14.41±1.4	4.0766
2b_2	(≥1000)	-4.14	-11.97	43.18	-45.63	13.88±1.2	-4.4178
2d_1	(≥1000)	-4.14	-9.71	33.86	-36.80	12.48±0.88	-1.348
2c_2	(≥1000)	-4.14	-20.9	40.82	-35.44	9.99±1.68	-4.4444

APPENDIX C: PREDICTING THE BINDING MODE FOR KINASE INHIBITORS

Table C.9: LR-MM-PBSA models generated to predict the free binding energy depending on MM-PBSA calculations with Inhibitor/CDK2 complex in binding mode 2. Calculations are performed including all the compounds. The compounds are ranked according to the descending order of P-loop/A-loop distance.

LR-MM-PBSA : $\Delta G_{\text{bind}} = \alpha \Delta E_{\text{vdw}} + \beta_1 \Delta E_{\text{coul}} + \beta_2 \Delta G_{\text{solv}} + \gamma$							
Cmp code	Ki	ΔG_{bind}				P-loop/A-loop	$\beta = 0.18$ $\beta_1 = -0.19$ $\alpha = 0.06$ $\gamma = +6.57$
(BM2)	CDK2	(obsrv.)	ΔE_{coul}	ΔG_{solv} (PB)	ΔE_{vdw}	Average dist (10 ns)	$r^2 = 0.41$, $\text{rmse} = 1.40$ $q^2 = 0.02$ (12.3)
2b	6.4	-7.17	-18.1	36.49	-41.08	11.11±1.6	-6.8433
2a	24	-6.38	-19.37	34.21	-31.21	9.93±1.36	-5.9365
2a_3	50	-5.94	-18.33	33.82	-32.41	9.43±1.2	-5.7525
2a_4	267	-4.93	-18.48	33.79	-32.56	9.56±1.2	-5.7841
2a_5	241	-4.99	-11.52	32.76	-39.58	8.1±0.85	-4.7983
2c	0.6	-8.5	-15.30	41.36	-38.04	7.51±0.70	-7.0974
2a_2	147	-5.3	-15.21	40.56	-39.15	7.6±0.6	-6.9979
	Ki	$\Delta G_{\text{bind}} \geq \Delta E_{\text{coul}}$		ΔG_{solv} (PB)	ΔE_{vdw}	Average dist.	Pred. ΔG_{bind}
2d_2	(≥1000)	-4.14	-2.99	32.99	-37.68	10.29±1.49	-2.4971
2d_1	(≥1000)	-4.14	-16.6	39.98	-39.4	9.45±0.80	-6.3782
2b_2	(≥1000)	-4.14	-10.47	33.97	-40.37	8.9±0.84	-4.1911
2c_2	(≥1000)	-4.14	-14.84	40.27	-39.62	7.6±0.8	-6.1297

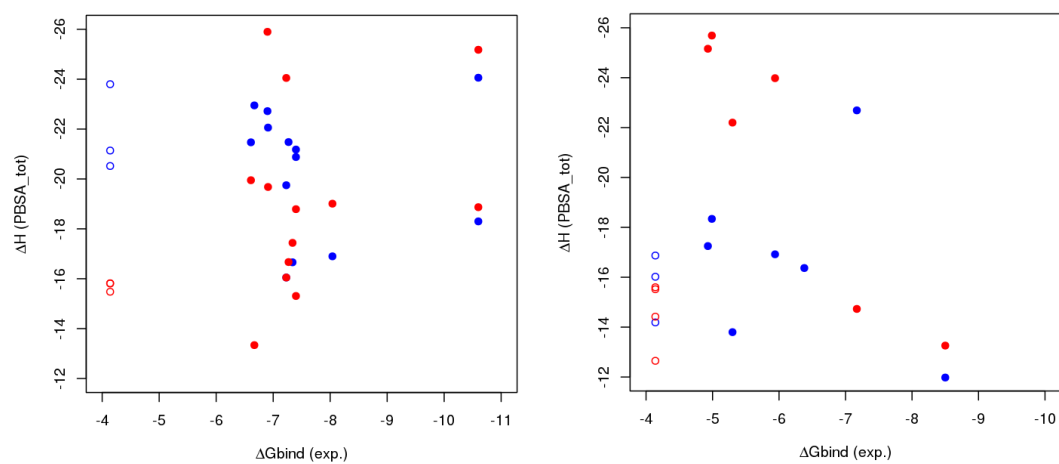


Figure C.3: Scatter plots of PBSA_{tot} vs. experimental binding free energies for GSK3 (left) and CDK2 (right). Red points (circles) represent the PBSA_{tot} for **binding mode 1**, while blue points (diamonds) represent PBSA_{tot} for **binding mode 2**. Non-filled points represent the inactive compounds plotted at ($\Delta G_{\text{bind}} = -4.14$).

C.5 LIE models of GSK3 β binding energy (Binding mode 1)

Table C.10: LIE model obtained for the whole dataset of 1-aza-9-oxafluorene derivatives using their binding affinities against GSK3 β , starting from docking solutions in binding mode 1. Compounds are ranked according to the descending order of P-loop/A-loop distance.

$\Delta G_{\text{bind}} (\text{pred.}) = \beta(\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle) + \alpha(\Delta \langle V^{\text{vdw}}_{\text{lig-surr}} \rangle) + \gamma$						
<i>Cmp code</i>	Ki	ΔG_{bind}			P-loop/A-loop	$\beta = -0.06$ $\alpha = 0.13$ $\gamma = -5.13$
BM1	GSK3 β	(obsrv.)	$\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta \langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	Average dist (10 ns)	$r^2 = 0.08$, $\text{rmse} = 1.4$
2d_5	4.4	-7.40	4.26	-15.72	14.5 \pm 0.9	-7.56
2b_2	0.02	-10.6	8.45	-18.02	14.2 \pm 0.7	-8.14
2d_2	1.5	-8.04	11.63	-14.95	14.09 \pm 0.6	-7.92
3a_3	5.4	-7.28	8.82	-14.10	13.89 \pm 0.8	-7.63
3a_2	5.8	-7.23	10.3	-15.3	13.75 \pm 0.5	-7.89
2d_6	4.8	-7.34	9.1	-17.25	13.77 \pm 0.6	-8.08
2b	5.8	-7.23	7.22	-17.57	13.5 \pm 0.5	-8.00
2a	14.8	-6.67	8.31	-11.80	12.69 \pm 0.8	-7.28
2a_5	16.3	-6.61	10.58	-11.85	12.53 \pm 1.3	-7.43
2a_3	9.9	-6.91	9.84	-11.82	12.05 \pm 1.4	-7.38
2d_3	9	-6.90	2.47	-12.91	11.76 \pm 1.7	-7.06
2c	0.02	-10.6	12.1	-11.55	11.36 \pm 0.8	-7.49
2d_1	4.1	-7.44	19.26	-14.40	11.25 \pm 0.4	-8.34

APPENDIX C: PREDICTING THE BINDING MODE FOR KINASE INHIBITORS

 Table C.11: LIE models obtained for subset 1 (average P-loop/A-loop distance between 13.5 to 14.5 Å) using their binding affinities against GSK3 β , starting from docking solutions in binding mode 1. Compounds are ranked according to the descending order of P-loop/A-loop distance.

Comp code		$\Delta G_{\text{bind}}(\text{pred.}) = \beta(\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle) + \alpha(\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle) + \gamma$			
GSK3- β , BM1	Avg. Dist	$\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	exp.	$\beta = -0.07$ $\alpha = 0.014$ $\gamma = -7.27$
2d_5	14.5 \pm 0.9	4.26	-15.72	-7.40	-7.17
2b_2	14.2 \pm 0.7	8.45	-18.02	-10.60	-8.72
2d_2	14.09 \pm 0.6	11.63	-14.95	-8.04	-7.70
3a_3	13.89 \pm 0.8	8.82	-14.1	-7.27	-6.98
3a_2	13.75 \pm 0.5	10.3	-15.3	-7.23	-7.71
2d_6	13.77 \pm 0.6	9.1	-17.25	-7.34	-8.45
2b	13.5 \pm 0.5	7.22	-17.57	-7.23	-8.36
Model r2 (rmse)					0.28 (1.28)

 Table C.12: LIE models obtained for subset 2 (average P-loop/A-loop distance between 11.25 to 12.45 Å) using their binding affinities against GSK3 β , starting from docking solutions in binding mode 1. Compounds are ranked according to the descending order of P-loop/A-loop distance.

Comp code		$\Delta G_{\text{bind}}(\text{pred.}) = \beta(\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle) + \alpha(\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle) + \gamma$			
GSK3 β , BM1	Avg. Dist	$\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	exp.	$\beta = -0.13$ $\alpha = -0.59$ $\gamma = -13.58$
2a	12.69 \pm 0.8	8.31	-11.80	-6.67	-7.59
2a_5	12.53 \pm 1.3	10.58	-11.85	-6.61	-7.86
2a_3	12.05 \pm 1.3	9.84	-11.82	-6.91	-7.78
2d_3	11.76 \pm 1.7	2.47	-12.91	-6.90	-6.17
2c	11.36 \pm 0.8	12.1	-11.55	-10.60	-8.23
2d_1	11.25 \pm 0.4	19.26	-14.4	-7.44	-7.45
Model r2 (rmse)					0.21 (1.76)

C.6 LIE models of CDK2 binding energy (Binding mode 1)

Table C.13: LIE model obtained for the whole dataset of 1-aza-9-oxafluorene derivatives using their binding affinities against CDK2, starting from docking solutions in binding mode 1. Compounds are ranked according to the descending order of P-loop/A-loop distance.

$\Delta G_{\text{bind}} (\text{pred.}) = \beta(\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle) + \alpha (\Delta \langle V^{\text{vdw}}_{\text{lig-surr}} \rangle) + \gamma$						
Cmp code	Ki	ΔG_{bind}			P-loop/A-loop	$\beta = 0.02$ $\alpha = -0.65$ $\gamma = -15.57$
BM1	CDK2	(obsrv.)	$\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta \langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	Average dist (10 ns)	$r^2 = 0.42$, $\text{rmse} = 0.91$
2a_4	267	-4.93	12.34	-13.36	11.52±0.60	-6.59
2a_3	50	-5.94	6.2	-15.36	11.41±0.94	-5.40
2a_5	241	-4.99	5.25	-15.87	11.25±0.83	-5.08
2a_2	147	-5.30	12.55	-14.93	11.23±0.78	-5.65
2c	0.6	-8.50	12.98	-12.22	11.22±0.69	-7.32
2a	24	-6.38	5.15	-12.91	10.0±0.66	-7.02
2b	6.4	-7.17	9.51	-14.01	9.78±1.3	-6.22

APPENDIX C: PREDICTING THE BINDING MODE FOR KINASE INHIBITORS

Table C.14: LIE models obtained for compounds' subset (average P-loop/A-loop distance between 11.22 to 11.52 Å) using their binding affinities against CDK2, starting from docking solutions in binding mode 1. Compounds are ranked according to the descending order of P-loop/A-loop distance.

$\Delta G_{\text{bind}} (\text{pred.}) = \beta(\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle) + \alpha (\Delta \langle V^{\text{vdw}}_{\text{lig-surr}} \rangle) + \gamma$						
Cmp code	Ki	ΔG_{bind}			P-loop/A-loop	$\beta = 0.2$ $\alpha = -1.08$ $\gamma = -23.41$
BM1	CDK2	(obsrv.)	$\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta \langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	Average dist (10 ns)	$r^2 = 0.57$, $\text{rmse} = 0.95$
2a_4	267	-4.93	12.34	-13.36	11.52±0.60	-6.50
2a_3	50	-5.94	6.20	-15.36	11.41±0.94	-5.57
2a_5	241	-4.99	5.25	-15.87	11.25±0.83	-5.21
2a_2	147	-5.30	12.55	-14.93	11.23±0.78	-4.76
2c	0.6	-8.50	12.98	-12.22	11.22±0.69	-7.60

C.7 Prediction of Inactive Compounds

Table C.15: Predictions of inactive compounds of GSK3

(a) Prediction of inactive compounds of GSK3 in BM2

Cmp. code	Ki	P-loop/A-loop	Applied Model		
(BM2)	(GSK3)	Avg. Dist.	$\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	$\Delta G_{\text{bind}}(\text{pred.}) = 0.63(\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle) + 0.32(\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle) - 8.22$
2a_4	≥ 1000	12.44	9.54	-11.5	-5.88
2a_2	≥ 1000	12.2	11.3	-13.87	-5.53
			$\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	$\Delta G_{\text{bind}}(\text{pred.}) = 0.86(\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle) + 0.23(\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle) - 7.89$
2c_2	≥ 1000	11.9	7.44	-11.87	-4.23

(b) Prediction of inactive compounds of GSK3 in BM1

Cmp. code	Ki	P-loop/A-loop	Applied Model		
(BM1)	(GSK3)	Avg. Dist.	$\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	$\Delta G_{\text{bind}}(\text{pred.}) = -0.13(\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle) - 0.59(\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle) - 13.58$
2a_4	≥ 1000	12.55	10.68	-11.79	-8.00
2c_2	≥ 1000	12.45	10.84	-11.82	-8.00
2a_2	≥ 1000	10.92	8.82	-13.71	-6.64

Table C.16: Predictions of inactive compounds of CDK2

(a) Prediction of inactive compounds of CDK2 in BM2

Cmp. code	Ki	P-loop/A-loop	Applied Model		
(BM2)	(CDK2)	Avg. Dist.	$\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	$\Delta G_{\text{bind}}(\text{pred.}) = 0.24(\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle) + 0.09(\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle) - 4.8$
2d_2	≥ 1000	10.29	5.42	-17.13	-4.9
2d_1	≥ 1000	9.45	9.49	-14.01	-3.71
2b_2	≥ 1000	8.90	4.90	-18.75	-5.3
			$\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	$\Delta G_{\text{bind}}(\text{pred.}) = 0.46(\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle) + 0.06(\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle) - 6.9$
2c_2	≥ 1000	7.6	6.79	-16.43	-4.76

(b) Prediction of inactive compounds of CDK2 in BM1

Cmp. code	Ki	P-loop/A-loop	Applied Model		
(BM1)	(CDK2)	Avg. Dist.	$\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	$\Delta G_{\text{bind}}(\text{pred.}) = 0.2(\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle) - 1.08(\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle) - 23.41$
2d_2	≥ 1000	14.41	9.59	-10.38	-10.28
2b_2	≥ 1000	13.88	6.67	-13.18	-7.84
2d_1	≥ 1000	12.48	12.5	-12.05	-7.89
			$\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	$\Delta G_{\text{bind}}(\text{pred.}) = 0.02(\Delta\langle V^{\text{el}}_{\text{lig-surr}} \rangle) - 0.65(\Delta\langle V^{\text{vdw}}_{\text{lig-surr}} \rangle) - 15.57$
2c_2	≥ 1000	9.99	5.25	-13.19	-6.89

Appendix D

Prediction of Binding Affinities (c-Kit D816V)

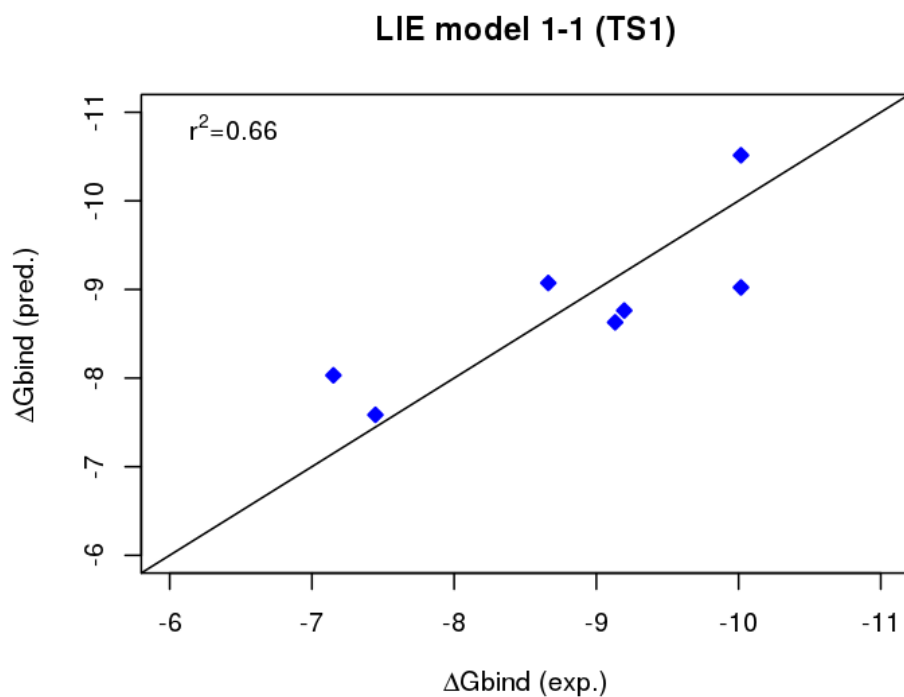
D.1 LIE Models for c-Kit D816V Inhibitors

D.1.1 LIE model 1-1**Training set 1 (TS1): (Compounds with distance between P_loop and A_loop = 8.4 – 8.7)**

