ASSESSING PROTEIN-LIGAND BINDING MODES, NOVEL DRUG SKELETON CANDIDATES FOR PDE4B AND CONFORMATIONAL REARRANGEMENTS OF EF-TU IN GTP HYDROLYSIS WITH COMPUTATIONAL TOOLS

by

Gülşah Çifci Bağatır

B.S., Chemistry, Boğaziçi University, 2008

M.S., Chemistry, Boğaziçi University, 2010

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Graduate Program in Chemistry

Boğaziçi University

2017

To My Family

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis advisor Prof. Viktorya Aviyente (Boğaziçi University) for her guidance, help, advice and patience as well as her continuous support to complete this work. I am very proud of being her student. I would also thank to my second advisor Assoc. Prof. Şaron Çatak for her supporting remarks.

Besides my advisors, I wish to extend my appreciation to the members of my committee: Assoc. Prof. Demet Akten (Kadir Has University), Assist Prof. Bülent Balta (Istanbul Technical University), Assoc. Prof. Rana Sanyal (Boğaziçi University), and Prof. Nurcan Tüzün for their support, valuable advices and comments. Special thanks to Assoc. Prof. Demet Akten and Assist. Prof. Bülent Balta for their contributions to the main chapters of this dissertation.

I would also like to acknowledge Prof Gerald Monard from Theoretical Chemistry and Biochemistry Group, SRSMC, Université de Lorraine in Nancy for his collaboration and continous help during the study of this dissertation.

The research has been supported by TÜBİTAK-CNRS project (Project No 113Z001). I acknowledge the TÜBİTAK BİDEB graduate student scholarship for the financial support and Bogazici University Foundation (BUVAK). Also, I acknowledge TÜBİTAK ULAKBİM High Performance Computing Center as well as the Istanbul Technical University National Center for high Performance Computing UHEM (Project number-1002822013) for computational resources. Finally, I also want to thank to Bogazici University Research Fund (12260 and 1856) for their financial support for my research.

I would like to thank all the former and current members of the Boğaziçi University Computational Chemistry group and the members of the Chemistry Department for always being kind and helpful. My special thanks go to my dear friends Tuğba Furuncuoğlu Özaltın and Sesil Agopcan Çınar for their endless support and friendship.

Finally, I would like to express my deepest gratitude and respect to my family, my mother, her support, advices and help every moment in my life, my father and my sister and of course my husband for his endless patience, continuous support, understanding and love throughout my life. I am very proud of being a member of this family in my life.

ABSTRACT

ASSESSING PROTEIN-LIGAND BINDING MODES, NOVEL DRUG SKELETON CANDIDATES FOR PDE4B AND CONFORMATIONAL REARRANGEMENTS OF EF-TU IN GTP HYDROLYSIS WITH COMPUTATIONAL TOOLS

In the first part of this dissertation, computer-aided drug design approaches, structure-based methodologies such as docking, molecular dynamics simulations and Gibbs free energy calculations with Molecular Mechanics-Generalized Born/Surface Area (MM-GB/SA) and ligand-based drug design methodologies like pharmacophore modeling are used to propose novel inhibitors for Phosphodiesterase4B (PDE4B) inhibitors.

Virtual screening based on structure-based pharmacophore models has been performed for PDE4B inhibitors. The free energy of binding ($\Delta G_{binding}$) as the total average of 40 independent simulations of each PDE4B inhibitors has been calculated with the MM-GB/SA method. The linear correlation between half maximal inhibitory concentration (IC_{50}) and MM-GB/SA results have been analyzed with the linear dependency between binding affinity (K_i) and IC_{50}, assuming that Michaelis-Menten constant (K_m), substrate concentrations [S], and experimental conditions are similar.

In the second part of this dissertation, the role of important amino acids in GTPase activity of EF-Tu-GTP for different organisms (*Thermos-Aquaticus (T.aquaticus)* and *Escheria Coli (E.coli)* complex have been determined by the aid of molecular dynamics (MD) simulations. The study has been carried out by comparing the experimental results with the results of the MD simulations. The conformational changes during the GTP hydrolysis in the Elongation factor-thermo unstable (EF-Tu) is explained with MD simulations.

ÖZET

PROTEİN-LİGAND BAĞLANMA MODU TAHMİNİ, PDE4B ENZİMİ İÇİN YENİ ADAY İLAÇ YAPILARI VE UZAMA FAKTÖRÜ TU'NUN GTP HİDROLİZİNDEKİ KONFORMASYONEL DEĞİŞİKLİKLERİNİN HESAPSAL YÖNTEMLERLE TESPİTİ

Bu tezin ilk kısmında, PDE4B enzimine yeni inhibitor önerileri için bilgisayar destekli ilaç tasarımı yaklaşımlarından, yapı-bazlı yöntemlerden doklama, moleküler dinamik benzetim ve serbest bağlanma enerjisi hesaplama yöntemlerinden Moleküler Mekanik/Genelleştirilmiş Born Yüzey Alan (MM-GB/SA) tekniği ile ligand-bazlı ilaç tasarımı yöntemlerinden olan farmakofor modelleme teknikleri kullanılmıştır.

PDE4B inhibitörleri için yapısal-bazlı farmakofor modelleri ile sanal eleme yapılmıştır. Her PDE4B inhibitörü için 40 bağımsız moleküler dinamik simülasyonun ortalaması Moleküler Mekanik/Genelleştirilmiş Born Yüzey Alan (MM-GB/SA) tekniği ile Gibbs serbest bağlanma enerjileri ($\Delta G_{binding}$) hesaplanmıştır. IC_{50} değerleri ile MM-GB/SA arasındaki doğrusal korelasyon Michaelis–Menten sabiti (K_m), sübstrat konsantrasyonu ve deney koşullarının aynı olduğu varsayılarak bağlanma isteği (K_i) ile IC_{50} arasındaki doğrusal bağıntıyla analiz edilmiştir.

Bu tezin ikinci kısmında, farklı organizmalardaki (*Thermos-Aquaticus (T.aquaticus)* ve *Escheria Coli (E.coli)* EF-Tu-GTP komplesinin GTPaz aktivitesindeki önemli amino-asitlerin moleküler dinamik simülasyonların yardımıyla tespit edilmiştir. Çalışma, moleküler dinamik benzetim sonuçlarının deneysel sonuçlarla kıyaslanmasıyla yürütülmüştür. Uzama faktörü Tu'nun GTP hidrolizi esnasındaki konformasyonel değişimi moleküler dinamik benzetimleriyle açıklanmıştır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

E_{MM}	Molecular Mechanic Energy
G_L	Free energy of ligand
G_R	Free energy of receptor
G_{RL}	Free energy of receptor-ligand complex system
G_{solv}	Solvation Energy
GB^{HCT}	Solvation model igb = 1
GB^{OBC}	Solvation model igb = 2
$GB^{OBC\text{-}2}$	Solvation model igb = 5
GB^{GBneck}	Solvation model igb = 7
$GB^{GBneck2}$	Solvation model igb = 8
IC_{50}^{est}	Estimated half maximal inhibitory concentration
IC_{50}^{exp}	Experimental half maximal inhibitory concentration
IC_{50}	Half maximal inhibitory concentration
K_i	Binding affinity
K_m	Michaelis-Menten constant
pK_a	Acid dissociation constant
R	Ideal gas constant
R^2	Pearson correlation coefficient
[S]	Substrate concentration
S_{MM}	Entropy term in MM-GB/SA method
$\Delta G_{binding}$	Free energy of binding
Å	Angstrom

LIST OF ACRONYMS/ABBREVIATIONS

ADMET	Absorption, Distribution, Metabolism, Excretion and Toxicity
AMBER	Assisted Model Building with Energy Refinement
AMP	5'-adenosine monophosphate
Asn	asparagine
Asp	aspartic acid
AUC	Area Under the Curve
B3LYP	Becke-3-parameter Lee-Yang-Parr functional
cAMP	Cyclic adenosine 3',5'-monophosphate
cGMP	Cyclic guanosine 3',5'-monophosphate
CHARMM	Chemistry at Harvard Molecular Mechanics
DSX	DrugScoreX
EF-Tu	Elongation Factor Thermo Unstable
ER	Enrichment Factor
FEP	Free Energy Perturbation
GA	Genetic Algorithm
GAFF	General AMBER Force Field
GDP	Guanosine diphosphate
GMP	5'-guanosine monophosphate
Gln	glutamine
GTP	Guanosine triphosphate
Ile	isoleucine
LBVS	Ligand-based Virtual Screening
LCPO	Linear Combinations of Pairwise Overlaps
LR	Linear Response
MAPE	Mean Absolute Percentage Error
MD	Molecular Dynamics
MM-GB/SA	Molecular Mechanics-Generalized Born/Surface Area
MM-PB/SA	Molecular Mechanics Poisson Boltzman Surface Area
NCI	National Cancer Institute
PDE	Phosphodiesterase

Phe	phenylalanine
PME	Particle Mesh Ewald
QSAR	Quatitative Structure Activity Relationship
RESP	Restrained Electrostatic Potential
Rmsd	Root mean square deviation
ROC	Receiver Operating Characteristic Curve
SASA	Solvent Accessible Surface Area
SBVS	Structure-based Virtual Screening
Thr	threonine
TI	Thermodynamic Integration
Tyr	tyrosine
UCR1	Upstream conserved region 1
UCR2	Upstream conserved region 2
VS	Virtual Screening

1. INTRODUCTION

1.1. Phosphodiesterase Enzyme (PDE)

Phosphodiesterase (PDE), is a large family of enzymes, which is responsible for the degradation of cyclic adenosine 3',5'-monophosphate (cAMP) and guanosine 3',5'-monophosphate (cGMP) into 5'-adenosine monophosphate (AMP) and 5'-guanosine monophosphate (GMP), respectively [1-3]. cAMP and cGMP (Figure 1.1) are known as intracellular second messengers that mediate various biological responses to a wide variety of hormones and neurotransmitters [2-4]. The roles of cyclic nucleotides in many metabolic processes are cardiac output, glycogenolysis, platelet aggregation, secretion, lipolysis, learning, ion channel conductance, apoptosis, and growth control [5-7]. The level of cAMP and cGMP in vivo is controlled by synthesis activity of adenylyl and guanylyl cyclases and by hydrolysis activity of PDE's [7]. The higher level of cyclic nucleotides by PDE's inhibition, has considerable effect on the activity of immune and inflammatory diseases [8].

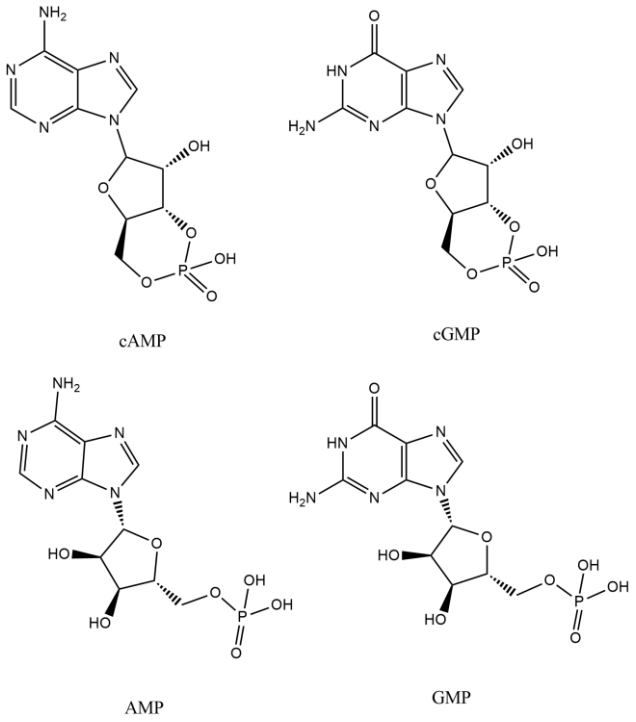


Figure 1.1. Cyclic nucleotides: cAMP and cGMP, and hydrolyzed products: AMP and GMP.

There are 11 families of PDE's (more than 60 mRNA splicing isoforms), which are categorized on the basis of substrate requirements, substrate specificity (cAMP and cGMP), tissue distribution and amino acid sequence (Table 1.1) [1,4,6,7,9-12]. 11 families, have the same general structure of a highly conserved catalytic site of about 300 amino acid with ~ 25-49 % sequence similarity, N-terminal regulatory domain [13].

Table 1.1. Tissue distribution and substrate specificity over 11 families of PDE's [14].

PDE Family	**Tissue Distribution**	**Substrate**
PDE1	Brain, heart, smooth muscle, lung	cGMP > cAMP
PDE2	Adrenal gland, lung, heart, platelets, brain, liver, corpus cavernosum	cGMP = cAMP
PDE3	Heart, liver, lung, platelets, vascular smooth muscle, corpus cavernosum	cAMP>cGMP
PDE4	Lung, mast cells, liver, kidney, brain	cAMP
PDE5	Corpus cavernosum, lung, vascular smooth muscle, platelets, brain, esophagus	cGMP
PDE6	Retina	cGMP > cAMP
PDE7	Skeletal muscle, T-cells, heart, kidney, brain, pancreas	cAMP > cGMP
PDE8	Testes, thyroid, eye, liver, kidney, heart, skeletal muscle, pancreas, T-cells	cAMP
PDE9	Brain, kidney, liver, lung	cGMP
PDE10	Brain, testes	cGMP > cAMP
PDE11	Prostate, skeletal muscle, kidney, liver, testes, pituitary, salivary glands	cAMP = cGMP

In the past three decades, various PDE inhibitors have been used as therapeutic agents like cardiotonic compounds, vasodilators, smooth muscle relaxants, antidepressants, antithrombotic compounds, antiasthma compounds and agents for improving cognitive functions such as memory [15-20].

1.1.1. PDE4 Gene Family and Its Isoforms

Among cAMP-specific isoenzyme of PDE's, PDE4 isozymes are found in many cell types and tissues as leukocytes, airway and vascular smooth muscle, vascular endothelium and brain [21]. Through the control of cAMP level, leukocyte responses as the pro-inflammatory actions of monocytes, T cells and neutrophils, airway and vascular smooth muscle construction, and neurotransmitter signaling are regulated by PDE4 [21].

The PDE4 family with 4 genes (A, B, C and D) comprises more than 20 different PDE4 isoforms which are categorized by an unique N-terminal domain [21-23]. Through chromosomal localization, PDE4A and PDE4C are encoded on different regions of Chromosome 19, PDE4B and PDE4D are found on Chromosome1 and 5, respectively [24].

The PDE4 isoforms are divided into three subgroups: long isoform, short isoform and super-short isoform [22] as shown in Figure 1.3. The long isoform has two regulatory domains: upstream conserved region 1 (UCR1) and upstream conserved region 2 (UCR2) in addition to the catalytic region [21]. The short isoform has only UCR2 and the super-short isoform has a truncated UCR2 [21]. The major role of UCR modules is the regulation of PDE4 by protein kinase A (PKA) and extracellular signal-regulated kinase (ERK) phosphorylation [22,23]. Long isoform is activated by the enzyme with PKA phosphorylation site on UCR1 module. This activation maintains the cellular desensitization machinery for cAMP signaling with the increase level for cAMP hydrolysis [25,26].

1.1.2. General Information of PDE's Crystal Structures

The crystal structure of PDE's in the protein databank (pdb) indicates some general features of the enzymes:

(i) PDE's generally have a compact arrangement of 16 α-helices into three subdomains as seen in Figure 1.2 [27].

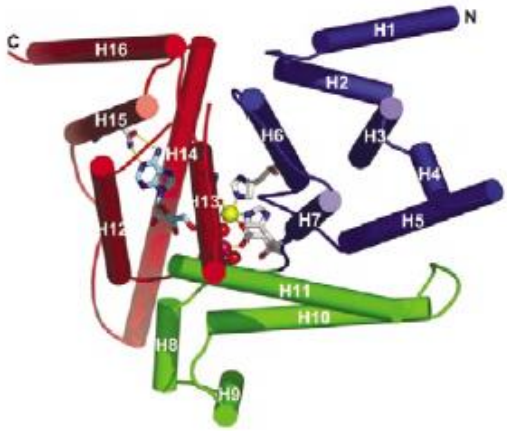


Figure 1.2. Compact arrangement of 16 α-helices in active site of crystal structure PDE4D [21].

(ii) The location of active site is at the junction of three subdomains with highly conserved residues among all the PDEs. The two divalent metal ions, Zn^{2+} and/or Mg^{2+}, which are at the wider part of active site, have six coordination sites with conserved, paired histidine, aspartate residues and two water molecules.

The active site is categorized into three sub-pockets [28]:

- A metal-binding pocket (M pocket): The divalent metal ions, zinc or magnesium, and highly conserved polar residues,
- A solvent-filled pocket (S pocket):hydrophilic amino-acids and a water network,
- Q pocket: purine-selective glutamine and hydrophobic clamp [28].

The PDE4 inhibitor's binding with various interactions in the active site prevents the cAMP metabolism. There is an indirect binding to the divalent metal ions with H-bonding

to the water. The hydrophobic interactions between the planar ring structure of the PDE4 inhibitors and hydrophobic residues like phenylalanine and isoleucine, act as a hydrophobic clamp.

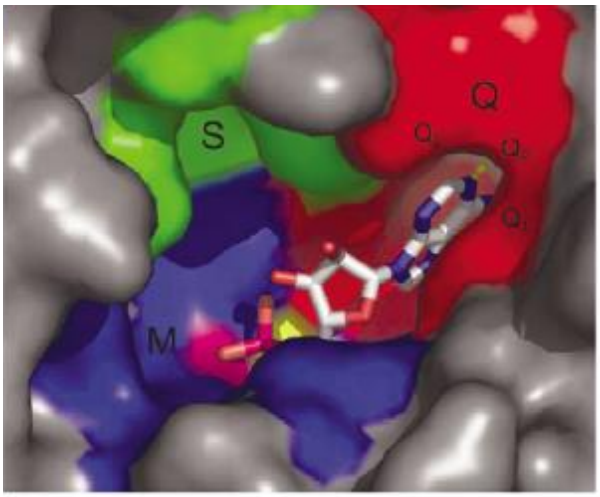


Figure 1.3. The subdomains of catalytic site of PDE4D. AMP was shown as a stick model and the yellow and magenta surfaces are Zn^{2+} and Mg^{2+} divalent metal ions, respectively [21].

For nucleotide selectivity, hydrogen bond interaction between the aromatic ring structure of inhibitors and the invariant glutamine residue in the Q pocket are important [8,28,29].

1.1.3. PDE4 Inhibitor Families

PDE4 inhibitors are proven as potent anti-inflammatory targets especially for the inflammatory and immune disorders like asthma, chronic obstructive pulmonary disease (COPD) and also therapeutic drugs for rheumatoid arthritis, multiple sclerosis, type II diabetes, septic shock, atopic dermatitis, and other immune diseases [18-20,30,31]. The clinical trials of PDE4 inhibitors have been limited due to side effects like vomiting, nausea and high level of gastric acid secretion [32]. Some PDE4 inhibitors are shown in Figure 1.4 [13].

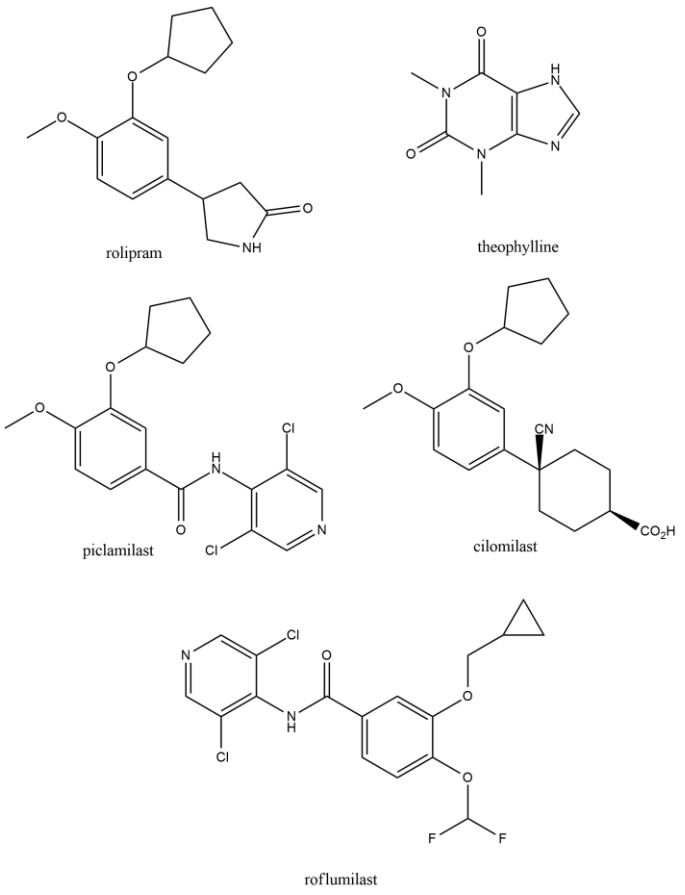


Figure 1.4. Some structures of PDE4 inhibitors [13].

There are many PDE4 inhibitor families which have been developed and some of these inhibitors are under clinical trials (Figure 1.5). Two important categories of these inhibitors:

(i) xanthine derivatives: theophylline, 3-isobutyl-methylxanthine (IBMX), arofylline, doxofylline and cipamfylline,

(ii) dialkoxyphenyl (catechol) family: rolipram, zardaverine, filaminast, mesopram, IC-485, piclamilast, atizoram, tetomilast, CC-1088, ONO-6126, cilomilast and roflumilast.

There are many other compounds which belong to new chemical families. These are AWD-12-28, an indole compound in Phase II trials for asthma; YM-976, a

pyridopyrimidinone derivative in Phase I clinical trials; Tofimilast, an indazole derivative in clinical trial; Ibudilast, a pyrazolopyridine compound in Phase II clinical trials for multiple schlerosis and Lirimilast, a benzofuran derivative under Phase II clinical trial for asthma [21].

In pharmaceutical industry, design of novel isoform-selective inhibitors with limited side effects for PDE4 family is a challenging task due to the high degree of sequence and structural similarity among different isoforms of this family.

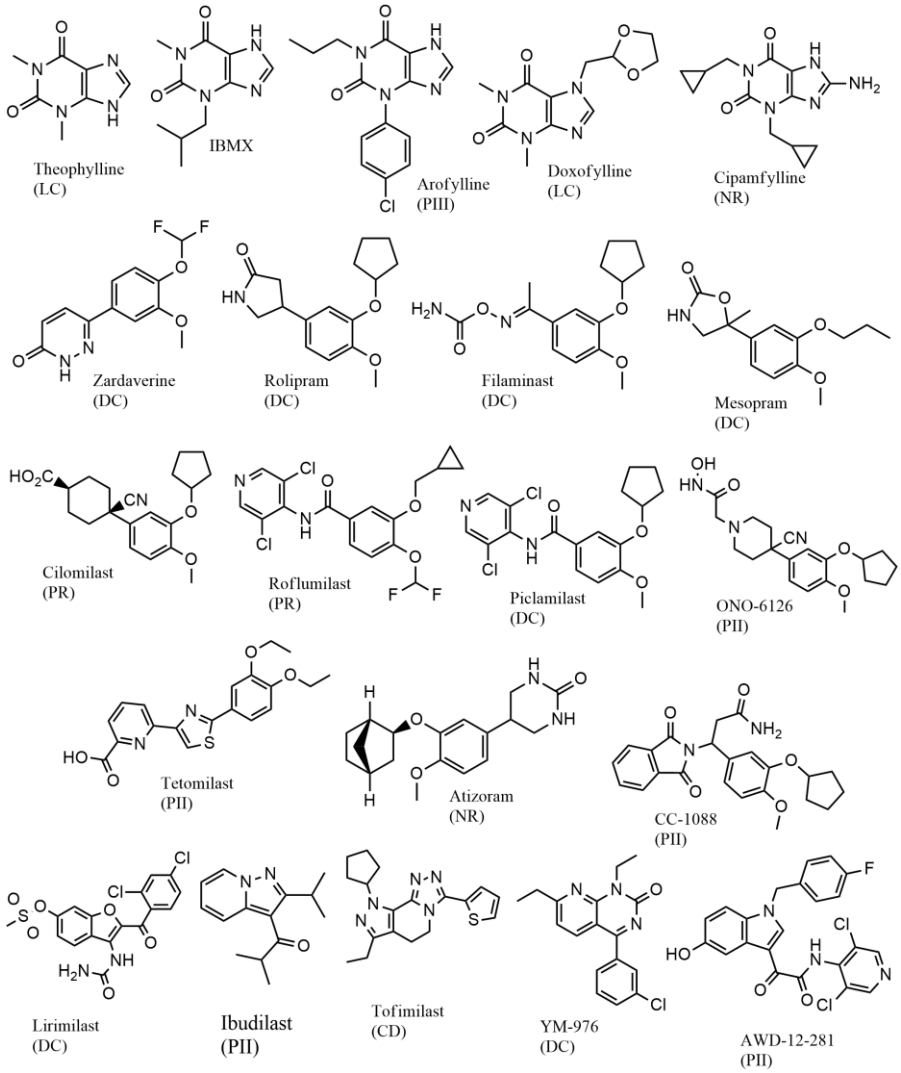


Figure 1.5. Chemical structures of various PDE4 inhibitors of different families. LC, launched; PR,pre-registration; PI, PII, PIII, PhaseI, PhaseII and PhaseIII clinical trial, respectively; CD, clinical development; DC, discountinued; NR, no development [21].

1.2. Elongation Factor Tu (EF-Tu)

1.2.1. EF-Tu Role in Elongation Cycle of Protein Synthesis

Protein synthesis on the ribosome is a combination of subprocesses, such as initiation of translation, protein elongation, termination and ribosome recycling [33,34]. These subprocesses are dependent on the guanosine triphosphate (GTP) hydrolysis which is controlled by the members of translational GTPases [35]. G-protein family members regulate various cellular activities as a carrier of information and biological components [36-39]. One of the common properties of the member of this superfamily is the cycle between an active and an inactive conformation which depend on the nucleotide binding either guanosine triphosphate (GTP: "on" state) (Figure 1.7) or guanosine diphosphate (GDP: "off" state) (Figure 1.8) [40-42].

Elongation factor Tu (EF-Tu), is one of the most important guanosine nucleotide binding proteins [43]. EF-Tu is a guanosine triphosphate (GTP)-bound complex which is in charge of binding of a new aminoacyl-tRNA (aa·tRNA) in the A site of the ribosome. In bacterial cells, EF-Tu·GTP is responsible to deliver aminoacyl-tRNA (aa·tRNA) to the A site of the m-RNA programmed ribosome carrying peptidyl-tRNA in the P-site. GTP-bound state of EF-Tu forms a high-affinity ternary complex with aa·tRNA [40]. The ternary complex, EF-Tu·GTP·aa·tRNA, binds to the ribosomal A site. Correct codon-anticodon interaction triggers GTP hydrolysis on EF-Tu. Major conformational rearrangement of EF-tu results in the release of EF-Tu·GDP, Pi and leaving the aa·tRNA (Figure 1.6) [44].

EF-Tu has a very low GTPase activity (intrinsic rate), the activity rate is increased when the interaction with the ribosome is observed. Ribosomal-sarcin ricin loop and invariant His84 are the important factors for the GTPase activation [40,45-48].

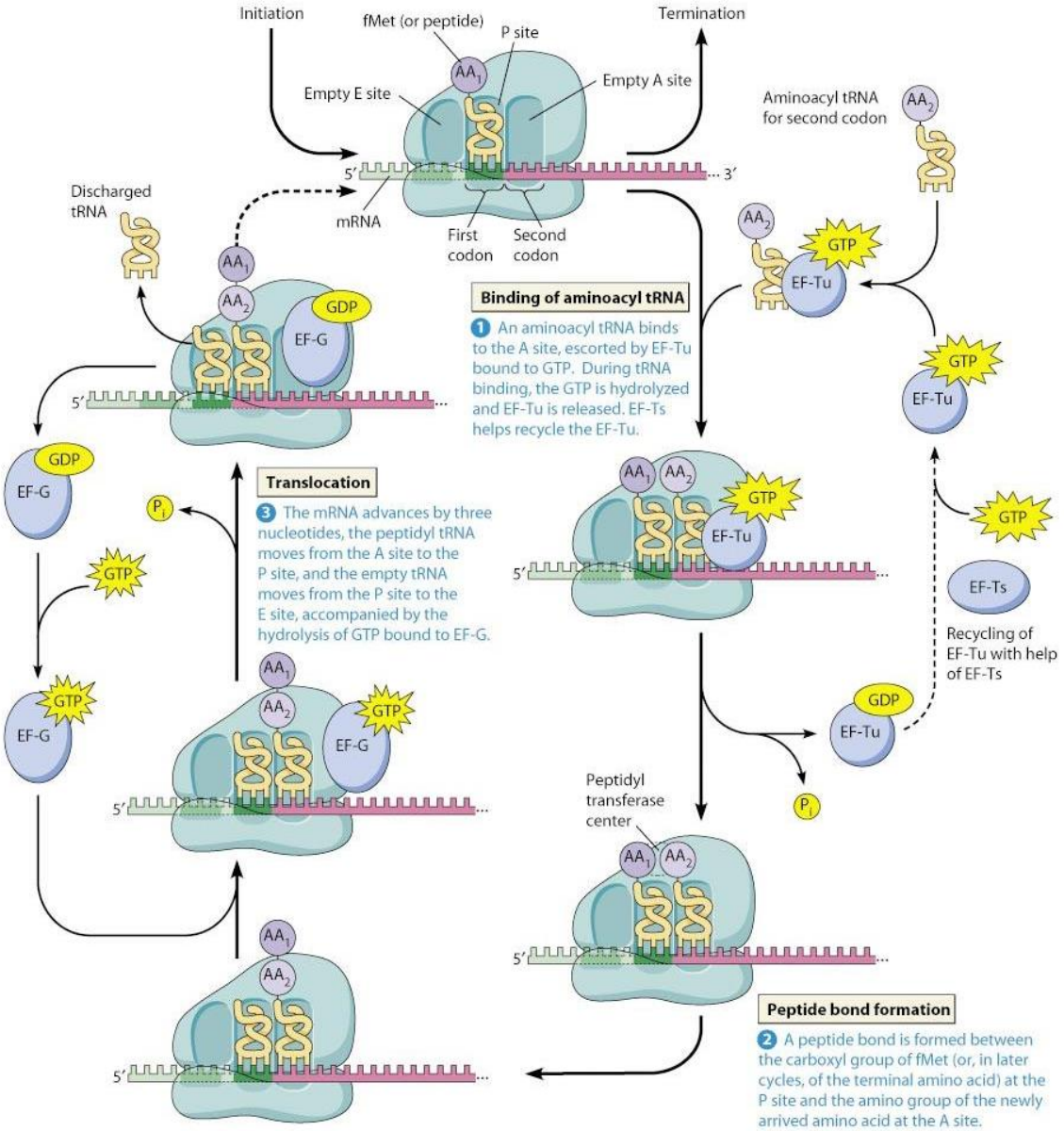


Figure 1.6. Elongation cycle of protein synthesis and the function of EF-Tu [44].

1.2.2. Structural Analysis of EF-Tu

The X-ray structure of bacterial EF-Tu from *Escherichia coli* (*E. coli*) was analyzed by Cour *et al.* and Jurnak in 1985. It is a monomeric protein composed of 400 aminoacids with a molecular mass of 40-52 kDa. The protein consists of three structural domains:

Domain I (Guanine nucleotide binding domain: residues 1-199), Domain II (residues 209-299) and Domain III (C terminal domain: residues 300-393) (Figure 1.7) [41,49].