Table D.1: Training set (TS1) for LIE model 1-1

$$\Delta G_{\text{bind}}(\text{pred.}) = 0.10 (\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle) + 0.16 (\Delta \langle V_{\text{lig-surr}}^{\text{vdw}} \rangle) - 5.01$$

$$(r^2=0.66, \text{rmse}=0.81, \text{adjusted } r^2=0.49, q^2=0.40, n=7)$$



Cmp code	IC50 (nM) (c-Kit_D816V)	ΔG_{bind} (observ.)	$\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle$	$\Delta \langle V_{\text{lig-surr}}^{\text{vdw}} \rangle$	Ploop/Aloop distance (Å)	ΔG_{bind} (pred.)
6367	70	-10.02	6.820	-36.933	8.54	-10.51
6020	300	-9.13	17.707	-32.605	8.74	-8.63
6614	270	-9.20	7.0335	-26.667	8.43	-8.76
6674	650	-8.66	7.367	-28.721	8.43	-9.07
6556	70	-10.02	12.856	-31.883	8.53	-9.02
6124	7800	-7.15	17.932	-29.196	8.42	-8.03
6151	4800	-7.45	17.943	-26.555	8.71	-7.59

Validation set 1 (VS1): (Compounds with distance between P_loop and A_loop = 8.4 – 8.7)

Table D.2: Validation set (VS1) for LIE model 1-1

Cmp code	IC50 (nM) (c-Kit_D816V)	ΔG_{bind} (observ.)	$\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta \langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	Ploop/Aloop distance (Å)	ΔG_{bind} (pred.)
6557	750	-8.57	16.48	-28.11	8.4	-7.85
5038	1500	-8.15	20.93	-29.97	8.55	-7.71
6275	2500	-7.84	9.35	-23.74	8.58	-7.87
5003	900	-8.46	16.38	-30.45	8.47	-8.24
6567	250	-9.24	11.5	-31.43	8.39	-8.89
6395	3000	-7.73	10.44	-23.77	8.75	-7.76
6275	2500	-7.28	9.35	-23.74	8.58	-7.87
6472	1300	-8.24	10.64	-23.57	8.3	-7.71
6377	350	-9.04	11.72	-33.43	8.57	-9.18
6376	150	-9.55	4.54	-29.83	8.2	-9.32

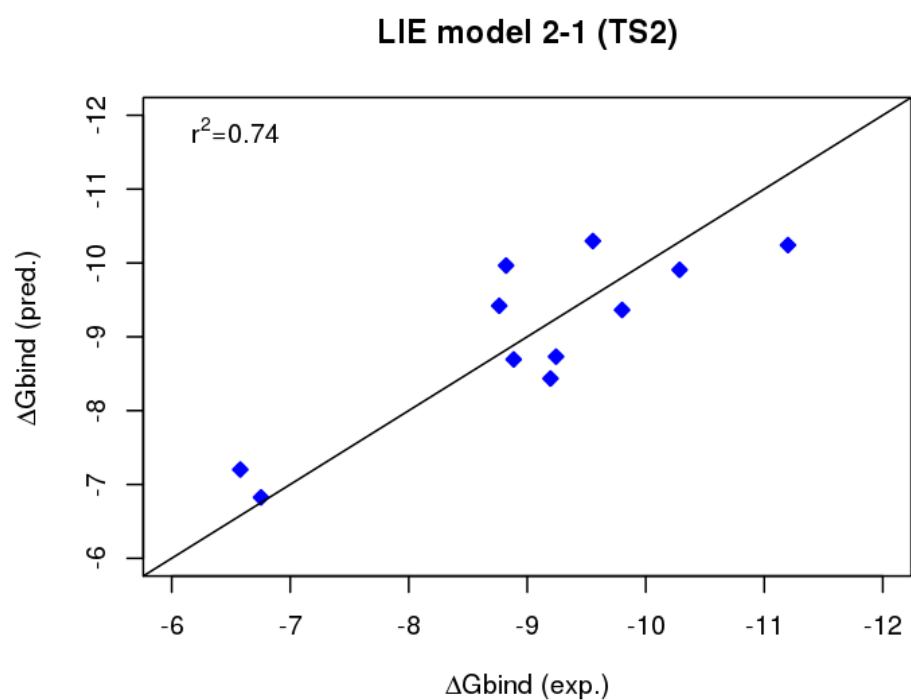
D.1.2 LIE model 2-1

Training set 2 (TS2): (Compounds with distance between P_loop and A_loop = 8.75 – 9.0)

Table D.3: Training set (TS2) for LIE model 2-1

$$\Delta G_{\text{bind}}(\text{pred.}) = 0.16 (\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle) + 0.45 (\Delta \langle V_{\text{lig-surr}}^{\text{vdw}} \rangle) + 1.56$$

$$(r^2=0.74, \text{rmse}=0.77, \text{adjusted } r^2=0.67, q_2 =0.55, n=11)$$



Cmp code	IC50 (nM) (c-Kit_D816V)	ΔG_{bind} (observ.)	$\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle$	$\Delta \langle V_{\text{lig-surr}}^{\text{vdw}} \rangle$	Ploop/Aloop distance (Å)	ΔG_{bind} (pred.)
6387	250	-9.24	10.19	-26.09	8.81	-8.73
6249	500	-8.82	13.77	-30.04	8.85	-9.96
6503	15000	-6.75	8.97	-21.50	8.95	-6.83
6620	550	-8.76	3.77	-25.36	8.98	-9.42
6561	450	-8.88	20.04	-29.44	9.0	-8.69
1869	20000	-6.58	2.49	-20.06	8.89	-7.20
6678	150	-9.55	14.04	-30.86	8.87	-10.30
1718	45	-10.28	0.51	-25.28	9.03	-9.91
6581	10	-11.20	3.69	-27.13	8.84	-10.24
6155	270	-9.20	16.22	-27.55	8.83	-8.43
6152	100	-9.80	13.11	-28.49	8.75	-9.36

Validation set 2 (VS2): (Compounds with distance between P_loop and A_loop = 8.75 – 9.2)

Table D.4: Validation set (VS2) for LIE model 2-1

$$\Delta G_{\text{bind}}(\text{pred.}) = 0.16 (\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle) + 0.45 (\Delta \langle V_{\text{lig-surr}}^{\text{vdw}} \rangle) + 1.56$$

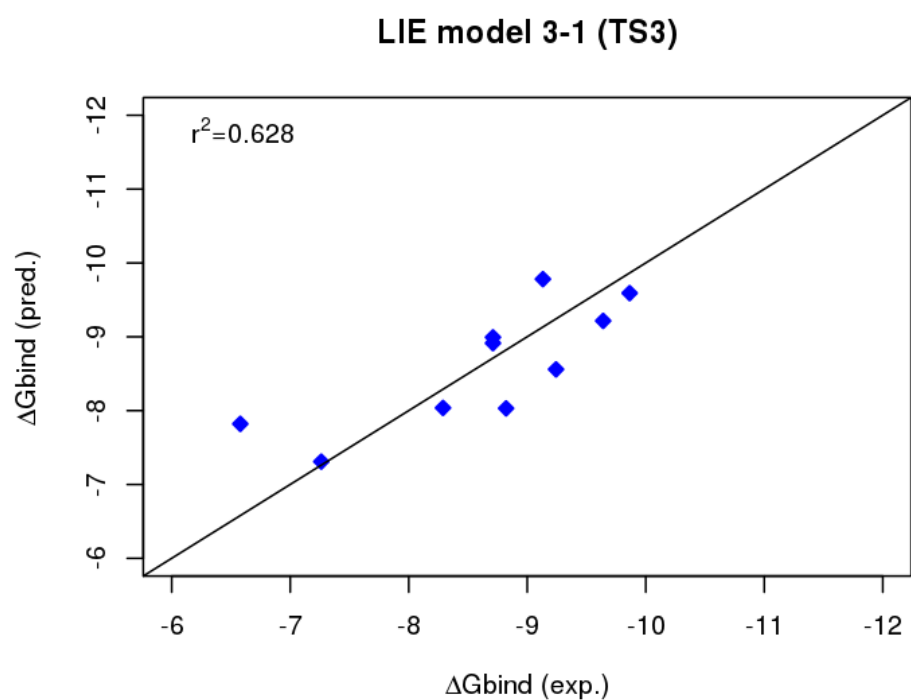
Cmp code	IC50 (nM) (c-Kit_D816V)	ΔG_{bind} (observ.)	$\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle$	$\Delta \langle V_{\text{lig-surr}}^{\text{vdw}} \rangle$	Ploop/Aloop distance (Å)	ΔG_{bind} (pred.)
6621	200	-9.38	9.81	-23.92	8.99	-7.87
6603	400	-8.96	6.26	-25.86	8.95	-9.33
6168	20000	-6.58	20.50	-25.57	9.08	-6.92
5145	20000	-6.58	3.52	-20.69	8.7	-7.395
6569	75	-9.97	10.96	-28.80	8.91	-9.94
5284	350	-9.04	18.34	-28.85	8.73	-8.77
5208	20	-10.78	9.25	-28.25	8.83	-9.96
5221	70	-10.02	12.90	-29.68	9.00	-10.03
5260	200	-9.38	2.89	-24.31	8.95	-9.16
5189	1500	-8.15	10.19	-23.68	9.2	-7.70
6601	300	-9.13	7.31	-27.69	8.92	-10.01
6255	480	-8.85	6.84	-26.06	8.91	-9.14

D.1.3 LIE model 3-1**Training set 3 (TS3): (Compounds with distance between P_loop and A_loop = 9.4 – 9.9)**

Table D.5: Training set (TS3) for LIE model 3-1

$$\Delta G_{\text{bind}}(\text{pred.}) = 0.047 (\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle) + 0.31 (\Delta \langle V_{\text{lig-surr}}^{\text{vdw}} \rangle) - 0.84$$

$$(r^2=0.628, \text{rmse}=0.70, \text{adjusted } r^2=0.52, q_2=0.48, n=10)$$



Cmp code	IC50 (nM) (c-Kit_D816V)	ΔG_{bind} (observ.)	$\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle$	$\Delta \langle V_{\text{lig-surr}}^{\text{vdw}} \rangle$	Ploop_Aloop distance (Å)	ΔG_{bind} (pred.)
5287	90	-9.86	13.16	-30.14	9.74	-9.59
6194	6500	-7.26	15.765	-23.21	9.59	-7.31
6213	1200	-8.29	23.92	-26.79	9.67	-8.04
6549	300	-9.13	12.67	-30.68	9.71	-9.78
6050	250	-9.24	15.76	-27.22	9.8	-8.56
6167	500	-8.82	14.02	-25.26	9.87	-8.03
6168	20000	-6.58	20.50	-25.57	9.9	-7.82
6528	600	-8.71	6.22	-26.90	9.47	-8.91
1860	130	-9.64	14.65	-29.16	9.58	-9.22
6612	600	-8.71	7.13	-27.29	9.56	-8.99

Validation set 3 (VS3): (Compounds with distance between P_loop and A_loop = 9.4 – 9.9)

Table D.6: Validation set (VS3) for LIE model 3-1

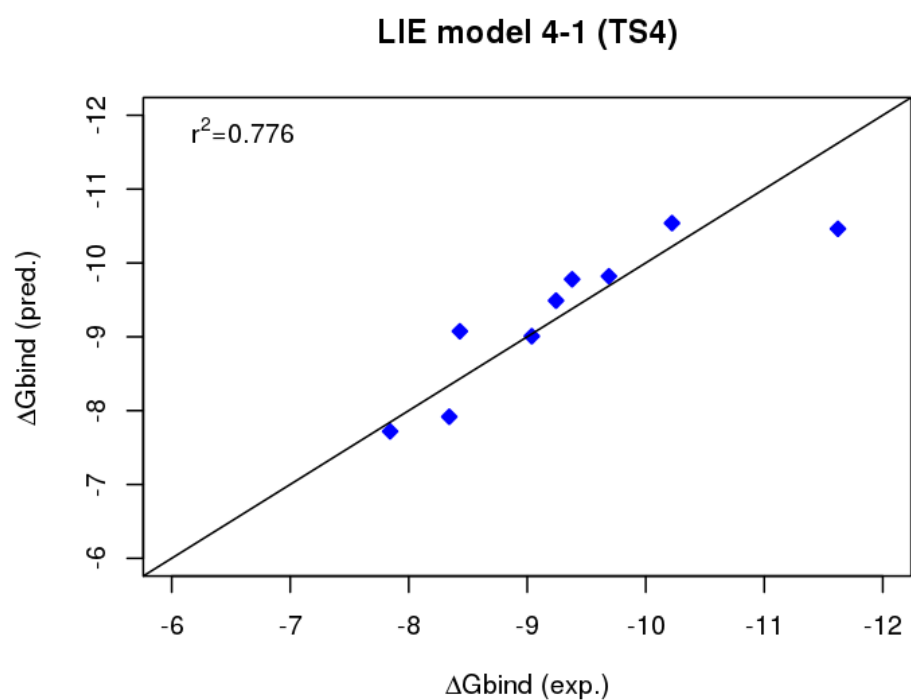
Cmp code	IC50 (nM) (c-Kit_D816V)	ΔG_{bind} (observ.)	$\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta \langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	Ploop/Aloop distance (Å)	ΔG_{bind} (pred.)
6418	90	-9.86	16.15	-33.73	9.7	-10.54
6330	100	-9.80	12.70	-30.37	9.56	-9.66
1860	130	-9.64	14.65	-29.16	9.58	-9.19
1805	400	-8.96	23.63	-27.06	9.7	-8.12
6108	670	-8.64	2.87	-24.16	9.83	-8.20
6514	300	-9.13	10.72	-29.70	9.76	-9.54
6671	20000	-6.58	14.87	-20.84	9.88	-6.60
5190	1300	-8.24	8.47	-26.73	9.6	-8.73
5196	150	-9.55	14.98	-28.22	9.52	-8.88
5197	60	-10.11	19.33	-29.13	9.63	-8.96
5246	200	-9.38	21.07	-29.96	9.7	-9.14
5186	1500	-8.15	8.63	-24.93	9.80	-8.16

D.1.4 LIE model 4-1**Training set 4 (TS4): (Compounds with distance between P_loop and A_loop = 10 - 10.5)**

Table D.7: Training set (TS4) for LIE model 4-1

$$\Delta G_{\text{bind}}(\text{pred.}) = 0.24 (\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle) + 0.21 (\Delta \langle V_{\text{lig-surr}}^{\text{vdw}} \rangle) - 7.5$$

$$(r^2=0.776, \text{rmse}=0.61, \text{adjusted } r^2=0.70, q_2=0.65, n=9)$$



Cmp code	IC50 (nM) (c-Kit_D816V)	ΔG_{bind} (observ.)	$\Delta \langle V_{\text{lig-surr}}^{\text{el}} \rangle$	$\Delta \langle V_{\text{lig-surr}}^{\text{vdw}} \rangle$	Ploop/Aloop distance (Å)	ΔG_{bind} (pred.)
6535	120	-9.69	13.41	-25.61	10.03	-9.82
6041	2500	-7.84	22.67	-26.36	10.03	-7.72
6071	350	-9.04	18.42	-27.51	10.12	-9.01
6431	250	-9.24	16.61	-27.70	10.5	-9.49
1823	1100	-8.34	25.48	-30.43	10.29	-7.92
1809	50	-10.22	10.51	-25.66	10.45	-10.54
1806	950	-8.43	18.20	-27.57	10.41	-9.07
4232	5	-11.62	9.28	-23.92	10.17	-10.46
1778	200	-9.38	10.14	-21.74	10.18	-9.78

Validation set 4 (VS4):(Compounds with distance between P_loop and A_loop = 10 - 10.5)

Table D.8: Validation set (VS4) for LIE model 4-1

Cmp code	IC50 (nM) (c-Kit_D816V)	ΔG_{bind} (observ.)	$\Delta \langle V^{\text{el}}_{\text{lig-surr}} \rangle$	$\Delta \langle V^{\text{vdw}}_{\text{lig-surr}} \rangle$	Ploop/Aloop distance (Å)	ΔG_{bind} (pred.)
6545	60	-10.11	11.24	-29.71	10.25	-11.04
6418	90	-9.86	11.69	-30.14	10.17	-11.02
6186	30	-10.53	12.68	-25.12	10.11	-9.73
6303	30	-10.532	9.33	-27.19	10.52	-10.97
1800	80	-9.94	9.92	-26.80	10.78	-10.75
1745	300	-9.13	16.66	-31.78	10.75	-10.17
1752	540	-8.77	14.78	-26.03	10.75	-9.42
6458	2700	-7.80	13.73	-23.21	10.75	-9.08
6581	10	-11.20	6.46	-27.99	10.67	-11.83

D.2 PBSA Calculations and LR-MM-PBSA Models

D.2.1 LR-MM-PBSA model 1-1

Training set 1 (TS1): (Compounds with distance between P_loop and A_loop = 8.4 – 8.7)

Table D.9: PBSA calculations for Training set 1

$$\text{PB 1-1 : } \Delta G_{\text{bind}} (\text{pred.}) = 0.45\Delta E_{\text{coul}} + 0.23 \Delta E_{\text{vdw}} - 0.37\Delta G_{\text{solv}} - 1.26$$

$$(r^2=0.47, \text{rmse}=0.81, q^2 =0.19)$$

Cmp code (TS 1)	IC ₅₀ (nM, cKit_D816V)	ΔG_{bind} (observ.)	PBSA _{tot}	ΔE_{coul}	ΔE_{vdw}	ΔG_{solv} (PB)	ΔG_{sa}	ΔG_{bind} (pred.)
6367	70	-10.02	-43.01	-36.46	-71	66	-1.55	-9.13
6020	300	-9.13	-39.82	-43.53	-68.62	73.83	-1.5	-8.81
6614	270	-9.20	-35.99	-36.4	-59.9	61.78	-1.47	-8.13
6674	650	-8.66	-38.97	-41.59	-65.55	69.67	-1.5	-8.80
6556	70	-10.02	-45.17	-38.81	-69.58	64.73	-1.51	-10.34
6124	7800	-7.15	-37.97	-39.18	-64.29	66.99	-1.49	-8.43
6151	4800	-7.45	-38.9	-31.49	-65.87	59.95	-1.49	-7.99

Validation set 1 (VS1): (Compounds with distance between P_loop and A_loop = 8.4 – 8.7)

Table D.10: PBSA calculations for Validation set 1

$$\text{PB 1-1 : } \Delta G_{\text{bind}} (\text{pred.}) = 0.45\Delta E_{\text{coul}} + 0.23 \Delta E_{\text{vdw}} - 0.37\Delta G_{\text{solv}} - 1.26 (\text{pred. } r^2=0.27)$$

Cmp code (VS 1)	IC ₅₀ (nM, cKit_D816V)	ΔG_{bind} (observ.)	PBSA _{tot}	ΔE_{coul}	ΔE_{vdw}	ΔG_{solv} (PB)	ΔG_{sa}	ΔG_{bind} (pred.)
6557	750	-8.57	-40.47	-24.78	-65.8	51.57	-1.46	-8.46
5038	1500	-8.15	-33.53	-23.84	-67.85	59.66	-1.5	-5.52
6275	2500	-7.84	-44.91	-25.85	-66.7	49.07	-1.43	-10.08
5003	900	-8.46	-37.47	-46.27	-67.82	78.09	-1.47	-8.79
6567	250	-9.24	-47.78	-38.6	-63.95	56.21	-1.44	-12.54
6395	3000	-7.73	-35.44	-36.4	-64.9	67.34	-1.48	-7.65
6275	2500	-7.28	-44.93	-25.85	-66.7	49.07	-1.45	-10.08
6472	1300	-8.24	-45.64	-34.79	-62.77	53.35	-1.43	-11.61
6377	350	-9.04	-47.6	-35.1	-68.44	57.4	-1.46	-11.56
6376	150	-9.55	-49.99	-42.72	-70.32	64.67	-1.62	-12.73

D.2.2 LR-MM-PBSA model 2-1**Training set 2 (TS2): (Compounds with distance between P_loop and A_loop = 8.75 – 9)**

Table D.11: PBSA calculations for Training set 2

$$\text{PB 2-1 : } \Delta G_{\text{bind}} (\text{pred.}) = 0.01\Delta E_{\text{coul}} + 0.08 \Delta E_{\text{vdw}} - 0.02\Delta G_{\text{solv}} - 2.32$$

$$(r^2=0.5, \text{rmse}=1.14, q^2 =0.15)$$

Cmp code (TS 2)	IC ₅₀ (nM, cKit_D816V)	ΔG_{bind} (observ.)	PBSA _{tot}	ΔE_{coul}	ΔE_{vdw}	ΔG_{solv} (PB)	ΔG_{sa}	ΔG_{bind} (pred.)
6387	250	-9.2429	-44.87	-37.09	-67.97	61.7	-1.51	-9.41
6249	500	-8.8214	-37.44	-27.96	-64.85	56.85	-1.48	-8.95
6503	15000	-6.7535	-136.46	-40.39	-56.55	-38.04	-1.48	-6.48
6620	550	-8.7635	-42.69	-59.64	-61.1	79.54	-1.49	-9.52
6561	450	-8.8855	-41.14	-38.59	-62.81	61.73	-1.47	-9.02
1869	20000	-6.5786	-36.2	-24.16	-59.47	48.88	-1.45	-8.31
6678	150	-9.5535	-50.04	-48.5	-69.86	69.97	-1.51	-9.87
1718	45	-10.2855	-28.73	-33.78	-58.94	65.47	-1.48	-8.74
6581	10	-11.20	-36.43	-41.92	-67.06	74.25	-1.57	-9.66
6155	270	-9.1961	-42.66	-50.82	-67.04	76.71	-1.51	-9.82
6152	100	-9.80	-34.89	-42.37	-63	71.99	-1.51	-9.30

Validation set 2 (VS2): (Compounds with distance between P_loop and A_loop = 8.75 – 9)

Table D.12: PBSA calculations for Validation set 2

$$\text{PB 2-1 : } \Delta G_{\text{bind}} (\text{pred.}) = 0.01\Delta E_{\text{coul}} + 0.08 \Delta E_{\text{vdw}} - 0.02\Delta G_{\text{solv}} - 2.32 (\text{pred. } r^2=0.082)$$

Cmp code (VS 2)	IC ₅₀ (nM, cKit_D816V)	ΔG_{bind} (observ.)	PBSA _{tot}	ΔE_{coul}	ΔE_{vdw}	ΔG_{solv} (PB)	ΔG_{sa}	ΔG_{bind} (pred.)
6621	200	-9.3786	-34.44	-36.47	-55.47	58.92	-1.44	-8.30
6603	400	-8.9571	-41.55	-29.66	-63.36	52.98	-1.51	-8.76
6168	20000	-6.5786	-42.39	-54.01	-58.73	71.83	-1.48	-8.99
5145	20000	-6.5786	-30.4	-30.33	-48.18	49.51	-1.4	-7.47
6569	75	-9.9749	-89.6	-38.48	-59.58	9.93	-1.47	-7.67
5284	350	-9.0383	-42.26	-40.24	-68.08	67.58	-1.52	-9.52
5208	20	-10.7786	-41.07	-44.36	-59.84	64.6	-1.47	-8.84
5221	70	-10.0169	-42.47	-45.26	-69.93	74.23	-1.51	-9.85
5260	200	-9.3786	-35.29	-44.99	-55.3	66.44	-1.44	-8.52
5189	1500	-8.1535	-40.97	-44.09	-56.34	60.9	-1.44	-8.49
6601	300	-9.1320	-40.21	-30.97	-64.82	57.09	-1.51	-8.96
6255	480	-8.8463	-42.57	-36.2	-62.63	57.77	-1.51	-8.85

D.2.3 LR-MM-PBSA model 3-1**Training set 3 (TS3): (Compounds with distance between P_loop and A_loop = 9.4 – 9.7)**

Table D.13: PBSA calculations for Training set 3

$$\text{PB 3-1 : } \Delta G_{\text{bind}} (\text{pred.}) = 0.01\Delta E_{\text{coul}} + 0.11 \Delta E_{\text{vdw}} + 0.02\Delta G_{\text{solv}} - 2.91$$

$$(r^2=0.23, \text{rmse}=1.09, q^2 =0.05)$$

Cmp code (TS 3)	IC ₅₀ (nM, cKit_D816V)	ΔG_{bind} (observ.)	PBSA _{tot}	ΔE_{coul}	ΔE_{vdw}	ΔG_{solv} (PB)	ΔG_{sa}	ΔG_{bind} (pred.)
5287	90	-9.86	-36.01	-26.54	-68.98	61.03	-1.52	-9.24
6194	6500	-7.26	-30.16	-24.47	-57.42	53.19	-1.46	-8.16
6213	1200	-8.29	-39.1	-23.31	-68.03	53.79	-1.5	-9.28
6549	300	-9.13	-49.21	-48.02	-56.78	57.05	-1.46	-8.36
6050	250	-9.24	-40.84	-42.03	-62.13	64.82	-1.5	-8.64
6167	500	-8.82	-44.9	-48.51	-59.5	64.59	-1.48	-8.46
6168	20000	-6.58	-42.39	-54.01	-58.73	71.83	-1.48	-8.26
6528	600	-8.71	-66.05	-37.05	-56.68	29.13	-1.45	-8.93
1860	130	-9.64	-39	-39.49	-67.67	69.66	-1.5	-9.07
6612	600	-8.71	-32.81	-29.13	-53.38	51.15	-1.45	-7.84

Validation set 3 (VS3): (Compounds with distance between P_loop and A_loop = 9.4 – 9.7)

Table D.14: PBSA calculations for Validation set 3

$$\text{PB 3-1 : } \Delta G_{\text{bind}} (\text{pred.}) = 0.01\Delta E_{\text{coul}} + 0.11 \Delta E_{\text{vdw}} + 0.02\Delta G_{\text{solv}} - 2.91 (\text{pred. } r^2=0.24)$$

Cmp code (VS 3)	IC ₅₀ (nM, cKit_D816V)	ΔG_{bind} (observ.)	PBSA _{tot}	ΔE_{coul}	ΔE_{vdw}	ΔG_{solv} (PB)	ΔG_{sa}	ΔG_{bind} (pred.)
6418	90	-9.86	-36.84	-28.62	-58.92	52.16	-1.46	-8.63
6330	100	-9.80	-44.78	-36.01	-67.23	59.95	-1.49	-9.46
1860	130	-9.64	-39	-39.49	-67.67	69.66	-1.5	-9.35
1805	400	-8.96	-36.07	-37.16	-61.63	64.18	-1.46	-8.78
6108	670	-8.64	-34.64	-40.55	-50.58	57.91	-1.42	-7.72
6514	300	-9.13	-48.38	-35.16	-69.11	57.39	-1.5	-9.72
6671	20000	-6.58	-24.43	-20.2	-54.66	51.89	-1.46	-8.09
5190	1300	-8.24	-39.26	-38.79	-56.98	57.96	-1.45	-8.41
5196	150	-9.55	-38.48	-28.8	-65.95	57.76	-1.49	-9.30
5197	60	-10.11	-39.51	-44.61	-60.54	67.14	-1.5	-8.67
5246	200	-9.38	-41.58	-36.79	-66.87	63.57	-1.49	-9.36
5186	1500	-8.15	-33.66	-14.53	-64.42	46.81	-1.52	-9.20

D.2.4 LR-MM-PBSA model 4-1**Training set 4 (TS4): (Compounds with distance between P_loop and A_loop = 10 - 10.5)**

Table D.15: PBSA calculations for Training set 4

$$\text{PB 4-1 : } \Delta G_{\text{bind}} (\text{pred.}) = 0.03\Delta E_{\text{coul}} - 0.13 \Delta E_{\text{vdw}} + 0.04\Delta G_{\text{solv}} - 18.9$$

$$(r^2=0.43, \text{rmse}=0.77, q^2 =0.23)$$

Cmp code (TS 4)	IC ₅₀ (nM, cKit_D816V)	ΔG_{bind} (observ.)	PBSA _{tot}	ΔE_{coul}	ΔE_{vdw}	ΔG_{solv} (PB)	ΔG_{sa}	ΔG_{bind} (pred.)
6535	120	-9.69	-42.5	-31.52	-60.25	50.73	-1.46	-9.66
6041	2500	-7.84	-37.55	-60.24	-60.84	84.99	-1.46	-8.92
6071	350	-9.04	-36.67	-29.25	-63.94	58.01	-1.49	-8.78
6431	250	-9.24	-38.72	-30.61	-63.41	56.78	-1.48	-8.95
1823	1100	-8.34	-37.22	-37.43	-67.03	68.75	-1.51	-8.14
1809	50	-10.22	-38.04	-36.15	-61.09	60.69	-1.49	-9.25
1806	950	-8.43	-38.13	-30.23	-61.35	54.91	-1.46	-9.29
4232	5	-11.62	-36.6	-57.87	-52.7	75.4	-1.43	-10.35
1778	200	-9.38	-36.56	-39.29	-53.63	57.78	-1.42	-10.46

Validation set 4 (VS4):(Compounds with distance between P_loop and A_loop = 10 - 10.5)

Table D.16: PBSA calculations for Validation set 4

$$\text{PB 4-1 : } \Delta G_{\text{bind}} (\text{pred.}) = 0.03\Delta E_{\text{coul}} - 0.13 \Delta E_{\text{vdw}} + 0.04\Delta G_{\text{solv}} - 18.9 (\text{pred. } r^2=0.20)$$

Cmp code (VS 4)	IC ₅₀ (nM, cKit_D816V)	ΔG_{bind} (observ.)	PBSA _{tot}	ΔE_{coul}	ΔE_{vdw}	ΔG_{solv} (PB)	ΔG_{sa}	ΔG_{bind} (pred.)
6545	60	-10.11	-40.43	-45.46	-59.56	66.08	-1.49	-9.42
6418	90	-9.86	-39.01	-21.58	-69.71	53.77	-1.49	-8.11
6186	30	-10.53	-36.45	-28.23	-61.74	54.99	-1.47	-9.24
6303	30	-10.53	-47.97	-36.35	-61.15	51.03	-1.5	-9.64
1800	80	-9.93	-37.12	-43.5	-68.89	76.79	-1.52	-7.74
1745	300	-9.13	-44.59	-37.19	-71.22	65.34	-1.52	-7.77
1752	540	-8.77	-36.08	-33.38	-61.22	59.99	-1.47	-9.21
6458	2700	-7.80	-27.23	-24.51	-54.98	53.68	-1.42	-10.09
6581	10	-11.2	-42.11	-45.27	-67.27	71.99	-1.56	-8.18

Bibliography

- [1] Hanahan, D., Weinberg, R. A., and Francisco, S. (2000) The Hallmarks of Cancer. *Cell* 100, 57–70.
- [2] Hanahan, D., and Weinberg, R. A. (2011) Hallmarks of cancer: the next generation. *Cell* 144, 646–74.
- [3] Blume-Jensen, P., and Hunter, T. (2001) Oncogenic kinase signalling. *Nature* 411, 355–65.
- [4] Sawyers, C. (2004) Targeted cancer therapy. *Nature* 432, 294–7.
- [5] Christoffersen, T., Guren, T. K., Spindler, K.-I. G., Dahl, O., Loenning, P. E., Gjertsen, B. T., Eystein, P., and Tore, B. (2009) Cancer therapy targeted at cellular signal transduction mechanisms: strategies, clinical results, and unresolved issues. *Eur. J. Pharmacol.* 625, 6–22.
- [6] Levitzki, A., and Klein, S. (2010) Signal transduction therapy of cancer. *Mol. Aspects Med.* 31, 287–329.
- [7] Dreesen, O., and Brivanlou, A. H. (2007) Signaling Pathways in Cancer and Embryonic Stem Cells. *Stem Cell Rev.* 3, 7–17.
- [8] Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The Protein Kinase Complement of the Human Genome. *Science* (80-.). 298, 1912–1934.
- [9] Huggins, D. J., Sherman, W., and Tidor, B. (2012) Rational Approaches to Improving Selectivity in Drug Design. *J. Med. Chem.* 55, 1424–1444.
- [10] Lennartsson, A., and Ekwall, K. (2009) Histone modification patterns and epigenetic codes. *Biochim. Biophys. Acta* 1790, 863–8.
- [11] Heery, D. M., and Fischer, P. M. (2007) Pharmacological targeting of lysine acetyltransferases in human disease: a progress report. *Drug Discov. Today* 12, 88–99.

BIBLIOGRAPHY

- [12] Arrowsmith, C. H., Bountra, C., Fish, P. V., Lee, K., and Schapira, M. (2012) Epigenetic protein families : a new frontier for drug discovery. *Nat. Rev. Drug Discov.* *11*, 384–400.
- [13] McLaughlin, F., Finn, P., and La Thangue, N. B. (2003) The cell cycle, chromatin and cancer: mechanism-based therapeutics come of age. *Drug Discov. Today* *8*, 793–802.
- [14] Dyda, F., Klein, D. C., and Hickman, A. B. (2000) Gcn5-related N-acetyltransferases : A Structural Overview. *Ann. Rev. Biophys. Biomol. Struct.* *29*, 81–103.
- [15] Furdas, S. D., Kannan, S., Sippl, W., and Jung, M. (2012) Small molecule inhibitors of histone acetyltransferases as epigenetic tools and drug candidates. *Arch. Pharm. Chem. Life Sci.* *345*, 7–21.
- [16] Holbert, M. A., and Marmorstein, R. (2005) Structure and activity of enzymes that remove histone modifications. *Curr. Opin. Struct. Biol.* *15*, 673–680.
- [17] Luger, K., Maeder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome particle at 2.8 Å resolution. *Nature* *389*, 251–260.
- [18] Hebbes, T. R., and Thorne, A. W. (1988) transcriptionally active chromatin. *EMBO J.* *7*, 1395–1402.
- [19] Lawless, M. W., Norris, S., O’Byrne, K. J., and Gray, S. G. (2009) Targeting histone deacetylases for the treatment of immune, endocrine & metabolic disorders. *Endocr. Metab. Immune Disord. Drug Targets* *9*, 84–107.
- [20] Sterner, D. E., and Berger, S. L. (2000) Acetylation of Histones and Transcription-Related Factors Acetylation of Histones and Transcription-Related Factors. *Microbiol. Mol. Biol. Rev.* *64*, 435–459.
- [21] Smith, B. C., and Denu, J. M. (2009) Chemical mechanisms of histone lysine and arginine modifications. *Biochim. Biophys. Acta* *1789*, 45–57.
- [22] Vetting, M. W., S de Carvalho, L. P., Yu, M., Hegde, S. S., Magnet, S., Roderick, S. L., and Blanchard, J. S. (2005) Structure and functions of the GNAT superfamily of acetyltransferases. *Arch. Biochem. Biophys.* *433*, 212–26.
- [23] Roth, S. Y., Denu, J. M., and Allis, C. D. (2001) Histone acetyltransferases. *Annu. Rev. Biochem.* *70*, 81–120.
- [24] Yang, X.-J. (2004) The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic Acids Res.* *32*, 959–76.