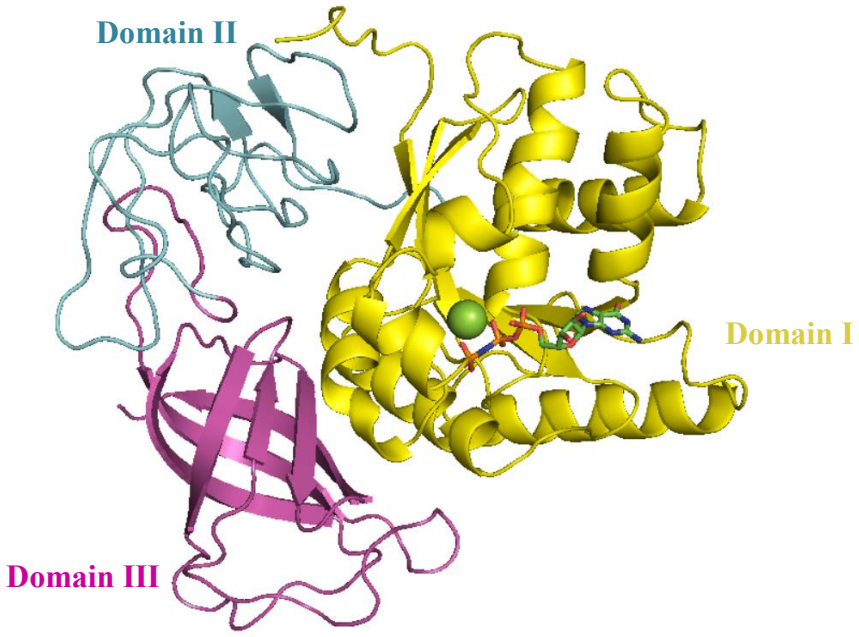


Figure 1.7. Structure of GTP bound EF-Tu (active form), (T. *aquaticus*, PDB code: 1EFT).

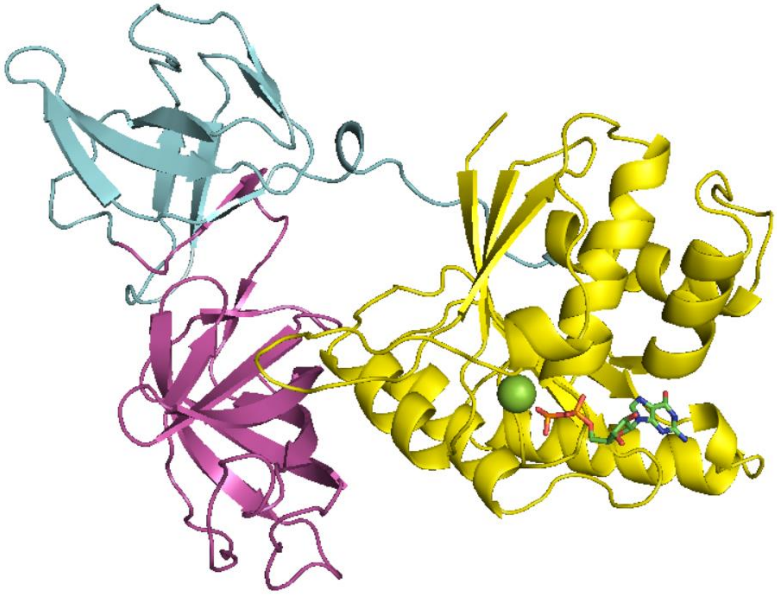


Figure 1.8. Structure of GDP bound EF-Tu (inactive form), (T. *aquaticus*, PDB code: 1TUI).

Domain I, an arrangement of parallel six-stranded β-sheet core surrounded by seven α-helices, includes the nucleotide binding site that binds to GTP or GDP [50]. It has sequence and structural similarity with other GTP/GDP binding proteins. The nucleotide binding pocket Domain I, has three important parts: Switch I, Switch II and P-loop which have been studied in the literature for the conformational changes. During the elongation cycle of EF-Tu in the protein synthesis, Domain II and C terminal domain have a rigid body behavior [51,52].

The change of the active form to inactive form of EF-Tu, EF-Tu·GTP to EF-Tu·GDP, the interaction between the three domains are weakened. In the active form of EF-Tu, closeness of the domains triggers the formation of the ternary complex: EF-Tu·GTP·aa·tRNA (Figure 1.9). Switch I in the GDP-bound EF-Tu forms a β-hairpin that extends into the aa·tRNA binding site. GTP binding to EF-Tu causes a local rearrangement: Switch I conformation turns to an α-helix from β-hairpin. The helix is flanked by D51 and T62. A Mg^{2+} ion, which is coordinated by T62, which in turn, interacts with β- and γ-phosphate groups of GTP. The other coordinations of Mg^{2+} are T25 and two water molecules. The side chain of D51 makes a hydrogen bond with one of these water molecules and another hydrogen bond with T62. Thus, both ends of the Switch I are connected to each other and to Mg^{2+} (via a water molecule in the case of D51). The backbone amide proton of T62 makes a hydrogen bond to the γ-phosphate group in *T. aquaticus* EF-Tu crystals but not in the *E. coli* EF-Tu crystal (PDB code: 1OB2).

1.2.3. Experimental Studies of EF-Tu

In the last fifteen years, there are many publications of both intrinsic and ribosome induced GTPase activity of EF-Tu to GTP and GDP by mutating some amino acid residues, especially in the GTP binding pocket.

The publication of a medium-resolution X-ray crystallographic structure of the ribosome·EF-Tu·aa·tRNA·GTP structure in its active conformation helps to understand the mechanism of ribosome-induced hydrolysis of GTP by EF-Tu.

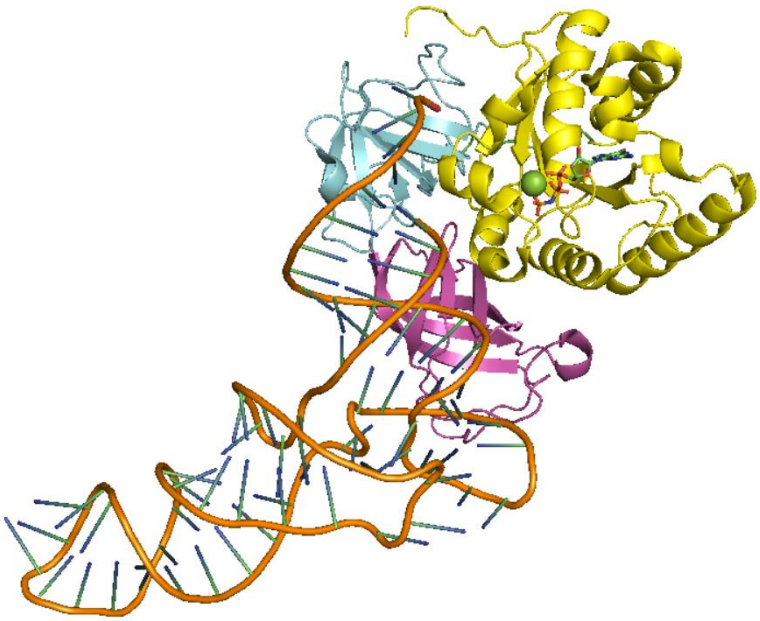


Figure 1.9. Structure of EF-Tu·GTP·aa·tRNA (ternary complex), (E. coli, PDB code: 1OB2).

An invariant histidine for the catalysis of GTP hydrolysis in translational GTPases was proposed by Vorhees *et al.* [53]. There are many alternatives for the explanation of the functions of H84 in the literature. The catalytic effect of H84 is explained by a general base mechanism [53]. The other functions are general acid catalysis [54], positioning of the nucleophilic water [40], positioning of the PGH motif [55] and acting as a conformational switch [56,57]. On the basis of computed pK_a of H85 by Wallin *et al.* and Adamczyk and Warshel [57], it is shown that H85 must be protonated in its active conformation, hence cannot act as a general base.

V20 and I61 were proposed to act as a hydrophobic gate to hinder the entrance of H85 into the active site in the absence of ribosomes. In order to accelerate the GTP hydrolysis in the presence of ribosomes, one or both wings of the hydrophobic gate should open, providing access of the catalytic histidine to the active site. However, neither V20G nor I60A mutations in *E. coli* increased the intrinsic GTPase activity, contradicting the idea of a hydrophobic gate.

2. OBJECTIVE AND SCOPE

In the previous section of the dissertation, a brief introduction about the PDE4 enzyme, its specific inhibitors, the importance of this enzyme subfamily in pharmaceutical industry and the active site orientation of EF-Tu during the GTPase activity are given. More detailed discussion of each one is shown in the following chapters.

In the next chapter (Chapter 3-Theoretical Background), the fundamental principles of the performed computations, including virtual screening based on structure-based pharmacophore models, docking, Molecular Dynamics and free binding energy calculations with Molecular Mechanics-Generalized Born/Surface Area (MM-GB/SA) methods are introduced. Detailed explanations on the methodologies are presented within the relevant chapters.

After this brief overview of the theoretical methods, the results of the studies are introduced in three chapters. Chapter 4 aims to provide novel skeletons for PDE4B subfamily inhibitors using a structure-based drug design protocol. In Chapter 5, with the information obtained by MD and free binding energy calculations by MM-GB/SA, theoretical IC_{50} values were calculated for some PDE4B inhibitors. In Chapter 6, the catalytic site orientation of EF-Tu have been analyzed using MD during the GTP hydrolysis mechanism. The crucial conclusions drawn from each chapter and concluding remarks are presented in Chapter 7.

3. THEORETICAL BACKGROUND

This chapter provides the basic principles of the theoretical approaches including Virtual Screening, Docking, Pharmacophore Modeling, Molecular Dynamics, Free Binding Energy Calculation method: MM-GB/SA. Because of the rich content of each theory, the most commonly used methods or the ones used in this dissertation are presented.

3.1. Virtual Screening

Today, medicinal chemists have serious problems such as the increased cost and enormous amount of time taken to discover a new drug, and competition amongst different drug companies. Therefore it is important to use multidisciplinary approaches which include various computational tools, methodologies for structure guided approach and also global gene expression data analysis by softwares.

In the computer-aided drug design, computational methods have speed up the discovery of a drug by reducing the number of iterations required and have often provided novel structures. For medicinal and computational chemists, Virtual Screening (VS) is a fast and cost-effective computational tool to screen large databases and score, rank or filter the compounds from the database [58].

VS methods can be categorized mainly into two groups: ligand-based (LBVS) and structure-based Virtual Screening (SBVS) as seen in Figure 3.1 [58] . In the LBVS method, structure-activity data of a set of active molecules is used to identify candidates for experimental evaluation. There are three different approaches in LBVS. These are: (1) similarity and substructure searching, (2) quantitative structure activity relationships (QSAR) and finally (3) pharmacophore matching [59]. In the second method of VS, SBVS, a database of candidates are docked into three-dimensional (3D) structure of the target protein (determined either experimentally through X-ray crystallography or NMR or computationally through homology modeling). According to docking results, the docked

molecules are ranked based on their binding affinity and this ranking can be used as a criteria for selection or combined with other evaluation methods for further analysis.

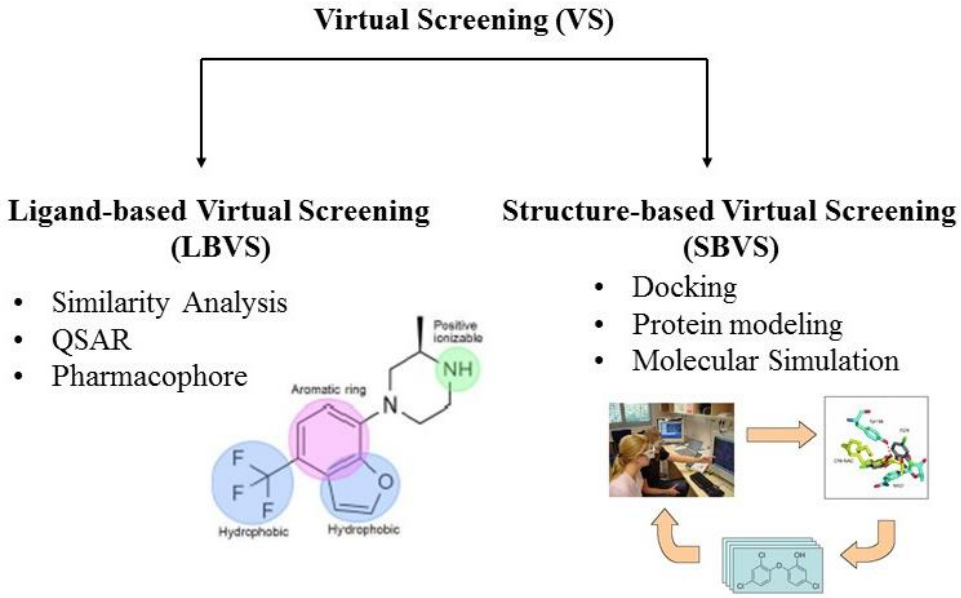


Figure 3.1. A virtual screening methods flowchart.

3.2. Molecular Docking

In the structure-based drug design, molecular docking is one of the most used computational tool for the prediction of the binding mode(s) of a ligand with a target (usually a protein). It can be used for virtual screening on databases of compounds and scoring different conformations in the active site of a receptor and filter out them [58,60]. Basically, molecular docking is based on two steps: prediction of the ligand conformation as well as its position and orientation within these sites (usually referred to as pose) and assessment of the binding affinity [61]. By this way, any docking program contains a conformational search algorithm to sample all degrees of freedom (translational, rotational, and conformational) and a scoring function to rank the poses of the ligand in a library [58].

3.2.1. Search Algorithm

The search algorithm provides the knowledge of positions of molecules in various locations, orientations and conformations in the active site of a target [60].

Conformational search algorithm can be either systematic or stochastic (Table 3.1). In systematic search methods, the energy landscape of the conformational space is analyzed and after evaluation cycle the convergency is achieved with the minimum energy the most likely binding mode. The second method, stochastic method, includes basically Monte Carlo search (MC) and Genetic Algorithms (GAs) [58].

MC methods (e.g. in ICM [62] and GLIDE [63]) change one parameter as position, orientation and conformation) at a time randomly and produce the energy of the new pose [58]. If the energy is lower than the previous one, it is kept. If the energy is not lower, then a selection process is applied to satisfy the probability of a defined pose according to its Boltzman weight factor [58]. The modifications of MC methods as tabu algorithm or simulated annealing algorithm, perform fewer iterations with less run time and have same accuracy as MC [58,60].

The second algorithm, GAs (e.g. in GOLD [64] and AutoDock [65]) is based on Darwin's biological evolution theory. In this algorithm, a ligand conformation with its different properties (dihedral angles and global rotation/translational vectors) are described as a chromosome [58]. The starting point of random search algorithm is this chromosome and a population of chromosomes is generated. The population is evaluated and the fittest parents with the lowest energy values are selected as a template for the new generation [58].

3.2.2. Scoring Functions

After the first step of the docking method, either systematic or stochastic method, many poses are generated for each ligand in a library of compounds.

Table 3.1. A list of various search algorithms for docking.

Systematic Search	***Random/Stochastic Search***
eHiTS [66]	AutoDock [65]
FRED [67]	Gold [64]
Surflex-Dock [68]	PRO_LEADS [69]
DOCK [70]	EADock [71]
GLIDE [63]	ICM [62]
EUDOC [72]	LigandFit [73]
FlexX [74]	Molegro Virtual Docker [75]
Hammerhead [76]	CDocker [77]

The incorrect poses among all poses of the each ligand can be determined in the scoring step of docking [61]. The main goal of a scoring function is to predict the binding affinity between a protein and ligand. The prediction of binding energy depends on physical-chemical filters which are intermolecular interactions, desolvation and entropic effects [78]. So, the accuracy of the scoring function is proportional with the number of physical-chemical filters. On the other hand, as the number of these filters increases, computational cost also increases. So, an ideal scoring function should have a balance between accuracy and speed especially for a large dataset [79].

Scoring functions can be categorized into three groups: (1) force-field based, (2) empirical and (3) knowledge-based functions [80].

Force-field based scoring functions depend on the sum of bonded (bond stretching, angle bending, and dihedral variation) and non-bonded terms of a classical molecular mechanics force field (e.g. AMBER and CHARMM) including electrostatic and van der Waals interactions [58]. A Lennard-Jones potential and the Coulomb energy are used to describe van der Waals interactions and the electrostatic component, respectively [58]. The main drawback of these methods is the lack of entropic component of the binding free energy [58].

The second approach is the empirical scoring function. There are several terms of the function that describe properties in the formation of the ligand-receptor complex. These terms generally define polar interactions like hydrogen-bonding and ionic interactions, apolar interactions such as lipophilic and aromatic interactions, loss of ligand flexibility (entropic effects) and also desolvation effects [58]. In the empirical scoring function, there is a developed model of a series of ligand-receptor complexes with known binding affinities and this training set is used to perform a multiple linear regression analysis [58]. Therefore, the major disadvantage of this approach is the accuracy of the data that is used in the training set [58].

The last approach is the knowledge-based scoring function. The general function is obtained from the pairwise energy potentials which are extracted from known ligand-receptor complexes [81]. These potentials are constructed by taking into account the frequency with which two different atoms are found within a given distance in the structural dataset. The different types of interactions observed in the dataset are categorized and weighted according to their frequency of occurrence. The sum of the individual interactions is the final score. The main advantage of this approach is the balance between accuracy and speed because this function does not depend on reproducing binding affinities as in empirical methods or *ab initio* calculations as in force-field methods [82].

3.2.3. Docking Process

The docking process includes several steps such as protein and ligand preparation, the bounding box setting, docking options, docking calculation and results analysis. A flowchart that shows the docking procedure is shown in Figure 3.2. The detailed procedure about the docking process will be discussed in Chapter 4, methodology part.

3.3. Pharmacophore Modeling and Pharmacophore-based Screening Methods

In computer-aided drug design, pharmacophore model is a set of features that is common to a series of active molecules.

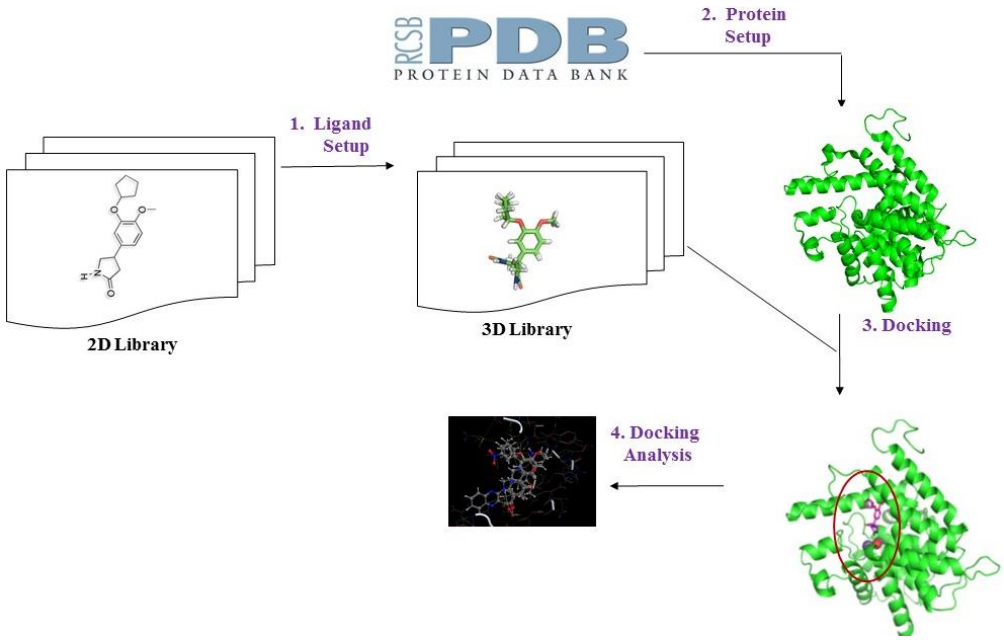


Figure 3.2. A docking based virtual screening flowchart [58].

These features can be hydrogen bond donors, hydrogen bond acceptors, aromatic rings, hydrophobic centers (also called neutral centers), positive charge centers, negative charge centers, acidic groups, basic groups, bulky groups engaged in steric interactions, planar atoms, CO_2 centroid (ester or carboxylic acid), NCN^+ centroid, metal (also called a metal ligator) and excluded volumes [60].

Pharmacophore-based screening is a method of virtual screening to assess millions of compounds by computer softwares. There are several computer programs for pharmacophore modeling such as Ds Catalyst [83], LigandScout [84], Phase [85], and pharmacophore module of MOE [86].

The types of pharmacophore modeling are structure-based and ligand-based methods [58]. In the structure based pharmacophore models, the formation of 3D structure of protein-ligand complex is used to generate the model [58]. On the other hand, in the ligand-based pharmacophore models, there is no crystal structure of protein-ligand complex, therefore the model can be derived from known ligands. In this method, either common features of a

training set of active molecules or a single compound is used as a template for pharmacophore model generation [58].

3.3.1. Pharmacophore Model Validation Methods

In pharmacophore-based screening methods, the goal is to enrich as many biologically active molecules as possible and to eliminate most of the inactive molecules from a compound database. There are many different quality metrics that help to evaluate the quality of the pharmacophore models. Some of the metrics are the chemical databases validation, enrichment metrics, Receiver Operating Characteristic Curve Analysis (ROC) and the area under the ROC curve (ROC-AUC) [58].

In the chemical database validation, it is important to build a proper set of active and inactive molecules in order to evaluate the enrichment of a pharmacophore model [58].

The enrichment metrics validation, two basic values, the sensitivity (Se) and the specificity (Sp) [58] are found. Se is the ratio of the number of selected active molecules and the number of all biological active database molecules [87,88]. Sp is the ability of a pharmacophore model to discard inactive compounds. It is found as the ratio of inactive molecules not selected by the model and all inactive molecules [87].

A pharmacophore-based screening has become a common technique for hit discovery over the past few years. It provides a set of essential chemical features of screened compounds and also new active chemical scaffolds as novel candidates in drug discovery.

3.4. Molecular Dynamics Simulation (MD)

Molecular dynamics (MD) simulations is one of the major tools for modeling proteins, nucleic acids and their complexes which enables us to predict how the positions and velocities of the particles in the system vary with time [89]. It is a model of motion of some group of particles by solving the classical equations of motion. The first molecular dynamic

simulation of a condensed phase system was performed by Alder and Wainwright in 1957 [90]. In 1970s, MD simulations were developed for more complex molecules and today MD provides insight into protein folding, refinement of homology modeling, and predicted structures, complex conformational changes and their relation to function, and computer-aided drug design.

For the MD simulations, first step is to specify a set of initial conditions (initial positions and velocities of all particles in the system) and interaction potential for deriving the forces among all the particles. Secondly, the evolution of the system in time can be continued by solving Newton's equation of motion for all particles in the system.

The MD simulation method which is simply based on the Newton's second law; the equation of motion is:

$$F_i = m_i a_i \tag{3.1}$$

where *F* is the force exerted of an atom in the case of molecular motion. As the gradient of the potential energy function with respect to the internal coordinates gives the force, the force acting on each atom, *i* could be found by the gradient of the forcefield:

$$F_i = -\frac{dU(r)}{dr_i} \tag{3.2}$$

Combination of these two equations gives:

$$-\frac{dU(r)}{dr_i} = m_i \frac{d^2r}{dt^2} \tag{3.3}$$

Potential energy is a function of positions of the atoms in the system. Hence the energy term can not be written as a function of time, this equation becomes very complicated and can not be solved analytically, it must be solved numerically [91].

There are many numerical algortihms such as Verlet Algorithm, Velocity-Verlet Algorithm, Leap-Frog Algorithm etc. to solve equations of motion. In many algorithms integration is partitioned into small steps, each of these steps is separated in specific time period Δt.

In molecular dynamics simulation, Periodic Boundary Conditions (PBC) enables macroscopic properties to be calculated from fewer particles of the systems. It is used to simulate processes in a small part of a large system. In PBCs, the primary cell is replicated in all simulated direction as image cells and primary and image cells have the same number, position, momentum of atoms, size and shape.

In molecular dynamics simulation, thermodynamic properties of the systems, can be obtained by statistical ensembles. Ensembles are formed based on number of particles (N), volume (V), energy (E), temperature (T) and pressure (P). The most popular ensembles are: (i) NVE ensemble (microcanonical ensemble): constant N, V and E, (ii) NVT ensemble (canonical ensemble): constant N, V and T, (iii) NTP ensemble (isothermal-isobaric ensemble): constant N, T and P. Among these ensembles, NVT is the most commonly preferred one for the biological systems due to its computational efficiency.

The other parameters such as the details of equilibration, heating, time step, cutoff for interatomic interactions etc. vary with respect to the systems. More detailed calculation schemes of molecular dynamics simulations in this study are introduced within the relevant chapters.

3.5. Free Binding Energy Calculation Methods

One of the most important goal of computational medicinal chemistry is to develop methods that can accurately estimate the free energy of binding, $\Delta G_{binding}$, which allows us to predict the binding strength of any drug candidate without synthesizing it [92]. This process has been described as the Holy Grail of structure-based drug design [93].

The process is as followed:

$$L + R \rightarrow LR \tag{3.4}$$

where L is the ligand, R is the receptor (the macromolecule target) and LR is the complex. The main objective is to develop methods that accurately can estimate the free energy of this reaction. Among the many methods that try to evaluate this free energy of binding, free energy perturbation (FEP) [94] and thermodynamic integration (TI) are commonly considered as the most rigorous and accurate methods. In these methods, the difference in binding free energy between two ligands is calculated by slowly changing one ligand into another, via a number of unphysical, intermediate states, using molecular dynamics (MD) or Monte Carlo simulations [95]. However, FEP and TI have found relatively little use in drug design, because they are computationally very expensive and they only properly converge for rather similar ligands. More simplified and faster methods have been developed, such as the linear interaction energy (LIE) [96], the Molecular Mechanics-Poisson Boltzmann/Surface Area (MM-PB/SA), and the Molecular Mechanics-Generalized Born/Surface Area (MM-GB/SA) methods [97]. These methods only simulate the end-points of the reaction, i.e., only physical states. They are mainly post-processing methods, i.e. they compute free energies of binding from an ensemble average that is usually obtained from a molecular dynamics (MD) or Monte-Carlo (MC) simulations using standard MM force field [97,98]. In MM/PBSA and MM/GBSA methods, the solvent is treated implicitly via a dielectric continuum. Compared with rigorous methods such as free energy perturbation (FEP) and thermodynamic integration (TI) methods, MM/PBSA and MM/GBSA are computationally more efficient [99].

3.5.1. Molecular Mechanics-Generalized Born/Surface Area (MM-GB/SA)

In MM-GB/SA or MM-PB/SA methods, the free energy of binding ($\Delta G_{\text{binding}}$) between a ligand (L) and a receptor (R) to form a complex RL is calculated as:

$$\Delta G_{binding} = < G_{RL} > - < G_R > - < G_L > \tag{3.5}$$

where each free energy is estimated as a sum, according to:

$$G = E_{MM} + G_{solv} - TS_{MM} \tag{3.6}$$

where E_{MM} is the molecular mechanics gas-phase energy of the reactant, consisting of the internal energy (from bonds, angles, and dihedral angles), as well as the non-bonded electrostatic and van der Waals energies:

$$E_{MM} = E_{int} + Eel + E_{vdW} \tag{3.7}$$

G_{solv} is the solvation energy, and is calculated with a continuum representation of the solvent for the polar part, and by a relation to the solvent-accessible surface area for the non-polar part. The polar part can either be calculated by using the Generalized Born method (MM-GB/SA) or by solving the Poisson-Boltzmann equation (giving MM-PB/SA) [100]. The last term TS_{MM} is the product of the absolute temperature and the entropy, which is calculated from a normal-mode analysis at the molecular-mechanics level [95,101]. The averages in the first equation are calculated from a set of snapshots obtained from MD or MC simulations.

MM-GB/SA and MM-PB/SA methods are promising methods that are now widely used and can give good results. However, they are crucially dependent on numerous factors:

- The quality of the force field used in the statistical simulations [102]
- The quality of the sampling, which should result in a statistically converged free energy of binding [92]
- The quality of the implicit solvation method (GB or PB) [99]
- The effect of the length of MD simulations [99]
- The number of independent MD simulations using different starting velocities [92]
- The suitable solute dielectric constant to calculate the polar solvation energies [99]
- The way by which frames may be selected from the MD simulation including clustering [92].

4. VIRTUAL SCREENING STUDY OF NOVEL SKELETON CANDIDATES FOR PDE4B

4.1. Molecular Docking Study Based on Pharmacophore Modeling for Novel Phosphodiesterase4 (PDE4) Inhibitors

In this study, pharmacophore modeling was carried out for novel Phosphodiesterase4 (PDE4) inhibitors. A pharmacophore-based virtual screening workflow, which resulted in 1959 hit compounds was performed with six chemical databases. Lipinski's rule of five was applied for physicochemical filtering of the hit molecules and this yielded 1840 compounds. Three docking software tools, AutoDock 4.0, AutoDock Vina, and Gold v5.1 were used for the docking process. All 1840 compounds and the known inhibitor rolipram were docked into the active site of the crystal structure of PDE4 taken from the Protein Data Bank (PDB code: 1RO6). A total of 234 compounds with all three scoring values higher than those of rolipram were determined with the three docking tools. The interaction maps of 14 potent inhibitors complexed with PDE4 B and D isoforms have been analyzed and seven key residues (Asn395, Gln443, Tyr233, Ile410, Phe446, Asp392, Thr407) were found to interact with more than 80% of the potent inhibitors. For each one of the 234 hit compounds, using the bound conformation with the highest AutoDock score, the interacting residues were determined. 117 out of 234 compounds are found to interact with at least five of the seven key residues and these were selected for further evaluation. The conformation with the highest AutoDock score for each 117 compounds were rescored using the DSX scoring function. This yielded a total of 101 compounds with better score values than the co-crystallized ligand rolipram. For ADME/TOX calculations, the FAF-Drugs2 server was used and 32 out of 101 compounds were found to be non-toxic.

Cyclic adenosine monophosphate (cAMP) and cyclic guanosine monophosphate (cGMP) are intracellular second messengers mediating the response of cells to a wide variety of hormones and neurotransmitters in signal transduction pathways [5-7]. The families of cyclic nucleotide phosphodiesterase (PDE) enzymes are responsible for the degradation of

cAMP or cGMP to 5’-AMP or 5’-GMP [1-3]. The concentration level of cAMP or cGMP in vivo is important for many pharmacological processes such as proinflammatory mediator production and action, ion channel function, muscle contraction, learning, differentation, apaptosis, lipogenesis, glycogenolysis and gluconeogenesis [5-7].

Up to now, 11 families of 21 human PDE genes were characterized by different substrate specificity (cAMP or cGMP) inhibition, substrate requirements, gene sequence and tissue distribution [1,4,6,7,9-12]. One of these families belongs to PDE4, a cAMP specific enzyme. The PDE4 isoform is a suitable drug target for a range of inflammatory and immune disorders like asthma and chronic obstructive pulmonary disease (COPD) and is also used as a theraupetic agent for rheumotoid arthritis, multiple sclerosis, type II diabetes, septic shock, atopic dermatitis, and other autoimmune diseases [103-108]. There are about 20 potent inhibitors (Figures 4.1- 4.2).