BIBLIOGRAPHY

- [25] Lee, K. K., and Workman, J. L. (2007) Histone acetyltransferase complexes: one size doesn't fit all. *Nat. Rev. Mol. Cell Biol.* 8, 284–95.
- [26] Chan, H. M., and La Thangue, N. B. (2001) p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *J. Cell Sci.* 114, 2363–73.
- [27] Polevoda, B., and Sherman, F. (2002) The diversity of acetylated proteins. *Genome Biol.* 3, 0006.1–6.
- [28] Liu, X., Wang, L., Zhao, K., Thompson, P. R., Hwang, Y., Marmorstein, R., and Cole, P. A. (2008) The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. *Nature* 451, 846–50.
- [29] Hodawadekar, S. C., and Marmorstein, R. (2007) Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. *Oncogene* 26, 5528–40.
- [30] Glozak, M. A., Sengupta, N., Zhang, X., and Seto, E. (2005) Acetylation and deacetylation of non-histone proteins. *Gene* 363, 15–23.
- [31] Poux, A. N., and Marmorstein, R. (2003) Molecular basis for Gcn5/PCAF histone acetyltransferase selectivity for histone and nonhistone substrates. *Biochemistry* 42, 14366–74.
- [32] Eliseeva, E. D., Valkov, V., Jung, M., and Jung, M. O. (2007) Characterization of novel inhibitors of histone acetyltransferases. *Mol. Cancer Ther.* 6, 2391–8.
- [33] Zhang, K., and Dent, S. Y. R. (2005) Histone modifying enzymes and cancer: going beyond histones. *J. Cell. Biochem.* 96, 1137–48.
- [34] Guidez, F., Howell, L., Isalan, M., Cebrat, M., Alani, R. M., Ivins, S., Hormaeche, I., McConnell, M. J., Pierce, S., Cole, P. A., Licht, J., and Zelent, A. (2005) Histone Acetyltransferase Activity of p300 Is Required for Transcriptional Repression by the Promyelocytic Leukemia Zinc Finger Protein. *Mol. Cell. Biol.* 25, 5552–5566.
- [35] Eichenbaum, K. D., Rodriguez, Y., Mezei, M., and Osman, R. (2010) The energetics of the acetylation switch in p53-mediated transcriptional activation. *Proteins, Struct., Funct., and Bioinf.* 78, 447–456.
- [36] Clements, A., Rojas, J. R., Trievel, R. C., Wang, L., Berger, S. L., and Marmorstein, R. (1999) Crystal structure of the histone acetyltransferase domain of the human PCAF transcriptional regulator bound to coenzyme A. *EMBO J.* 18, 3521–32.

BIBLIOGRAPHY

- [37] Kim, C. M., and Cole, P. A. (2001) Bisubstrate ketone analogues as serotonin N-acetyltransferase inhibitors. *J. Med. Chem.* *44*, 2479–85.
- [38] Jones, G., and Willett, P. (1995) Docking small-molecule ligands into active sites. *Pharm. Biotechnol.* *6*, 652–656.
- [39] Irwin, J. J., and Shoichet, B. K. (2005) ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* *45*, 177–82.
- [40] Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., and Coleman, R. G. (2012) ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.*
- [41] Irwin, J., and Shoichet, B. ZINC - A Free Database of Commercially Available Compounds for Virtual Screening (zinc.docking.org).
- [42] Lipinski, C. a. (2001) Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* *44*, 235–49.
- [43] Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., and Kollman, P. (2003) A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase. *J. Comput. Chem.* *24*, 1999–2012.
- [44] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. a., and Case, D. a. (2004) Development and testing of a general amber force field. *J. Comput. Chem.* *25*, 1157–74.
- [45] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* *79*, 926.
- [46] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* *103*, 8577.
- [47] Miyamoto, S., and Kollman, P. a. (1992) Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* *13*, 952–962.
- [48] Humphrey, W., Dalke, a., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* *14*, 33–8, 27–8.
- [49] Stimson, L., Rowlands, M. G., Newbatt, Y. M., Smith, N. F., Raynaud, F. I., Rogers, P., Bavetsias, V., Gorsuch, S., Jarman, M., Bannister, A., Kouzarides, T., McDonald, E., Workman, P., and Aherne, G. W. (2005) Isothiazolones as inhibitors of PCAF and p300 histone acetyltransferase activity. *Mol. Cancer Ther.* *4*, 1521–32.

BIBLIOGRAPHY

- [50] Furdas, S. D., Shekfeh, S., Bissinger, E.-M., Wagner, J. M., Schlimme, S., Valkov, V., Hendzel, M., Jung, M., and Sippl, W. (2011) Synthesis and biological testing of novel pyridoisothiazolones as histone acetyltransferase inhibitors. *Bioorg. Med. Chem.* *19*, 3678–89.
- [51] Szewczuk, L. M., Saldanha, S. A., Ganguly, S., Bowers, E. M., Javoroncov, M., Karanam, B., Culhane, J. C., Holbert, M. a., Klein, D. C., Abagyan, R., and Cole, P. a. (2007) De novo discovery of serotonin N-acetyltransferase inhibitors. *J. Med. Chem.* *50*, 5330–8.
- [52] Furdas, S. D., Shekfeh, S., Kannan, S., Sippl, W., and Jung, M. (2012) Rhodanine carboxylic acids as novel inhibitors of histone acetyltransferases. *Medchemcomm* *3*, 305–311.
- [53] Jie-Lou Liao, J. (2007) Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J. Med. Chem.* *50*, 409–24.
- [54] Zuccotto, F., Ardini, E., Casale, E., and Angiolini, M. (2010) Through the "gatekeeper door": exploiting the active kinase conformation. *J. Med. Chem.* *53*, 2681–94.
- [55] Kufareva, I., and Abagyan, R. (2008) Type-II Kinase Inhibitor Docking, Screening, and Profiling Using Modified Structures of Active Kinase States. *J. Med. Chem.* *51*, 7921–7932.
- [56] Backes, A. C., Zech, B., Felber, B., Klebl, B., and Müller, G. (2008) Small-molecule inhibitors binding to protein kinases . Part I : exceptions from the traditional pharmacophore. *Expert Opin. Drug Discov.* *3*, 1409–1425.
- [57] Backes, A. C., Zech, B., Felber, B., Klebl, B., and Müller, G. (2008) Small-molecule inhibitors binding to protein kinase . Part II : the novel pharmacophore approach of type II and type III inhibition. *Expert Opin. Drug Discov.* *3*, 1427–1449.
- [58] Liu, Y., and Gray, N. S. (2006) Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.* *2*, 358–64.
- [59] Morphy, R. (2010) Selectively Nonselective Kinase Inhibition : Striking the Right Balance. *J. Med. Chem.* *53*, 1413–1437.
- [60] Knight, Z. a., and Shokat, K. M. (2005) Features of selective kinase inhibitors. *Chem. Biol.* *12*, 621–37.
- [61] Knight, Z. A., Lin, H., and Shokat, K. M. (2010) Targeting the cancer kinome through polypharmacology. *Nat. Rev. Cancer* *10*, 130–137.

BIBLIOGRAPHY

- [62] Dar, A. C., and Shokat, K. M. (2011) The evolution of protein kinase inhibitors from antagonists to agonists of cellular signaling. *Annu. Rev. Biochem.* 80, 769–95.
- [63] Scapin, G. (2002) Structural biology in drug design: selective protein kinase inhibitors. *Drug Discov. Today* 7, 601–11.
- [64] Levitzki, A., and Mishani, E. (2006) Tyrphostins and other tyrosine kinase inhibitors. *Annu. Rev. Biochem.* 75, 93–109.
- [65] Mol, C. D., Dougan, D. R., Schneider, T. R., Skene, R. J., Kraus, M. L., Scheibe, D. N., Snell, G. P., Zou, H., Sang, B.-c., and Wilson, K. P. (2004) Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *J. Biol. Chem.* 279, 31655–63.
- [66] Schindler, T., Bornmann, W., Pellicena, P., Miller, W. T., Clarkson, B., and Kuriyan, J. (2000) Structural Mechanism for STI-571 Inhibition of Abelson Tyrosine Kinase. *Science* 289, 1938–42.
- [67] Jain, R. K., Duda, D. G., Clark, J. W., and Loeffler, J. S. (2006) Lessons from phase III clinical trials on anti-VEGF therapy for cancer. *Nat. Clin. Pract. Oncol.* 3, 24–40.
- [68] Ivy, S. P., Wick, J. Y., and Kaufman, B. M. (2009) An overview of small-molecule inhibitors of VEGFr signaling. *Nat. Rev. Clin. Oncol.* 6, 569–579.
- [69] Remsing Rix, L. L., Rix, U., Colinge, J., Hantschel, O., Bennett, K. L., Stranzl, T., Mueller, A., Baumgartner, C., Valent, P., Augustin, M., Till, J. H., and Superti-Furga, G. (2009) Global target profile of the kinase inhibitor bosutinib in primary chronic myeloid leukemia cells. *Leukemia* 23, 477–485.
- [70] Clark, J. D., Flanagan, M. E., and Telliez, J.-b. (2014) Discovery and Development of Janus Kinase (JAK) Inhibitors for Inflammatory Diseases. *J. Med. Chem.* 57, 5023–5038.
- [71] Atwell, S. et al. (2004) A Novel Mode of Gleevec Binding Is Revealed by the Structure of Spleen Tyrosine Kinase. *J. Biol. Chem.* 279, 55827–55832.
- [72] Huang, Y.-M. M., Chen, W., Potter, M. J., and Chang, C.-E. a. (2012) Insights from free-energy calculations: protein conformational equilibrium, driving forces, and ligand-binding modes. *Biophys. J.* 103, 342–51.
- [73] Wang, Z., Canagarajah, B. J., Boehm, J. C., Kassisà, S., Cobb, M. H., Young, P. R., Abdel-Meguid, S., Adams, J. L., and Goldsmith, E. J. (1998) Structural basis of inhibitor selectivity in MAP kinases. *Structure* 6, 1117–28.

BIBLIOGRAPHY

- [74] Simard, J. R., Getlik, M., Grütter, C., Pawar, V., Wulfert, S., Rabiller, M., and Rauh, D. (2009) Development of a fluorescent-tagged kinase assay system for the detection and characterization of allosteric kinase inhibitors. *J. Am. Chem. Soc.* *131*, 13286–96.
- [75] Gajiwala, K. S. et al. (2009) KIT kinase mutants show unique mechanisms of drug resistance to imatinib and sunitinib in gastrointestinal stromal tumor patients. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 1542–7.
- [76] Bixby, D., and Talpaz, M. (2009) Mechanisms of resistance to tyrosine kinase inhibitors in chronic myeloid leukemia and recent therapeutic strategies to overcome resistance. *Hematology Am. Soc. Hematol. Educ. Program* 461–76.
- [77] Daub, H., Specht, K., and Ullrich, A. (2004) Strategies to overcome resistance to targeted protein kinase inhibitors. *Nat. Rev. Drug Discov.* *3*, 1–10.
- [78] Bikker, J. a., Brooijmans, N., Wissner, A., and Mansour, T. S. (2009) Kinase domain mutations in cancer: implications for small molecule drug design strategies. *J. Med. Chem.* *52*, 1493–509.
- [79] Guimarães, C. R. W., Rai, B. K., Munchhof, M. J., Liu, S., Wang, J., Bhattacharya, S. K., and Buckbinder, L. (2011) Understanding the Impact of the P-loop Conformation on Kinase Selectivity. *J. Chem. Inf. Model.* *51*, 1199–204.
- [80] Donato, N. J., and Talpaz, M. (2000) Clinical Use of Tyrosine Kinase Inhibitors : Therapy for Chronic Myelogenous Leukemia and Other Cancers. *Clin. Cancer Res.* *6*, 2965–2966.
- [81] Gounder, M. M., and Maki, R. G. (2011) Molecular basis for primary and secondary tyrosine kinase inhibitor resistance in gastrointestinal stromal tumor. *Cancer Chemother. Pharmacol.* *67 Suppl 1*, S25–43.
- [82] Milojkovic, D., and Apperley, J. (2009) Mechanisms of Resistance to Imatinib and Second-Generation Tyrosine Inhibitors in Chronic Myeloid Leukemia. *Clin. Cancer Res.* *15*, 7519–7527.
- [83] McLean, S. R., Gana-Weisz, M., Hartzoulakis, B., Frow, R., Whelan, J., Selwood, D., and Boshoff, C. (2005) Imatinib binding and cKIT inhibition is abrogated by the cKIT kinase domain I missense mutation Val654Ala. *Mol. Cancer Ther.* *4*, 2008–15.
- [84] Roberts, K. G., Odell, A. F., Byrnes, E. M., Baleato, R. M., Griffith, R., Lyons, A. B., and Ashman, L. K. (2007) Resistance to c-KIT kinase inhibitors conferred by V654A mutation. *Mol. Cancer Ther.* *6*, 1159–66.

BIBLIOGRAPHY

- [85] Hsueh, Y.-S., Lin, C.-L., Chiang, N.-J., Yen, C.-C., Li, C.-F., Shan, Y.-S., Ko, C.-H., Shih, N.-Y., Wang, L.-M., Chen, T.-S., and Chen, L.-T. (2013) Selecting tyrosine kinase inhibitors for gastrointestinal stromal tumor with secondary KIT activation-loop domain mutations. *PLoS One* 8, e65762.
- [86] Schittenhelm, M. M., Shiraga, S., Schroeder, A., Corbin, A. S., Griffith, D., Lee, F. Y., Bokemeyer, C., Deininger, M. W. N., Druker, B. J., and Heinrich, M. C. (2006) Dasatinib (BMS-354825), a dual SRC/ABL kinase inhibitor, inhibits the kinase activity of wild-type, juxtamembrane, and activation loop mutant KIT isoforms associated with human malignancies. *Cancer Res.* 66, 473–81.
- [87] Shah, N. P., Lee, F. Y., Luo, R., Jiang, Y., Donker, M., Akin, C., and Dc, W. (2006) Dasatinib (BMS-354825) inhibits KIT D816V , an imatinib-resistant activating mutation that triggers neoplastic growth in most patients with systemic mastocytosis. *Blood* 108, 286–291.
- [88] Tokarski, J. S., Newitt, J. a., Chang, C. Y. J., Cheng, J. D., Wittekind, M., Kiefer, S. E., Kish, K., Lee, F. Y. F., Borzilleri, R., Lombardo, L. J., Xie, D., Zhang, Y., and Klei, H. E. (2006) The structure of Dasatinib (BMS-354825) bound to activated ABL kinase domain elucidates its inhibitory activity against imatinib-resistant ABL mutants. *Cancer Res.* 66, 5790–7.
- [89] Mellor, H. R., Bell, A. R., Valentin, J.-P., and Roberts, R. R. a. (2011) Cardiotoxicity associated with targeting kinase pathways in cancer. *Toxicol. Sci.* 120, 14–32.
- [90] Force, T., and Kolaja, K. L. (2011) Cardiotoxicity of kinase inhibitors: the prediction and translation of preclinical models to clinical outcomes. *Nat. Rev. Drug Discov.* 10, 111–26.
- [91] Karaman, M. W. et al. (2008) A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* 26, 127–132.
- [92] Cozzini, P., Fornabaio, M., Marabotti, A., Abraham, D. J., Kellogg, G. E., and Mozzarelli, A. (2004) Free Energy of Ligand Binding to Protein : Evaluation of the Contribution of Water Molecules by Computational Methods. *Curr. Med. Chem.* 11, 1345–1359.
- [93] Gohlke, H., and Klebe, G. (2002) Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angew. Chem. Int. Ed. Engl.* 41, 2644–76.
- [94] Ladbury, J. E. (1996) Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.* 3, 973–80.

BIBLIOGRAPHY

- [95] Grubmüller, H., Heymann, B., and Tavan, P. (1996) Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* 271, 997–9.
- [96] De Beer, S. B. a., Vermeulen, N. P. E., Oostenbrink, C., and Oostenbrink, C. (2010) The role of water molecules in computational drug design. *Curr. Top. Med. Chem* 10, 55–66.
- [97] Santos, R., Hritz, J., and Oostenbrink, C. (2010) Role of water in molecular docking simulations of cytochrome P450 2D6. *J. Chem. Inf. Model.* 50, 146–54.
- [98] Michel, J., Tirado-rives, J., and Jorgensen, W. L. (2009) Prediction of the Water Content in Protein Binding Sites. *J. Phys. Chem. B* 113, 13337–13346.
- [99] Luccarelli, J., Michel, J., Tirado-rives, J., and Jorgensen, W. L. (2010) Effects of Water Placement on Predictions of Binding Affinities for p38r MAP Kinase Inhibitors. *J. Chem. Theory Comput.* 6, 3850–3856.
- [100] Michel, J., Tirado-rives, J., and Jorgensen, W. L. (2009) Energetics of Displacing Water Molecules from Protein Binding Sites : Consequences for Ligand Optimization. *JACS* 9, 15403–15411.
- [101] Levy, Y., Onuchic, J. N., and Onuchic, N. (2006) Water Mediation in Protein Folding and Molecular Recognition. *Ann. Rev. Biophys. Biomol. Struct.* 35, 389–415.
- [102] Huggins, D. J., Marsh, M., and Payne, M. C. (2011) Thermodynamic Properties of Water Molecules at a Protein-Protein Interaction Surface. *J. Chem Theory Comput.* 7, 3514–3522.
- [103] Gilson, M. K., Given, J. A., Bush, B. L., and Mccammon, J. A. (1997) The Statistical-Thermodynamic Basis for Computation of Binding Affinities : A Critical Review. *Biophys. J.* 72, 1047–1069.
- [104] Gustchina, a., Sansom, C., Prevost, M., Richelle, J., Wodak, S. Y., Wlodawer, a., and Weber, I. T. (1994) Energy calculations and analysis of HIV-1 protease-inhibitor crystal structures. *Protein Eng.* 7, 309–17.
- [105] Li, Z., and Lazaridis, T. (2006) Thermodynamics of Buried Water Clusters at a Protein-Ligand Binding Interface. *J. Phys. Chem. B* 110, 1464–1475.
- [106] Liu, J., He, X., and Zhang, J. Z. H. (2013) Improving the Scoring of Protein-Ligand Binding Affinity by Including the Effects of Structural Water and Electronic Polarization. *J. Chem. Inf. Model.* 53, 1306–1314.

BIBLIOGRAPHY

- [107] Deng, Y., and Roux, B. (2009) Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B* 113, 2234–46.
- [108] Miyamoto, S., and Kollman, P. A. (1993) Absolute and Relative Binding Free Energy Calculations of the Interaction of Biotin and Its Analogs With Streptavidin Using Molecular Dynamics / Free Energy Perturbation Approaches. *Proteins Struct. Funct. Genet.* 245, 226–245.
- [109] Aqvist, J. (1990) Ion-Water Interaction Potentials Derived from Free Energy Perturbation Simulations. *J. Phys. Chem* 8021–8024.
- [110] Straatsma, T. P., and Berendsen, H. J. C. (1988) Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *J. Chem. Phys.* 89, 5876.
- [111] Guimarães, C. R. W., and Cardozo, M. (2008) MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J. Chem. Inf. Model.* 48, 958–70.
- [112] Lyne, P. D., Lamb, M. L., and Saeh, J. C. (2006) Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* 49, 4805–8.
- [113] Hou, T., Wang, J., Li, Y., and Wang, W. (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods: II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.* 32, 866–877.
- [114] Pearlman, D. A. (2005) Evaluating the Molecular Mechanics Poisson-Boltzmann Surface Area Free Energy Method Using a Congeneric Series of Ligands to p38 MAP Kinase. *J. Med. Chem.* 48, 7796–7807.
- [115] Page, C. S., and Bates, P. A. (2006) Can MM-PBSA Calculations Predict the Specificities of Protein Kinase Inhibitors? *J. Comput. Chem.* 27, 1990–2007.
- [116] Kríz, Z., Otyepka, M., Bártová, I., and Koca, J. (2004) Analysis of CDK2 active-site hydration: a method to design new inhibitors. *Proteins* 55, 258–74.
- [117] Zhang, B., Tan, V. B. C., Lim, K. M., and Tay, T. E. (2007) Significance of Water Molecules in the Inhibition of Cylin-Dependent Kinase 2 and 5 Complexes. *J. Chem. Inf. Model* 47, 1877–1885.
- [118] de Courcy, B., Piquemal, J.-p., and Garbay, C. (2010) Polarizable Water Molecules in Ligand-Macromolecule Recognition . Impact on the Relative Affinities of Competing Pyrrolopyrimidine Inhibitors for FAK Kinase. *JACS* 132, 3312–3320.

BIBLIOGRAPHY

- [119] Spackova, N., Cheatham, T. E., Ryjacek, F., Lankas, F., Meervelt, L. V., Hobza, P., and Sponer, J. (2003) Molecular Dynamics Simulations and Thermodynamics Analysis of DNA-Drug Complexes. Minor Groove Binding between 4',6-Diamidino-2-phenylindole and DNA Duplexes in Solution. *Biochemistry* 1759–1769.
- [120] Makarov, V. A., Andrews, B. K., Smith, P. E., and Pettitt, B. M. (2000) Residence Times of Water Molecules in the Hydration Sites of Myoglobin. *Biophys. J.* 79, 2966–74.
- [121] Henchman, R. H., and McCammon, J. A. (2002) Extracting Hydration Sites around Proteins from Explicit water simulations. *J. Comput. Chem.* 23, 861–869.
- [122] Pearlman, D. A., and Charifson, P. S. (2001) Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System. *J. Med. Chem.* 44, 3417–3423.
- [123] Lazaridis, T. (1998) Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *J. Phys. Chem. B* 102, 3531–3541.
- [124] Li, Z., and Lazaridis, T. (2003) Thermodynamic Contributions of the Ordered Water Molecule in HIV-1 Protease. *JACS* 125, 6636–6637.
- [125] Young, T., Abel, R., Kim, B., Berne, B. J., and Friesner, R. A. (2007) Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *PNAS* 104, 808–13.
- [126] Abel, R., Young, T., Farid, R., Berne, B. J., and Friesner, R. a. (2008) Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* 130, 2817–31.
- [127] Beuming, T., Farid, R., and Sherman, W. (2009) High-energy water sites determine peptide binding affinity and specificity of PDZ domains. *Protein Sci.* 18, 1609–1619.
- [128] Robinson, D. D., Sherman, W., and Farid, R. (2010) Understanding Kinase Selectivity Through Energetic Analysis of Binding Site Waters. *ChemMedChem* 5, 1–11.
- [129] Huggins, D. J., Mckenzie, G. J., Robinson, D. D., Narva, A. J., Hardwick, B., Robertstomson, M., Venkitaraman, A. R., Grant, G. H., and Payne, M. C. (2010) Computational Analysis of Phosphopeptide Binding to the Polo-Box Domain of the Mitotic Kinase PLK1 Using Molecular Dynamics Simulation. *PLoS Comput. Biol.* 6.
- [130] Guimarães, C. R. W., and Mathiowetz, A. M. (2010) Addressing limitations with the MM-GB/SA scoring procedure using the WaterMap method and free energy perturbation calculations. *J. Chem. Inf. Model.* 50, 547–59.

BIBLIOGRAPHY

- [131] Kohlmann, A., Zhu, X., and Dalgarno, D. (2012) Application of MM-GB/SA and WaterMap to SRC Kinase Inhibitor Potency Prediction. *ACS Med. Chem. Lett.* 3, 94–99.
- [132] Beglov, D., and Roux, B. (1997) An Integral Equation To Describe the Solvation of Polar Molecules in Liquid Water. *J. Phys. Chem. B* 101, 7821–7826.
- [133] Fernández, A., and Scott, R. (2003) Dehydron: a structurally encoded signal for protein interaction. *Biophys. J.* 85, 1914–28.
- [134] Fernández, A., and Crespo, A. (2008) Protein wrapping: a molecular marker for association, aggregation and drug design. *Chem. Soc. Rev.* 37, 2373–82.
- [135] Fernández, A. et al. (2007) An anticancer C-Kit kinase inhibitor is reengineered to make it more active and less cardiotoxic. *J. Clin. Invest.* 117, 4044–4054.
- [136] Chène, P. (2008) Challenges in design of biochemical assays for the identification of small molecules to target multiple conformations of protein kinases. *Drug Discov. Today* 13, 522–9.
- [137] Smyth, L. a., and Collins, I. (2009) Measuring and interpreting the selectivity of protein kinase inhibitors. *J. Chem. Biol.* 2, 131–51.
- [138] Gould, T. D. (2006) Targeting glycogen synthase kinase-3 as an approach to develop novel mood-stabilising medications. *Expert Opin. Ther. Targets* 10, 377–92.
- [139] Mazanetz, M. P., and Fischer, P. M. (2007) Untangling tau hyperphosphorylation in drug design for neurodegenerative diseases. *Amyotroph. Lateral Scler.* 6, 464–479.
- [140] Iqbal, K., Alonso, A. D. C., Chen, S., Chohan, M. O., El-Akkad, E., Gong, C.-X., Khatoon, S., Li, B., Liu, F., Rahman, A., Tanimukai, H., and Grundke-Iqbal, I. (2005) Tau pathology in Alzheimer disease and other tauopathies. *Biochim. Biophys. Acta* 1739, 198–210.
- [141] Avila, J. (2006) Tau phosphorylation and aggregation in Alzheimer’s disease pathology. *FEBS Lett.* 580, 2922–7.
- [142] Mazanetz, M. P., Withers, I. M., Loughton, C. a., and Fischer, P. M. (2008) Exploiting glycogen synthase kinase 3beta flexibility in molecular recognition. *Biochem. Soc. Trans.* 36, 55–8.
- [143] Leroy, K., Boutajangout, A., Authelet, M., Woodgett, J. R., Anderton, B. H., and Brion, J.-P. (2002) The active form of glycogen synthase kinase-3beta is associated with granulovacuolar degeneration in neurons in Alzheimer’s disease. *Acta Neuropathol.* 103, 91–9.

BIBLIOGRAPHY

- [144] Hernández, F., de Barreda, E. G., Fuster-Matanzo, A., Goñi Oliver, P., Lucas, J. J., and Avila, J. (2009) The role of GSK3 in Alzheimer disease. *Brain Res. Bull.* 80, 248–50.
- [145] Bhat, R. V., Budd Haeberlein, S. L., and Avila, J. (2004) Glycogen synthase kinase 3: a drug target for CNS therapies. *J. Neurochem.* 89, 1313–7.
- [146] Plyte, S. E., Hughes, K., Nikolakaki, E., Pulverer, B. J., and Woodgett, J. R. (1992) Glycogen synthase kinase-3: functions in oncogenesis and development. *Biochim. Biophys. Acta* 1114, 147–62.
- [147] Kypta, R. M. (2005) GSK-3 inhibitors and their potential in the treatment of Alzheimers disease. *Expert Opin. Pharmacother.* 15, 1315–1331.
- [148] Luo, J. (2009) Glycogen synthase kinase 3beta (GSK3beta) in tumorigenesis and cancer chemotherapy. *Cancer Lett.* 273, 194–200.
- [149] Dorronsoro, I., Castro, A., and Martinez, A. (2002) Inhibitors of glycogen synthase kinase-3: future therapy for unmet medical needs? *Expert Opin. Ther. Pat.* 12, 1527–1536.
- [150] Dominguez, I., and Green, J. B. (2001) Missing links in GSK3 regulation. *Dev. Biol.* 235, 303–13.
- [151] Toogood, P. L. (2001) Cyclin-dependent kinase inhibitors for treating cancer. *Med. Res. Rev.* 21, 487–98.
- [152] Matthews, D. J., and Mary E. Gerritsen, *Targeting Protein Kinases for Cancer Therapy*; 2008; p 798.
- [153] Klein, S., and Levitzki, A. (2006) Signal Transduction Therapy for Cancer - Whither Now? *Curr. Signal Transduct. Ther.* 1, 1–12.
- [154] Sridhar, J., Akula, N., and Pattabiraman, N. (2006) Selectivity and potency of cyclin-dependent kinase inhibitors. *AAPS J.* 8, E204–21.
- [155] Kramer, T., Schmidt, B., and Lo Monte, F. (2012) Small-Molecule Inhibitors of GSK-3: Structural Insights and Their Application to Alzheimer's Disease Models. *Int. J. Alzheimers. Dis.* 2012, 381029.
- [156] Bain, J., Plater, L., Elliott, M., Shpiro, N., Hastie, C. J., McLauchlan, H., Klevernic, I., Arthur, J. S. C., Alessi, D. R., and Cohen, P. (2007) The selectivity of protein kinase inhibitors: a further update. *Biochem. J.* 408, 297–315.

BIBLIOGRAPHY

- [157] Gentles, R. G., Hu, S., and Dubowchik, G. M. *Recent Advances in the Discovery of GSK3 Inhibitors and a Perspective on their Utility for the Treatment of Alzheimer Disease; in Annual reports of medicinal chemistry*; Academic Press, 2009; Vol. 44; pp 3–26.
- [158] Bertrand, J., Thieffine, S., Vulpetti, A., Cristiani, C., Valsasina, B., Knapp, S., Kalisz, H., and Flocco, M. (2003) Structural Characterization of the GSK-3 β Active Site Using Selective and Non-selective ATP-mimetic Inhibitors. *J. Mol. Biol.* 333, 393–407.
- [159] Bhat, R. et al. (2003) Structural insights and biological effects of glycogen synthase kinase 3-specific inhibitor AR-A014418. *J. Biol. Chem.* 278, 45937–45.
- [160] Berg, S., Bergh, M., Hellberg, S., Hogdin, K., Lo-alfredsson, Y., Soderman, P., von Berg, S., Weigelt, T., Ormo, M., Xue, Y., Tucker, J., Neelissen, J., Jerning, E., Nilsson, Y., and Bhat, R. (2011) Discovery of Novel Potent and Highly Selective Glycogen Synthase Kinase-3 β (GSK3 β) Inhibitors for Alzheimers Disease: Design, Synthesis, and Characterization of Pyrazines. *J. Med. Chem.* 55, 9107–9119.
- [161] Voigt, B., Krug, M., Schächtele, C., Totzke, F., and Hilgeroth, A. (2008) Probing novel 1-aza-9-oxafluorenes as selective GSK-3beta inhibitors. *ChemMedChem* 3, 120–6.
- [162] Tell, V., Mahmoud, K. A., Wichapong, K., Schächtele, C., Totzke, F., Sippl, W., and Hilgeroth, A. (2012) Novel aspects in structure activity relationships of profiled 1-aza-9-oxafluorenes as inhibitors of Alzheimers disease-relevant kinases cdk1, cdk5 and gsk3beta. *Medchemcomm* 3, 1413.
- [163] Tell, V., Holzer, M., Herrmann, L., Mahmoud, K. A., Schächtele, C., Totzke, F., and Hilgeroth, A. (2012) Multitargeted drug development: Discovery and profiling of dihydroxy substituted 1-aza-9-oxafluorenes as lead compounds targeting Alzheimer disease relevant kinases. *Bioorg. Med. Chem. Lett.* 22, 6914–8.
- [164] Brasca, M. G. et al. (2009) Identification of N,1,4,4-tetramethyl-8-[[4-(4-methylpiperazin-1-yl)phenyl]amino]-4,5-dihydro-1H-pyrazolo[4,3-h]quinazoline-3-carboxamide (PHA-848125), a potent, orally available cyclin dependent kinase inhibitor. *J. Med. Chem.* 52, 5152–63.
- [165] Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003) Improved protein-ligand docking using GOLD. *Proteins* 52, 609–23.
- [166] Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., and Mee, R. P. (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided. Mol. Des.* 11, 425–45.

BIBLIOGRAPHY

- [167] Friesner, R. a., Banks, J. L., Murphy, R. B., Halgren, T. a., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* *47*, 1739–49.
- [168] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins Struct. Funct. Bioinforma.* *65*, 712–725.
- [169] Wang, W., and Kollman, P. A. (2000) Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.* *303*, 567–82.
- [170] Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Case, D. A., Cheatham, T. E., Duan, Y., Wang, W., Donini, O., Cieplak, P., and Srinivasan, J. (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* *33*, 889–97.
- [171] Kollman, P. (1993) Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* *93*, 2395–2417.
- [172] Gohlke, H., and Case, D. a. (2004) Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.* *25*, 238–50.
- [173] Homeyer, N., and Gohlke, H. (2012) Free Energy Calculations by the Molecular Mechanics Poisson - Boltzmann Surface Area Method. *Mol. Inform.* *31*, 114–122.
- [174] Aqvist, J., and Marelus, J. (2001) The Linear Interaction Energy Method for Predicting Ligand Binding Free Energies. *Comb. Chem. High Throughput Screen.* *4*, 613–626.
- [175] Aqvist, J., Medina, C., and Samuelsson, J.-e. (1994) A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* *7*, 385–391.
- [176] Hansson, T., Marelus, J., and Åqvist, J. (1998) Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aided. Mol. Des.* *12*, 27–35.
- [177] Doudou, S., Sharma, R., Henchman, R. H., Sheppard, D. W., and Burton, N. a. (2010) Inhibitors of PIM-1 kinase: a computational analysis of the binding free energies of a range of imidazo [1,2-b] pyridazines. *J. Chem. Inf. Model.* *50*, 368–79.
- [178] Nervall, M., Hanspers, P., Carlsson, J., Boukharta, L., and Aqvist, J. (2008) Predicting Binding Modes from Free Energy Calculations. *J. Med. Chem.* *51*, 2657–2667.