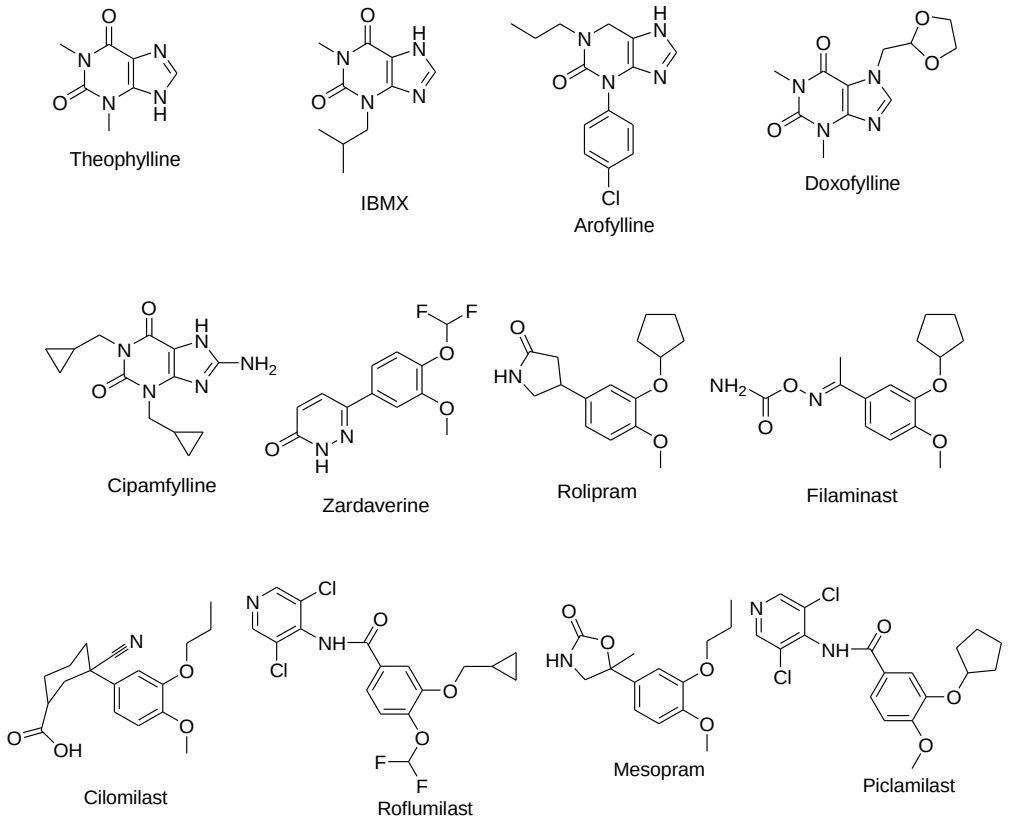


Figure 4.1. Potent inhibitors for PDE4 [13].

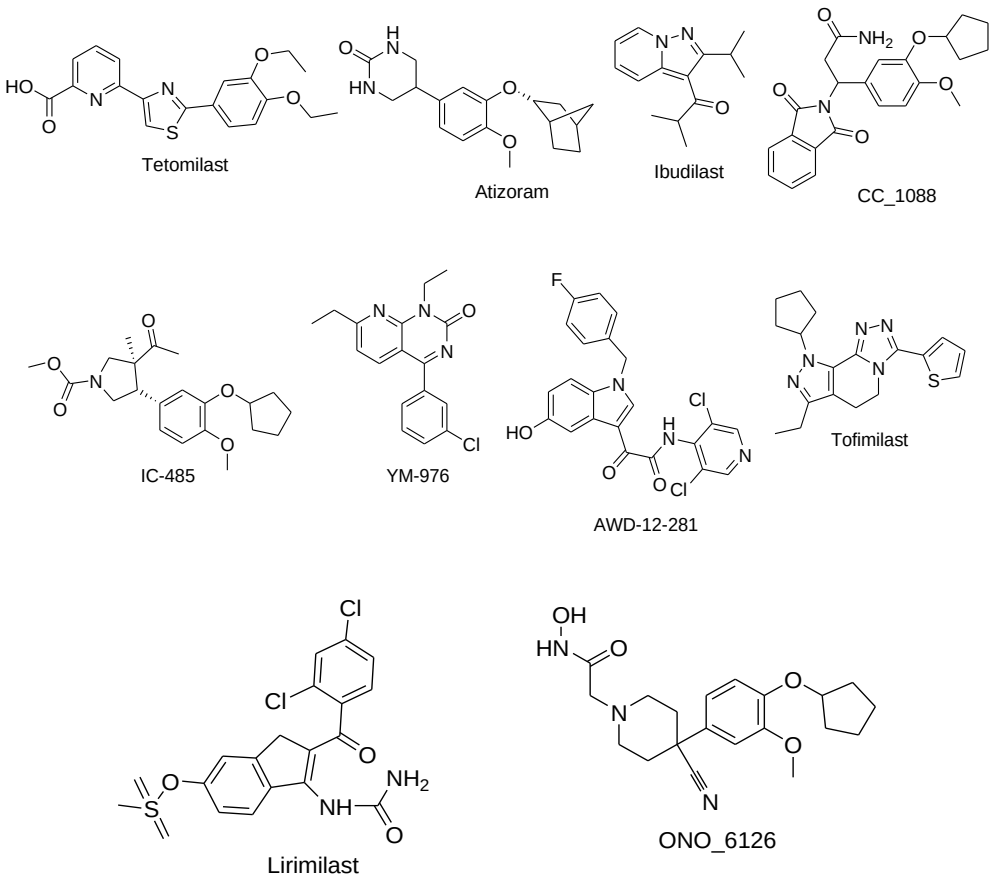


Figure 4.2. Different potent inhibitors for PDE4 [13].

Rolipram is one of the earliest and most extensively studied PDE4 inhibitors. However it is known to be the cause of common side effects, such as nausea, headache and diarrhea [109-111]. The potentially important clinical benefits of PDE4 inhibition, coupled with the limitations of current PDE4 inhibitors, highlight the need for novel PDE4 inhibitor chemotypes [112].

The aim of this study is to suggest novel potent inhibitor structures for PDE4 with limited side-effects. For this purpose, structure-based pharmacophore models for PDE4 inhibitors were prepared and a library of compounds that fit most the pharmacophore model is created. For virtual screening experiments, the docking process was carried out with three docking software tools, Autodock 4.0 [65], AutoDock Vina [113] and Gold v5.1 [64]. The

known active inhibitor, rolipram, was also docked and its score value was set as the lower limit. The compounds that had score values higher than the lower limit in all three docking experiments were selected. Based on the MOE interaction maps, seven residues (Asn395, Gln443, Tyr233, Ile410, Phe446, Asp392, Thr407) interacting with more than 80% of the known potent PDE4 inhibitors displayed in Figure 4.1-4.2, is created. Then, for each selected compound's most favorable binding pose, all the interacting residues are monitored and compared with the residues interacting with the potent inhibitors. The ligands that have interactions with at least five of the seven key residues are reselected and their highest-score conformations obtained from AutoDock are rescored using the DSX scoring function. This yielded a total of 101 compounds with DSX [82] score values higher than that of rolipram. Finally, 101 compounds were filtered based on ADME/TOX calculation on the FAF-Drugs2 [114] server and 32 compounds were found to satisfy the ADME properties and to be non-toxic.

4.1.1. Methodology

4.1.1.1. Pharmacophore Model and Database Generation. In this study, the pharmacophore models were generated by the software LigandScout [84] with its default parameters. LigandScout yielded a total of eight pharmacophore models for four different PDE4 crystal structures taken from the Protein Data Bank (PDB codes: 1RO6, 1Y2J, 1XM4 and 2FM0). Then, pharmacophore-based screenings were carried out using these models against six distinct chemical databases (Chembridge Library, Maybridge Library, Asinex-Gold Library, Asinex-Platinum Library, SPECS database and NCI (National Cancer Institute) with the software Catalyst [115]. The pharmacophore model, which gave the highest number of hits upon screening, was chosen. The compounds with a fit value -number of matching pharmacophore points, the maximum being five- higher than 2.5 were selected for further evaluation for Lipinski's rule of five with MOE [86]. The compounds that satisfy Lipinski's rule of five were subjected to the docking process with AutoDock 4.0, AutoDock Vina and Gold v5.1.

4.1.1.2. Docking. The crystal structure of PDE4 (PDB code:1RO6) taken from the PDB databank was chosen because of its highest resolution and relatively intact structure [116]. The target protein is a homodimer. At the binding site, the water molecule which is found near the ligand and the two metal ions Zn^{2+} and Mg^{2+}, were kept fixed during the docking process. Using MOE, the polar hydrogens were added to the protein based on the protonation states of the ionizable side chains. Using AutoDockTools (or ADT, the GUI of AutoDock), Gasteiger charges were assigned to both the ligand and the macromolecule and non-polar hydrogens were merged to the bonded heavy atom based on the united atom model. For AutoDock 4.0, the docking process for screening was composed of 10 independent runs which yielded 10 docked conformations for each compound in the library. For conformational search, the simulated annealing methodology [117] was used with its default parameters. A total of 150 distinct ligand conformers are initially generated and positioned randomly in the binding pocket. The number of energy evaluations are adjusted with respect to the number of torsions in the ligand based on the general guidelines of AutoDock. The target protein was held rigid during docking. A pre-calculated three-dimensional energy grid of equally spaced discrete points is generated prior to docking, for a rapid energy evaluation, using the program AutoGrid [118]. The grid box with dimensions of 32 Å x 72 Å x 31 Å covers the entire binding site and its neighboring residues. The distance between two grid points is set to 0.375 Å. AutoDock Vina uses the same molecular structure file format used by AutoDock 4.0. A pre-calculated three-dimensional energy grid map and assigning atom charges are not needed for AutoDock Vina. The two major distinctions between two docking tools are the scoring function and the search algorithm.

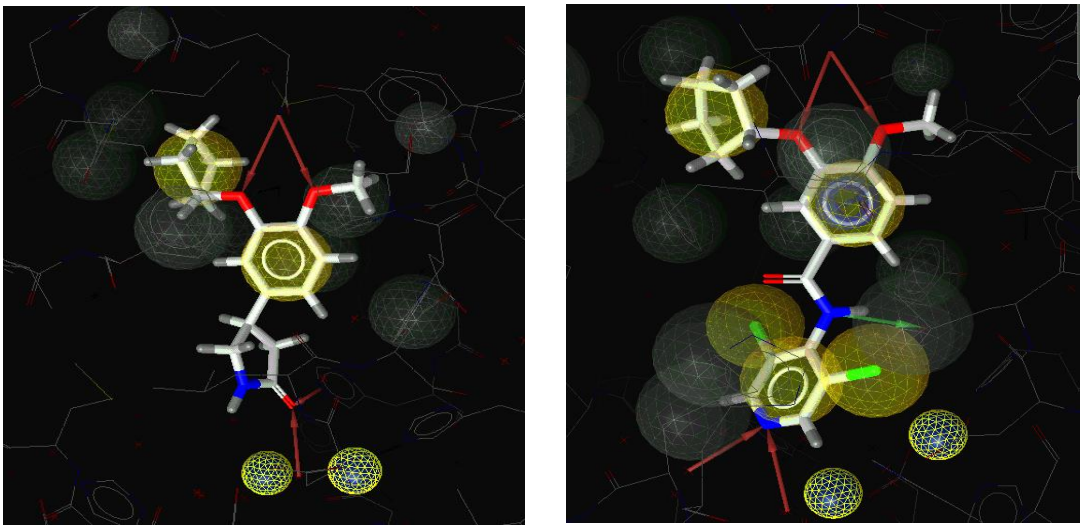
The docking process was repeated with another docking tool, Gold v5.1. In Gold, a genetic algorithm (GA) was used to explore the full range of ligand conformational flexibility and the rotational flexibility of receptor hydrogens [64]. The mechanism for ligand placement is based on fitting points. The CHEMPLP, GoldScore, ChemScore, and ASP scoring functions were used to score the poses and the best poses obtained with CHEMPLP were chosen. For defining the binding site, a spherical region surrounding the ligand with a radius of 12 Å was chosen. All other variables were held fixed at their default values. 50 ga runs were carried out for each ligand.

Following the docking process, a total of 234 compounds with all three scoring values higher than those of rolipram were determined from AutoDock 4.0, AutoDock Vina and Gold v5.1. For each one of the 234 hit compounds, using the bound conformation with the highest AutoDock score, the interacting residues were determined and 117 out of 234 compounds are found to interact with at least five of the seven key residues. All 117 compounds were rescored using the DSX v.088 scoring function. This yielded a total of 101 compounds with better score values than that of the ligand rolipram. As a final step, toxicity calculation was carried out on FAF-Drugs2 server. FAF-Drugs2 (Free ADMET Filtering-Drugs2) is the first free web-based package (http://bioserv.rpbs.univ-paris-diderot.fr/FAF-Drugs/) capable of preparing compound libraries through physicochemical rules, functional groups and Pan Assay Interference Compounds (PAINS) detection [119]. Finally, 32 compounds passed the filtering test for ADME and toxicity.

4.1.2. Results

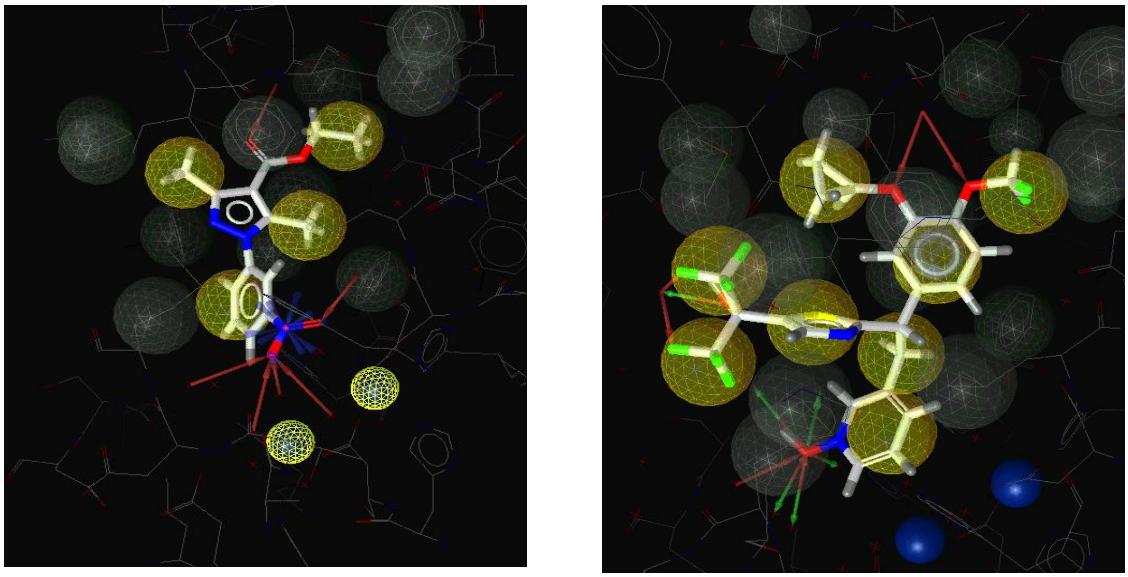
4.1.2.1. Pharmacophore Models. Figures 4.3-4.4. show the pharmacophore points determined by LigandScout for four different crystal structures. In each figure, based on the residues surrounding the ligand and the excluded volume, potential interaction points on the ligand are highlighted as follows: a hydrophobic feature colored as a yellow sphere is the most common feature of all four models. Two other features are represented with red and green arrows representing the hydrogen bond acceptor and donor groups, respectively. Finally, the gray sphere represents the excluded volume. Two metal ions are represented with two small yellow spheres situated at the lower part of the Figures 4.3- 4.4.

In the 1RO6 complex crystal structure (Figure 4.3-a), the co-crystallized ligand is rolipram. The aromatic ring and the five-membered ring at the top portion of the molecule represent the two hydrophobic features. In addition, the two red arrows represent the interactions between two oxygen atoms and Gln443. At the lower portion, the carbonyl oxygen of heterocyclic ring interacts with the water molecule positioned between two metal ions, Zn^{2+} and Mg^{2+}.



a) 1RO6 b) 1XM4

Figure 4.3. Pharmacophore models for a) 1RO6 and b) 1XM4 crystal structures.



a) 1Y2J b) 2FM0

Figure 4.4. Pharmacophore models for a) 1Y2J and b) 2FM0 crystal structures.

In the 1XM4 complex crystal structure (Figure 4.3-b), the ligand is piclamilast. As in 1RO6 complex, the two hydrophobic features and the hydrogen bond between two oxygen

atoms and Gln443 are present. In 1Y2J, 3,5-Dimethyl-1-(3-Nitro-Phenyl)-1H-Pyrazole-4-Carboxylic Acid Ethyl Ester is the ligand (Figure 4.4-a). The hydrophobic aromatic feature is the same as in the other two models, one of them being on the aromatic ring and the others on the methyl group attached to the five-membered ring. Also, as in 1RO6 model, the water molecule positioned between two metal ions, Zn^{2+} and Mg^{2+}, interacts with the oxygen at the lower portion of the molecule. In the last model, 2FM0 (Figure 4.4-b), the ligand is (S)-3-(2-(3-Cyclopropoxy-4-(Difluoromethoxy)Phenyl)-2-(5-(1,1,1,3,3,3-Hexafluoro-2-Hydroxypropan-2-yl)Thiazol-2-yl)Ethyl)Pyridine 1-Oxide (L-269298). As in 1RO6 complex, the two hydrophobic features and the hydrogen bond between two oxygen atoms and Gln443 are present.

In order to display the common pharmacophore features of these four complex models, their ligands were structurally aligned with MOE as shown in Figure 4.5.

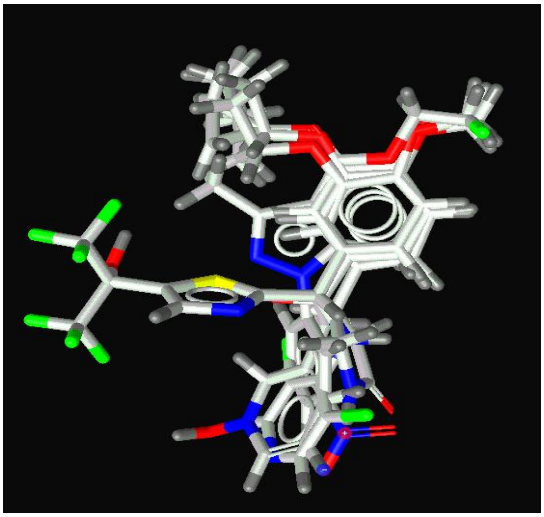


Figure 4.5. Ligand alignment with MOE.

We have also examined six different crystal structures (PDE4) which have the co-crystallized ligand rolipram: two are with the PDE4 B isoform (PDB codes: 1XN0 and 1XMY) and four are with the PDE4 D isoform (PDB codes: 1TBB, 1Q9M, 1OYN, and 3G4K). These six structures and the 1RO6 complex were aligned and superimposed with

respect to the backbone (the root-mean-square deviation (rmsd) of each structure with respect to 1RO6 is found to be less than 1.7 Å) (Figure 4.6). The ligand rolipram from the crystal structure 1RO6 is shown in purple. Only in 1RO6, the five-membered ring in the ligand rolipram, is able to interact with the water molecule (HOH_788) and the two divalent metal ions (Zn^{2+} and Mg^{2+}). Because of its highest resolution and intact structure in the B isoform, the docking process was carried out with 1RO6.

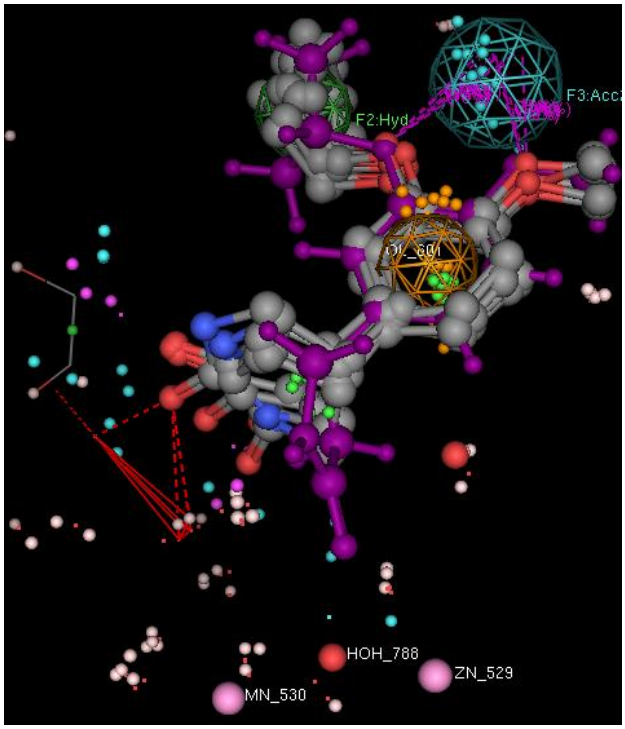


Figure 4.6. The alignment of seven crystal structures co-crystallized with rolipram.

4.1.2.2. Virtual Screening via Docking. As our reference molecule, rolipram was first docked to the active site of the enzyme. The results are sorted based on the binding energy, which is the score value of AutoDock (Table 4.1). The conformation with the lowest binding energy (or the highest score value) has an rmsd value of 0.81 Å with respect to the known conformation in the crystal structure. Also, 8 out of 10 predicted poses have an rmsd value under 1 Å, with binding energies between -8.17 and -8.00 kcal/mol. Only one out of 10 poses has failed to predict the native state and it has the lowest score value (or highest energy) and the highest rmsd value. The strong correlation between the score value and the rmsd value

proves AutoDock 4.0 to be a satisfactory tool for molecular recognition studies of the PDE4 enzyme.

Table 4.1. Docking results for ligand rolipram with AutoDock 4.0.

Rank (Sub-Rank)	***Run***	***Binding Energy (kcal/mol)***	***Reference RMSD (Å)***
1 (1)	2	-8.17	0.81
1 (2)	5	-8.16	0.79
1 (3)	10	-8.16	0.79
1 (4)	1	-8.16	0.78
1 (5)	9	-8.16	0.82
1 (6)	8	-8.16	0.80
1 (7)	4	-8.16	0.78
1 (8)	3	-8.00	0.86
1 (9)	6	-7.27	1.53
2 (1)	7	-6.13	18.85

The docking process was also carried out for the potent inhibitors displayed in Figure 4.1- 4.2. Seven of these potent inhibitors are the co-crystallized ligands for different PDE4 complex structures extracted from the Protein Databank. Five of them correspond to PDE4 B isoform (PDB codes: 1XM4 with piclamilast, 1XLZ with filaminast, 1XLX with cilomilast, 1XM6 with mesopram and 1XMU with roflumilast) and two of them belong to PDE4 D isoform (PDB codes: 1MKD with zardaverine and 1ZKN with IBMX). These potent inhibitors were individually docked into their complex crystals with AutoDock 4.0, AutoDock Vina and Gold v5.1. Furthermore, they were also docked into the 1RO6 active site. In Table 4.2, the second column represents the binding energy values with AutoDock when the ligands were docked into the 1RO6 active site and the last column shows the corresponding values when the ligands were docked into their own co-crystallized structures. As seen in Figure 4.7, AutoDock (1RO6), AutoDock (co-crystallized) and AutoDock Vina

yield the same trend in binding energies; e.g. cilomilast and IBMX have the highest and lowest binding energies respectively. This trend is also justified with Gold v5.1 demonstrating the validity of the methodologies used in this study.

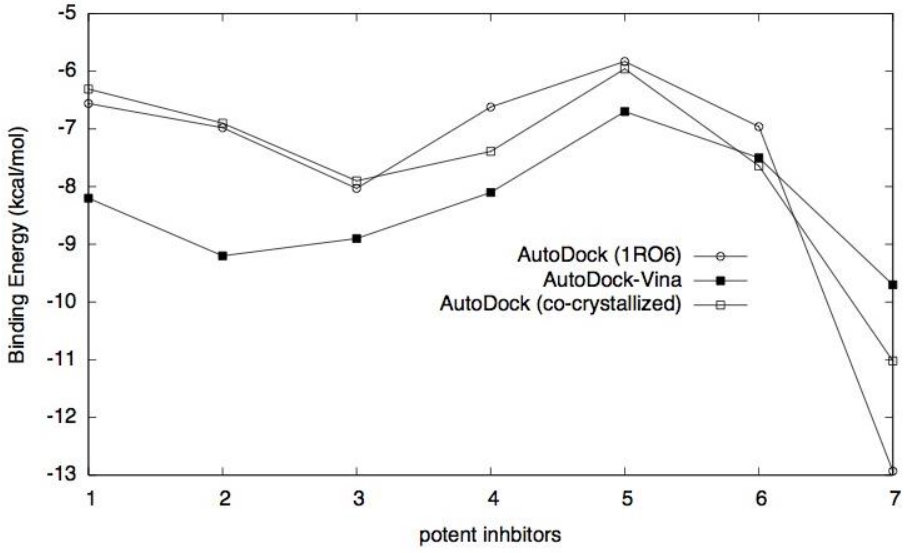


Figure 4.7. The correlation between the AutoDock scores docked into 1RO6 and the co-crystallized complexes.

Table 4.2. Docking results of potent ligands with AutoDock 4.0, AutoDock Vina and Gold v5.1.

Potent inhibitors	**Binding Energy AutoDock 4.0[a]**	**Binding Affinity AutoDock Vina[a]**	**Gold CHEMPLP Score[a]**	**Binding Energy AutoDock 4.0[b]**
1.Zardaverine (1MKD)	-6.56	-8.2	64.7323	-6.31
2.Roflumilast(1XMU)	-6.98	-9.2	70.6268	-6.90
3.Piclamilast(1XM4)	-8.03	-8.9	73.1025	-7.90
4.Mesopram(1XM6)	-6.62	-8.1	62.8033	-7.39
5.IBMX(1ZKN)	-5.83	-6.7	55.278	-5.96
6.Filaminast(1XLZ)	-6.96	-7.5	76.1991	-7.64
7.Cilomilast(1XLX)	-12.93	-9.7	92.3694	-11.02

a: Target is the PDEIV in 1RO6. b: Target is the PDEIV in their own co-crystallized complexes.

Following rolipram, 1840 candidate molecules that satisfy the criteria described in Methodology, were put to a virtual screening via AutoDock v 4.0, using the same docking parameters used for rolipram. For each ligand, the lowest binding energy (highest score) value was extracted. A total of 635 molecules were found to have a binding energy lower than that of rolipram.

Similarly, all 1840 molecules were subjected to a docking process via AutoDock Vina. It was proposed that Vina significantly improves the average accuracy of the binding mode predictions compared to AutoDock v4.0, judging by the tests on the training set used in AutoDock v4.0 development. Similar to AutoDock v4.0, rolipram was first docked with Vina, to check the performance in prediction. Vina was several orders of magnitude faster than AutoDock v4.0 and the conformation with the lowest binding energy (or the highest score) had an rmsd value of 0.73 Å, which is slightly closer to the native state (0.81 Å in AutoDock v4.0). As a result of docking 1840 compounds with AutoDock Vina, 489 compounds that had better binding affinities than that of rolipram were found.

Finally, the docking process was carried out by using Gold v5.1. Similar to AutoDock v4.0 and Vina, rolipram was first docked with different scoring functions: CHEMPLP, GoldScore, ChemScore and ASP. Among the scoring functions, CHEMPLP gave the best score. Thus, we decided to use Gold v5.1 with CHEMPLP scoring function to dock all 1840 compounds in the training set into the active site of 1RO6. 681 compounds that had better CHEMPLP score values than that of rolipram, were selected as a result of this docking process.

The next step consists of combining the results from AutoDock v4.0, AutoDock Vina and Gold v5.1. A total of 234 compounds were found to have stronger binding energies (higher scores) than rolipram in all three docking experiments, thus were selected for further analysis.

4.1.2.3. Protein-Ligand Interaction Maps. In this part of the study, two-dimensional protein-ligand interaction maps for known potent inhibitors (Figures 4.1- 4.2) and for the 234 hit

molecules were created with MOE. First, the ligand interaction maps for the known potent inhibitors were analyzed and five key amino acid residues (Asn395, Gln443, Tyr233, Ile410, Phe446) were observed to interact with more than 90% of the 14 known potent inhibitors (Figure 4.8). In addition to five key residues, in almost all potent molecules, Asp275 and Asp392 were observed in the active site to coordinate with Mg^{2+}, Zn^{2+} and water, as seen in Figure 4.10.

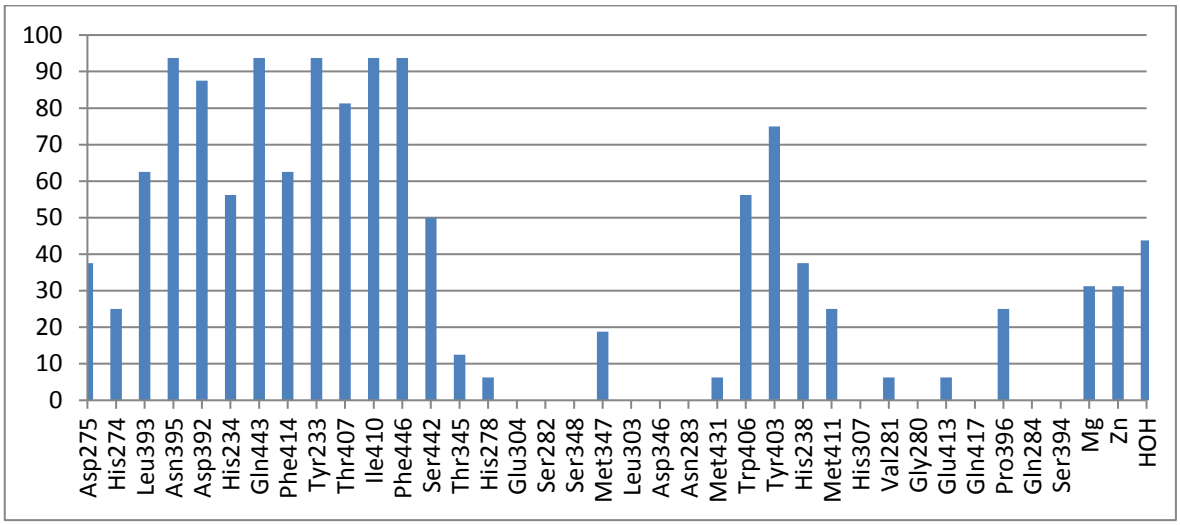


Figure 4.8. Percentage values of amino acid residues in the active site of the enzyme for known potent inhibitors.

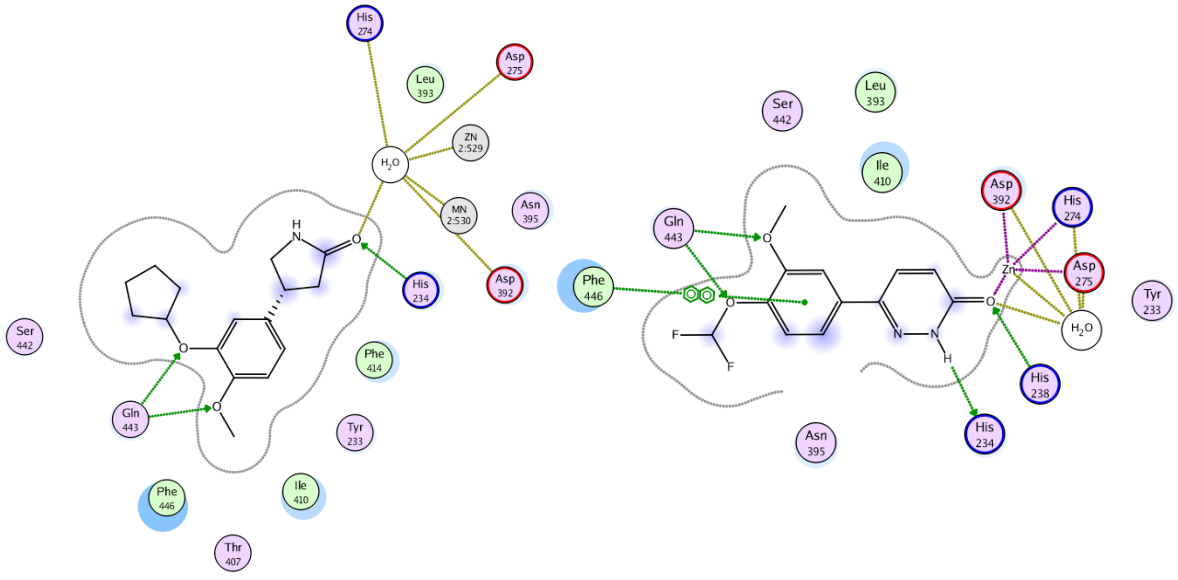


Figure 4.9. Ligand interaction maps for some potent inhibitors created with MOE.

For the 234 hit molecules, similar protein-ligand interaction maps were prepared and the percentage of occurrences of residues in the active site of the enzyme were calculated (Figure 4.10). As seen in Figure 4.9, a total of five residues interact with more than 80% of the hit compounds and they are Asp392, His234, Tyr233, Ile410, and Phe446. Three of these five residues which are Tyr233, Ile410, and Phe446 also interact with more than 90% of the known potent inhibitors as shown in Figure 4.8.

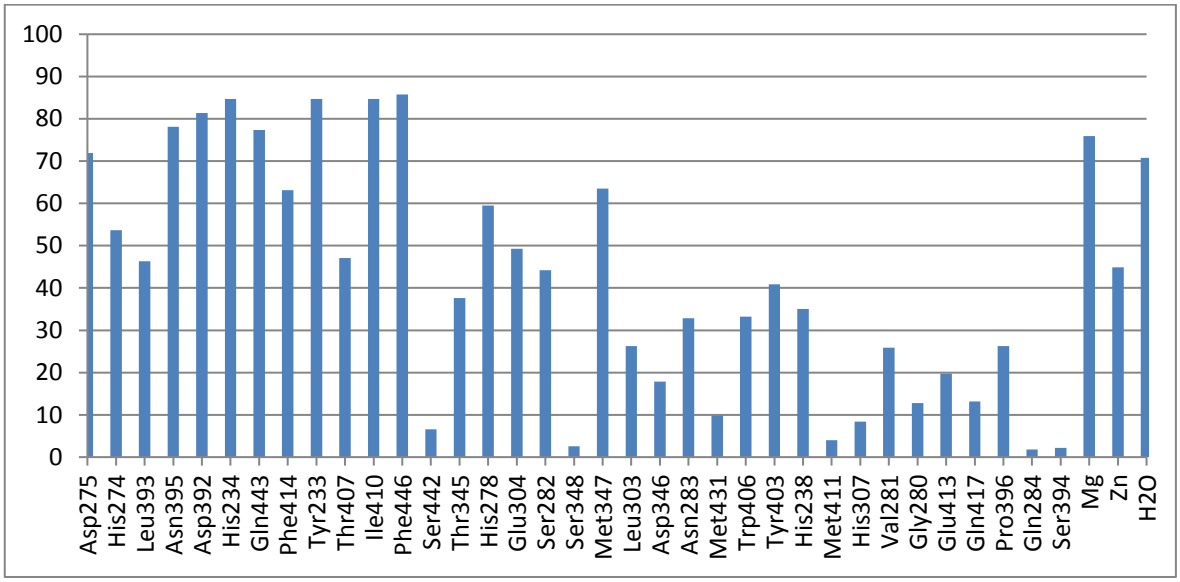


Figure 4.10. Percentage values of interactions observed in the active site of the enzyme with docked ligands.

Furthermore, two residues which are Asp392 and Thr407 are observed to interact with more than 80%, but less than 90% of the known potent inhibitors. Hence, they are selected as key residues in the binding site in addition to five key residues for the potent inhibitors mentioned above. Consequently, a total of seven selected key residues were inspected in the interaction maps of the hit molecules for which the bound conformation with the highest AutoDock score was used, Among the 234 hit compounds, 117 molecules were found to interact with at least five of the seven key residues and thus were selected for the next round of analysis where they were subjected to a knowledge-based scoring function called DSX.

For each of the 117 compounds, the conformation with the highest score obtained previously from AutoDock v4.0 was selected and its score value is determined using DSX. Rolipram's known conformation in the crystal structure was also rescored with DSX and taken as the lower limit. A total of 101 compounds that had higher DSX score values than the lower limit were selected for further tests. Analysis of the structure of the hit compounds shows that the trifluoromethyl (-CF_3) and the 1,3-dimethoxybenzene (-$C_6H_4(OCH_3)_2$) groups are the common functional groups shared by most of the compounds. The trifluoromethyl group (-CF_3) interacts with two key residues Tyr233, Asp392 and the 1,3-dimethylbenzene interacts with another key residue Gln443 within a proximity of about 4 Å. These residues are previously shown to be present in both known potent inhibitors and hit compounds with high percentages.

As the final step of screening, 101 compounds were subjected to ADME/toxicity filtering on FAF-Drugs2 server and 32 of them which passed the test were proposed as candidates for the novel PDE4 inhibitors as shown in Figure 4.11. The flowchart that illustrates each step of the screening used in this study is displayed as in Figure 4.12.

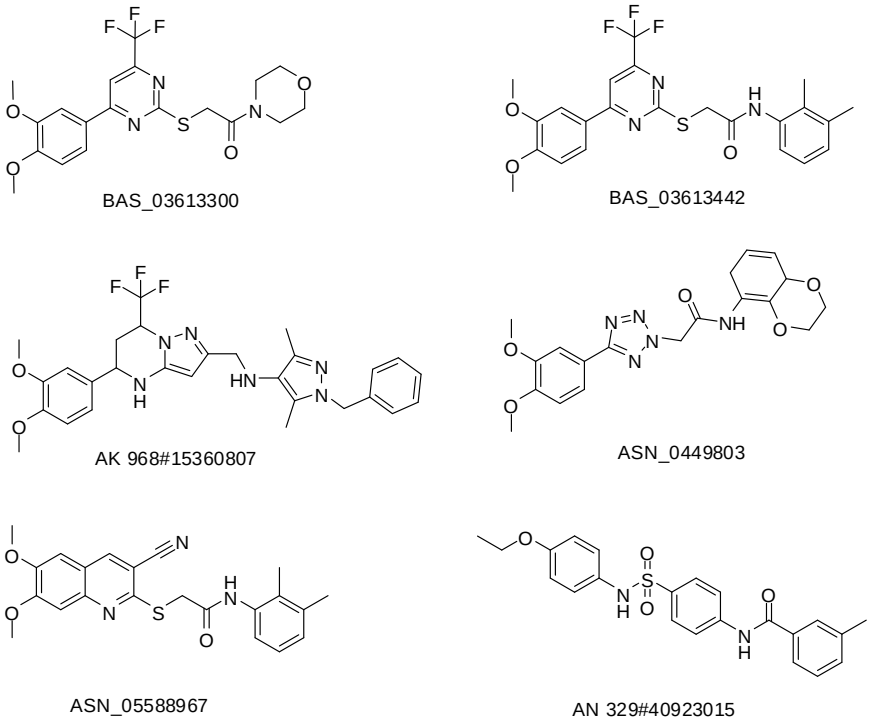


Figure 4.11. Candidate inhibitor molecules which have the highest pharmacophore fit values (greater than 3.8).

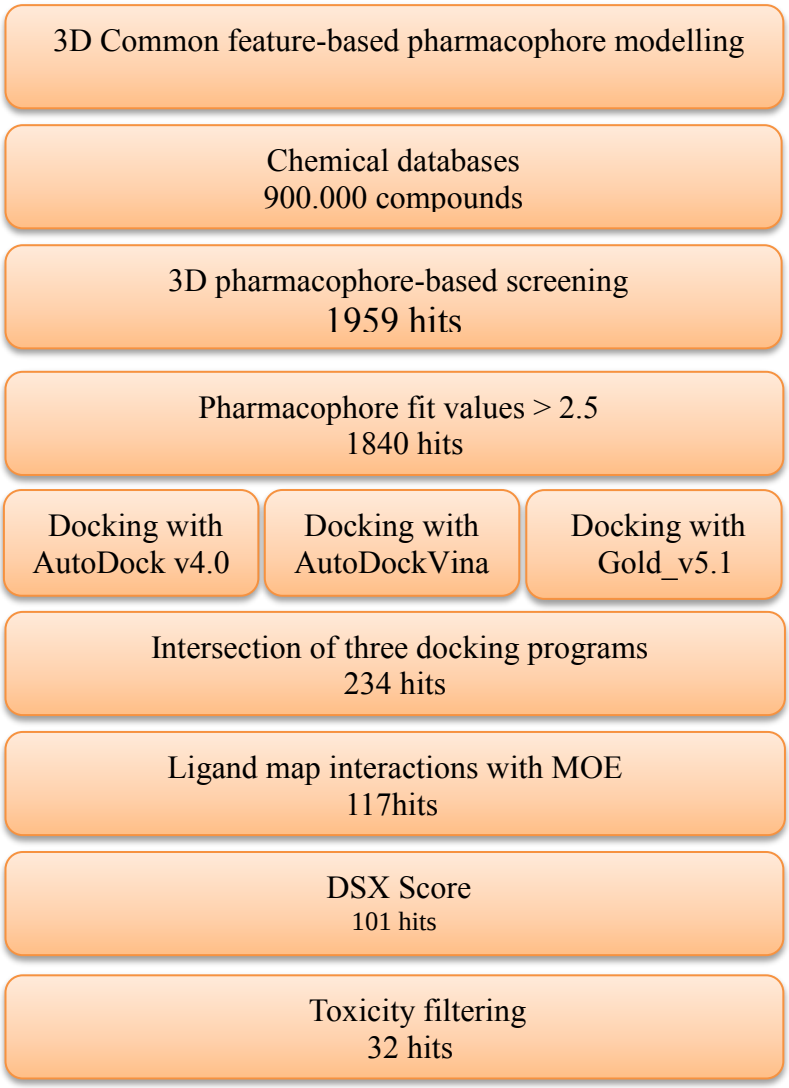


Figure 4.12. A flowchart depicting the procedure followed in this study.

4.1.3. Conclusion

In this study, pharmacophore models were derived from four different protein-ligand crystal structures and they were employed to screen Chembridge Library, Maybridge Library, Asinex-Gold Library, Asinex-Platinum Library, SPECS database and NCI for the detection of PDE4 inhibitors. After the docking process with AutoDock v4.0, AutoDock

Vina, and Gold v5.1, 234 inhibitors with score values higher than those of rolipram in all three docking experiments were determined as initial hit compounds. The ligand interaction maps for the 14 known potent PDE4 inhibitors reveal seven key residues: Asn395, Gln443, Tyr233, Ile410, Phe446, Asp392 and Thr407 which interact with more than 80% of the potent inhibitors. The analysis of 2D protein-ligand interaction maps revealed that 117 out of 234 compounds interact with at least five of the seven key residues. Therefore, they are selected for further analysis which consists of rescoring with DSX scoring function. Rescoring yielded 101 compounds with DrugScore values less than -95.56 which belongs to rolipram. All 101 compounds were subjected to ADME/toxicity and only 32 of them which passed the filtering were proposed as the most promising compounds for novel PDE4 inhibitors.

4.2. An improvement of the Pharmacophore Model and Docking with Gold v5.1 Scoring Functions for Novel Phosphodiesterase4 Inhibitors

In this study, virtual screening based on pharmacophore models and molecular docking for Phosphodiesterase4B (PDE4B) inhibitors were carried out to propose novel inhibitors for PDE4B family. The pharmacophore model from the previous work was refined based on conformational change of known PDE4B inhibitor rolipram in the crystal structure (PDB code: 1RO6) with software MOE. The database that was screened against six chemical databases from previous work, was resulted in 1840 compounds, combined with twenty-six known potent PDE4B co-crsytallized inhibitors from protein databank. A conformational search based on stochastic method for this combined dataset was proceeded with MOE. Virtual screening based on first refined pharmacophore model yielded 12114 and for the second one, 14740 conformers, were merged in a new dataset containing totally 26854 conformers for docking and binding test. For each conformers of each hit compound, docking process was carried out with Gold v5.1 with four different scoring functions: ASP, ChemScore, ChemPLP, and GoldScore. Two dimensional ligand interaction maps for the twenty-six known PDE4B inhibitors were analyzed with MOE. According to interaction maps, seven key residues (Tyr233, His234, Met347, Ile410, Phe414, Gln443, Phe446) in the active sites within 5 Å, with occurrence more than 85% of known potent inhibitors were

found. For the docked conformers of each compound for each scoring functions were tested if they had interaction with these key residues or not. The conformers of each scoring functions were kept as hits if they have passed the binding test. After the binding test, enrichment factors for each scoring functions were determined by the threshold values and the conformers that have had better scores than thresholds in all of four scoring functions were selected. For taking only one conformer for each different compound, the best conformer with the best docking score were chosen. The final step is to classify these 263 hits based on Tanimato coefficient similarity with MOE. 263 hits were clustered with more than 60% similarity and this yielded 84 different classes. Among these classes, 4 hit compounds were found that had better interaction residue energies than known inhibitor rolipram.

The goal of this study is to propose novel potent inhibitor structures for PDE4B family with limited side effects. The study started with virtual screening based on structure-based pharmacophore models. The docking process was carried out with four different scoring functions; ASP, ChemScore, CHEMPLP and GoldScore of Gold v5.1. Based on two-dimensional ligand map interactions, created by MOE, for known inhibitors from pdb, seven key residues that give interaction with more than 85 % of the known inhibitors, were determined and a binding test was applied for each conformation of each docked compound. The docked poses that didn't have interaction with these key residues were eliminated. For filtering based on docking, enrichment factors were calculated with the threshold values for each scoring function and the compounds that were found above the thresholds in all four scoring functions were chosen. At the final step, clustering based on Tanimato coefficient similarity was proceeded with MOE and 4 hit compounds that have better interaction energy values with the residues in the active site than known inhibitor rolipram are proposed as novel scaffolds for PDE4B selective inhibitors.

4.2.1. Methodology

4.2.1.1. Structure-Based Pharmacophore Modeling and Database Generation. This study starts with the pharmacophore model (1RO6 model) which was created with software

LigandScout and validated its discriminative power between active and inactive inhibitors for a decoy set with a ROC curve generated with the same software as in our previous work. The database that was obtained after screening against six chemical databases (Chembridge Library, Maybridge Library, Asinex-Gold Library, Asinex-Platinum Library, SPECS database and NCI) with Catalyst with its default parameter is the starting database in this study.

In the 1RO6 crystal structure, there are two conformations of ligand rolipram in the same chain, chain A (Figure 4.14) In the previous work, only the conformation 1 of ligand rolipram was discussed, but in this study, the pharmacophore model was refined based on the conformational change of rolipram with software MOE. A new database, which contains 1840 hits from the previous work and 26 co-crystallized known inhibitors for PDE4 from pdb, were merged and a conformational search was proceeded with MOE. For the conformational search, stochastic method was chosen. The iteration limit was set to 100, MM iteration limit was 200, RMS Gradient was 0.05, RMSD Limit was 0.5 and the conformational limit was 10. The other parameters were kept as default. The merged database was screened based on both first and second model then the hit compounds were combined in a new dataset.

4.2.1.2. Docking and Binding Test. For the docking process, 1RO6 crystal structure was chosen because of its highest resolution (2 Å), intact structure and also co-crystallized structure with known ligand rolipram. The target protein is a homodimer and contains two divalent metal ions Zn^{2+} and Mg^{2+}. During the docking process, a water molecule which is between two metal ions in the active site was kept. For the protonation, polar hydrogens were added with MOE.

Docking was carried out using Gold v5.1 with four different scoring functions: ASP, CHEMPLP, ChemScore and GoldScore. In Gold v5.1, the genetic algorithm (GA) parameters were used to explore the full range of ligand conformational flexibity and the rotational flexibility of receptor hydrogens. For the ligand placement, fitting points option was chosen and the binding site was defined as a spherical region surrounding the ligand

with a radius of 12 Å. For the other parameters (ga run: 10 and ga search: slow) were held fixed as default. For each conformer, 10 ga runs were performed.

For the binding test based on the 26 co-crystallized ligands native binding mode (from Protein Databank), the key residues in the active site within 5 Å of the each crystal structures for each ligand were determined with MOE. Two dimensional protein-ligand interaction maps were prepared with MOE again and the percentage of occurrences of residues in the active site of the enzyme were calculated.

According to the binding test, the docked pose of each conformer of each compound has to be interact with the key residues otherwise it was eliminated from the dataset.

For further elimination, the threshold values for each scoring function were determined with enrichment factor method.

All the selected molecules were simply combined and a molecule was counted as a hit if it was found in above of all the four threshold for docking. After elimination based on threshold values, only the best conformers with best docking scores for each compounds were chosen.

As a final step, a clustering based on Tanimato coefficient similarity was carried out with MOE, 84 families which belong to different compounds were obtained. According to two-dimensional interaction maps, the interaction energies between ligand and the residues within 5Å in the active site were calculated and four representative scaffolds which have better interaction energies than the known ligand rolipram, were proposed as novel structures for PDE4B inhibitors.

4.2.2. Results

4.2.2.1. Structure-based Pharmacophore Modeling and Conformational Change. Figure 4.13 represents the pharmacophore model for PDE4B selective inhibitor rolipram which was

created in our previous work with LigandScout. In this model, important pharmacophoric features were represented with yellow spheres for hydrophobic property, red arrows for hydrogen bond acceptors and the green ones for hydrogen bond donors. At the bottom part of active site, in the M pocket, two metal ions were shown as two small yellow spheres.

In the 1RO6 crystal structure, the known inhibitor rolipram, has two conformations in the same chain, chain A as shown in Figure 4.14. In the first conformation of rolipram, the oxygen atoms give interaction with Gln443 residue at the top portion of the active site and at the M sub-pocket, the five-membered ring has an orientation which provides an interaction between carbonyl oxygen atom on the ring and water molecule between two metal ions. Comparing with the first conformation; in the second conformation of rolipram, the five-membered ring on the left part, the interaction between carbonyl oxygen and the water molecules was lost and the coordination with the metal ions also was not observed (Figure 4.14) So, to see the conformational change on the results, starting with the previous pharmacophore model (Figure 4.13), MOE was used to refine the model based on conformational change.

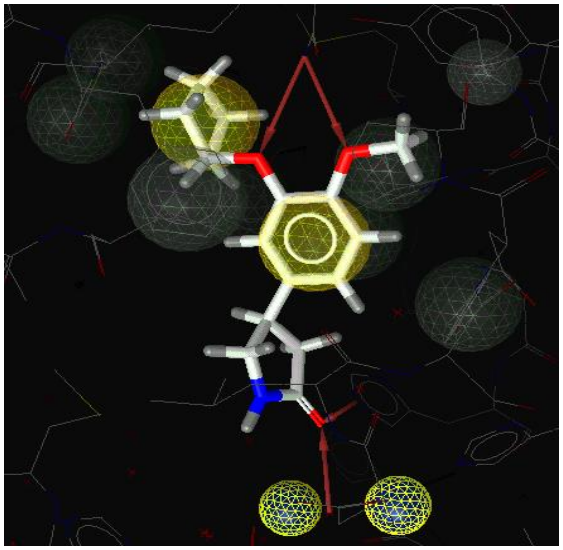


Figure 4.13. 1RO6 pharmacophore model with co-crystallized ligand rolipram.

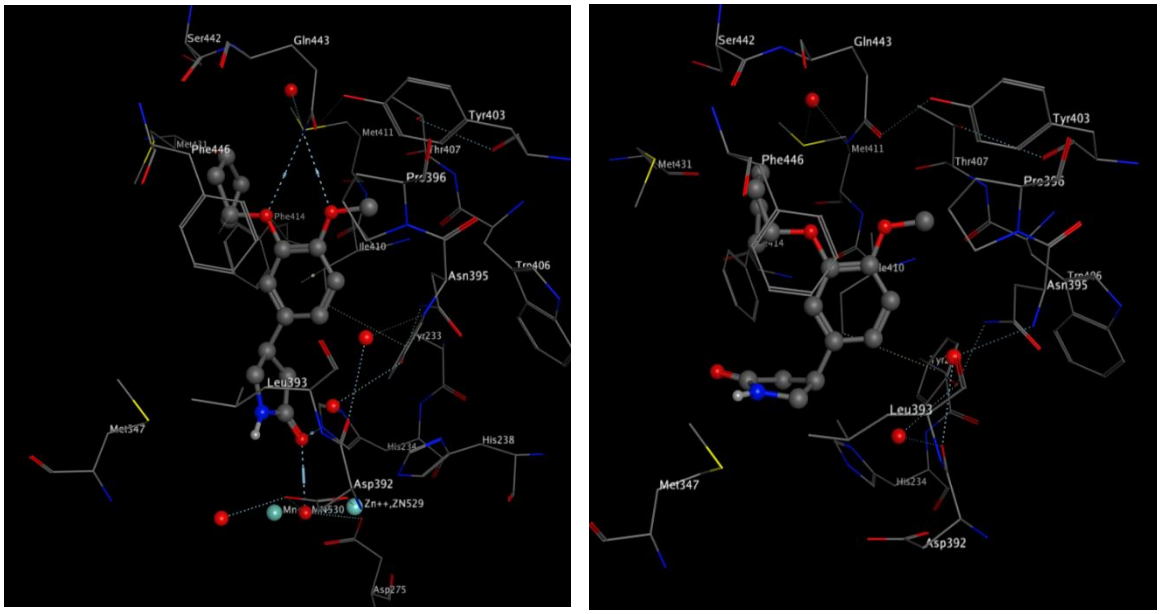
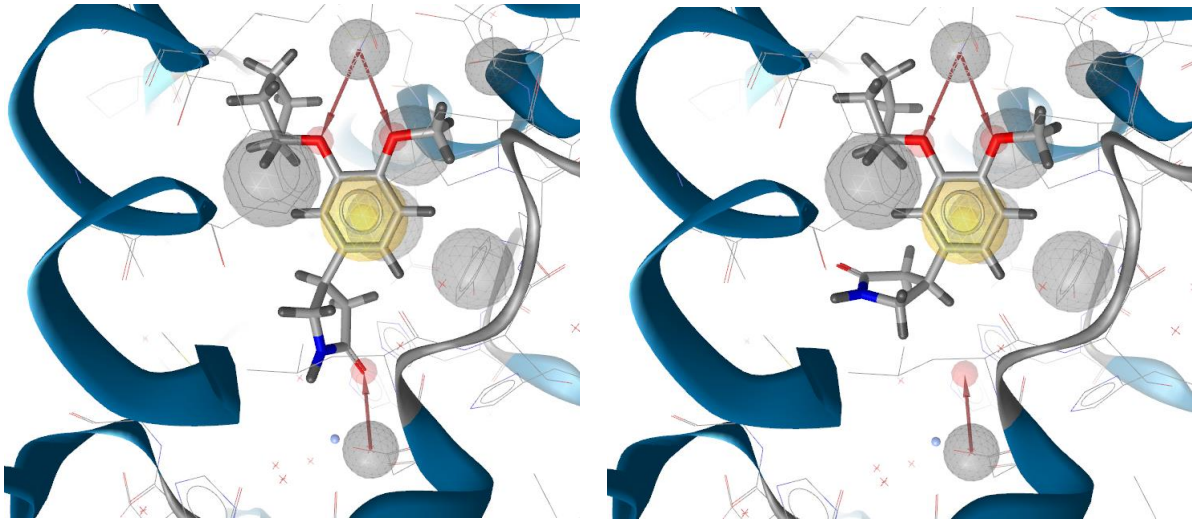


Figure 4.14. Different conformations of known inhibitor rolipram in the 1RO6 crystal structure.



.

Figure 4.15. Refined pharmacophore models with MOE based on conformational change of known inhibitor rolipram.

As seen in Figure 4.15, refined pharmacophore models have four and three important pharmacophoric features, respectively. Grey spheres show hydrogen acceptors and the yellow ones represent aromatic rings.

4.2.2.2. Virtual Screening Based on Pharmacophore Models. For the database, twenty-six known inhibitors from pdb and the database from our previous work were combined in a new database (totally 1867 compounds). For this new database, a conformational search for each compound was carried out with MOE. In the conformational search, the stochastic method was chosen and the other parameters were set as the iteration limit 100, MM iteration limit 200, RMS Gradient 0.05, RMSD Limit 0.5 and the conformational limit 10. After the conformational search, totally 17488 conformers were obtained for further study.

The virtual screening was carried out on both pharmacophore models and the first one yielded 12114 hits and the second one gave 14740 hits, which were merged in a database totally 26854 hits for binding test and docking process.

The key residues in the active site within 5 Å of the each crystal structures for each ligands were determined based on the twenty-six co-crystallized ligands native binding mode. Two dimensional protein-ligand interaction maps were prepared and the percentage of occurrences of residues in the active site of the enzyme were calculated as shown in Figure 4.18.

As seen in Figure 4.16, a total of seven residues interact more than 85 % of the co-crystallized ligands. Seven key residues are: Tyr233, His234, Met347, Ile410, Phe414, Gln443, and Phe446.

With this binding test, a decoy set that contains 67 non-inhibitors and 62 inhibitors from co-crystallized ligands and CHEMBL database, was tested. According to this test, a compound will be as a hit for further study if the best pose of the compound will be interacting with these seven key residues. The results when the compounds were docked with four different scoring functions are shown in Table 4.3. As seen in Table 4.3, more than

70% of the inhibitors have passed the binding test when they are docked with all of four different scoring functions. The percentage more than 70% value represents that it can be used as a powerful tool to filter out non-inhibitors from the database.

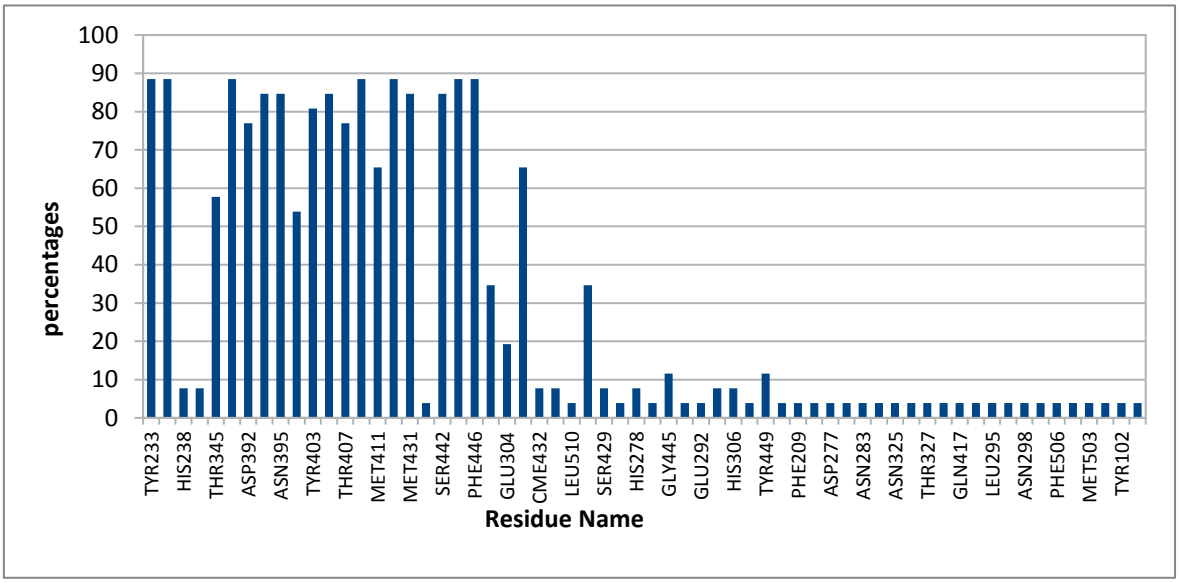


Figure 4.16. Percentage occurrences of residues of known inhibitors in the active site of enzyme.

Table 4.3. The percentage of decoy set that passed filtering test based on four different scoring functions with Gold v5.1.

Scoring Function Method	# of Inhibitors passed binding test	# of non-inhibitors passed binding test	% of Filtering Test
ASP	58	10	58/67*100= 86%
CHEMPLP	49	14	49/67*100=73%
ChemScore	49	20	49/67*100=73%
GoldScore	51	4	51/67*100=76%

After the binding test, the docking results were analyzed with the enrichment factor. To determine the enrichment factor, the Equation 4.1 was used:

$$ER = \frac{(TP/n)}{(A/N)} \quad (4.1)$$

TP is the number of true positives, n is the total number of compounds with a score value above threshold, A is the number of inhibitors and N is the total number of compounds screened in the study.

For example, for the CHEMPLP scoring function, a total of 37 inhibitors from the decoy set (67 inhibitors) and 7925 conformers from the large database (26854 conformers) have passed the binding test with a threshold score value larger than 70.9. These results correspond to an enrichment factor of 1.87 for 29.5 % database coverage. % database coverage belongs to n/N x 100.

For the other scoring functions, ASP, ChemScore and GoldScore, the same procedure was repeated. The enrichment factor and % database coverage values were calculated as shown in Table 4.4.

Table 4.4. The calculated enrichment factors and % database coverage values for each scoring functions.

Scoring Function Method	**Enrichment Factor**	**% database coverage**
CHEMPLP	(37/67) : (7925/26854) = 1.87	(7925/26854) x 100 = 29.5 %
ASP	(48/67) : (16170/26854) = 1.18	(16170/26854) x 100 = 60.2 %
ChemScore	(32/67) : (2994/26854) = 4.2	(2994/26854) x 100 = 11.15 %
GoldScore	(40/67) : (8337/26854) = 1.92	(8337/26854) x 100 = 31.04 %

After the enrichment factor calculations, the conformers that have found in all of four scoring functions were selected and the others were eliminated from the database. In the database, a compound has different conformers with different poses from docking but it is enough to select a compound as a hit if its best conformer with best docking score provides all the requirements such as passing the binding test and being in all of four scoring functions

after docking. So for each compound, only the best conformer with the best score value was chosen and finally 263 different compounds were selected. The last step is to classify these compounds according to Tanimoto coefficient with more than 60 % similarity. According to classification, 84 different families were obtained, and 4 different scaffolds have better interaction residue energy than known inhibitor rolipram (Table 4.5).

4.2.3. Conclusion

In this part of the study, an improvement pharmacophore model study was performed for PDE4 enzyme. As seen in Table 4.5, the new scaffolds have similar functional parts with the known inhibitor rolipram. At the top portion of active site, which is known as Q pocket, the oxygen atoms can give interaction with Gln443 residue as similar with rolipram. And the other two oxygens, one of them is carbonyl oxygen, are found in the M pocket and have been coordinated with metal ion Mg^{2+} (Figure 4.17). So, these new scaffolds show that these interactions are important and most promising parts of the novel compounds for PDE4B inhibition.