BIBLIOGRAPHY

- [179] Stjernschantz, E., Marelius, J., Medina, C., Jacobsson, M., Vermeulen, N. P. E., and Oostenbrink, C. (2006) Are Automated Molecular Dynamics Simulations and Binding Free Energy Calculations Realistic Tools in Lead Optimization ? An Evaluation of the Linear Interaction Energy (LIE) Method. *J. Chem. Inf. Model.* *46*, 1972–1983.
- [180] Bissantz, C., Kuhn, B., and Stahl, M. (2010) A Medicinal Chemists Guide to Molecular Interactions. *J. Med. Chem.* *53*, 5061–5084.
- [181] Rönstrand, L. (2004) Signal transduction via the stem cell factor receptor/c-Kit. *Cell. Mol. Life Sci.* *61*, 2535–48.
- [182] Lennartsson, J., and Rönstrand, L. (2006) The stem cell factor receptor/c-Kit as a drug target in cancer. *Curr. Cancer Drug Targets* *6*, 65–75.
- [183] Masson, K., and Rönstrand, L. (2009) Oncogenic signaling from the hematopoietic growth factor receptors c-Kit and Flt3. *Cell. Signal.* *21*, 1717–26.
- [184] Ashman, L. K., and Griffith, R. (2013) Therapeutic targeting of c-KIT in cancer. *Expert Opin. Investig. Drugs* *22*, 103–15.
- [185] Chan, P. M., Ilangumaran, S., La Rose, J., Chakrabarty, A., and Rottapel, R. (2003) Autoinhibition of the Kit Receptor Tyrosine Kinase by the Cytosolic Juxtamembrane Region. *Mol. Cell. Biol.* *23*, 3067–3078.
- [186] Sun, J., Pedersen, M., Rönstrand, L., and Ro, L. (2009) Mechanisms of Signal Transduction : The D816V mutation of c-Kit circumvents a requirement for Src family kinases in c-Kit. *J. Biol. Chem.* *284*, 11039–47.
- [187] Ma, Y. (2002) The c-KIT mutation causing human mastocytosis is resistant to STI571 and other KIT kinase inhibitors; kinases with enzymatic site mutations show different inhibitor sensitivity profiles than wild-type kinases and those with regulatory-type mutations. *Blood* *99*, 1741–1744.
- [188] Piao, X., Paulson, R., van der Geer, P., Pawson, T., and Bernstein, a. (1996) Oncogenic mutation in the Kit receptor tyrosine kinase alters substrate specificity and induces degradation of the protein tyrosine phosphatase SHP-1. *Proc. Natl. Acad. Sci. U. S. A.* *93*, 14665–9.
- [189] Casteran, N., De Sepulveda, P., Beslu, N., Aoubala, M., Letard, S., Lecocq, E., Rottapel, R., and Dubreuil, P. (2003) Signal transduction by several KIT juxtamembrane domain mutations. *Oncogene* *22*, 4710–22.

BIBLIOGRAPHY

- [190] Mol, C. D., Lim, K. B., Sridhar, V., Zou, H., Chien, E. Y. T., Sang, B.-C., Nowakowski, J., Kassel, D. B., Cronin, C. N. C. N., and McRee, D. E. (2003) Structure of a c-Kit Product Complex Reveals the Basis for Kinase Transactivation. *J. Biol. Chem.* 278, 31461–31464.
- [191] Roskoski, R. (2003) STI-571: an anticancer protein-tyrosine kinase inhibitor. *Biochem. Biophys. Res. Commun.* 309, 709–717.
- [192] Roskoski, R. (2005) Signaling by Kit protein-tyrosine kinase—the stem cell factor receptor. *Biochem. Biophys. Res. Commun.* 337, 1–13.
- [193] Demetri, G. D. et al. (2006) Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomised controlled trial. *Lancet* 368, 1329–38.
- [194] Ribatti, D. (2010) Tyrosine Kinase Inhibitors as Antiangiogenic Drugs in Multiple Myeloma. *Stem Cells* 1225–1231.
- [195] Jensen, B. M., Metcalfe, D. D., and Gilfillan, A. M. (2007) Targeting kit activation: a potential therapeutic approach in the treatment of allergic inflammation. *Inflamm. Allergy Drug Targets* 6, 57–62.
- [196] Park, H., Hong, S., and Hong, S. (2012) Nocodazole is a High-Affinity Ligand for the Cancer-Related Kinases ABL, c-KIT, BRAF, and MEK. *ChemMedChem* 7, 53–6.
- [197] Manley, P. W., Buchdunger, E., Fabbro, D., Fendrich, G., Furet, P., Meyer, T., and Zimmermann, J. (2002) Imatinib : a selective tyrosine kinase inhibitor. *Eur. J. Cancer* 38, 19–27.
- [198] Nagar, B., Bornmann, W. G., Pellicena, P., Pd, M. I., Sti, I., Schindler, T., Veach, D. R., Miller, W. T., Clarkson, B., and Kuriyan, J. (2002) Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res.* 62, 4236–43.
- [199] Foster, R., Griffith, R., Ferrao, P., and Ashman, L. (2004) Molecular basis of the constitutive activity and STI571 resistance of Asp816Val mutant KIT receptor tyrosine kinase. *J. Mol. Graph. Model.* 23, 139–52.
- [200] Torrent, M., Rickert, K., Pan, B.-S., and Sepp-lorenzino, L. (2004) Analysis of the activating mutations within the activation loop of leukemia targets Flt-3 and c-Kit based on protein homology modeling. *J. Mol. Graph. Model.* 23, 153–165.

BIBLIOGRAPHY

- [201] Valent, P., Akin, C., Sperr, W. R., Mayerhofer, M., Födinger, M., Fritsche-Polanz, R., Sotlar, K., Escribano, L., Arock, M., Horny, H.-P., and Metcalfe, D. D. (2005) Mastocytosis: pathology, genetics, and current options for therapy. *Leuk. Lymphoma* 46, 35–48.
- [202] Laine, E., Chauvot de Beauchêne, I., Perahia, D., Auclair, C., and Tchertanov, L. (2011) Mutation D816V alters the internal structure and dynamics of c-KIT receptor cytoplasmic region: implications for dimerization and activation mechanisms. *PLoS Comput. Biol.* 7, e1002068.
- [203] Laine, E., Auclair, C., and Tchertanov, L. (2012) Allosteric communication across the native and mutated KIT receptor tyrosine kinase. *PLoS Comput. Biol.* 8, e1002661.
- [204] Heinrich, M. C., Marino-Enriquez, A., Presnell, A., Gastrointestinal, D.-r., Tumors, S., Donsky, R. S., Griffith, D. J., McKinley, A., Patterson, J., Taguchi, T., Liang, C.-W., and Fletcher, J. a. (2012) Sorafenib Inhibits Many Kinase Mutations Associated with Drug-Resistant Gastrointestinal Stromal Tumors. *Mol. Cancer Ther.* 11, 1770–1780.
- [205] Namboodiri, H. V., Bukhtiyarova, M., Ramcharan, J., Karpusas, M., Lee, Y., and Springman, E. B. (2010) Analysis of imatinib and sorafenib binding to p38alpha compared with c-Abl and b-Raf provides structural insights for understanding the selectivity of inhibitors targeting the DFG-out form of protein kinases. *Biochemistry* 49, 3611–8.
- [206] Caenepeel, S., Renshaw-Gegg, L., Baher, A., Bush, T. L., Baron, W., Juan, T., Manoukian, R., Tasker, A. S., Polverino, A., and Hughes, P. E. (2010) Motesanib inhibits Kit mutations associated with gastrointestinal stromal tumors. *J. Exp. Clin. Cancer Res.* 29, 96.
- [207] Dubreuil, P. et al. (2009) Masitinib (AB1010), a potent and selective tyrosine kinase inhibitor targeting KIT. *PLoS One* 4, e7258.
- [208] Wang, R., Lai, L., and Wang, S. (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided. Mol. Des.* 16, 11–26.
- [209] Harris, P. A., Cheung, M., Hunter, R. N., Brown, M. L., Veal, J. M., Nolte, R. T., Wang, L., Liu, W., Crosby, R. M., Johnson, J. H., Epperly, A. H., Kumar, R., Luttrell, D. K., and Stafford, J. a. (2005) Discovery and evaluation of 2-anilino-5-aryloxazoles as a novel class of VEGFR2 kinase inhibitors. *J. Med. Chem.* 48, 1610–9.
- [210] Aleksandrov, A., and Simonson, T. (2010) Molecular dynamics simulations show that conformational selection governs the binding preferences of imatinib for several tyrosine kinases. *J. Biol. Chem.* 285, 13807–15.

BIBLIOGRAPHY

- [211] Lin, Y.-l., Meng, Y., Jiang, W., and Roux, B. (2013) Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. *PNAS* *110*, 1664–1669.
- [212] Seeliger, M. a., Ranjitkar, P., Kasap, C., Shan, Y., Shaw, D. E., Shah, N. P., Kuriyan, J., and Maly, D. J. (2009) Equally potent inhibition of c-Src and Abl by compounds that recognize inactive kinase conformations. *Cancer Res.* *69*, 2384–92.
- [213] Simonson, T. (2002) Free Energy Simulations Come of Age : Protein-Ligand Recognition. *Acc. Chem. Res.* *35*, 430–437.
- [214] Jorgensen, W. L. (1989) Free Energy Calculations: A Breakthrough for Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* *22*, 184–189.
- [215] Genheden, S., and Ryde, U. (2012) Comparison of end-point continuum-solvation methods for the calculation of protein-ligand binding free energies. *Proteins* *80*, 1326–42.
- [216] Swanson, J. M. J., Henchman, R. H., and Mccammon, J. A. (2004) Revisiting Free Energy Calculations : A Theoretical Connection to MM / PBSA and Direct Calculation of the Association Free Energy. *Biophys. J.* *86*, 67–74.
- [217] Sulea, T., Corbeil, C. R., and Purisima, E. O. (2010) Rapid Prediction of Solvation Free Energy. 1. An Extensive Test of Linear Interaction Energy (LIE). *J. Chem. Theory Comput.* *6*, 1608–1621.
- [218] Sulea, T., Cui, Q., and Purisima, E. O. (2011) Solvated Interaction Energy (SIE) for Scoring Protein- Ligand Binding Affinities . 2 . Benchmark in the CSAR- 2010 Scoring Exercise. *J. Chem. Inf. Model.* *51*, 2066–81.
- [219] Deng, Y., and Roux, B. (2008) Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations. *J. Chem. Phys.* *128*, 115103.
- [220] Michel, J., and Essex, J. W. (2010) Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J. Comput. Aided. Mol. Des.* *24*, 639–58.
- [221] Wang, J., Morin, P., Wang, W., and Kollman, P. A. (2001) Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA. *Antivir. Ther.* *123*, 5221–5230.
- [222] Weis, A., Katebzadeh, K., Söderhjelm, P., Nilsson, I., and Ryde, U. (2006) Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field. *J. Med. Chem.* *49*, 6596–606.

BIBLIOGRAPHY

- [223] Kuhn, B., and Kollman, P. A. (2000) Binding of a Diverse Set of Ligands to Avidin and Streptavidin : An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. *J. Med. Chem.* *43*, 3786–3791.
- [224] Haider, M. K., Bertrand, H.-O., and Hubbard, R. E. (2011) Predicting Fragment Binding Poses Using a Combined MCSS MM-GBSA Approach. *J. Chem. Inf. Model.* *51*, 1092–1105.
- [225] Foloppe, N., and Hubbard, R. (2006) Towards predictive ligand design with free-energy based computational methods? *Curr. Med. Chem.* *13*, 3583–3608.
- [226] Huang, N., Kalyanaraman, C., Irwin, J. J., and Jacobson, M. P. (2006) Physics-based scoring of protein-ligand complexes: enrichment of known inhibitors in large-scale virtual screening. *J. Chem. Inf. Model.* *46*, 243–53.
- [227] Huang, N., Kalyanaraman, C., Bernacki, K., and Jacobson, M. P. (2006) Molecular mechanics methods for predicting protein-ligand binding. *Phys. Chem. Chem. Phys.* *8*, 5166–77.
- [228] Thompson, D. C., Humblet, C., and Joseph-McCarthy, D. (2008) Investigation of MM-PBSA rescoring of docking poses. *J. Chem. Inf. Model.* *48*, 1081–91.
- [229] Greenidge, P. a., Kramer, C., Mozziconacci, J.-C., and Wolf, R. M. (2013) MM/GBSA Binding Energy Prediction on the PDBbind Data Set: Successes, Failures, and Directions for Further Improvement. *J. Chem. Inf. Model.* *53*, 201–209.
- [230] Hou, T., and Yu, R. (2007) Molecular dynamics and free energy studies on the wild-type and double mutant HIV-1 protease complexed with amprenavir and two amprenavir-related inhibitors: mechanism for binding and drug resistance. *J. Med. Chem.* *50*, 1177–88.
- [231] Yang, T., Wu, J. C., Yan, C., Wang, Y., Luo, R., Michael, B., Dalby, K. N., and Ren, P. (2011) Virtual Screening Using Molecular Simulations. *Proteins* *79*, 1940–1951.
- [232] Yang, C.-y., Sun, H., Chen, J., Nikolovska-coleska, Z., and Wang, S. (2009) Importance of Ligand Reorganization Free Energy in Protein-Ligand Binding-Affinity Prediction. *J. Am. Chem. Soc.* *131*, 13709–13721.
- [233] Kongsted, J., Soederhjelm, P., and Ryde, U. (2009) How accurate are continuum solvation models for drug-like molecules ? *J. Comput. Aided. Mol. Des.* *23*, 395–409.

BIBLIOGRAPHY

- [234] Numata, J., Wan, M., and Knapp, E.-W. (2007) Conformational entropy of biomolecules: beyond the quasi-harmonic approximation. *Genome Inf. Ser.* 18, 192–205.
- [235] Numata, J. Conformational entropy from molecular simulations: statistical mechanics using the tools of information theory. Ph.D. thesis, 2012.
- [236] Massova, I., and Kollman, P. A. (2000) Combined molecular mechanical and continuum solvent approach (MM-PBSA / GBSA) to predict ligand binding. *Perspect. Drug Discov. Des.* 18, 113–135.
- [237] Baron, R., Hünenberger, P. H., and McCammon, J. A. (2009) Absolute Single-Molecule Entropies from Quasi-Harmonic Analysis of Microsecond Molecular Dynamics: Correction Terms and Convergence Properties. *J. Chem. Theory Comput.* 5, 3150–3160.
- [238] Kongsted, J., and Ryde, U. (2009) An improved method to predict the entropy term with the MM / PBSA approach. *J. Comput. Aided Mol. Des.* 23, 63–71.
- [239] Singh, N., and Warshel, A. (2010) Absolute binding free energy calculations: On the accuracy of computational scoring of protein-ligand interactions. *Proteins* 78, 1724–1735.
- [240] Sham, Y. Y., Muegge, I., and Warshel, A. (1998) The effect of protein relaxation on charge-charge interactions and dielectric constants of proteins. *Biophys. J.* 74, 1744–53.
- [241] Ravindranathan, K., Tirado-Rives, J., Jorgensen, W. L., Guimar, C. R. W., and Guimarães, C. R. W. (2011) Improving MM-GB / SA Scoring through the Application of the Variable Dielectric Model. *J. Chem. Theory Comput.* 7, 3859–3865.
- [242] Singh, N., and Warshel, A. (2010) Absolute binding free energy calculations: On the accuracy of computational scoring of protein-ligand interactions. *Proteins* 78, 1705–1723.
- [243] Huo, S., Massova, I., and Kollman, P. A. (2002) Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *J. Comput. Chem.* 23, 15–27.
- [244] Nurisso, A., Blanchard, B., Audfray, A., Rydner, L., Oscarson, S., Varrot, A., and Imberty, A. (2010) Role of water molecules in structure and energetics of *Pseudomonas aeruginosa* lectin I interacting with disaccharides. *J. Biol. Chem.* 285, 20316–27.
- [245] Wong, S., Amaro, R. E., and Mccammon, J. A. (2009) MM-PBSA Captures Key Role of Intercalating Water Molecules at a Protein-Protein Interface. *J. Chem. Theory Comput.* 5, 422–429.
- [246] Spacková, N., Cheatham, T. E., Ryjáček, F., Lankas, F., Van Meervelt, L., Hobza, P., and Sponer, J. (2003) Molecular dynamics simulations and thermodynamics analysis

BIBLIOGRAPHY

- of DNA-drug complexes. Minor groove binding between 4',6-diamidino-2-phenylindole and DNA duplexes in solution. *J. Am. Chem. Soc.* *125*, 1759–69.
- [247] Freedman, H., Huynh, L. P., Le, L., Cheatham, T. E., Tuszynski, J. a., and Truong, T. N. (2010) Explicitly solvated ligand contribution to continuum solvation models for binding free energies: selectivity of theophylline binding to an RNA aptamer. *J. Phys. Chem. B* *114*, 2227–37.
- [248] Aqvist, J., and Hansson, T. (1996) On the Validity of Electrostatic Linear Response in Polar Solvents. *J. Phys. Chem.* *100*, 9512–9521.
- [249] Roux, B., and Karplus, M. (1994) SIMULATIONS OF THE GRAMICIDIN CHANNEL. *Ann. Rev. Biophys. Biomol. Struct.* *23*, 731–61.
- [250] Hansson, T., and Aqvist, J. (1995) Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.* *8*, 1137–1144.
- [251] Wang, J., Dixon, R., and Kollman, P. A. (1999) Ranking Ligand Binding Affinities With Avidin : A Molecular Dynamics-Based Interaction Energy Study. *Proteins Struct. Funct. Genet.* *34*, 69–81.
- [252] Ljungberg, K. B., Marelius, J., Musil, D., Svensson, P., Norden, B., and Aqvist, J. (2001) Computational modelling of inhibitor binding to human thrombin. *Eur. J. Pharm. Sci.* *12*, 441–6.
- [253] Wall, I. D., Leach, A. R., Salt, D. W., Ford, M. G., and Essex, J. W. (1999) Binding Constants of Neuraminidase Inhibitors : An Investigation of the Linear Interaction Energy Method. *J. Med. Chem.* 5142–5152.
- [254] Lamb, M. L., Tirado-Rives, J., and Jorgensen, W. L. (1999) Estimation of the binding affinities of FKBP12 inhibitors using a linear response method. *Bioorg. Med. Chem.* *7*, 851–60.
- [255] Jones-hertzog, D. K., and Jorgensen, W. L. (1997) Binding Affinities for Sulfonamide Inhibitors with Human Thrombin Using Monte Carlo Simulations with a Linear Response Method. *J. Med. Chem.* *40*, 1539–1549.
- [256] Paulsen, M. D., and Ornstein, R. L. (1996) Binding free energy calculations for P450cam-substrate complexes. *Protein Eng.* *9*, 567–571.
- [257] Wang, W., Wang, J., and Kollman, P. A. (1999) What Determines the van der Waals Coefficient β in the LIE (Linear Interaction Energy) Method to Estimate Binding Free