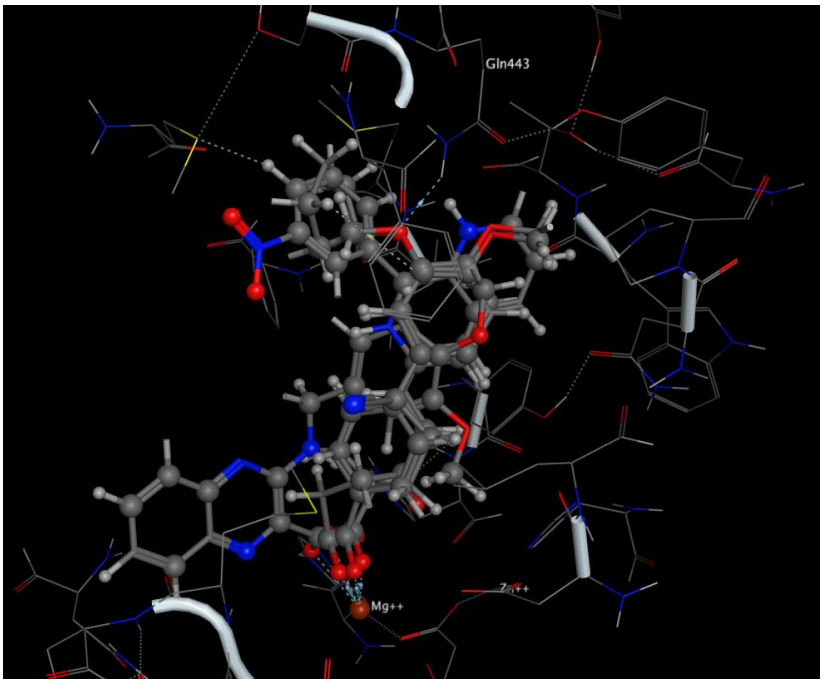
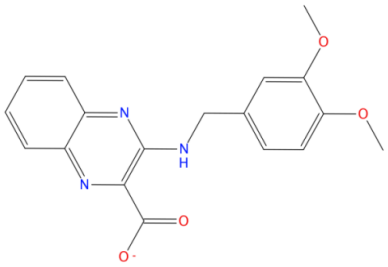
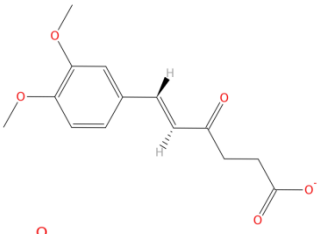
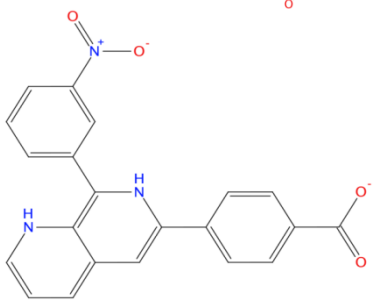
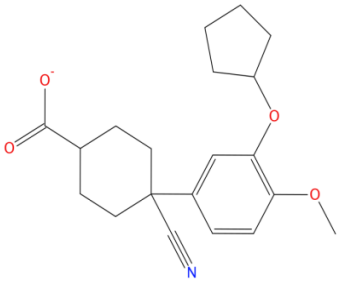


Figure 4.17. Alignment of proposed scaffolds with MOE.

Table 4.5. The structures of 4 candidate inhibitors for PDE4B.

Compound Name	Interaction Residue Energy (kcal/mole)	Structure
ZINC01645531	-68.3	
ZINC01574265	-64.8	
4-[8-(3-nitrophenyl)-1,7-naphthyridin-6-yl]benzoic acid	-63	
Cilomilast	-60.8	

5. ASSESSING PROTEIN-LIGAND BINDING MODES FOR KNOWN PDE4B INHIBITORS

In a first step in the discovery of novel potent inhibitor structures for the PDE4B family with limited side effects, we present a protocol to evaluate the potentiality of newly designed molecules through the estimation of their IC_{50} values. Our protocol is based on reproducing the linear relationship between the logarithm of experimental IC_{50} values ($\log(IC_{50})$) and their calculated binding free energies ($\Delta G_{binding}$). From 13 known PDE4B inhibitors, we show here that 1) binding free energies obtained after a docking process by AutoDock are not accurate enough to reproduce this linear relationship; 2) MM-GB/SA post-processing of molecular dynamics (MD) trajectories of the top ranked AutoDock pose improves the linear relationship; 3) by taking into account all representative structures obtained by AutoDock and by averaging MM-GB/SA computations on a series of 40 independent MD trajectories, a linear relationship between $\log(IC_{50})$ and the lowest $\Delta G_{binding}$ is achieved with $R^2 = 0.944$.

5.1. Introduction

The cyclic nucleotide phosphodiesterase (PDE) is an enzyme responsible for the degradation of the second messengers cyclic adenosine 3',5'-monophosphate (cAMP) and guanosine 3',5'-monophosphate (cGMP) into 5'-adenosine monophosphate (5'-AMP) and 5'-guanosine monophosphate (5'-GMP) respectively in many cell types [1-3].

The second messengers, cAMP and cGMP, are essential for many metabolic processes such as vision, muscle contraction, neurotransmission, exocytosis, cell growth, differentiation, learning, apoptosis, lipogenesis, glycogenolysis, ion channel functions and gluconeogenesis [5-7,120]. The regulation of the level of second messengers in vivo by synthesis activity of the receptor-linked enzymes (adenyl and guanylyl cyclases) and hydrolysis into 5'-nucleotide monophosphates by PDEs is therefore of crucial importance [8,21,28,121].

Up to now, 11 families of PDE enzyme with different substrate specificity, inhibition, substrate requirements, gene sequence and tissue distribution have been reported [1,4,6,7,12]. Among these families, the cAMP specific one is PDE4, which is encoded by four different isoforms as A, B, C and D. These isoforms are characterized by unique N-terminal regions [21]. The PDE4 subfamily has attracted much attention for its usage in the treatment of inflammatory and immune disorders such as asthma, chronic obstructive pulmonary disease (COPD), rhinitis and also as therapeutic agent for rheumatoid arthritis, multiple sclerosis, type II diabetes, septic shock, atopic dermatitis, and other autoimmune diseases [18-20,30].

In the PDE4 subfamily, among the four isoforms A, B, C and D, PDE4B has a specific importance especially in the inflammatory responses of lymphocytes [28]. The design of novel inhibitors for PDE4B is of significant interest to the pharmaceutical industry due to its usage as an attractive target for anti-inflammatory diseases. There are many PDE4 inhibitors that have been under clinical trials [13,21,28] however their clinical utility has often been limited due to their side effects like vomiting, nausea and increased gastric secretion [122]. It is thus important to design a novel PDE4B selective inhibitor with reduced side effects and improved pharmacological profile.

Designing small molecules with desirable binding affinity and biological activity is one of the major goals in computational biology [95,98,123,124]. Molecular docking is a popular method used to identify the orientations of molecules into the active site of a target protein structure. In the last years, docking methods have been improved by adding energy contributions or by refining the parameters for scoring functions but there are still some limitations especially like sometimes poor correlation between docking score values and experimental results [98,123]. Up to now, many studies involving molecular docking, molecular modeling, pharmacophore modeling, the investigation of the hydrolysis mechanism and the description of the structure-activity relationships for PDE4 inhibitors have been published. Different series of PDE4 selective inhibitors have been studied by Alexander *et al.* [125], Kuang *et al.* [126], Ke *et al.* [121], and Guay *et al.* [127], Xu *et al.* [128], Wierzbicki *et al.* [129], and Zhan *et al.* [130] have focused on the hydrolysis

mechanism of PDE4 enzyme. In 2002, Colicelli *et al.* [117] have carried out a molecular docking study of competitive PDE inhibitors. Another molecular docking study with development of an empirical binding free energy for PDE4 inhibitors in 2006 was performed by Barreiro *et al.* [131]. Zhu *et al.* [116] have combined multiple pharmacophore modeling and molecular docking process to suggest novel PDE4 inhibitors. Another pharmacophore modeling study for PDE4 was carried out by Gu *et al.* [132]. However, to the best of our knowledge, no study based on performing molecular dynamics simulations and calculating free binding energies with different methods for PDE4 family has been reported so far.

In this context, an important goal of computational medicinal chemistry is to develop methods that can accurately estimate the free energy of binding, $\Delta G_{\text{binding}}$, and that could allow the estimation of the binding strength of any drug candidate prior to its synthesis. The free binding energies can be represented as:

$$\Delta G_{binding} = -RTlogK_i \tag{5.1}$$

where R is the ideal gas constant, T is the temperature, and K_i is the binding affinity of the inhibitor. The K_i constant can be related to experimental IC_{50} values based on the following equation [133]:

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_m}} \tag{5.2}$$

From Eq. 5.2 the binding affinity K_i depends on the IC_{50} value, the substrate concentration [S] and the Michaelis-Menten constant K_m. For a set of ligands and their experimentally measured IC_{50}, there should therefore be a linear dependency between K_i and IC_{50} provided that the experimental conditions for all ligands are similar: the substrate concentration should be identical for all experiments and the thermodynamical conditions should remain similar (i.e., temperature, pressure, pK_a, etc.). From this point of view, a linear trend between $\Delta G_{\text{binding}}$ and log(IC_{50}) values should be expected.

There are many computational approaches for free energy calculation such as free energy perturbation (FEP) [94], thermodynamic integration (TI) [134], linear response (LR) [96], Molecular Mechanics-Generalized Born/Surface Area (MM-GB/SA) and Molecular Mechanics-Poisson Boltzmann/Surface Area (MM-PB/SA) methods [135,136]. Among these methods, the most accurate and rigorous ones are FEP and TI [137]. Despite their accuracy, they have found little use in drug design [138] due to their convergence only for rather similar ligands and computational cost [92]. The MM-GB/SA and MM-PB/SA methods, that combine molecular mechanics energy and implicit solvation models, are simple and faster than FEP [123]. Therefore, they have been widely used in free energy calculations in computational medicinal chemistry [95,98]. It is important to achieve statistically fully converged results and statistical estimates in order to test how well the methods reproduce the experimental data. As Ryde and coworkers have shown, converged results using MM-GB/SA method can be achieved by running multiple independent short molecular dynamics simulations starting with different initial velocities and a same initial structure rather than by running a single (very) long simulation [92].

In this project, the aim is to evaluate binding energies with the MM-GB/SA method and show the correlation between the binding energies and half maximal inhibitory concentration (IC_{50}) values of the ligands. The study includes i) building a database of experimental IC_{50} values that include a training and a test set; ii) performing docking process for each ligand, iii) carrying out independent molecular dynamics simulations for the top ranked poses of each ligand and calculating the free binding energy using the MM-GB/SA approach, iv) analyzing the role of the possible alternative poses of each ligand from MM-GB/SA calculations and finally v) applying a linear regression method on the training set to establish a relationship between calculated $\Delta G_{binding}$ and experimental $\log(IC_{50})$ and verifying the reliability of our approach with the test set.

5.2. Methodology

5.2.1. Training and Test Set

For the dataset preparation, the ligands with known IC_{50} values from experimental studies of Dal Piaz *et al.* [31] and Zhang *et al.* [28] were chosen due to their selectivity for PDE4B and their large range of different IC_{50} values. These ligands were also searched in the Binding Database [139] and it was found that some of them have more than one IC_{50} value reported, as the ligands cilomilast and npv (Tables 5.1 and 5.2). The training set has been designed to contain IC_{50} values from a single source: those of Dal Piaz *et al.* [31]. It contains 8 ligands for which experimental IC_{50} values range from 0.6 to 9.0 μM (Table 5.1). The test set contains 7 molecules: rolipram, tadalafil, filaminast, mesopram, zardaverine, cilomilast and npv. Their experimental IC_{50} values range from 0.025 to 9.2 μM (Table 5.2).

5.2.2. Protein and Dataset Preparation

The starting structure for the protein is the human PDE4B enzyme (PDB code: 1RO6, 2 Å resolution, Figure 5.1). The X-ray structure contains two identical chains with rolipram as a co-crystallized ligand and two metal ions: Zn^{2+} and Mg^{2+}. All our calculations were carried out on one single active chain which includes the two metal ions, Zn^{2+} and Mg^{2+}, and the water molecule (residue #788 in 1RO6) positioned between these two atoms. The choice of using the 1RO6 X-ray structure over other available PDE4B X-ray structures like the apo one (PDB code: 1F0J) was dictated by the fact that the two structures are very similar (the RMSD between the backbones of 1RO6 and 1F0J is only 0.13 Å) and that the docking procedure always yielded lower binding energies for 1RO6 than for 1F0J.

The ligand dataset is a combination of training and test sets (Tables 5.1 and 5.2). The IC_{50} values of the ligands are known from different experimental studies [28,29,140,141]. The training set contains molecules that have been experimentally tested using a single source: guinea pig ventricular tissue [140]. The test set contains ligands which have been tested for inhibition using PDE4B proteins from various sources: guinea pig [140], human

[28,29], or rat [141]. All these protein sources share a strong sequence homology (> 95% of identity). For example, the sequence alignment between guinea pig and human PDE4B in UniProt has shown that they differ by only five residues that are out of the active site.

Table 5.1. Ligand names, 2D chemical sketches and experimental IC_{50} values for the training set [140].

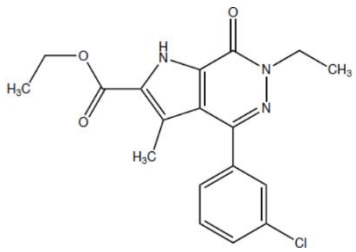
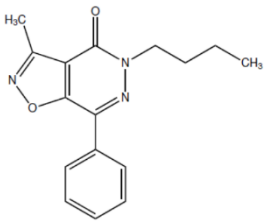
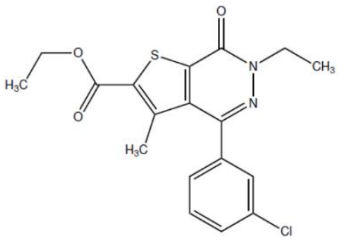
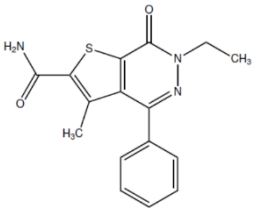
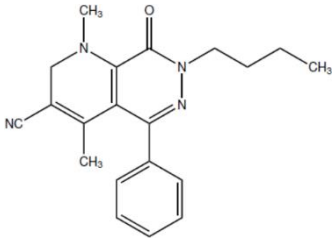
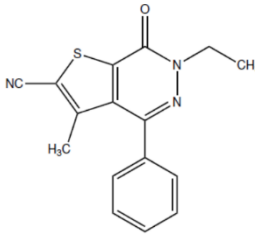
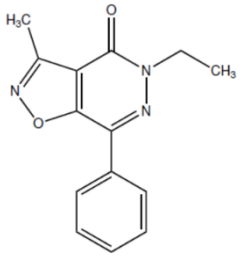
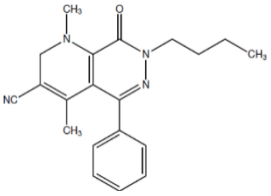
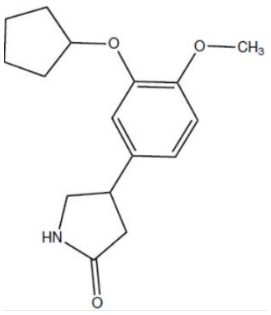
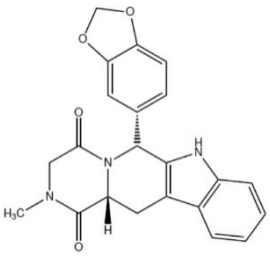
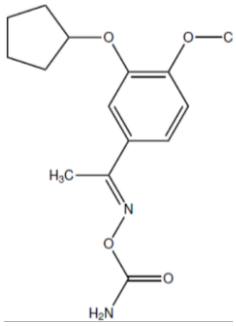
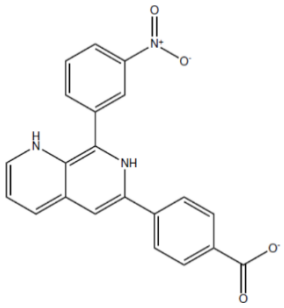
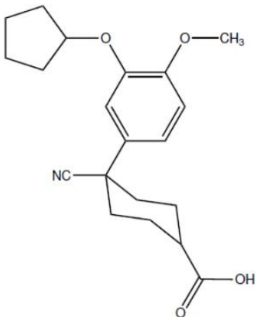
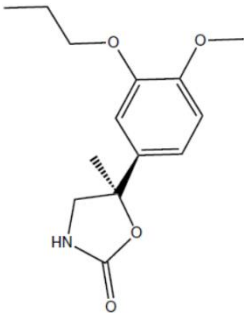
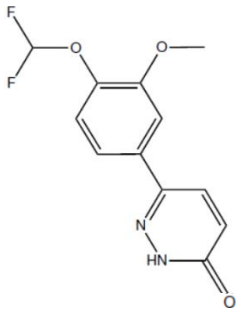
Training Set					
Ligand	2D Structure	IC_{50}(μM)	Ligand	2D Structure	IC_{50}(μM)
Ligand 3		0.6 ±0.1	Ligand 7		6 ±0.5
Ligand 4		0.9 ±0.2	Ligand 8		3 ± 0.5
Ligand 5		1.1 ±0.4	Ligand 9		4 ± 0.5
Ligand 6		9 ± 0.8	Ligand 10		2 ± 0.5

Table 5.2. Ligand names, 2D structures and experimental IC_{50} values for the test set.

Test Set					
Ligand	2D Structure	IC_{50}(μM)	Ligand	2D Structure	IC_{50}(μM)
rolipram		0.32 ±0.09 [140]	tadalafil		9.2 [28]
filaminast		0.96 [28]	npv		0.049 ± 0.007 [141] (0.650) [29]
cilomilast		0.025 [28] (0.31) [141]	mesopram		0.42 [28]
zardaverine		0.93 [28]			

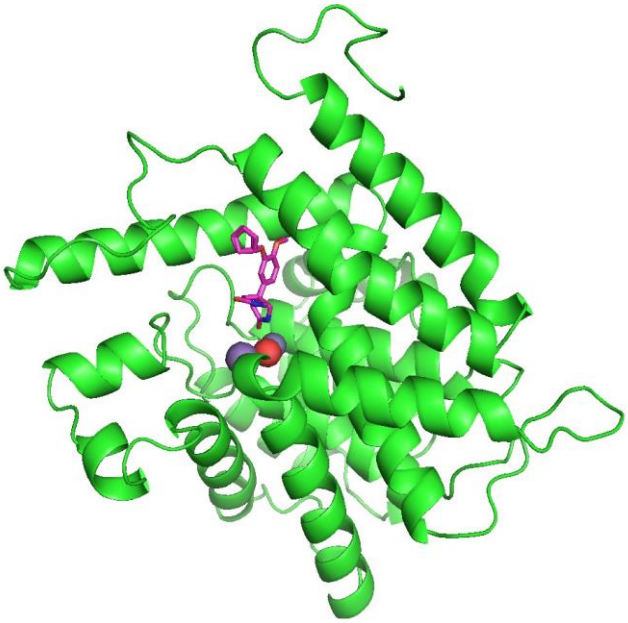


Figure 5.1. Cartoon representation of PDE4B X-ray structure. Chain A is represented in green as cartoon, the co- crystallized ligand as ball & stick, Zn^{2+} and Mg^{2+} are in purple, the water molecule is in red.

5.2.3. Docking Procedure

The docking process was carried out with AutoDock v4.2 [142]. For each ligand, ten independent runs were performed. A pre-calculated three-dimensional energy grid of equally spaced discrete points was generated prior to the docking using the program AutoGrid [142]. The grid box (32Å x 72Å x 31Å) contains the active site and several key residues important for the protein-ligand interaction. The distance between two grid points was set to 0.375 Å. The grid map files were calculated by AutoGrid for the ligand atom types: A, NA, C, OA, and N. For conformational search, Lamarckian Genetic Algorithm, which combines a local search and a genetic algorithm to provide both efficient global space coverage and local search optimization, was chosen. During the process, the protein was held rigid. The population size was set to 150, the maximum number of energy evaluations was set to

2500000, the maximum number of generations was 27000, the mutation rate was 0.02 and the crossover rate was 0.8. The remaining parameters were set as the default values.

Ligand atomic charges were calculated with the restrained electrostatic potential fit (RESP) method at the B3LYP/cc- pVTZ level after a full geometrical optimization carried out at the B3LYP/6-31G* level. This procedure is compatible with the charges obtained for the Amber force field [143] used in the subsequent molecular dynamics (MD) runs. Thus, the same set of atomic charges will be used in the docking process, in the production of the MD trajectories and in the calculations of the binding free energy during the MM-GB/SA post-process.

At the end of each docking process, the 10 docked poses of each ligand were clustered based on their RMSD values using a cluster RMSD threshold of 0.5Å. For each cluster of each ligand, a representative pose with the lowest $\Delta G_{\text{binding}}$ value was selected and incorporated in our analysis in order to take into account the diversity of the binding modes.

5.2.4. Molecular Dynamics Simulations

Hydrogen atoms were added to the system with the tleap module of AMBER 12 [144]. For histidines, the protonation state was determined based on PROPKA [145] calculations and hydrogen bond pattern analysis. Counter sodium ions were added to neutralize the system. Waters from the crystal structure were deleted except for the water molecule that is located between the two metal ions Zn^{2+} and Mg^{2+} and is hydrogen bonded to the co-crystallized ligand rolipram. The system was solvated with TIP3P [146] water molecules extending to at least 10 Å from the protein. The system was cubic with edge length 74.50 Å and had an initial density of 1.0 g/cm^3.

The MD simulations were performed using the CUDA [147,148] version of the PMEMD module of AMBER 12. The Amber ff03 [143] force field was used to model the PDE4B protein while the general AMBER force field (GAFF) [149] force field parameters were used to model the ligands. The SHAKE [150] algorithm was chosen to constrain bond

lengths involving hydrogen atoms. The Andersen temperature coupling algorithm was applied to ensure a constant temperature (NVT) ensemble. The time step was set to 2 fs.

In gas phase, before the solvation of the system, a short minimization followed by one MD run was carried out for 100 ps at 10 K to optimize the hydrogen atom positions: all heavy atoms were restrained to their crystallographic positions using a harmonic potential with a force constant of 100 kcal.mol^{-1}.A$^{\circ-2}$. After solvation, the equilibration of the system was performed in five stages [151]. First, only the hydrogen atoms of the system were allowed to move during 100 ps at 10 K (i.e., by applying a force constant of 50 kcal.mol^{-1}.A$^{\circ-2}$ on all heavy atom positions). Second, the water molecules were allowed to move for the next 100 ps at the same temperature. Third, the force constant on the protein heavy atom positions was decreased to 5 kcal.mol^{-1}.A$^{\circ-2}$ for another 100 ps. Then the whole system was free to move during 100 ps at 10 K. Finally, the thermostat temperature was smoothly increased from 10 K to 300 K for another 2 ns.

After equilibration, for each ligand representative of its cluster, forty independent simulations were performed up to 1 ns at 300 K with different initial velocities. During the production runs, coordinates were saved every 2 ps for the subsequent MM-GB/SA calculations. Using NVIDIA Tesla M2090 GPU, one 1 ns simulation takes in average 1.2 hours for a speed of about 20 ns/day.

5.2.5. MM-GB/SA Post-Processing

The free energy of binding for each ligand is calculated using the equation:

$$\Delta G_{binding} = < G_{RL} > - < G_R > - < G_L > \quad (5.3)$$

where RL, R and L stand for receptor-ligand complex, receptor and ligand, respectively. The average free energy of each system is estimated as a sum of three terms:

$$G = E_{MM} + G_{solv} - TS_{MM} \quad (5.4)$$

where E_{MM} is the molecular mechanics energy of each system, including internal, non-bonded electrostatics, and van der Waals energies. G_{solv} is the solvation energy which consists of a polar and a nonpolar part. The polar solvation free energy is calculated by a Generalized Born (GB) approach. The nonpolar solvation free energy is computed by a relation to the solvent-accessible surface area (SASA). The last term TS_{MM} is the product of absolute temperature and the entropy.

In this study, the first two terms were calculated using the MMPBSA.py module of AMBER 12 with all water molecules stripped off [101]. To evaluate the polar solvation free energy, different solvation models have been evaluated: GB^{HCT} [152-154], GB^{OBC} [155], GB^{OBC-2} [144,155], GB^{GBneck} [156], and $GB^{GBneck2}$ [157]. The hydrophobic contribution has been approximated by the Linear Combinations of Pairwise Overlaps (LCPO) method [158].

In this study, the entropy term was not included in our calculations although it could have been evaluated through a usual normal-mode analysis [159]. There have been much debate in the literature about the entropy term in MM-GB/SA calculations and whether it should be systematically included or not to improve the accuracy of the results [99,160,161]. In our case, given the high computational cost of its calculation and the good prediction that we have obtained without including it, we have neglected the entropy term component. We recall here that the aim of this study was not to estimate the experimental $\Delta G_{binding}$, but to relate the computed free energy values to experimental inhibition concentrations.

Finally, the calculated $\Delta G_{binding}$ values are averaged over 40 independent simulations for each ligand.

5.3. Results and Discussion

5.3.1. Best Docking Scores vs. Experimental IC_{50} Values

The study has started with the docking process of all ligands in both datasets into the target PDE4B enzyme using AutoDock v4.2. For each ligand, ten poses are obtained from a

total of 10 docking runs. The best (i.e., top ranked) pose with the lowest AutoDock $\Delta G_{binding}$ value is recorded and a linear correlation between the $\Delta G_{binding}$ and log(IC_{50}) is searched for.

In Figure 5.2, the correlation between the lowest AutoDock $\Delta G_{binding}$ values and the corresponding log(IC_{50}) values is represented for the training set. Only a weak linear correspondence exists between $\Delta G_{binding}$ and experimental log(IC_{50}) with R^2 value of 0.135. That means that, while AutoDock is capable of discriminating between different poses and of finding true positive hits, its scoring function is not capable of estimating experimental $\Delta G_{binding}$ values in the case of PDE4B.

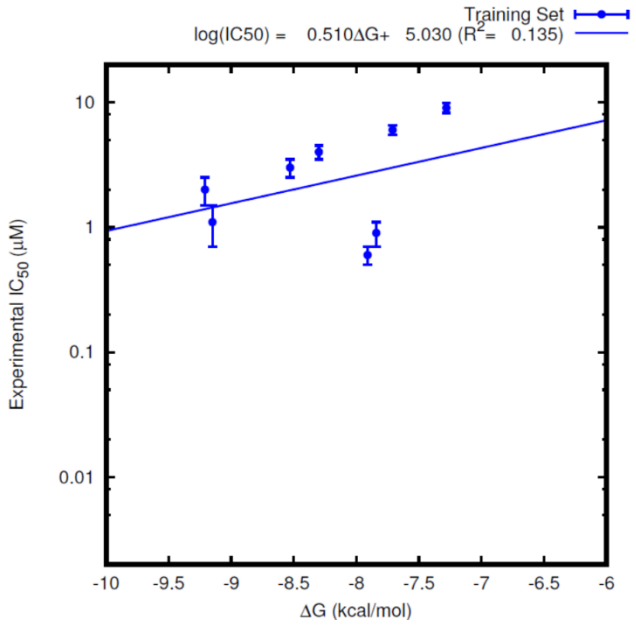


Figure 5.2. Correlation between experimental IC_{50} values and the lowest ΔG scores (kcal/mol) obtained by a series of AutoDock docking computations of the training set (in blue). Vertical error bars: standard experimental deviations. Blue dashed line: linear fit between lowest AutoDock $\Delta G_{binding}$ values and experimental log(IC_{50}).

5.3.2. Convergence of the Free Energy Results

Another way to obtain binding free energies is to use the MM-GB/SA approach. Here, $\Delta G_{binding}$ energies are obtained by post-processing MD trajectories of complexed

protein:ligand structures. In our cases, we have used as starting structures for the MD runs,the complexed structures obtained by AutoDock. For each docked pose, we have performed 40 independent 1 ns MD runs. The convergence of $\Delta G_{binding}$ calculations for two independent runs corresponding to the ligand rolipram is represented in Figure 5.3. It shows that a 1 ns trajectory is enough to ensure the convergence of $\Delta G_{binding}$ for that run. However, two independent runs can give rather different results: one MD yields $\Delta G_{binding}$ = -54.6 +/- 3.6 kcal/mol while the other yields $\Delta G_{binding}$ = -60.5 +/- 3.4 kcal/mol. As suggested by Genheden and Ryde, converged MM-GB/SA results can be obtained by averaging multiple independent trajectories. Figure 5.3 represents the convergence of MM-GB/SA $\Delta G_{binding}$ energies for rolipram as a function of the number of independent trajectories. Convergence is obtained after 40 trajectories (-57.6 +/- 1.6 kcal/mol). Adding more trajectories do not change the picture beyond: $\Delta G_{binding}$ = -57.8 +/- 1.6 kcal/mol after 80 runs.

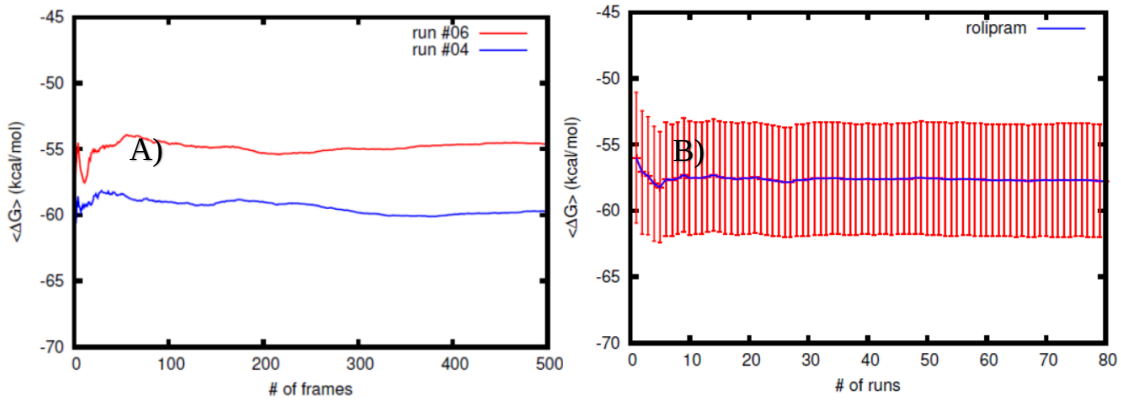


Figure 5.3. Convergence of the ΔGbinding MM-GB/SA computations for rolipram using multiple MD trajectories. A) Convergence of the averaged ΔGbinding, in kcal/mol, for two independent runs of 1 ns (500 frames each); B) Convergence of the averaged ΔGbinding, in kcal/mol, as a function of the number of independent 1 ns long MD trajectories, the error bars represent the standard deviation in kcal/mol.

In the following steps, all MM-GB/SA free energies will be calculated for every distinct ligand pose representative of each cluster using the same protocol: the MM-GB/SA

post-processing of 40 independent MD runs using different random initial velocities associated to the structure coordinates of the corresponding pose as obtained by AutoDock.