BIBLIOGRAPHY

- Energies Using Molecular Dynamics Simulations ? *Proteins Struct. Funct. Genet.* 402, 395–402.
- [258] Gilson, M. K., and Zhou, H.-X. (2007) Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* 36, 21–42.
- [259] Chang, C.-e., and Gilson, M. K. (2004) Free Energy, Entropy, and Induced Fit in Host-Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. 13156–13164.
- [260] Chen, W., Chang, C.-e., and Gilson, M. K. (2004) Calculation of Cyclodextrin Binding Affinities : Energy , Entropy , and Implications for Drug Design. *Biophys. J.* 87, 60.
- [261] Aqvist, J. (1996) Calculation of Absolute Binding Free Energies for Charged Ligands and Effects of Long-Range Electrostatic Interactions. *J. Comput. Chem.* 17, 1587–1597.
- [262] Aqvist, J., Luzhkov, V. B., and Brandsdal, B. O. (2002) Ligand binding affinities from MD simulations. *Acc. Chem. Res.* 35, 358–65.
- [263] Carlsson, J., Ande, M., Nervall, M., and Åqvist, J. (2006) Continuum Solvation Models in the Linear Interaction Energy Method. *J. Phys. Chem. B.* 110, 12034–12041.
- [264] Naïm, M., Bhat, S., Rankin, K. N., Dennis, S., Chowdhury, S. F., Siddiqi, I., Drabik, P., Sulea, T., Bayly, C. I., Jakalian, A., and Purisima, E. O. (2007) Solvated interaction energy (SIE) for scoring protein-ligand binding affinities. 1. Exploring the parameter space. *J. Chem. Inf. Model.* 47, 122–33.
- [265] Su, Y., Gallicchio, E., Das, K., Arnold, E., and Levy, R. M. (2007) Linear Interaction Energy (LIE) Models for Ligand Binding in Implicit Solvent : Theory and Application to the Binding of NNRTIs to HIV-1 Reverse Transcriptase. *J. Chem. Theory Comput.* 3, 256–277.
- [266] Zhou, Z., Bates, M., and Madura, J. D. (2006) Structure Modeling , Ligand Binding , and Binding Affinity Calculation (LR-MM-PBSA) of Human Heparanase for Inhibition and Drug Design. *Proteins Struct. Funct. Bioinforma.* 592, 580–592.
- [267] Rizzo, R. C., Tirado-rives, J., and Jorgensen, W. L. (2001) Estimation of Binding Affinities for HEPT and Nevirapine Analogues with HIV-1 Reverse Transcriptase via Monte Carlo Simulations. *J. Med. Chem.* 145–154.
- [268] Zhou, R., Friesner, R. A., Ghosh, A., Rizzo, R. C., Jorgensen, W. L., and Levy, R. M. (2001) New Linear Interaction Method for Binding Affinity Calculations Using a Continuum Solvent Model. *J. Phys. Chem.* 105, 10388–10397.

BIBLIOGRAPHY

- [269] Kolb, P., Huang, D., Dey, F., and Caffisch, A. (2008) Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. *J. Med. Chem.* 51, 1179–88.
- [270] Wichapong, K., Lawson, M., Pianwanit, S., Kokpol, S., and Sippl, W. (2010) Postprocessing of Protein-Ligand Docking Poses Using Linear Response MM-PB / SA : Application to Wee1 Kinase Inhibitors. *J. Chem. Inf. Model.* 50, 1574–1588.

Nomenclature

A-loop : Activation loop of kinase domains

AANAT : Aryl Alkyl N-AcetylTransferase or Serotonin AcetylTransferase; a member of GNAT family of Acetyltransferases.

ABL : Abelson Kinase or Abelson murine leukemia viral oncogene homolog 1 also known as ABL1; is a protein that, in humans, is encoded by the ABL1 gene located on chromosome 9. ABL1 proto-oncogene encodes a cytoplasmic and nuclear protein tyrosine kinase that has been implicated in processes of cell differentiation, cell division, cell adhesion, and stress response. Mutations in the ABL1 gene are associated with chronic myelogenous leukemia (CML). In CML, the gene is activated by being translocated within the BCR (breakpoint cluster region) gene on chromosome 22.

Ac-CoA : Acetyl Coenzyme-A; the cofactor for Acetyltransferases.

ACK1 : activated Cdc42 kinase; a non-receptor tyrosine kinase, ACK1, that binds to multiple receptor tyrosine kinases encoded by TNK2 gene. It interacts with Cdc42Hs in its GTP-bound form and inhibits both the intrinsic and GTPase-activating protein (GAP)-stimulated GTPase activity of Cdc42Hs, using unique sequence of 47 amino acids C-terminal to an SH3 domain. ACK1 is a survival kinase and shown to be associated with tumor cell survival, proliferation, hormone-resistance and radiation resistance. The activation of ACK1 has been observed in prostate, breast, pancreatic, lung and ovarian cancer cells.

AD : Alzheimer disorder

ADA : Adaptorprotein.

AGC : a group of kinases; named after the Protein Kinase A, G, and C families (PKA, PKC, PKG), this group contains many core intracellular signaling kinases which are modulated by cyclic nucleotides, phospholipids and calcium.

Akt : or Protein Kinase B; a serine/threonine-specific protein kinase that plays a key role in multiple cellular processes such as glucose metabolism, apoptosis, cell proliferation, transcription and cell migration.

AMBER : Assisted Model Building with Energy Refinement; a family of force fields for molecular dynamics of biomolecules; and also the name for the molecular dynamics software package that simulates these force fields.

AML : Acute Myeloid Leukaemia

NOMENCLATURE

Aurora : a family of mitotic serine/threonine kinases, classified as part of other kinases. They are implicated with important processes during mitosis and meiosis whose proper function is integral for healthy cell proliferation.

CAMK : a group of protein kinases; best known for the Calmodulin/Calcium regulated kinases (CAMK) in CAMK1 and CAMK2 families, this also has several families of non-calcium regulated kinases (CHK1, CHK2, PKD, and PIM kinases family).

CaMKII : Ca²⁺/calmodulin-dependent protein kinase II or CaM Kinase II. a serine/threonine-specific protein kinase that is regulated by the Ca²⁺/calmodulin complex. CaMKII is involved in many signaling cascades and is thought to be an important mediator of learning and memory.

CBP : CREB-binding protein or CREBBP.

CDK : Cyclin-Dependent Kinase; a family of protein kinases participating in regulating the cell cycle. They are also involved in regulating transcription, mRNA processing, and the differentiation of nerve cells. CDKs become active upon binding to a regulatory proteins called cyclins.

CK1 : Casein kinase 1 family; a family of serine/threonine-selective protein kinases that function as regulators of signal transduction pathways in most eukaryotic cell types. CK1 isoforms are involved in Wnt signaling, circadian rhythms, nucleo-cytoplasmic shuttling of transcription factors, DNA repair, and DNA transcription.

CK2 : Casein kinase 2; a serine/threonine-selective protein kinase. Casein kinase 2 is involved in Wnt signaling, and has been implicated in cell cycle control, DNA repair, regulation of the circadian rhythm and other cellular processes.

CLK : CDK-like kinases; a serine-threonine kinases which act as major regulators of mRNA splicing by phosphorylation of Serine/Arginine-rich (SR) proteins, which function in the RNA processing pathway.

CMGC : a group of protein kinases; named after another set of families (CDK, MAPK, GSK3 and CLK), this group has a diversity of functions in cell cycle control, MAPK signaling, splicing and other unknown functions.

CML : Chronic Myeloid Leukemia

CSFR : Colony stimulating factor-receptor

CYP : Cytochrome P450; superfamily of proteins containing a heme cofactor, acts as the terminal oxidase enzymes in electron transfer chains, broadly categorized as P450-containing systems. CYP enzymes have been identified in all domains of life - animals, plants, fungi, protists, bacteria, archaea, and even in viruses, and they are the major enzymes involved in drug metabolism, accounting for about 75 per cent of the total metabolism.

NOMENCLATURE

- DFG** : a famous motif Aspartate(D)-Phenylalanine(F)-Glycine(G) in the Kinases activation loop used to distinguish between the two functional states of the kinases: the active and inactive conformations.
- DNMT** : DNA-methyltransferase
- EGFR** : Epidermal Growth Factor Receptor; cell-surface receptor for members of the epidermal growth factor family. It is a member of the ErbB family of receptors, a subfamily of four closely related receptor tyrosine kinases: EGFR (ErbB-1), HER2/c-neu (ErbB-2), Her 3 (ErbB-3) and Her 4 (ErbB-4).
- ERK** : Extracellular-signal-Regulated Kinases (ERKs) or classical MAP kinases; the last key regulator of MAPK-ERK pathway; Many different stimuli, including growth factors, cytokines, virus infection, ligands for heterotrimeric G protein-coupled receptors, transforming agents, and carcinogens, activate the ERK pathway.
- FAK** : Focal Adhesion Kinase, known also as Protein Tyrosine Kinase 2 (PTK2).
- FEP** : Free Energy Perturbation; a rigorous computational method to calculate the relative binding energy.
- FGFR** : Fibroblast Growth Factor Receptors; a family of tyrosine kinase receptors that bind to members of the fibroblast growth factor family of proteins. Some of these receptors are involved in pathological conditions.
- FLT3** : Fms-like tyrosine kinase 3, Cluster of differentiation antigen 135 (CD135), receptor-type tyrosine-protein kinase FLT3, or fetal liver kinase-2 (Flk2). FLT3 is a cytokine receptor which belongs to the receptor tyrosine kinase class III. CD135 is the receptor for the cytokine Flt3 ligand (FLT3L).
- FXa** : Stuart-Prower factor known also as prothrombinase, thrombokinese or thromboplastin; a serine endopeptidase which is a member of the coagulation cascade.
- GB** : Generalized Born model of implicit solvent
- GIST** : GastroIntestinal Stromal Tumor
- GNAT** : GCN5-related N-AcetylTransferase; A big family of Acetyltransferases.
- GOLD** : Genetic Optimizaion of Ligand Docking; docking software by the Cambridge Crystallographic Data Centre (CCDC).
- GPCR** : G Protein-Coupled Receptors; a large protein family of receptors that sense molecules outside the cell and activate inside signal transduction pathways and, ultimately, cellular responses. They are involved basically in c-AMP pathway and PI3K pathway. Their ligands include light-sensitive compounds, odors, pheromones, hormones, and neurotransmitters, and vary in size from small molecules to peptides to large proteins.
- Gromacs** : GRONingen MAchine for Chemical Simulations; a molecular dynamics package primarily designed for simulations of proteins, lipids and nucleic acids, originally developed in the Biophysical Chemistry department of University of Groningen. It is also

NOMENCLATURE

free, open source released under the GNU General Public License. Gromacs can use many force-fields like GROMOS, Amber, CharmM, and OPLS.

GSK3 : Glycogen Synthase Kinase 3; a serine/threonine protein kinase from the CMGK group. In mammals GSK-3 is encoded by two known genes, GSK-3 alpha (GSK3A) and GSK-3 beta (GSK3B). it is implicated in a number of diseases, including Type II diabetes (Diabetes mellitus type 2), Alzheimer Disease, inflammation, cancer, and bipolar disorder.

HAT : Histone AcetylTransferases

HCK : hemopoietic cell kinase; a member from Src tyrosine kinases. It plays a role in neutrophil migration and in the degranulation of neutrophils.

HDAC : Histone DeAcetylases

hERG : human Ether-à-go-go-Related Gene; a gene (KCNH2) that codes for a protein known as Kv11.1, the alpha subunit of a potassium ion channel.

IGF1 : Insulin-like Growth Factor 1.

IKK : IκB Kinase; an enzyme complex that is involved in propagating the cellular response to inflammation by the upstream NF-κB signal transduction cascade.

ILK : Integrin-Linked Kinase; a serine/threonine protein kinase with 5 ankyrin-like repeats, which associates with the cytoplasmic domain of beta integrins and acts as a proximal receptor kinase regulating integrin-mediated signal transduction. ILK was found to localize to the centrosome and regulate mitotic spindle organization.

IRAK : Interleukin-1-Receptor-Associated Kinases; a family of serine/threonine kinases that become associated with the interleukin-1 receptor (IL1R) upon stimulation. They are reported to participate in the IL1-induced upregulation of NF-kappaB. these kinases associate with TRAFs and signal through the NFκB and Jnk pathways. What appear to be plant orthologs account for the vast majority of plant kinases, with over 500 members in higher plant genomes, most of which are receptors, possibly analogous to the receptor tyrosine kinases of metazoans.

IRK : Insulin Receptor Kinase; a transmembrane receptor that is activated by insulin, IGF-I, IGF-II and belongs to the large class of tyrosine kinase receptors.

IRS : insulin receptor substrate protein

JAK : Janus kinase; a family of intracellular, nonreceptor tyrosine kinases that transduce cytokine-mediated signals via the JAK-STAT pathway. The name is taken from the two-faced Roman god of beginnings and endings, Janus, because the JAKs possess two near-identical phosphate-transferring domains. One domain exhibits the kinase activity, while the other negatively regulates the kinase activity of the first.

JM : JuxtaMembrane; the part of the cytoplasmic region which is adjacent to the transmembrane domain of Receptor kinases.

NOMENCLATURE

- KDR** : Kinase insert Domain Receptor; also known as Vascular Endothelial Growth Factor Receptor 2 (VEGFR-2): a type III receptor tyrosine kinase.
- KIT** : Mast/stem cell growth factor receptor (SCFR), also known as proto-oncogene c-Kit or tyrosine-protein kinase Kit or CD117. CD117 is an important cell surface marker used to identify certain types of hematopoietic (blood) progenitors in the bone marrow. To be specific, hematopoietic stem cells (HSC), multipotent progenitors (MPP), and common myeloid progenitors (CMP) express high levels of CD117. In addition, mast cells, and melanocytes in the skin express CD117.
- Lck** : Lymphocyte-specific protein tyrosine Kinase; a member of the Src family of tyrosine kinases. It regulates intracellular signaling pathways inside lymphocytes, mostly T cells.
- LIE** : Linear Interaction Energy
- LIECE** : Linear Interaction Energy with Continuum Electrostatics
- LR-MM-PBSA** : Linear-Response-Molecular Mechanics-Poisson Boltzmann/ Solvent-accessible Surface Area.
- LRA** : Linear Response Approximation.
- MAPK** : Mitogen-Activated Protein Kinases or ERKs; a family of protein kinases in CMGC group; they are involved in directing cellular responses to a diverse array of stimuli, such as mitogens, osmotic stress, heat shock and proinflammatory cytokines, regulating proliferation, gene expression, differentiation, mitosis, cell survival, and apoptosis - among many others. They include ERKs, p38 MAPKs, and JNKs, beside atypical MAPKs.
- MC** : Monte Carlo simulation; a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results, resembling the act of playing and recording results in a real gambling casino. They are useful for simulating systems with many coupled degrees of freedom, such as fluids, disordered materials, strongly coupled solids, and cellular structures.
- MD** : Molecular Dynamics; a computer simulation of physical movements of atoms and molecules in the context of N-body simulation. Movements of atoms or particles are determined by numerically solving the Newton's equations of motion for a system of interacting particles, where forces between the particles and potential energy are defined by molecular mechanics force fields.
- MEK** : Mitogen/Extracellular signal-regulated Kinase, also known as MAPKK or MAP2K: Mitogen-Activated Protein Kinase Kinase; a kinase enzyme which phosphorylates mitogen-activated protein kinase (MAPK). They works as activators of ERK (MAPKs), JNK, and p38 MAPK.
- MET** or c-MET: hepatocyte growth factor receptor; a membrane receptor that is essential for embryonic development and wound healing, activated by Hepatocyte growth factor (HGF). Abnormal MET activation in cancer correlates with poor prognosis, where aberrantly active MET triggers tumor growth, formation of new blood vessels (angiogenesis) that supply the tumor with nutrients, and cancer spread to other organs (metastasis).