5.3.3. MM-GB/SA Binding Free Energies of Top Ranked AutoDock Poses vs. Experimental IC_{50} Values

The $\Delta G_{binding}$ values have been calculated using the MM-GB/SA approach for the top ranked poses of all ligands in the training set and the test sets. Figure 5.4 represents the correlation between $\Delta G_{binding}$ and the logarithm of the experimental IC_{50}. For the training set, the linearity of the trend is more pronounced ($R^2 = 0.788$) than when using the AutoDock scores ($R^2 = 0.135$). This shows that using a molecular force field as the AMBER force field yields more accurate results.

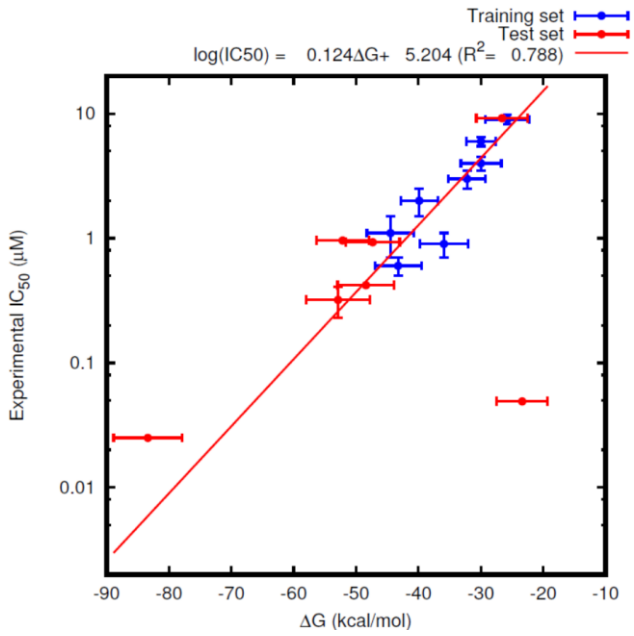


Figure 5.4. Correlation between experimental IC_{50} values and MM-GB/SA averaged $\Delta G_{binding}$ free energies computed from the top ranked AutoDock poses of the training set (in red) and the test set (in blue). Vertical error bars: standard experimental deviation. Horizontal error bars: computed standard $\Delta G_{binding}$ deviations. Red dashed line: linear fit between $\Delta G_{binding}$ values for the top ranked AutoDock poses and experimental $\log(IC_{50})$.

When the test set is assessed (Figure 5.4, error bars in blue), most $\Delta G_{binding}$ values are correlated to their experimental IC_{50} counterparts as in the training set. However, one value is off the linear region by more than 50 kcal/mol. This corresponds to the npv ligand for which two IC_{50} values have been reported: 0.049 [141] and 0.650 [29]. Given the linear trend of the binding free energies found for the training set, from these two IC_{50} values should correspond two possible $\Delta G_{binding}$: one around -66.3 kcal/mol, the other around -45.4 kcal/mol. Using the top ranked AutoDock pose, the MM-GB/SA binding free energy is computed at -23.4 +/- 4.1 kcal/mol instead.

5.3.4. Minimum MM-GB/SA Binding Free Energies vs. Experimental IC_{50} Values

If MM-GB/SA $\Delta G_{binding}$ values are better correlated to experimental IC_{50} values than AutoDock $\Delta G_{binding}$ values, one can wonder whether alternative poses obtained by AutoDock would be ranked similarly if the docking score was obtained from a MM-GB/SA computation instead. While we cannot change the way AutoDock optimizes the poses during molecular docking, we have performed MM-GB/SA calculations on a more diverse set of poses: one representative pose of each cluster for each ligand in the training set was chosen and MM-GB/SA $\Delta G_{binding}$ was computed using the same multiple MD trajectory approach than for the top ranked AutoDock pose. The number of alternative poses per ligand in the training set varies from 1 (e.g., ligand5) to 5 (e.g., ligand7).

In Figure 5.5, the correlation between the calculated $\Delta G_{binding}$ and the experimental IC_{50} values is represented. For some ligands, a lower $\Delta G_{binding}$ value than for the top ranked AutoDock pose is found. When the minimum averaged $\Delta G_{binding}$ values are used (blue filled circles in Figure 5.5), a better linear trend is found than when only top ranked AutoDock poses are considered (red filled circles in Figure 5.5). The relationship between computed averaged $\Delta G_{binding}$ and experimental $\log(IC_{50})$ is expressed as:

$$log(IC)_{50} = 0.110\Delta G_{binding} + 5.060 \quad (5.5)$$

with a correlation coefficient $R^2 = 0.944$.

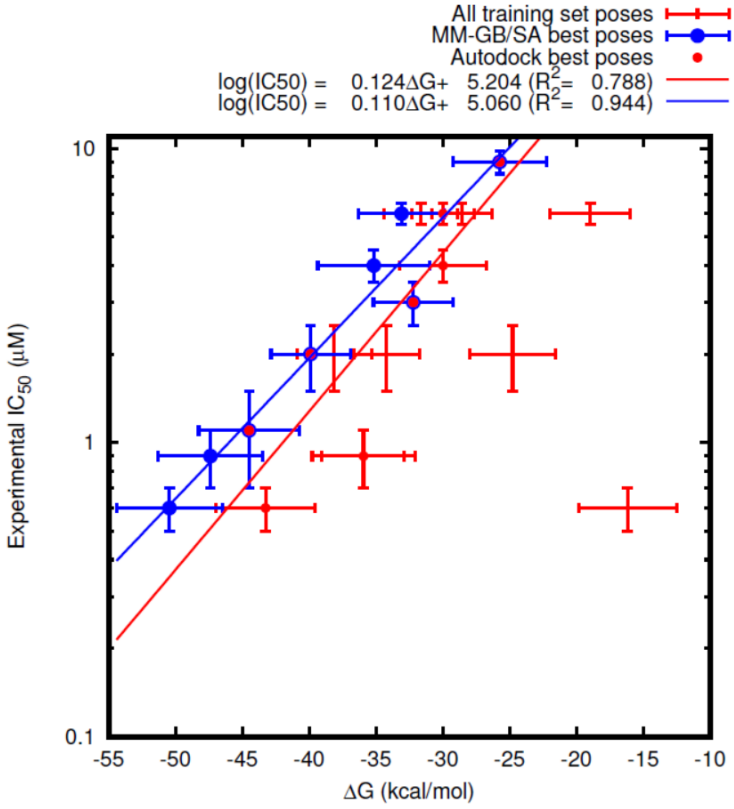


Figure 5.5. Correlation between experimental IC_{50} values of the training set and MM-GB/SA averaged $\Delta G_{binding}$ free energies computed for all AutoDock poses (one representative pose per AutoDock family). Vertical error bars: standard experimental deviations. Horizontal error bars: computed standard $\Delta G_{binding}$ deviations. Red filled circles: $\Delta G_{binding}$ values corresponding to the top ranked AutoDock poses. Blue filled circles: minimum $\Delta G_{binding}$ values. Red dashed line: linear fit between $\Delta G_{binding}$ values for the top ranked AutoDock poses and experimental $\log(IC_{50})$. Blue dashed line: linear fit between minimum $\Delta G_{binding}$ values and experimental $\log(IC_{50})$.

The improvement of the correlation coefficient shows that while AutoDock is capable of discriminating between bad and good binding poses, its docking scores are not quantitative enough to be used directly to evaluate the binding affinity of a ligand for PDE4B. However, by using the many different poses extracted from AutoDock runs and by applying a protocol that involves MM-GB/SA calculations on multiple independent

trajectories, it is possible to recover correct $\Delta G_{binding}$ values that are in quantitative agreement with experimental values.

5.3.5. Estimation of IC_{50} Values

Using Equation 5, it is now possible to estimate IC_{50} values from MM-GB/SA $\Delta G_{binding}$ values. Table 5.3 summarizes all the results that have been obtained for the test set and the training set when applying one of the three computational approaches presented here: i) linear fitting using the AutoDock ΔG scores of the top ranked poses; ii) linear fitting using averaged MM-GB/SA values for the top ranked AutoDock poses; iii) linear fitting using the lowest averaged MM-GB/SA values among representative poses of all AutoDock clusters. Because IC_{50} values are spread in an exponential range from 0.025 μM to 9.2 μM, we use mean absolute percentage error (MAPE) as a criteria to evaluate the error between experimental IC_{50} values and estimated IC_{50} values. MAPE numbers, expressed as percentage, are calculated using the following expression:

$$MAPE = \frac{1}{N}\sum_{i=1} N \left| \frac{IC_{50}^{est} - IC_{50}^{exp}}{IC_{50}^{exp}} \right| \tag{5.6}$$

where IC_{50}^{est} and IC_{50}^{exp} are the estimated and the experimental IC_{50} values for molecule i, respectively.

Figure 5.6 shows the correlation between estimated IC_{50} values using the GB^{OBC} model and experimental IC_{50} values for both the training set used to define Eq. 5 and the test set. By using all AutoDock clusters, the estimated IC_{50} values from the test set are within 38% of relative error (see Table 5.3). Like in the training set, the use of distinct AutoDock poses improves the estimation significantly and no ligand from the test set are wrongly estimated as it was the case when only the top ranked AutoDock poses were considered (Figure 5.4).

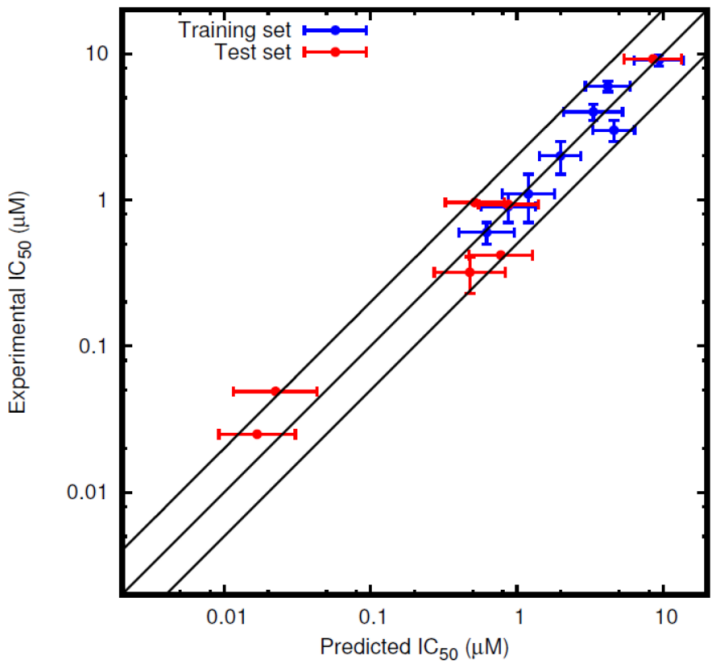


Figure 5.6. Correlation between experimental IC_{50} values and estimated IC_{50} values obtained after fitting averaged MM- GB/SA free energies computed from all AutoDock poses (1 representent per family) using the GB^{OBC} model. red: training set, blue: test set. Vertical error bars: standard experimental deviations. Horizontal error bars: standard estimated deviations. Black line: an ideal estimation while dashed lines: an error factor of 2 (upper dashed line) or 0.5 (lower dashed line) in the estimated IC_{50} value, respectively.

From Table 5.3, IC_{50} prediction using the AutoDock scores gives a MAPE of 101.7% for the training set and 553.6% for the test set, respectively. As stated above, AutoDock ΔG values show a linear trend but the correlation is not strong (R^2 = 0.135). When using the GB^{OBC} model on the top ranked AutoDock poses, the agreement between experimental and predicted IC_{50} values is improved(R^2 = 0.788, MAPE= 38.5% for the training set). However, some ligands like npv are wrongly predicted. This yields a MAPE of 2925.5% for the test set. By adding alternative poses, the agreement for the training set is much better, yielding R^2 = 0.944 and MAPE= 15.1%. Test ligands, including npv (see below), are now correctly predicted with a MAPE of around 40%. The GB^{OBC} model is the one that leads to the best prediction. But other GB/SA models like GB^{HCT}, GB^{OBC-2}, and GB^{GBneck} also give reliable

predictions and are always superior than the approach which uses only the top ranked AutoDock poses. Surprisingly, the $GB^{GBneck2}$ model yields the worst results among all GB/SA models. This was not expected since it is one of the most recent GB/SA model and it has proved to be accurate in modeling solvent effects in protein folding studies [162]. At the same time, $GB^{GBneck2}$ is the GB/SA model that yields to the smallest standard error when using multiple MD trajectories.

Finally, one important question that arises from those results is to check if our current protocol is capable of discriminating between experimental values when several are available in the literature. This is the case for cilomilast and npv. Surprisingly, these two molecules are the only two of our sets that contain a carboxylate group. The results reported Table 5.3 have been obtained when the carboxylate form was considered. We have recomputed predicted IC_{50} values for the carboxylic acid form for both molecules. For cilomilast and using the GB^{OBC} model, the predicted IC_{50} values for the carboxylate and the carboxylic acid forms are 0.017 ± 0.008 and 0.278 ± 0.106 μM, respectively. These two values are both in good agreement with the two reported experimental values: 0.025 μM and 0.31 μM. A possible interpretation of this agreement could be that subtle differences in the two experimental protocols yielded to the measurement of the two different acidic forms of cilomilast. This is somewhat confirmed in the case of npv. The two predicted IC_{50} values are 0.022 ± 0.011 and 1.256 ± 0.392 μM for the basic and the acidic forms of the carboxylic acid group, respectively. The predicted IC_{50} value of the carboxylic acid form again resembles more the experimental value (0.650) of Ref. [29] while the basic form resembles more the experimental value from Ref. [141]. It would be of course hazardous to generalize such findings, but, in our case, two main points can be drawn: 1) the change of protonation of ionizable residues can greatly affect the computed binding energies and great care should be taken to assess such effects; 2) when multiple experimental values are available, it does not necessarily mean that some of them are ”correct” or ”wrong”, but they can represent different states or be the results of applying different measurement protocols.

Table 5.3. Linear fitting results, estimated IC_{50}, in μM, for all approaches and comparison with experimental values.

Method	**Autodock**			**MM-GB/SA**				
Pose	**top**	**top ranked**			**min ΔG**			
GB		**OBC**	**HCT**	**OBC**	**OBC-2**	**GBneck**		**GBneck2**
a	0.510	0.124	0.106	0.110	0.104	0.085		0.042
b	5.030	5.204	5.214	5.060	4.982	4.361		2.975
R^2	0.135	0.788	0.929	0.944	0.945	0.892		0.780
Molecule				**Estimated IC_{50} for the training set**				**exp.**
ligand3	2.7	0.8 ± 0.3	0.6 ± 0.2	0.6 ± 0.2	0.7 ± 0.2	0.7 ± 0.2	0.8 ± 0.2	0.6 ± 0.1 [140]
ligand4	2.8	2.1 ± 0.8	0.9 ± 0.3	0.9 ± 0.3	1.0 ± 0.3	1.1 ± 0.4	1.1 ± 0.2	0.9 ± 0.2 [140]
ligand5	1.4	0.7 ± 0.3	1.3 ± 0.4	1.2 ± 0.4	1.1 ± 0.4	0.9 ± 0.3	0.9 ± 0.2	1.1 ± 0.4 [140]
ligand6	3.7	7.4 ± 2.6	8.7 ± 2.7	9.4 ± 3.0	10.0 ± 3.1	9.4 ± 2.3	6.7 ± 0.9	9.0 ± 0.8 [140]
ligand7	3.0	4.4 ± 1.1	4.1 ± 1.1	4.2 ± 1.2	4.2 ± 1.2	4.4 ± 1.1	4.9 ± 0.6	6.0 ± 0.5 [140]
ligand8	2.0	3.3 ± 1.0	4.7 ± 1.2	4.6 ± 1.3	4.4 ± 1.2	4.8 ± 1.1	4.7 ± 0.6	3.0 ± 0.5 [140]
ligand9	2.2	4.4 ± 1.5	3.2 ± 1.1	3.3 ± 1.2	3.3 ± 1.2	2.4 ± 0.8	1.9 ± 0.5	4.0 ± 0.5 [140]
ligand10	1.4	1.3 ± 0.4	2.3 ± 0.6	2.0 ± 0.6	1.9 ± 0.5	2.2 ± 0.6	3.4 ± 0.5	2.0 ± 0.5 [140]
MAPE (%)	101.7	38.5	19.5	15.1	16.5	24.4	38.4	
Molecule				**Estimated IC_{50} for the test set**				**exp.**
tadalafil	2.3	6.7 ± 2.7	7.7 ± 2.8	8.5 ± 3.1	7.2 ± 2.7	5.6 ± 1.7	4.1 ± 0.8	9.2 [28]
rolipram	2.38	0.25 ± 0.12	0.36 ± 0.12	0.48 ± 0.20	0.42 ± 0.18	0.98 ± 0.37	0.29 ± 0.06	0.32±0.09 [140]
filaminast	3.15	0.28 ± 0.11	0.35 ± 0.12	0.52 ± 0.19	0.45 ± 0.16	0.33 ± 0.12	0.36 ± 0.07	0.96 [28]
mesopram	4.59	0.44 ± 0.19	0.62 ± 0.21	0.78 ± 0.31	0.70 ± 0.27	0.89 ± 0.37	0.51 ± 0.12	0.42 [28]
zardaverine	5.41	0.51 ± 0.21	0.87 ± 0.31	0.88 ± 0.33	0.86 ± 0.31	1.11 ± 0.36	0.59 ± 0.11	0.93 [28]
cilomilast	0.239	0.006	0.006	0.017	0.013	0.035	1.2×10^{-04}	0.025 [28]
		±0.003	±0.003	±0.008	±0.006	±0.014	$\pm 4.9 \times 10^{-05}$	0.31 [141]
npv	0.342	9.962	0.023	0.022	0.011	0.012	0.002	0.049 ± 0.007 [141]
		±3.964	±0.010	±0.011	±0.006	±0.006	±0.001	0.650 [29]
MAPE (%)	553.6	2925.5	38.8	40.1	43.6	79.5	54.4	

5.4. Conclusion

In this study, the Molecular Mechanics-Generalized Born/Surface Area (MM-GB/SA) method was used to estimate the free energy of binding, $\Delta G_{binding}$, of 15 PDE4B inhibitors. Since there exists a linear dependency between binding affinity (K_i) and IC_{50}, assuming that Michaelis-Menten constant (K_m), substrate concentrations [S], and experimental conditions are similar, the goal was to obtain a linear correspondence between log(IC_{50}) values and $\Delta G_{binding}$.

The first step of this study was the database preparation with a combination of training and test ligand sets categorized based on their IC_{50} values. As a second step, a molecular docking study was performed. This yielded poor correlations between the docking scores, expressed as ΔG values, and the experimental IC_{50} ones. The results indicated that docking scores are not reliable enough to provide a linear dependency between IC_{50} values and $\Delta G_{binding}$.

After the docking process, 40 independent 1 ns long MD simulations were performed for the all representative poses of each AutoDock cluster. Our results show that, instead of a single long simulation, running multiple independent runs starting from the same structure but with different initial velocities can yield to statistically converged MM-GB/SA free energies of binding.

The binding free energy calculations were repeated for different solvation models: GB^{OBC}, GB^{OBC-2}, GB^{HCT}, GB^{GBneck}, and $GB^{GBneck2}$. The best results were obtained with the GB^{OBC} model, but other GB/SA models, except $GB^{GBneck2}$, lead to similar results. After checking the results according to best docked poses for each inhibitor, the linear trend was improved when all different clusters for each ligand were considered. A linear relationship between estimated IC_{50} vs experimental ones with $R^2 = 0.944$ was achieved. The reliability of our approach was verified with the test set that is here correctly predicted.

Overall, our study indicates that, to obtain a linear dependency between log(IC_{50}) and MM-GB/SA results, it is important to take into account all different poses obtained after a docking process and not the best ones only. Such approach will be used in future studies to serve as benchmark for putative PDE4B ligands when no experimental value is available.

6. CONFORMATIONAL REARRANGEMENTS TRIGGERING GTP HYDROLYSIS IN EF-TU

Wild type and mutant EF-Tu·GTP complexes have been studied via molecular dynamics simulations. In *T. aquaticus*, Switch I explores various conformations, including one where R57 enters the active site like the catalytic arginine in other G-proteins. When R57 is out of the active site, a Na^+ or K^+ ion coordinates to GTP in parts of the simulations. These findings suggest that R57 plays a catalytic role alternately with Na^+/K^+ in the intrinsic GTP hydrolysis. The conformational changes positioning this residue in the active site occur more slowly in *E. coli* with respect to *T. aquaticus* due to the differences in temperature and sequence on Switch I. In addition, pK_a calculations show that an important fraction of H85 is doubly protonated and H85 spends a considerable time in the active site even in the absence of ribosomes. Our simulations also indicate that the T62-γ-phosphate and T62-Mg^{2+} interactions are necessary to maintain the GTP bound conformation of Switch I, whereas D51 is not crucial.

6.1. Introduction

Elongation factor Tu (EF-Tu), responsible of the delivery of the aminoacyl-tRNA to the ribosome exhibits a cognate codon-anticodon pairing dependent GTPase activity [51,52,163,164]. The intrinsic GTP hydrolysis rate of EF-Tu is low [165-167] but can be accelerated by ≈ 6 - 7 orders of magnitude by programmed ribosomes [167,168]. In none of the ribosome bound EF-Tu crystals (e. g. 2XQD/2XQE [53] or 3FIH/3FIK [169]), the ribosome interacts directly with GTP or the hydrolytic water. Instead, the sarcin-ricin loop (SRL) of the ribosomal 23S RNA interacts with either H85 (2XQD and 2XQE) or H19 (3FIH and 3FIK) of EF-Tu (*Thermus aquaticus* amino acid numbering is used unless otherwise stated). Similarly, electron microscopy studies indicated interactions between SRL and EF-Tu, but not between any part of the ribosome and GTP or hydrolytic water [169]. Hence, the GTP hydrolysis is thought to be triggered by a conformational rearrangement of EF-Tu which is in turn caused by a conformational rearrangement of the ribosome and tRNA

[40,51,52,163,164,168-171]. However, proposed conformational changes on EF-Tu are controversial and highly debated.

EF-Tu consists of three domains [51,52,163,164,172]. GTP binds to Domain I (also called the G-domain) while tRNA binds to the cavity between the three domains [173]. Studies of the catalytic residues and conformational changes in the literature, as well as this work, focus on three regions of Domain I: Switch I, Switch II and P-loop.

Switch I in the GDP-bound EF-Tu forms a β-hairpin that extends into the aa·tRNA binding site. GTP binding to EF-Tu causes a local rearrangement: Switch I conformation turns to an α-helix from β-hairpin [49,172,174]. The helix is flanked by D51 and T62. T62 coordinates to a Mg^{2+} ion which in turn, coordinates to the β- and γ-phosphate groups of GTP [172]. The remaining ligands of Mg^{2+} are T25 and two water molecules [172]. The side chain of D51 makes a hydrogen bond with one of these water molecules and another hydrogen bond with T62 [172]. Thus, both ends of the Switch I are connected to each other and to Mg^{2+} (via a water molecule in the case of D51). The backbone amide proton of T62 makes a hydrogen bond to the γ-phosphate group in *T. aquaticus* EF-Tu crystals [172] but not in the *E. coli* EF-Tu crystal (PDB code: 1OB2). The T62S mutant has similar properties as the wild type whereas the alanine mutant has low intrinsic and ribosome induced GTPase activity, and low tRNA affinity [46]. Mutation of D50 in *E. coli* does not alter the GTP hydrolysis rate, but decreases the tRNA affinity. Fluorescence studies indicate that a rearrangement in Domain I, probably involving Switch I, contributes to GTPase activation [40,166,168] although the exact mechanism remains to be explained.

In many G-proteins, a catalytic arginine, such as R178 of $G_{i\alpha1}$ [175] was found to be crucial for GTP hydrolysis. The arginine finger, situated on the GTPase activating protein (GAP), stimulates GTP hydrolysis by Ras [176]. However, previous studies were not successful in identifying an equivalent arginine on EF-Tu. A potential candidate, the conserved R58 of Switch I on *E. coli* EF-Tu (R59 on *T. aquaticus*) was determined to be involved in ribosome binding but not in catalysis [177]. In *T. aquaticus* EF-Tu, there is a second arginine (R57) on Switch I. In *E. coli*, the same position is occupied by a lysine

(K56). A sequence alignment search indicates that at this position there is either a lysine or an arginine in a broad range of species from bacteria and archaea to multicellular eukaryotes, suggesting that the positive charge is conserved. No experimental data exists on R57 in *T. aquaticus*. On the other hand, in *E. coli*, it has been observed that K56 is monomethylated during the growth phase and dimethylated in the stationary phase. In *E. coli*, K56 is monomethylated during the growth phase and dimethylated in the stationary phase and methylation of K56 decreases the rate of tRNA-dependent GTP hydrolysis. Mutations at these positions K56A and K56Q slows GTP hydrolysis (Baş and Bilgin, in preparation). As discussed in the next sections, our study demonstrates that R57 can enter the active site upon a reorientation of Switch I, hence can play a catalytic role. Thus, the movement of Switch I and, in particular, the position of R57 are at the focus of our study.

The stimulation of GTP hydrolysis by programmed ribosomes in translational GTPases needs an invariant histidine (H85) located at the N-terminus of Switch II. This residue is oriented away from the active site in EF-Tu·GTP binary complex crystals [172]. However, in some crystals with kirromycin (PDB code: 1OB2) or ribosome [53], H85 is oriented into the active site. This observation leads to the idea that the stimulation by the ribosome arises from the reorientation of the side chain of H85. H84A mutation in *E. coli* reduced the intrinsic GTPase activity to 10% according to Scarano *et al.* [46] whereas Maracci *et al.* [171] did not observe a significant change in the intrinsic GTPase rate upon this mutation [171]. H84A was completely inactive in poly(Phe) synthesis [46,171] indicating that this residue is crucial for stimulation by the ribosome.

The function of H85 is highly controversial in the literature. Among the proposed functions are general base catalysis [43,53], general acid catalysis [54], positioning of the nucleophilic water [40], positioning of the PGH motif [55] and acting as a conformational switch [56,57]. Wallin *et al.* [35] and Adamczyk *et al.* [57] computed the pK_a of H85 and found that it must be protonated in its active conformation, hence cannot act as a general base.

V20 and I61 were proposed to act as a hydrophobic gate to hinder the entrance of H85 into the active site in the absence of ribosomes. In order to accelerate the GTP hydrolysis in the presence of ribosomes, one or both wings of the hydrophobic gate should open, providing access of the catalytic histidine to the active site. However, neither V20G nor I60A mutations in *E. coli* increased the intrinsic GTPase activity, contradicting the idea of a hydrophobic gate.

The P-loop consists of a GHXDXGKT motif and is involved in GTP binding. D21 was proposed to be involved in catalysis, probably by hydrogen bonding to the β-γ bridging oxygen [178]. Mutation of the corresponding A30 in Rab5 to any amino acid but proline yielded a functional protein, indicating that the backbone amide proton is involved [178]. In addition, the backbone -NH groups of G23, K24 and T25 make hydrogen bonds to the non-bridging oxygens of β-phosphate [172]. The side chain of K24 interacts with both β- and γ-phosphates [172]. Warshel and coworkers suggest that H84 (in *E. coli*) allosterically relocates the P-loop into a catalytic position in the presence of ribosome and the major transition state stabilization arises from the electrostatic effect of the P-loop [56,57].

Counter ions, in particular K^+, are suggested to have important roles in the structural stability and GTPase activity of EF-Tu. Aqvist proposed that a K^+ ion ligated by D50 and E55 (in *E. coli*) was crucial for the local structural integrity of the GTPase catalytic centre [35]. Rodnina and coworkers found that K^+ ions moderately accelarated the GTPase rate in the wild type EF-Tu, but this effect was lost in the presence of programmed ribosomes or in the D21A mutant [171].

Several molecular dynamics studies on EF-Tu exist in the literature. Dynamics of recognition between tRNA and EF-Tu was studied by Eargle *et al.* [179]. EF-Tu·GDP and EF-Tu·tRNA complexes as well as free EF-Tu were considered in the simulation of Kulczycka *et al.* [180]. Wallin *et al.* [35] focused on the positions and protonation states of H84 (in *E. coli*) and nucleophile (H_2O or OH^-) [35]. Warshel and coworkers [56,57] carried out EVB simulations to explore the catalytic roles of H84 and P-loop.

As discussed below, many conformational rearrangements related to the GTP hydrolysis occur even in the absence of programmed ribosomes. In the present study, we investigate these conformational rearrangements through molecular dynamics simulations on EF-Tu·GTP complexes from *E. coli* or *T. aquaticus*. In particular, we focus on i) the repositioning of R57, ii) the Switch I movement controlled by T62 and D51, iii) H85 and hydrophobic gate, and iv) the effect of counter ions.

6.2. Methodology

All simulations have been carried out using the ff03 force field [143,181] as implemented in the Amber11 and Amber14 program packages [182,183]. The GTP parameters have been taken from Meagher *et al.* [184].

The starting structure for the *T. aquaticus* simulations has been taken from the crystallographic structure of EF-Tu complexed with a non hydrolyzable GTP analog (PDB code: 1EFT) [172]. The GTP analog has been replaced by GTP. For the *E. coli* binary complex simulations, the tRNA has been removed from the crystallographic structure of the ternary complex (PDB code: 1OB2) and the GTP analog has been replaced by GTP. For mutant simulations, mutations have been done manually.

All simulations have been performed under periodic boundary conditions, using a truncated octahedron solvent box containing TIP3P water molecules. The distance between the box edge and the closest solute atom has been taken to be at least 10 Å. Na^+ or K^+ ions have been added to neutralize the charge of the simulated systems. Electrostatic interactions have been evaluated using the Particle Mesh Ewald (PME) method [147] with a cut off value of 9 Å. A cut off distance of 9 Å has been applied for van der Waals interactions.

The initial structures have been energy until the root-mean-square force is less than 0.01 kcal/mol·Å. Following the energy minimization, the systems have been simulated in the NPT ensemble. Then, the systems have been heated by using the Langevin thermostat [185] from 10 K to 310 K in the case of *E. coli* and to 343 K in the case of *T. aquaticus* (the

thriving temperature of each species). The pressure has been set to 1 bar. All hydrogen containing bonds have been constrained by applying the SHAKE algorithm [150], allowing a time step of 2 fs.

The pK_a values of Nδ and Nε of H85 in *T. aquaticus* EF-Tu, its D85E and I61A mutants as well as H84 in *E. coli* EF-Tu have been computed using thermodynamic integration simulations [186] as described by Simonson *et al.* [187].

6.3. Results

A total of 15 simulations have been carried out on wild type and mutant *E. coli* and *T. aquaticus* EF-Tu structures. Different protonation states of H85 (H84 in *E. coli*) have been considered. In all *T. aquaticus* and some *E. coli* simulations, the system has been neutralized with Na^+ ions, except some *E. coli* simulations in the presence of K^+ ions. We have designed 3 *E. coli* EF-Tu simulations with different initial positions of K^+ ions: i) all K^+ ions in the solvent, ii) one of the K^+ ions between the side chain of D21, backbone O of G59 and GTP (Site 1), iii) one of the K^+ ions between D50 and E55 (Site 2). We have preferred using *E. coli* EF-Tu and K^+ because experimental data [171,188] exist for this system and ion coordinations are more stable due to the lower temperature. The nomenclature and descriptions of all simulations are given in Table 6.1.

6.3.1. Conformation and pK_a of H85

The pK_a of H85 has been calculated via thermodynamic integration simulations for the wild type, D87E and I61A mutants of *T. aquaticus* EF-Tu as well as the wild type *E. coli* EF-Tu (Table 6.2). In both species Nε is always protonated at physiological pH. Nδ in *T. aquaticus* EF-Tu is mostly unprotonated. Nevertheless, a significant amount of protonated H85 exists (≈ 24 % from Henderson-Hasselbach equation). Depending on the length of the equilibration period, the pK_a of Nδ in *E. coli* EF-Tu is estimated between 7.6 and 8.1, indicating that it is mostly protonated, even though an important fraction of unprotonated H84 exists.

Table 6.1. Nomenclature and descriptions of the simulations. Histidines protonated at the Nδ or Nε positions are denoted as HID or HIE, respectively and a doubly protonated histidine as HIP. All simulations are 200 ns long, except ecHIE343, 300 ns long.

simulation name	***organism***	***temperature (K)***	***counter ion***	***Description***
taHIE	*T. aquaticus*	343	Na^+	HIE85
taHID	*T. aquaticus*	343	Na^+	HID85
taHIP	*T. aquaticus*	343	Na^+	HIP85
taT62A	*T. aquaticus*	343	Na^+	T62A mutant, HIP85
taR59A	*T. aquaticus*	343	Na^+	R59A mutant, HIP85
taR57K	*T. aquaticus*	343	Na^+	R57K mutant, HIE85
taK52N	*T. aquaticus*	343	Na^+	K52N mutant, HIE85
ecHIP	*E. coli*	310	Na^+	HIP84
ecHIE	*E. coli*	310	Na^+	HIE84
ecHIE343	*E. coli*	343	Na^+	HIE84 at 343 K
ecHIPK	*E. coli*	310	K^+	K^+ instead of Na^+, HIP84
ecHIPK1	*E. coli*	310	K^+	K^+ initially at Site 1, HIP84
ecHIPK2	*E. coli*	310	K^+	K^+ initially at Site 2, HIP84
ecK56R	*E. coli*	310	Na^+	K56R mutant, HIE84
ecN51K	*E. coli*	310	Na^+	N51K mutant, HIE84

In the D87E mutant, the increase in the side chain length of residue at position 87 allows the carboxylic acid group to coordinate H85. The presence of a nearby negative charge results in an elevated pK_a value for both Nδ and Nε. This mutant is known to mimic the effect of the unprogrammed ribosome [177]. On the other hand, the I61A mutation lowers the pK_a of both nitrogen atoms of the imidazole ring by ≈ 1 pK_a unit. Probably, due to the larger side chain of isoleucine in the wild type, a more hydrophobic environment between H85 and GTP favors the electrostatic interactions and yields a higher pK_a than in the I61A mutant. Hence, the role of the "hydrophobic gate" [165] seems to set the pK_a of H85, rather than controlling its entrance into the active site. Several groups including Wallin *et al.* [35], Adamczyk and Warshel [57], Aleksandrov and Field [54] found the pK_a of H85 as 9, 11 and 17, respectively.

In order to understand whether the conformational behavior of H85/H84 depends on its protonation states, simulations have been performed with three different protonation states of this histidine. From Figure 6.1, it can be seen that in the doubly protonated (taHIP,

ecHIP, ecHIPK1) and Nδ-protonated (taHID) forms, H85 spends ≈ 50%, 68%, 75% and 10% of its time in the active site, respectively. In the Nε-protonated form (taHIE, ecHIE), H85/H84 is almost always oriented towards the solvent. The unprotonated Nδ of its side chain makes hydrogen bonds to the backbone amide protons of some or all of A86/A85, D87/D86 and Y88/Y87 (Figure 6.2).

Table 6.2. pK_a values of Nδ and Nε of H85/H84 in *T. aquaticus*/*E.coli*.

	Nδ	**Nε**
T. aquaticus		
Wild type	6.9	9.1
D87E	7.9 (7.9)*	10.7 (9.8)*
I61A	6.1	7.9
E. coli		
Wild type	**7.6 (8.1)****	**9.8**

* no part of the simulations is omitted as equilibration period.
** no part of the λ = 0.11270 simulation is omitted as equilibration period.

In taHID, hydrogen bonds between Nε and these backbone protons are very rare. Also, the hydrogen bond lengths and angles are less favorable. This fact is likely to be the origin of the pK_a difference of ≈ 2 units between Nδ and Nε. When the doubly protonated H85/H84 side chain is rotated towards the active site, either Nδ or Nε can face the active site. If Nδ faces the active site, Nδ-H makes a hydrogen bond with the nucleophilic water molecule (Figure 6.2). Such a hydrogen bond is not possible when Nε faces the active site. In taHIP, ecHIP and ecHIPK1, the former situation has been mostly observed (Figure 6.2). In ecHIPK, H84 is always in the active site but Nδ faces the solvent whereas Nε-H makes a hydrogen bond with GTP (Figure 6.2). Such a hydrogen bond is rare in ecHIP and ecHIPK1, absent in taHIP. In ecHIPK1 where a K^+ ion is always at Site 1, H84 spends 75% of its time in the active site. In addition, in some parts of ecHIP, ecHIPK and taHIP, the H84/H85 and a Na^+ or K^+ ion have been simultaneously found in the active site (Figure 6.2). Hence, their presence in the active site is not mutually exclusive.

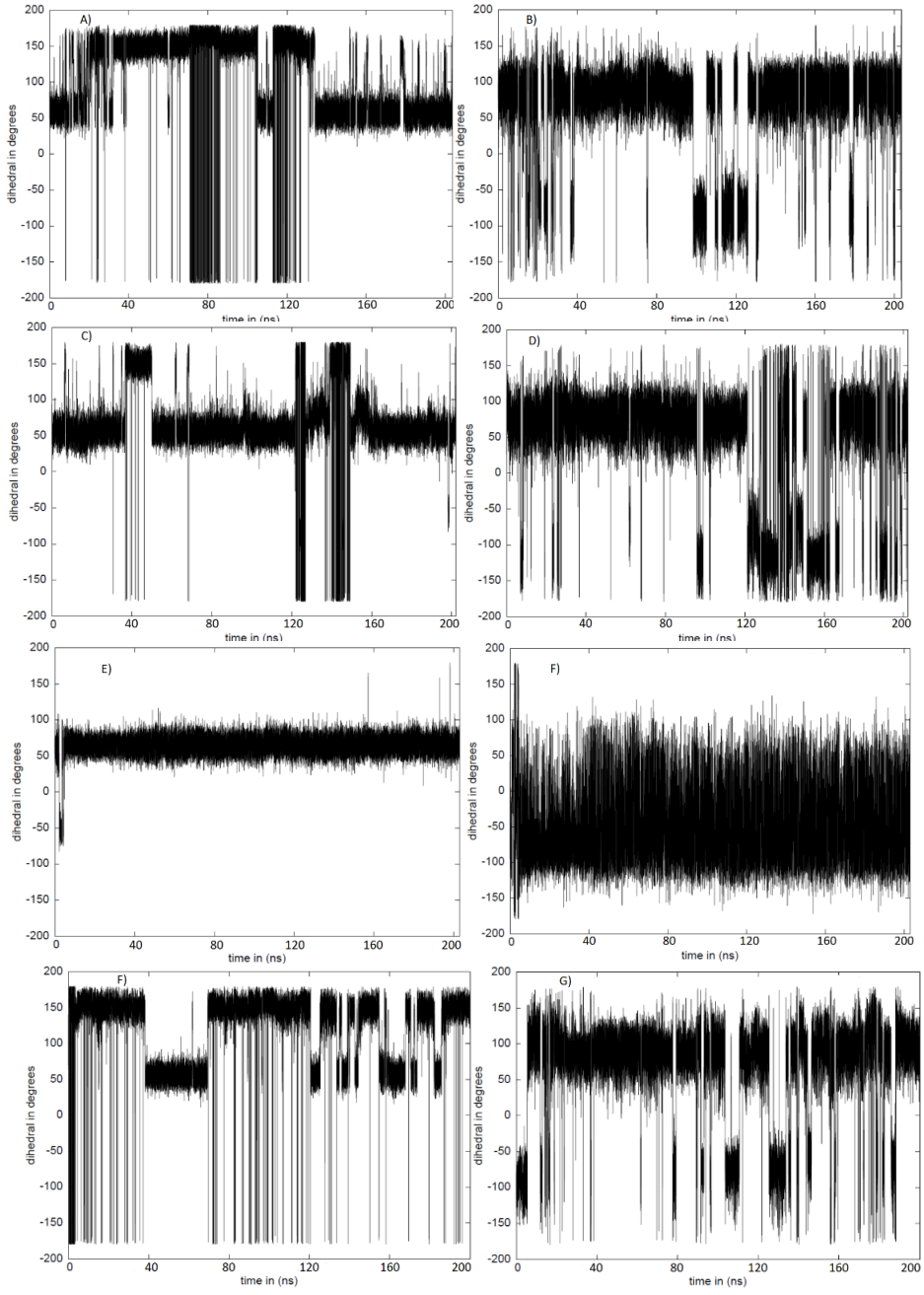


Figure 6.1. Conformations of the H85/H84 side chain. Positive or negative values of the Cα-Cβ-Cγ-Nδ dihedral correspond to conformations where Nδ or Nε is oriented towards the active site, respectively. A and B) taHIP, C and D) taHID, E and F) taHIE, G) and H) ecHIP.

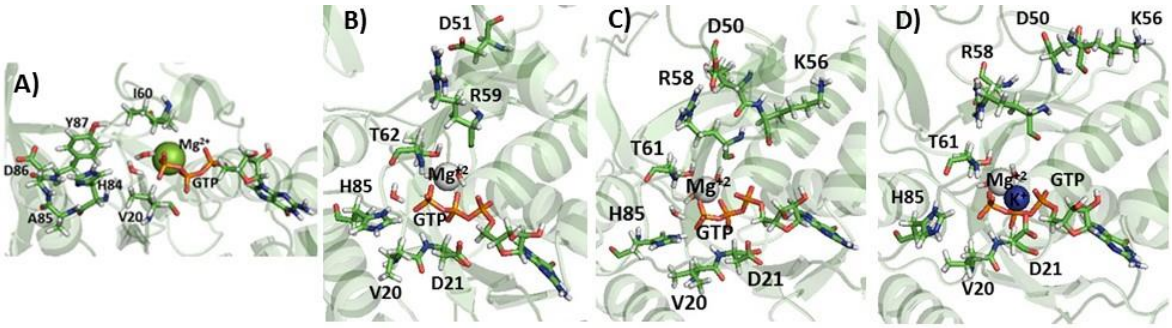


Figure 6.2. Representative views of conformations where H85/H84 side chain is A) away from the active site and the unprotonated Nδ makes hydrogen bonds to A86, D87 and Y88 backbone amide groups, B) in the active site and Nδ-H makes a hydrogen bond with the nucleophilic water, C) in the active site and Nε-H makes a hydrogen bond with GTP, D) in the active site together with a Na^+ ion.

It is not possible to deduce the exact role of H85 from the present results. In addition, our simulations may not be completely equilibrated, therefore our data may be only semiquantitative. Nevertheless, some remarks can be made. Since in *T. aquaticus* EF-Tu, H85 is protonated for 24% of the time and spends ≈ 50% of its time in the active site when protonated, the overall time spent in the active site can be estimated as 12%. In *E. coli*, H84 is mostly protonated and remains in the active site 68-75% of the time when protonated. In the literature, the stimulating effect of the ribosome is suggested to be mediated by the reorientation of the histidine towards the active site. In the presence of programmed ribosomes, the wild type *E. coli* EF-Tu hydrolyzes GTP 10^6-fold faster than the H84A mutant [171]. However, according to our results, since H85 is situated in the active site 12% of the time in the absence of ribosome, the stimulating effect of the latter would be only ≈ 8-fold in *T. aquaticus*, and even less in *E. coli*, if this reorientation was the only effect of the ribosome. Moreover, the H84A mutation should have a bigger impact on the intrinsic GTPase rate than experimentally observed [171,176]. Since our pK_a calculations on both *T. aquaticus* and *E. coli* EF-Tu consistently yield an elevated pK_a for the histidine, we do not expect a high error from sampling issues. In addition, all simulations with a doubly protonated histidine consistently indicate that this residue stays long time in the active site. Therefore, the difference between the ribosome induced 10^6-fold rate acceleration mediated

by H84 found experimentally and the much smaller effect of the reorientation of the histidine predicted here cannot be attributed to computational errors. It must be noted, however, that the estimation of a ≈ 8-fold rate acceleration assumes that even the neutral form of histidine is always in the active site in the presence of the ribosome. If only the protonated form of H85 is catalytically competent, the function of the ribosome must be more subtle, as it must increase the pK_a of this residue and reposition it at the same time. The D87E mutation, which is supposed to mimic the effect of unprogrammed ribosomes [46], increases the pK_a. But, still the pK_a of Nδ is close to the physiological pH and a mixture of protonated and neutral forms is expected. The neutral form is stabilized by the unprotonated Nδ and the amide protons of A86, D87 and Y88. The experimental structures in the presence of the programmed ribosome show that access to these protons is prevented by the D87 side chain (PDB codes: 2XQD, 3FIH), presumably because of a small shift in the tRNA position with respect to the ternary complex (PDB codes: 1B23, 1OB2). This fact is expected to increase the pK_a further. When H85 is protonated, Nε is usually oriented towards the active site in the D87E mutant. However, in the experimental ribosome bound structures Nε interacts with the ribosome. Thus, at least part of the effect of the programmed ribosome may be to make H85 fully protonated and orient Nδ towards the active site. Nevertheless, this effect is unlikely to explain completely the experimentally observed stimulatory effect of the ribosome via histidine, especially in *E. coli* where H84 is mostly protonated and Nδ is in the active site.

One may figure out that the ribosome not only reorients the histidine but also locates it at a deeper position than can be reached in the binary complex. Indeed, in the crystal structure 2XQD, the distance between Nδ of the histidine and Pγ of GTP is 4.59 Å, but in taHIP, ecHIP, ecHIPK and ecHIPK1 simulations, this distance is below 4.6 Å only 0.3%, 0.7%, 0% and 0.2% of the time, respectively.

Warshel and coworkers state that H84 is a switch that repositions the P-loop to trigger the catalysis [54,56]. However, the distances between the β-phosphate oxygen and H22, G23, K24, T25 amide protons as well as the distance between the β-γ-bridging oxygen and the amide proton of D21 exhibit no correlation with the orientation of H85 in taHIP (Figure

6.3-6.4). Moreover, different protonation states of H85 do not have significant effects on these distances. However, it is possible that the ribosome itself positions the P-loop directly; and H85, in turn, positions the ribosome. Since our calculations do not consider the ribosome, this point cannot be addressed here. Aqvist and coworkers suggested that, in *E. coli*, the orientation of H84 towards the active site brings the backbone -NH groups of H84 and G83 in a catalytic position where they stabilize the nucleophilic OH^- ion [55]. However, the distances between the backbone protons of these residues and the Pγ atom of GTP in taHIP, ecHIPK, ecHIPK1 and ecHIPK2, i.e. simulations where the histidine moves in and out of the active site, reveal no correlation with the orientation of the histidine.

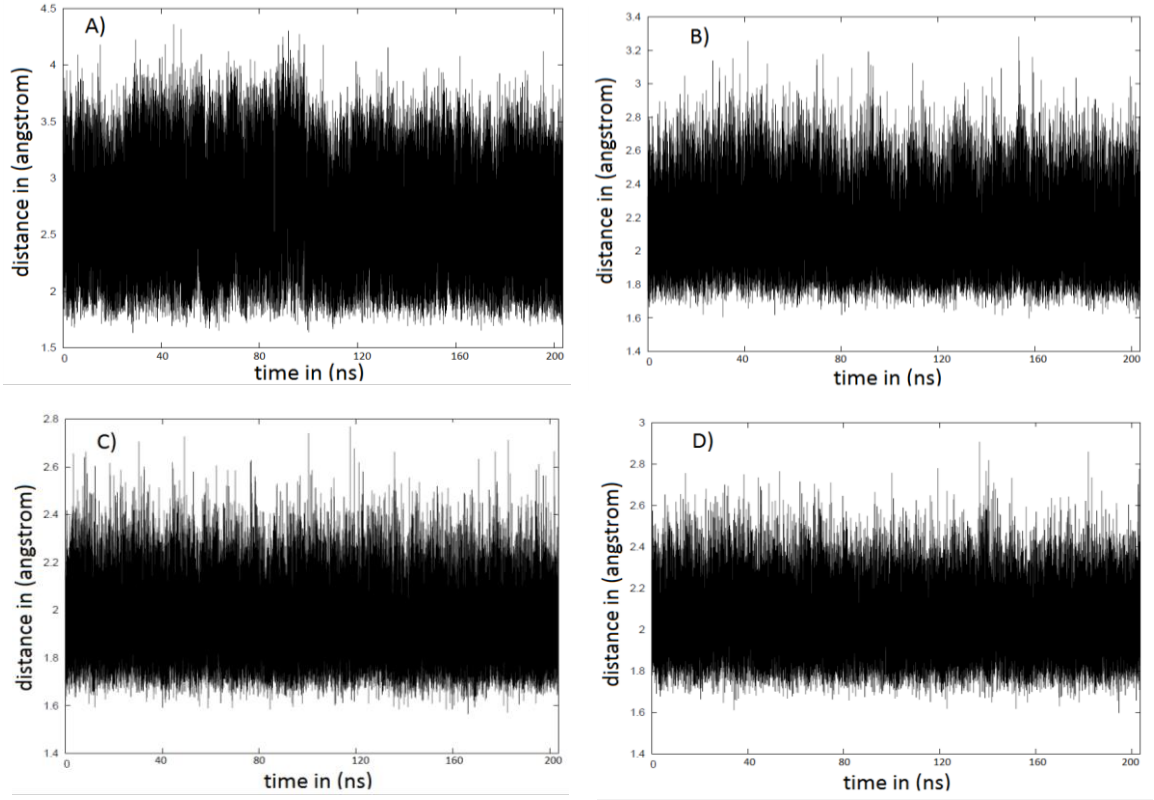


Figure 6.3. The distances between the β-phosphate oxygen and H22, G23, K24, T25 amide protons (A-D) in taHIP.

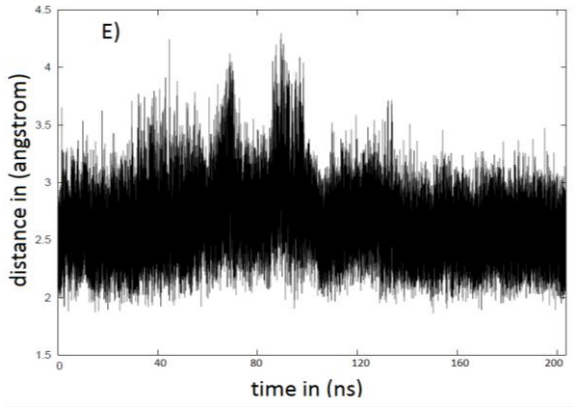


Figure 6.4. The distance between the β-γ-bridging oxygen and the amide proton of D21 in taHIP.

6.3.2. Switch I Movement and Positioning of R57 into the Active Site

In most simulations, Switch I has been found to be very mobile, reorienting away from Domain II towards the GTP binding site by rotating along a virtual axis passing through I61 and T62. This motion is depicted in Figures 6.5-6.8 for taHIP, taHIE, taHID and ecHIP, respectively. The deviation from the crystallographic position of Switch I is especially large in taHID where it becomes almost parallel to GTP. However, all sampled conformations have very short life times and Switch I continuously oscillates between different positions.

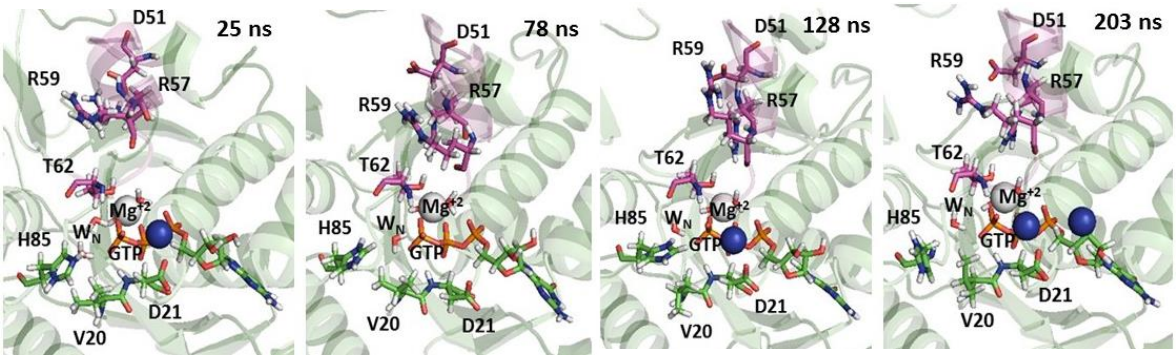


Figure 6.5. Snapshots illustrating the course of the taHIP simulations.

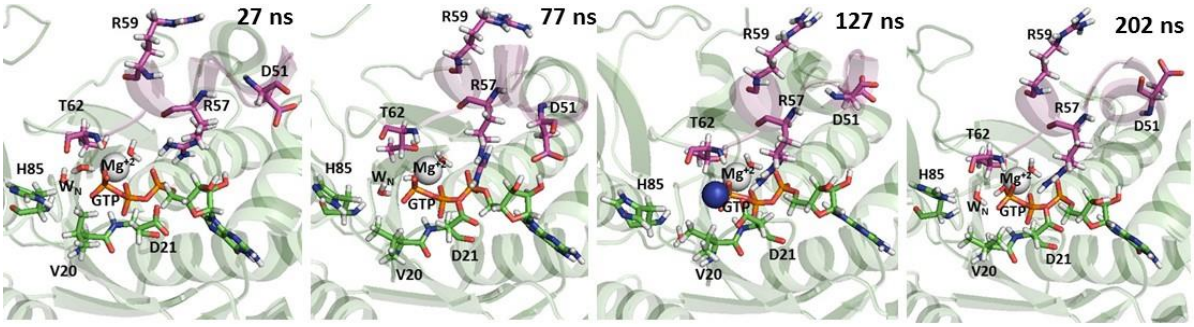


Figure 6.6. Snapshots illustrating the course of the taHIE simulation

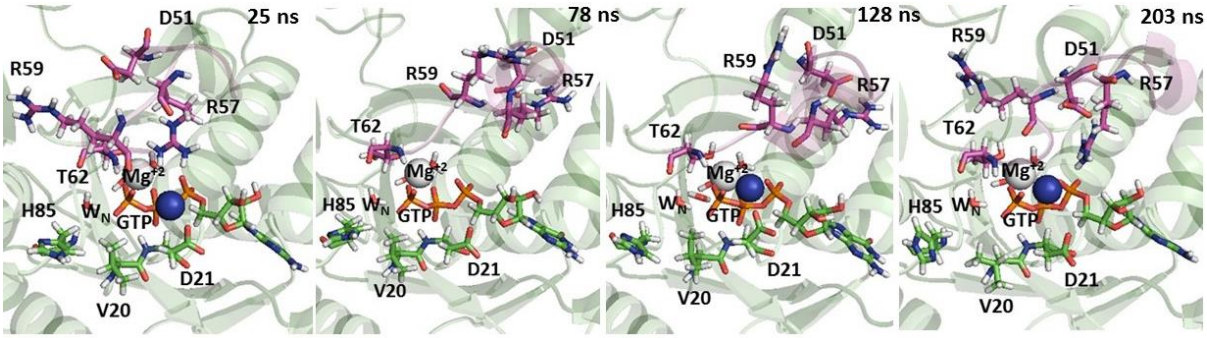


Figure 6.7. Snapshots illustrating the course of the taHID simulation.

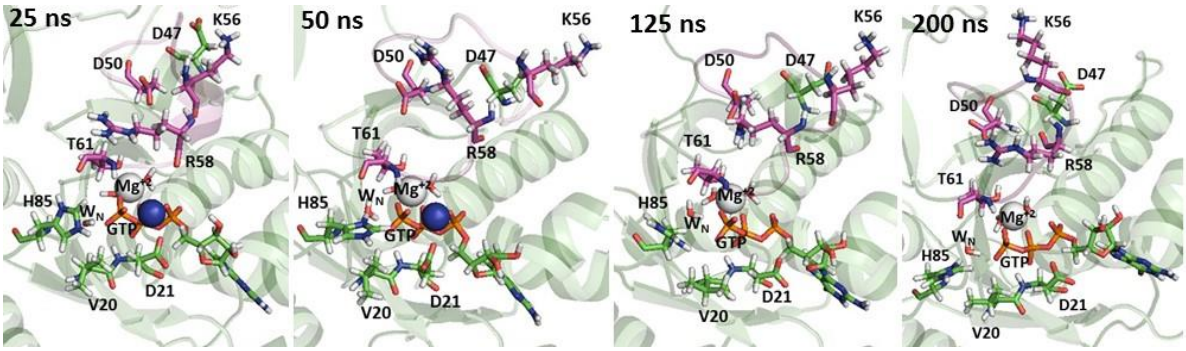


Figure 6.8. Snapshots illustrating the course of the ecHIP simulation.

Figure 6.9 shows the time evolution of the distance between Cζ of R57 and the β-γ-bridging oxygen of GTP in taHIP, taHIE and taHID. In taHIP, R57 never comes close to GTP (Figure 6.9). In the first 9 nanoseconds of taHIE, R57 makes a hydrogen bond to the γ-phosphate, then it loses contact with GTP. In the last ≈ 35 ns of the simulation, it approaches GTP again and makes a hydrogen bond with the α-phosphate (Figure 6.9). During the last ≈ 120 ns of taHID, R57 interacts with the γ- and α-phosphates as well as with the β-γ-bridging oxygen (Figure 6.9). The positioning of R57 in the active site evokes a catalytic role similar as the arginine finger in the Ras-GAP complex or the catalytic arginine in heterotrimeric G-proteins.

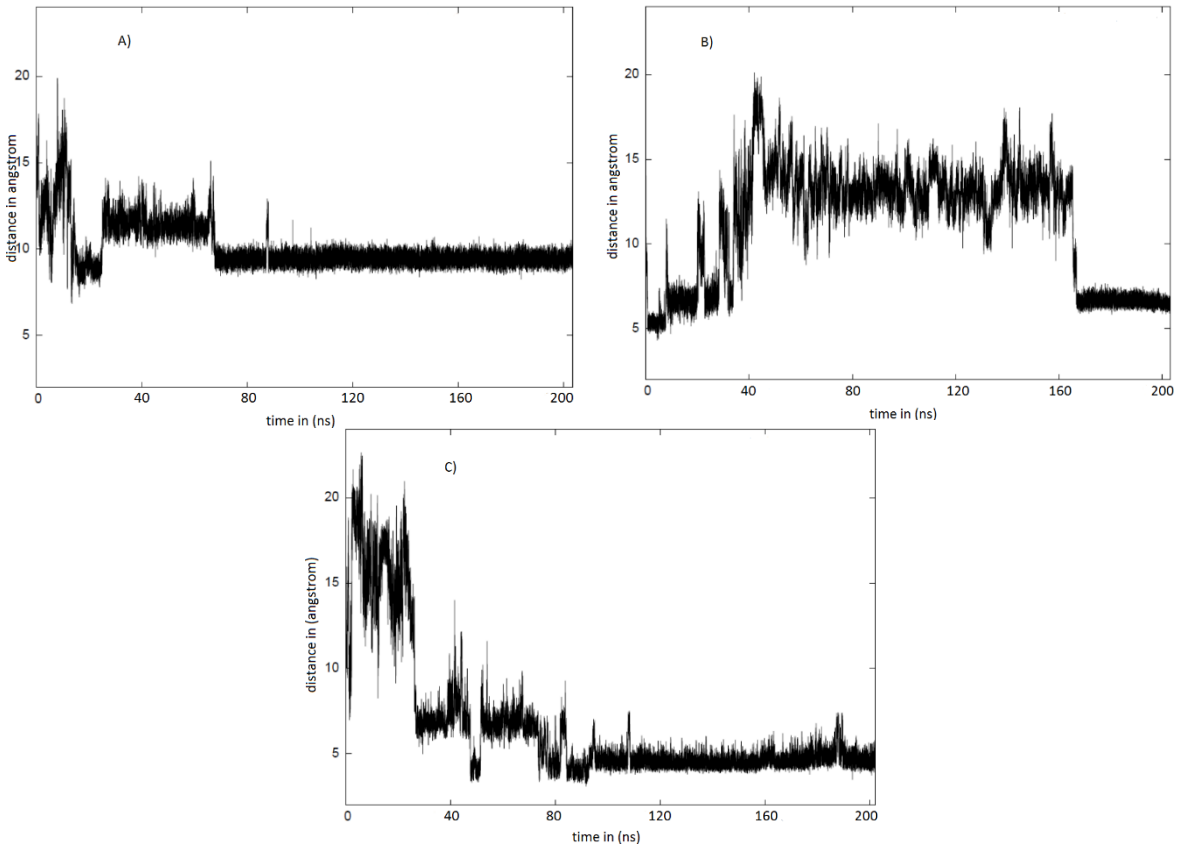


Figure 6.9. The time evolution of the distance between Cζ of R57 and the β-γ-bridging oxygen of GTP in taHIP (A), taHIE (B) and taHID (C), respectively.

Since R57 enters the active site even in the absence of ribosome, it is likely that it contributes to the intrinsic GTP hydrolysis. However, it is not situated there throughout all the simulations. One may assume that the interaction with the ribosome would anchor Switch

I in an orientation suitable to locate R57 in a catalytic position permanently. In none of the experimentally determined ribosome bound EF-Tu structures, an interaction between R57 and the nucleotide is observed. But, in one of them (3FIH), there is a GDP in the active site. In another one (2XQD), GTP is replaced by its analog containing a methylene group instead of the β-γ-bridging oxygen. In both situations, the interaction between R57 and the nucleotide is expected to be weak. Consequently, R57 is situated away from the active site, as opposed to our simulations.

As Switch I and R57 do not behave identically in taHIP, taHID and taHIE simulations, one may ask whether their movement is coupled with the protonation state of H85. An electrostatic interaction between H85 and charged residues on Switch I is unlikely to have a significant effect because they are situated quite far from each other, separated by high dielectric solvent environment. On the other hand, when H85 is rotated towards the active site, it makes a hydrophobic contact with I61 which, in turn, might affect the position of Switch I. However, we have not observed any correlation between the positions of H85 and Switch I. For instance, in taHIP, H85 moves in and out of the active site but R57 does not come close to GTP. In taHIE, H85 is almost always in the solvent while Switch I is very mobile and R57 enters the active site. In taHID, at the onset of the entrance of R57 into the active site, H85 is oriented into the solvent. But later, when it rotates into the active site, R57 still remains there. Thus, it seems that the difference in the behavior of Switch I and R57 in these three simulations is simply coincidental. Switch I oscillates independently of the protonation state of H85. During these oscillations, R57 becomes trapped in the negatively charged regions for a while (between D51 and D81 in taHIP, Figure 6.10: A-B; between D51 and E180 in taHID, Figure 6.10C-D and taHIE, Figure 6.11E-F; near GTP in taHID and taHIE, Figures 6.9-B and 6.9-C). Of course, the difference between the three simulations also indicate that none of them is completely equilibrated. Nevertheless, the fact that R57 can spend a considerable time in the active site provides evidence that this residue can be the equivalent of the catalytic arginine in other G-proteins.