NOMENCLATURE

- MET is deregulated in many types of human malignancies, including cancers of kidney, liver, stomach, breast, and brain.
- MM-GBSA : Molecular Mechanics-Generalized Born/Solvent-accessible Surface Area.
- MM-PBSA : Molecular Mechanics-Poisson Boltzmann/ Solvent-accessible Surface Area.
- MOZ : MYST histone acetyltransferase (monocytic leukemia) 3, also known as MYST3.
- MYST : Histone Acetyltransferase family which include: include MOZ, YBF2/SAS3, SAS2 and TIP60.
- NF κ B : Nuclear Factor kappa-light-chain-enhancer of activated B cells, a protein complex that controls transcription of DNA. NF- κ B is found in almost all animal cell types and is involved in cellular responses to stimuli such as stress, cytokines, free radicals, ultraviolet irradiation, oxidized LDL, and bacterial or viral antigens
- NFT : Neuro-Fibrillary Tangle; are aggregates of hyperphosphorylated tau protein that are most commonly known as a primary marker of Alzheimer's Disease. Their presence is also found in numerous other diseases known as tauopathies.
- NRTKs : Non-Receptor Tyrosine Kinases or cytoplasmic tyrosine kinases; include 32 tyrosine kinases which regulate cellular processes like cell's growth, proliferation, differentiation, adhesion, migration and apoptosis and they are critical components in the regulation of the immune system. They are classified into: ABL family, ACK family, CSK family, FAK family, FES family, FRK family, JAK family, SRC-A family, SRC-B family, TEC family, and SYK family.
- P-loop : Glycine-rich Phosphate-binding-loop or Glycine-rich P-loop; a flexible loop in the N-lobe of kinase domains, which binds mainly to phosphate groups of ATP.
- p300-CBP : p300-CBP coactivator family is composed of two closely related transcriptional co-activating proteins (or coactivators): p300 (also called EP300 or E1A binding protein p300), and CBP (also known as CREB-binding protein or CREBBP).
- p38 or p38 mitogen-activated protein kinases (p38 MAPKs): also called RK or CSBP (Cytokinin Specific Binding Protein); serine/threonine/tyrosine-specific protein kinases belonging to the CMGC (CDK/MAPK/GSK3/CLK) kinase group. They regulate proliferation, gene expression, differentiation, mitosis, cell survival, and apoptosis. They form a class of mitogen-activated protein kinases that are responsive to stress stimuli, such as cytokines, ultraviolet irradiation, heat shock, and osmotic shock, and are involved in cell differentiation, apoptosis and autophagy. Four p38 MAP kinases, p38- α (MAPK14), - β (MAPK11), - γ (MAPK12 / ERK6), and - δ (MAPK13 / SAPK4), have been identified. Similar to the SAPK/JNK pathway, p38 MAP kinase is activated by a variety of cellular stresses including osmotic shock, inflammatory cytokines, lipopolysaccharides (LPS), Ultraviolet light, and growth factors.
- p53 or Tumor protein p53: also known as p53, cellular tumor antigen p53, phosphoprotein p53, or tumor suppressor p53, is a protein that in humans is encoded by the TP53 gene.

NOMENCLATURE

It is crucial in multicellular organisms, where it regulates the cell cycle and, thus, functions as a tumor suppressor, preventing cancer; described as the guardian of the genome because of its role in conserving stability by preventing genome mutation.

PB : Poison-Boltzmann model of implicit solvent

PCAF : p300/CBP-associated factor; also known as K(lysine) acetyltransferase 2B (KAT2B), a human gene and transcriptional coactivator (with histone acetyltransferase activity) associated with p53. PCAF has separate acetyltransferase and E3 ubiquitin ligase domains as well as a bromodomain for interaction with other proteins.

PDGFR : Platelet-Derived Growth Factor Receptors; cell surface tyrosine kinase receptors for members of the platelet-derived growth factor (PDGF) family. They include extracellular region of the receptor consists of five immunoglobulin-like domains while the intracellular part is a tyrosine kinase domain. Dimerization is a prerequisite for the activation of the kinase, and it happens only upon binding with their ligands/growth factors. They include: PDGFR α , PDGFR β , c-KIT, FLT3, and CSF1R; which are involved in important signaling pathways: MAPK-ERK, and PI3K pathways.

PDZ domain : the first letters of three proteins: post synaptic density protein (PSD95), Drosophila disc large tumor suppressor (Dlg1), and zonula occludens-1 protein (zo-1); which share a common structural domain of 80-90 amino-acids found in the signaling proteins of bacteria, yeast, plants, viruses and animals.

PI3K : Phosphoinositide-3-Kinases; a family of enzymes involved in cellular functions such as cell growth, proliferation, differentiation, motility, survival and intracellular trafficking, which in turn are involved in cancer.

PIM : Proto-oncogene serine/threonine-protein kinases; a family of serine/threonine kinases; first described in relation to murine T-cell lymphomas. Three isoforms of PIM kinases exist in humans, involved in multiple human cancers, including prostate cancer, acute myeloid leukemia and other hematopoietic malignancies.

PKAc : Protein Kinase Ac

PKC : Protein kinase C; a family of serine/threonine protein kinase enzymes. PKC is involved in receptor desensitization, in modulating membrane structure events, in regulating transcription, in mediating immune responses, in regulating cell growth, and in learning and memory. Effects of PKC are cell-type-specific.

PKL : Protein kinase-like; Contains a number of diverse families that share a PKL fold and catalytic mechanism with the ePKs but do not have substantial sequence similarity. This group also contains a number of lipid, sugar, and other small-molecule kinases.

PKs : Protein Kinases; kinase enzymes that modify other proteins by chemically adding phosphate groups to them (phosphorylation). Phosphorylation usually results in a functional change of the target protein (substrate) by changing enzyme activity, cellular location, or association with other proteins. The human genome contains about 500 protein kinase genes and they constitute about 2

NOMENCLATURE

PIGF : Placenta Growth Factor

PLK : Polo-Like Kinases; Protein kinases from group other kinases, they are important regulators of cell cycle as they are involved in the formation of the mitotic spindle and in the activation of CDK/cyclin complexes during M-phase of the cell cycle.

PMF : Potential of Mean Force

Pyk2 or PTK2B: Protein tyrosine kinase 2 beta; a cytoplasmic protein tyrosine kinase and member of FAK subfamily, that is involved in calcium-induced regulation of ion channels and activation of the map kinase signaling pathway.

RAF kinase : Rapidly Accelerated Fibrosarcoma kinase; a family of three serine/threonine-specific protein kinases (from TKL group) that are related to retroviral oncogenes. Activation of RAF kinases requires interaction with RAS-GTPases, to participate in the RAS-RAF-MEK-ERK signal transduction cascade (acting as MAP4K).

Ras : Rat sarcoma; the name given to a family of related proteins (small GTPase) which is ubiquitously expressed in all cell lineages and organs.

RET : RET is an abbreviation for rearranged during transfection, a receptor tyrosine kinase for members of the glial cell line-derived neurotrophic factor (GDNF) family of extracellular signalling molecules; encoded by the RET proto-oncogene gene on chromosome 10. RET loss of function mutations are associated with the development of Hirschsprung's disease, while gain of function mutations are associated with the development of various types of human cancer, including medullary thyroid carcinoma, multiple endocrine neoplasias type 2A and 2B, pheochromocytoma and parathyroid hyperplasia.

RGC : Receptor Guanylate Cyclases; a small group contains an active guanylate cyclase domain, which generates the cGMP second messenger, and a catalytically inactive kinase domain, which appears to have a regulatory function.

Rhodanine : 2-Thioxo-4-thiazolidinone or 2-Thioxo-1,3-thiazolidin-4-one

RIPK : Receptor-Interacting Protein Kinase; a family of serine/threonine-protein kinases, which have function in a variety of cellular pathways including the NF- κ B pathway and programmed necrotic cell death (necroptosis).

RMSD : Root-Mean-Squared-Deviation

ROC curve : Receiver Operating Characteristic curve

ROCK : RhO-associated Kinase; a kinase belonging to the AGC (PKA/ PKG/PKC) family of serine-threonine kinases. It is involved mainly in regulating the shape and movement of cells by acting on the cytoskeleton.

ROR : RAR-related orphan receptor; a member of the nuclear receptor family of intracellular transcription factors.

NOMENCLATURE

- ROR2** : Tyrosine-protein kinase transmembrane receptor; known as neurotrophic tyrosine kinase, receptor-related 2, responsible for aspects of bone and cartilage growth.
- RTK** : Receptor Tyrosine Kinase; high-affinity cell surface receptors for many polypeptide growth factors, cytokines, and hormones. Out of 90 tyrosine kinase genes, 58 encodes receptor tyrosine kinase. They are classified to 20 classes or families: EGF receptor family or ErbB family, Insulin receptor family, PDGF receptor family, FGF receptor family, VEGF receptors family, HGF receptor family, Trk receptor family, Eph receptor family, AXL receptor family, LTK receptor family, TIE receptor family, ROR receptor family, DDR receptor family, RET receptor family, KLG receptor family, RYK receptor family, MuSK receptor family, ROS receptor family, and AATYK receptor family.
- SAGA** : Spt, Ada, GCN5 Acetyltransferase.
- SAR** : Structure-Activity Relationship
- SAS2** : Something about Silencing 2; yeast gene which encodes one of MYST histone acetyltransferases.
- SCLC** : Small Cell Lung Cancer
- SM** : Systematic Mastocytosis
- Src** : (pronounced sarc as it is short for sarcoma) is a proto-oncogene encoding a non-receptor protein tyrosine kinase family called Src family kinases.
- STAT protein** : Signal Transducer and Activator of Transcription; a transcription factor regulates many aspects of growth, survival and differentiation in cells.
- STE** : a group of protein kinases; Homologs of the yeast STE7, STE11 and STE20 genes, which form the MAPK cascade, transducing signals from the surface of the cell to the nucleus.
- SYK** : Spleen tyrosine kinase; a family of non-receptor cytoplasmic tyrosine kinases which includes SYK kinase and ZAP-70. Both kinases Syk and Zap-70 transmit signals from the B-Cell receptor and T-Cell receptor. Syk plays a similar role in transmitting signals from a variety of cell surface receptors. SYK is implicated in several instances of hematopoietic malignancies.
- TAF** : TATA-binding protein associated factor
- TGF** : Transforming Growth Factor; two classes of polypeptide growth factors, TGF-alpha and TGF-beta. Both classes are upregulated in some types of cancer and disorders.
- TI** : Thermodynamic Integration; a rigorous computational method to calculate the relative binding energy.
- TIE** : tyrosine kinase with immunoglobulin-like and EGF-like domains; an angiopoietin receptor which are cell-surface receptors that bind and are activated by the angiopoietins, (Ang1, Ang2, Ang3, Ang4).

NOMENCLATURE

- TIP60 : HIV-1 TAT-interactive protein; a histone acetyl transferase (HAT) of the MYST family.
- TK : Tyrosine kinases group; This group phosphorylates almost exclusively on tyrosine residues, as opposed to most other kinases that are selective for serine or threonine.
- TKL : Tyrosine kinase-like group; The group most similar to tyrosine kinases, but whose activities are generally on serine/threonine substrates.
- Trk : Trk receptors; a family of tyrosine kinases that regulates synaptic strength and plasticity in the mammalian nervous system, affecting neuronal survival and differentiation through several signal cascades. The common ligands of trk receptors are neurotrophins, a family of growth factors critical to the functioning of the nervous system.
- vdW : van der Waals interaction or potential; the sum of the attractive or repulsive forces between molecules (or between parts of the same molecule) other than those due to covalent bonds or the electrostatic interaction; including force between neutral atoms, force between two permanent dipoles (Keesom force), a permanent dipole and a corresponding induced dipole (Debye force), between two instantaneously induced dipoles (London dispersion force). usually, it is simplified to Lennard-Jones potential, that approximates the interaction between a pair of neutral atoms or molecules.
- VEGFR : Vascular Endothelial Growth Factor Receptor. a family of important tyrosine kinase receptors, involved in both vasculogenesis (the formation of the circulatory system) and angiogenesis (the growth of blood vessels from pre-existing vasculature).
- WEE1 : a nuclear serine/threonine kinase and a key regulator of cell cycle progression. It is a kinase determining the timepoint of entry into mitosis, through inhibiting Cdk1, and consequently influencing the size of the daughter cells.

Curriculum Vitae

Personal Information

Name: Suhaib Shekfeh
Date of birth: 15th August 1976
Place of birth: Homs (Syria)
Nationality: Syrian
E-mail: suhaib.shekfeh(at)pharmazie.uni-halle.de
Alternative E-Mail: s.shekfeh@gmail.com

Education

1. Diplom of Pharmacy (Post-Graduate) : Martin-Luther-University Halle-Wittenberg June 2009 Diploma Thesis: Molecular modeling and discovery of novel inhibitors for Histone AcetylTransfearse (Pcaf).
2. 1993–1998 B.SC degree in pharmacy : University of Damascus, Syria.

Work History

Drug Design Researcher

- July 2009 – PhD student and scientific coworker in Prof. Wolfgang Sippl Research Group.
- 2008 Diplom study in Martin-Luther-University Halle Molecular modeling and Docking studies for the discovery of Histone AcetylTransferase (Pcaf).
- Virtual screening for Inhibitors for Histone Acetyltransferases (Pcaf) : Diploma work in MLU Halle-Wittenberg, under supervision of Prof. Dr. Wolfgang Sippl
- 2001–2003 Jordan University. Amman, Jordan
- Building a chemical compounds database using Catalyst.
- Building pharmacophore models for pharmaceutical targets using Hypogen in Catalyst.
- Qsar studies for improving the pharmacophore prediction.

Teaching Experience

- Analytical Chemistry (according to Pharmacopoeia); Arzneibuch Untersuchung Praktikum; 2010-1012 (Institute für Pharmazie, MLU Halle-Wittenberg)

Other experience

- Polymer modifications for using it in microcapsules production. Pharmaceutical Industry and Marketing
- 2007–February 2008 Al-Farouk Pharmaceuticals Damascus, Syria Technical support , and Marketing Formulation developer using Maltrin™ & Spreps™.
- Official pharmacist.
- Health service, Drug Formulation for special purpose

PROFESSIONAL SOCIETY

1. American Chemical Society, member

Professional Achievements

Publications

1. B. Maurer, U. Mathias, S. Shekfeh, P. Papatheodorou, J. Orth, T. Jank, C. Schwan, W. Sippl, K. Aktories, M. Jung. From cosubstrate similarity to inhibitor diversity-inhibitors of ADP-ribosyltransferases from kinase inhibitor screening. *Mol. BioSyst.*, 2011, 7, 799-808.
2. S. D. Furdas, S. Shekfeh, E. Bissinger, J. Wagner, C. H. Arrowsmith, M. M. Mangos, V. Valkov, M. Hendzel, M. Jung, W. Sippl. Synthesis and biological testing of novel pyridoisothiazolones as histone acetyltransferase inhibitors. *Bioorganic & Medicinal Chemistry* 19(12), 2011, 3678- 3689.
3. S. D. Furdas, S. Shekfeh, S. Kannan, W. Sippl, M. Jung, Rhodanine carboxylic acids as novel inhibitors of Histone acetyltransferases. *Med. Chem. Commun.* 2012, 3, 305-311.

Posters

1. Shekfeh, S.; Hilgeroth, A. and Sippl, W., Computational analysis of conformational changes in GSK3- β /CDK2 kinases for understanding inhibitor selectivity. 3rd International Meeting - Conformational transitions in macromolecular interactions. 4th - 6th November 2013, Halle (Saale), Germany.

2. S. Shekfeh, S.D. Furdas, M. Jung, W. Sippl. Docking study and binary classification model of isothiazolones as irreversible inhibitors of the histone acetyltransferase PCAF. in : (International Workshop – New Approaches in Drug Design & Discovery, Rauschholzhausen, 22.-25.03.2010, Jahrestagung der Deutschen Pharmazeutischen Gesellschaft, Jena, 29.09 – 01.10.2009, 5th Summer School Medicinal Chemistry., 13.-15. 09.2010, Regensburg.)
3. S. D. Furdas, J. M. Wagner, S. Shekfeh, P. Brown, W. Sippl, M. Jung. Histone acetyltransferases inhibition by pyridoisothiazolones. EMBO Conference Series, Chemical Biology 2010 22.-25.09.2010, EMBL Heidelberg.
4. B. Maurer, P. Papatheodorou, C. Schwan, S. Shekfeh, W. Sippl, K. Aktories, M. Jung. Development of an enzymatic assay for ADP-ribosylating enzymes via the chemical quantitation of the co-substrate NAD. 5th Summer School Medicinal Chemistry. 13.-15. 09.2010, Regensburg
5. S. D. Furdas, J. M. Wagner, S. Shekfeh, P. Brown, W. Sippl, M. Jung Pyridoisothiazolones as a novel class of inhibitors of histone acetyltransferase activity: synthesis and biological testing. XXXth European School of Medicinal Chemistry (ESMEC) 04.-09. 07. 2010, Urbino, Italien, and Jahrestagung der Deutschen Pharmazeutischen Gesellschaft, Jena, 29.09 – 01.10.2009.

Erklärung (Declaration of Authorship)

Hiermit erkläre ich, dass ich die vorliegende Dissertationsschrift selbständig und ohne fremde Hilfe angefertigt, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und die aus ihnen wörtlich oder inhaltlich entnommen Stellen als solche kenntlich gemacht habe.

Die Arbeit wurde ausschließlich der Mathematisch-Naturwissenschaftlichen Fakultät der Martin-Luther-Universität Halle-Wittenberg vorgelegt und an keiner anderen Universität oder Hochschule weder im In- oder Ausland zur Erlangung des Doktorgrades eingereicht.

Halle (Saale), den 11.11.2014