In the simulations on *E. coli* EF-Tu, K56 (corresponding to R57 in *T. aquaticus*) has not entered the active site and stayed mostly in the solvent in both ecHIP and ecHIE. From

time to time, it makes a salt bridge with D47 or hydrogen bonds with backbone carbonyl groups of residues 47-52. In addition, the deviation of Switch I helix from its crystallographic position is modest as in taHIP (Figures 6.5 and 6.8). To test the stability of a configuration with K56 in the active site, we have placed this residue next to GTP and carried out a 25 ns simulation with its position restrained. Then, we have removed the restraint and simulated for 175 ns. K56 remained in the active site. Thus, this conformation is not unstable but the forward and reverse conformational barriers are high to be crossed within our simulation time at 310 K.

There are several differences between *E. coli* and *T. aquaticus* EF-Tu that may explain their different behaviors in our simulations. First, because of the lower temperature in *E. coli* simulations with respect to *T. aquaticus* (310 and 343 K, respectively) conformational fluctuations may take place more slowly and our simulation may not be long enough to observe the entrance of K56 into the active site. To test this hypothesis, we have carried out a new simulation at 343 K (ecHIE343) for 300 ns. K56 has come near the α-phosphate to make a water mediated interaction although it has not entered the active site (Figure 6.10).

Secondly, there are variations at two positions on Switch I. The equivalent of R57 and K52 of *T. aquaticus* EF-Tu are K56 and N51 in *E. coli*, respectively. Lysine is slightly shorter than arginine and more flexible. It may be more difficult for lysine to reach the active site. Instead, it spends more time near more accessible negatively charged groups such as D47. We have performed R57K and K56R mutant simulations for *T. aquaticus* and *E. coli* EF-Tu, respectively (taR57K and ecK56R). In both cases, the arginine/lysine has made a contact with the α-phosphate, but has not moved further towards the reaction center. The taR57K simulation has been performed at 343 K. Hence, even though R57 has been substituted by lysine, the high temperature has provided sufficient energy to bring this residue next to the α-phosphate. On the other hand, though ecK56R has been carried out at 310 K, the arginine that substitutes lysine has been able to come next to GTP. We have also performed K52N and N51K mutant simulations for *T. aquaticus* and *E. coli* EF-Tu, respectively (taK52N and ecN51K). In part of the wild type simulations, K52 in *T. aquaticus* interacts electrostatically with nearby negatively charged groups such as D49.

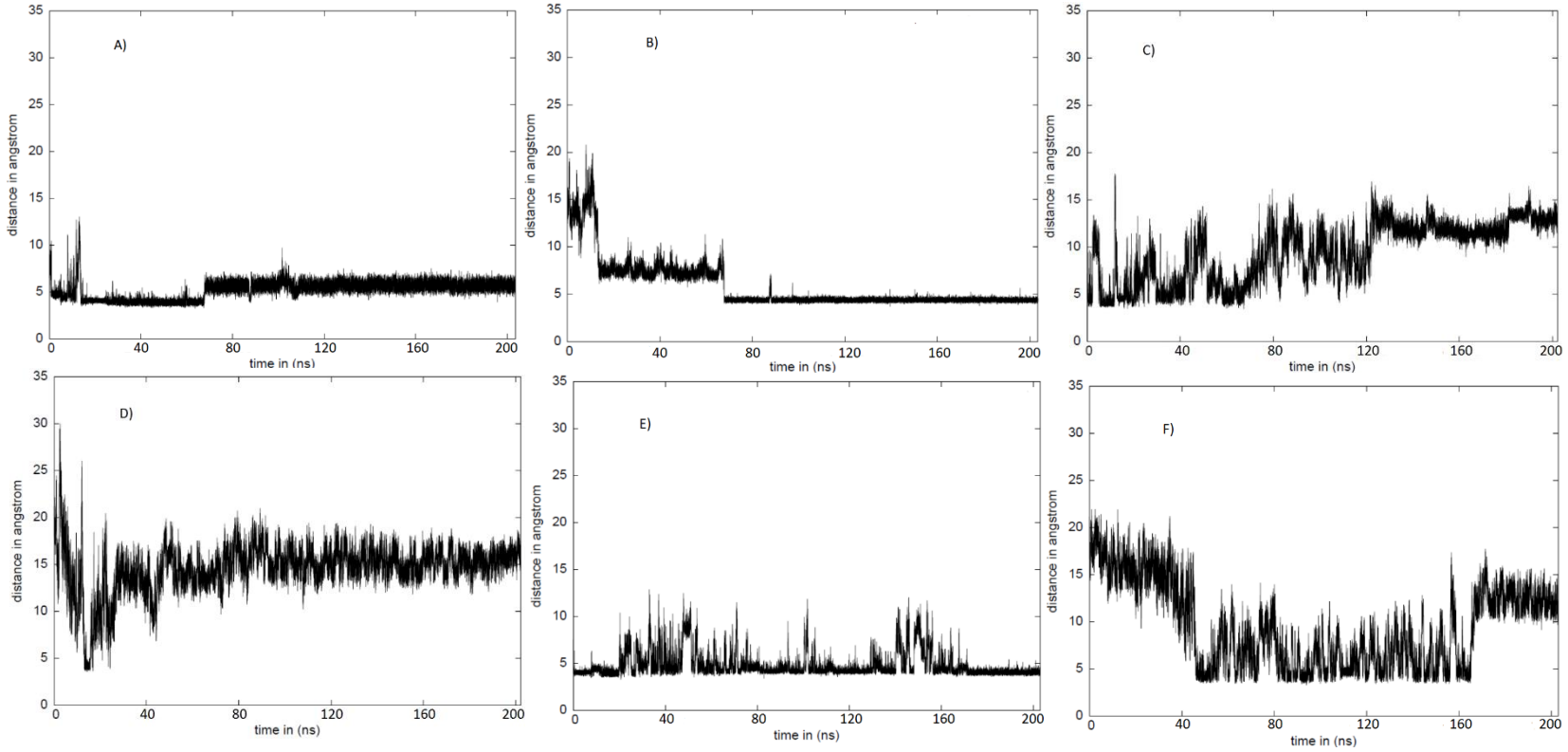


Figure 6.10. The distance R57 (Cζ) - D51 (Cγ), R57 (Cζ) - D81 (Cγ) in taHIP (A-B); R57 (Cζ) - D51 (Cγ),R57 (Cζ) - E180 (Cε) in taHID (C-D) and taHIE (E-F), respectively.

This interaction may serve as a built-in ionic strength around these groups, preventing R57 to be trapped by their electrostatic effects. In none of taK52N or ecN51K, the R57/K56 has not entered the active site. Though the exact effect of K52/N51 cannot be deduced, it seems that K52 in *T. aquaticus* facilitates the movement of Switch I, but solely its effect is not enough to bring R57 next to GTP. Apparently, the differences in the temperature and in the sequences of *T. aquaticus* and *E. coli* Switch I account for the discrepancy in the behavior of EF-Tu from the two organisms. Because of these differences, conformational rearrangements involving Switch I and R57 in *T. aquaticus* occur fast enough to be observed in our simulations, whereas they require a longer time in *E. coli*. Representative snapshots from taR57K, ecK56R, taK52N and ecN51K are given in Figure 6.12.

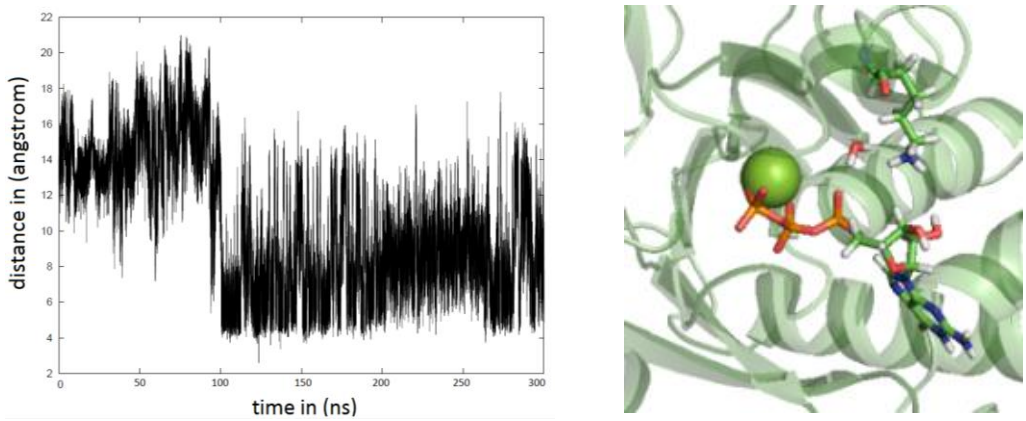


Figure 6.11. The distance between K56 and α-phosphate of GTP in ecHIE343 during 300 ns (left) and a representative view of K56 and GTP when they are involved in a water mediated interaction (right).

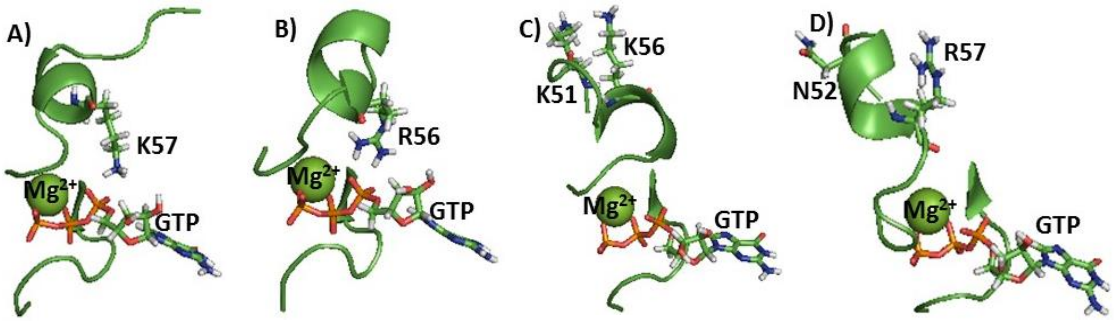


Figure 6.12. Representative snapshots from A) taR57K, B) ecK56R, C) ecN51K and D) taK52N simulations.

During the entire ecHIPK1, 71% of ecHIE and approximately half of taHIP, taHID, taHIE and ecHIP, the putative position of lysine in the active site (Site 1) is occupied by a Na^+ or K^+ ion. With its positive charge, this cation may substitute K56/R57. While K^+ always makes direct contacts with GTP, because of its smaller size Na^+ is more mobile in the active site and displays either direct or water mediated interactions with the nucleotide interchangeably. In taHID where R57 is mostly in the active site, Na^+ shifts towards the γ-phosphate rather than being next to the β-γ-bridging oxygen. A K^+ ion at Site 1 was observed experimentally [171,188]. It was shown to have 3-fold catalytic effect on the intrinsic GTP hydrolysis, but no effect in the presence of ribosomes [171]. We hypothesize that R57/K56 and Na^+/K^+ occupy the active site alternately during the intrinsic hydrolysis, whereas the ribosome anchors the catalytic residue next to GTP permanently, dispensing with the need for Na^+/K^+. It is also possible that R57/K56 cooperates with Na^+/K^+ in intrinsic hydrolysis, at least from time to time, as they can be simultaneously situated in the active site in taHID. On the other hand, simultaneous presence of H85 and R57 in the active site in the presence of ribosome would leave no room for a Na^+ or K^+ ion, in agreement with the experimentally observed K^+ independence of the latter reaction.

6.3.3. Interactions between Switch I and GTP

In all simulations, Mg^{2+} maintains its coordination shell that includes T25, T62 and two water molecules. However, D51 (which makes hydrogen bonds to T62 and a water coordinated to Mg^{2+} in the crystals) is very mobile, scanning various positions shown in Figures 6.6-6.9. Notably, in some parts of taHIP, D51 makes a salt bridge with R57 near the tRNA binding site. In most part of taHID, it makes hydrogen bonds with 2' and 3' –OH groups of the ribose moiety of GTP. In some parts of taHIE, it makes a salt bridge with R57 which is mostly in the active site in contact with GTP. In ecHIPK2 and part of ecHIPK, D50 (in *E. coli*) coordinates to a K^+ ion. The mobility of this aspartate arises probably from the electrostatic repulsion between its side chain and GTP. In ecHIPK2 where a K^+ ion has been placed initially at Site 2, between D50 and E55, this ion has conserved its position during the entire simulation. D50 and Switch I have remained at their positions in the crystal structure (Figure 6.13), presumably because this ion screens the electrostatic repulsion

between D50 and GTP. The presence of a K^+ ion at Site 2 was previously suggested by Wallin *et al.* [35] to be essential for the stability of the crystallographic structure. On the other hand, our catalytic R57 hypothesis requires that Switch I oscillates between various conformations. Thus, we have checked the occupancy of Site 2 and the position of D50 in various simulations.

When we have increased the simulation temperature of ecHIPK2 to 343 K in order to accelerate conformational transitions, K^+ has left Site 2 and D50 has moved away from its position in the crystal structures within 20 ns. In ecHIE343, even though D50 has moved away from its crystallographic position, it has remained near E55 and captured a Na^+ ion. Afterwards, it has returned to its crystallographic position for a short time. Then, both D50 and Na^+ have left this position. In ecHIPK, ecK56R, taK52N and taR57K, after moving away from its crystallographic position, D50/D51, together with E55/E56, has coordinated a cation for a few nanoseconds (Figure 6.13). But, this has not brought it back to its initial position. Later, the cation has gone away. In the rest of the simulations, Site 2 is not occupied. Thus, it seems that the presence of a Na^+ or K^+ ion at Site 2 stabilizes Switch I at its crystallographic position. But the ion does not stay there permanently, especially at 343 K, allowing Switch I oscillate between different conformations. In ecHIPK2, simultaneous occupation of both ion binding sites has been observed during 7% of the simulation (Figure 6.13).

In all crystallographic structures of *T. aquaticus* EF-Tu, the backbone amide hydrogen of T62 makes a hydrogen bond with the γ-phosphate group of GTP (Figure 6.14). However, in all *E. coli* EF-Tu crystals, the amide proton of the corresponding threonine (T61) is oriented away from GTP and none of the Switch I residues is involved in a hydrogen bond with the nucleotide. In all our *T. aquaticus* simulations, the interaction between T61 and GTP has remained stable. Interestingly however, in *E. coli* simulations ecHIP, ecHIE and ecHIPK2, the amide proton of T61 has rotated towards GTP, like in *T. aquaticus*. In ecHIPK, the amide proton of I60 has been reoriented to interact with the γ-phosphate (Figure 6.14).

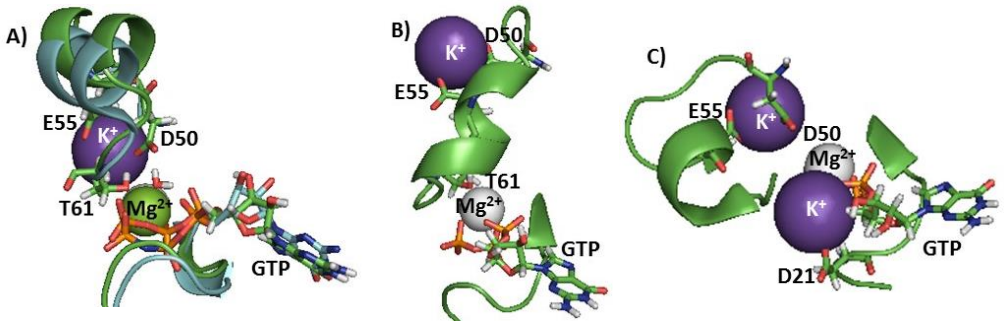


Figure 6.13. A) A snapshot from ecHIPK2 superimposed to the crystal structure 1OB2; B) a snapshot from ecHIPK where a K^+ ion is bound to D50 and E55 after D50 leaves its crystallographic position; C) a snapshot from ecHIPK2 showing that K^+ ions can bind to Site 1 and Site 2 simultaneously.

We have carried out an additional simulation using the same initial geometry as ecHIP, but with different initial velocities. In this simulation, neither a hydrogen bond between Switch I and GTP, nor a cation in the active site has been observed. The side chain of T61 has lost its coordination with Mg^{2+}. One of the water molecules in the coordination shell of Mg^{2+} has shifted to replace T61 while the α-phosphate group has filled the position of this water molecule (Figure 6.14). Thus, it appears that the lack of a cation in the active site or a hydrogen bond between Switch I and GTP destabilizes the active site organization. The reason why the hydrogen bond between Switch I and GTP is seen in *T. aquaticus* EF-Tu crystals, but not in *E. coli* crystals remains to be investigated.

We have carried out a simulation on the T62A mutant. In this simulation, the amide proton of the alanine that replaces threonine has lost the hydrogen bond with GTP. Like in the other simulations, D51 has left its crystallographic position. Having lost all its interactions with GTP and Mg^{2+}, Switch I has moved towards its position in EF-Tu·GDP complexes (i.e. towards the tRNA binding site) while keeping its helical secondary structure. Two configurations have been mainly sampled: one where R57 makes hydrogen bonds to the γ-phosphate oxygens and one where Switch I is farther away from GTP (Figure 6.14). The mobility of Switch I towards the tRNA binding site can account for the slow hydrolysis

rate and low tRNA affinity in this mutant [165]. Also, I61 and V20 are too far apart to constitute a hydrophobic gate.

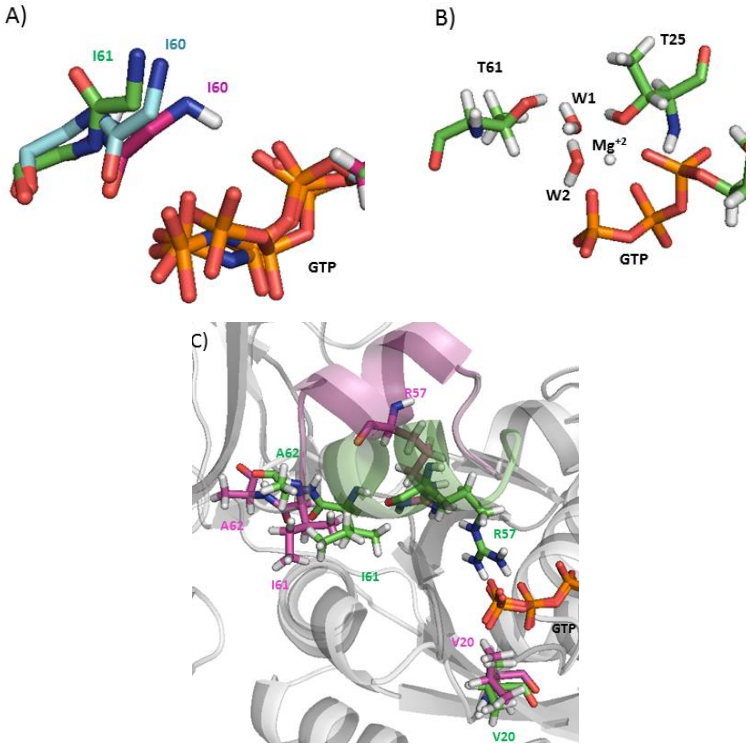


Figure 6.14. A) Interactions between I60/61 and GTP, green: the crystal structure of the *T. aquaticus* EF-Tu (1EFT), cyan: the crystal structure of the *E. coli* EF-Tu (1OB2), pink: a snapshot from the ecHIPK simulation; B) a snapshot from the second ecHIP simulation after the break down of the coordination shell of Mg^{2+}; C) two representative snapshots from the taT62A simulation, green: R57 makes hydrogen bonds to GTP, pink: R57 and Switch I are away from GTP.

There are three interactions that anchor Switch I to the active site: water mediated interaction between D51 and Mg^{2+}, coordination between T62 side chain and Mg^{2+}, hydrogen bond between T62 backbone and GTP. D51 is very mobile in most simulations, nevertheless Switch I remains attached to the active site through T62, although it oscillates between various positions mentioned above and depicted in Figures 6.5-6.8. However, in simulation periods where D51 retains its crystallographic interactions (especially when

coordinated to a cation, as discussed above), Switch I oscillates less and remains nearly at its crystallographic positon. These observations suggest that the interaction involving D51 is not crucial for anchoring Switch I to the active site but reduces its mobility. On the other hand, both in *T. aquaticus* T62A mutant and *E. coli* EF-Tu where the T61 backbone amide proton is oriented away from GTP, Switch I moves towards its position in EF-Tu·GDP crystals. Hence, both the coordination of T62 side chain with Mg^{2+} and the hydrogen bond between T62 backbone and GTP are necessary to anchor Switch I. Interestingly, the only nucleotide-EF-Tu interaction removed after GTP hydrolysis is the T62-γ-phosphate hydrogen bond. It seems that the loss of this hydrogen bond due to hydrolysis, leads to the loss of the T62-Mg^{2+} coordination. As a result, Switch I moves towards the tRNA binding site.

6.4. Conclusion

In this study, molecular dynamics simulations on the wild type and mutant EF-Tu proteins from *T. aquaticus* and *E. coli*, in their GTP bound form, have been performed. In some *T. aquaticus* EF-Tu simulations, a subtle movement of Switch I repositions R57 in the active site, possibly giving rise to a catalytic function. However, in *E. coli* EF-Tu simulations, the equivalent residue K56 has not entered the active site. Based on the following observations, we suggest that the entrance of R57 into the active site is not coincidental, but catalytically important:

In T. aquaticus EF-Tu·GTP binary complexes, at least in some simulations, R57 can take a catalytic position, similar to that in other G-proteins. Thus, it should accelerate, to some extent, the intrinsic hydrolysis in this organism.

Our results are consistent with the experimental finding that the movement of Switch I is involved in GTPase activation.

In many parts of the simulations, a Na^+ or K^+ ion is found at the catalytic position instead of R57/K56, suggesting that Na^+/K^+ can substitute this residue by providing a

positive charge. Moreover, R57 and Na^+/K^+ can be situated in the active site simultaneously, with a shift of Na^+/K^+ towards the γ-phosphate, suggesting that they can cooperate. This is consistent with experiments indicating that K^+ ions play a catalytic role in the intrinsic hydrolysis.

A Na^+ or K^+ ion can move in and out of Site 2, allowing Switch I oscillate between different conformations, especially at 343 K.

When *E. coli* EF-Tu is simulated at the thriving temperature of *T. aquaticus*, K56 approaches the active site. Similarly, when K56 in *E. coli* EF-Tu is mutated to arginine, the equivalent residue in *T. aquaticus*, this residue approaches the active site. Thus, conformational transitions in *T. aquaticus* occur faster than in *E. coli* because of differences in the sequence and thriving temperature. We presume that in longer simulations, the entrance of K56 into the active site would be observed.

We have computed the pK_a of H85 (H84) in both *T. aquaticus* and *E. coli* and investigated its conformational behavior. We have shown that, at physiological pH, an important fraction of this residue is in doubly protonated form. The hydrophobic gate and a nearby negative charge increase the pK_a. When doubly protonated, this residue spends a considerable time in the active site.

Our study provides information about the relative importance of the three interactions between Switch I and GTP. The interaction involving D51 is not crucial for positioning Switch I, whereas both the T62-γ-phosphate and the T62-Mg^{2+} coordination are necessary to maintain the GTP bound conformation of Switch.

7. CONCLUDING REMARKS

This dissertation presents a scaffold-based drug discovery paradigm of a protein-ligand complex system and molecular dynamics simulations analysis for conformational rearrangements of a hydrolysis mechanism in the protein environment.

In the first part of this thesis, the modeling of pharmacophores for PDE4B inhibitors and pharmacophore-based virtual screening of drug-like compound libraries was performed to propose novel compounds. The procedure was followed by docking process of co-crystallized ligand rolipram and other candidate molecules. After the filtering based on docking, enrichment factors, ligand map interactions and clustering, the most promising scaffolds for novel PDE4B inhibitors are similar to rolipram structure and give interactions with the Gln443 residue and Mg^{2+} metal ion coordination.

The study was followed by an original study based on how to correlate experimental IC_{50} values to computed free energies of binding for PDE4B enzyme system. In this study, it was shown that while a docking software like AutoDock is capable of discriminating between different poses and of finding true positive hits, its scoring function is not capable of estimating experimental $\Delta G_{binding}$ values in PDE4B case. Secondly, it was found that averaged binding free energies obtained by the MM-GB/SA post-processing of many independent molecular dynamics trajectories gave much more reliable results. Especially, when all clusters obtained by AutoDock (and not only the best ones) are taken into account, a linear relationship between experimental $\log(IC_{50})$ and $\Delta G_{binding}$ values can be obtained with a very good correlation ($R^2 = 0.944$).

In the second part of the dissertation, the residues which are involved in GTP hydrolysis and their roles in GTPase activity were analyzed via molecular dynamics simulations. It was shown that R57 on the Switch I region can enter the active site to play a catalytic role, similar to the catalytic arginine residues in other enzymatic GTP hydrolysis

reactions. Also, it can be said that a conformational change of H85 cannot account for the ribosome induced rate acceleration. It is likely that coupled effects of R57 and H85 are required for the ribosome induced catalysis.

Overall, throughout this dissertation we tried to address specific issues such as pharmacophore modeling, docking studies and free energy of binding ($\Delta G_{binding}$) of PDE4B system and have analyzed the role of important amino acids in GTPase activity of EF-Tu-GTP for different organisms (*Thermos-Aquaticus (T.aquaticus)* and *Escheria Coli (E.coli)* complex by the aid of molecular dynamics (MD) simulations. The modeling approaches used in this dissertation are advantageous in terms of the investigation of challenging problems such as novel scaffolds in computer-aided drug design area by saving time for experiments. The detailed conclusions related with the specific problems are drawn within the relevant chapters.

As a future work, different free energy of binding calculation (Thermodynamic Integration and MM-PB/SA) methods can be used for the evaluation of selected free energies of binding.

Also, various parameters such as the convergence of the statistical ensemble, the effects of the length of the MD simulations, the usage of independent trajectories that can lead to a proper evaluation of the free energy of binding between a protein receptor and a small organic ligand may be investigated.

It is expected to test experimentally the novel candidates that are obtained from the pharmacophore-based virtual screening study.

Finally, from the PDE4B study part, the same procedure of $\Delta G_{binding}$ calculation and IC_{50} value estimation may be repeated for the complexes with ligands proposed from the virtual screening part, and the knowledge of calculating binding Gibbs Free energies may be extended to further studies.

For the conformational rearrangement of GTP hydrolysis in EF-Tu, the MD simulations may be extended to observe the entrance of K56 into the active site of EF-Tu. Also, the MD simulations may be carried out for ternary complex and the role of tRNA will be investigated.

It is also desirable to elucidate the enhancing role of ribosome (with a rate of ~10^5 times faster than that of its intrinsic activity) in the hydrolysis mechanism of GTP binded EF-tu by carrying out MD simulations on EF-Tu·GTP complex on the ribosome.

REFERENCES

1. Torphy, T. J., "Phosphodiesterase isozymes - Molecular targets for novel antiasthma agents", *American Journal of Respiratory and Critical Care Medicine*, Vol. 157, pp. 351-370, 1998.

2. Conti, M. and S. L. C. Jin, "The molecular biology of cyclic nucleotide phosphodiesterases", *Progress in Nucleic Acid Research and Molecular Biology, Vol 63*, Vol. 63, pp. 1-38, 2000.

3. Soderling, S. H. and J. A. Beavo, "Regulation of cAMP and cGMP signaling: new phosphodiesterases and new functions", *Current Opinion in Cell Biology*, Vol. 12, pp. 174-179, 2000.

4. Mehats, C., C. B. Andersen, M. Filopanti, S. L. C. Jin and M. Conti, "Cyclic nucleotide phosphodiesterases and their role in endocrine cell signaling", *Trends in Endocrinology and Metabolism*, Vol. 13, pp. 29-35, 2002.

5. Antoni, F. A., "Molecular diversity of cyclic AMP signalling", *Frontiers in Neuroendocrinology*, Vol. 21, pp. 103-132, 2000.

6. Houslay, M. D., M. Sullivan and G. B. Bolger, "The multienzyme PDE4 cyclic adenosine monophosphate-specific phosphodiesterase family: intracellular targeting, regulation, and selective inhibition by compounds exerting anti-inflammatory and antidepressant actions", *Adv Pharmacol*, Vol. 44, pp. 225-342, 1998.

7. Houslay, M. D. and G. Milligan, "Tailoring cAMP-signalling responses through isoform multiplicity", *Trends in Biochemical Sciences*, Vol. 22, pp. 217-224, 1997.

8. Xu, R. X., A. M. Hassell, D. Vanderwall, M. H. Lambert, W. D. Holmes, M. A. Luther, W. J. Rocque, M. V. Milburn, Y. D. Zhao, H. M. Ke and R. T. Nolte, "Atomic structure

of PDE4: Insights into phosphodiesterase mechanism and specificity", *Science*, Vol. 288, pp. 1822-1825, 2000.

9. Corbin, J. D. and S. H. Francis, "Cyclic GMP phosphodiesterase-5: Target of sildenafil", *Journal of Biological Chemistry*, Vol. 274, pp. 13729-13732, 1999.

10. Muller, T., P. Engels and J. R. Fozard, "Subtypes of the type 4 cAMP phosphodiesterases: Structure, regulation and selective inhibition", *Trends in Pharmacological Sciences*, Vol. 17, pp. 294-298, 1996.

11. Manganiello, V. C., M. Taira, E. Degerman and P. Belfrage, "Type-Iii cGMP-Inhibited Cyclic-Nucleotide Phosphodiesterases (PDE-3 Gene Family)", *Cellular Signalling*, Vol. 7, pp. 445-455, 1995.

12. Thompson, W. J., "Cyclic-Nucleotide Phosphodiesterases - Pharmacology, Biochemistry And Function", *Pharmacology & Therapeutics*, Vol. 51, pp. 13-33, 1991.

13. Huai, Q., H. C. Wang, Y. J. Sun, H. Y. Kim, Y. D. Liu and H. M. Ke, "Three-dimensional structures of PDE4D in complex with roliprams on inhibitor selectivity", *Structure*, Vol. 11, pp. 865-873, 2003.

14. Bischoff, E., "Potency, selectivity, and consequences of nonselectivity of PDE inhibition", *International Journal of Impotence Research*, Vol. 16, pp. S11-14, 2004.

15. Rotella, D. P., "Phosphodiesterase 5 inhibitors: Current status and potential applications", *Nature Reviews Drug Discovery*, Vol. 1, pp. 674-682, 2002.

16. Corbin, J. D. and S. H. Francis, "Pharmacology of phosphodiesterase-5 inhibitors", *International Journal of Clinical Practice*, Vol. 56, pp. 453-459, 2002.

17. Reilly, M. P. and E. R. Mohler, "Cilostazol: Treatment of intermittent claudication", *Annals of Pharmacotherapy*, Vol. 35, pp. 48-56, 2001.

18. Huang, Z., Y. Ducharme, D. Macdonald and A. Robichaud, "The next generation of PDE4 inhibitors", *Current Opinion in Chemical Biology*, Vol. 5, pp. 432-438, 2001.

19. Souness, J. E., D. Aldous and C. Sargent, "Immunosuppressive and anti-inflammatory effects of cyclic AMP phosphodiesterase (PDE) type 4 inhibitors", *Immunopharmacology*, Vol. 47, pp. 127-162, 2000.

20. Giembycz, M. A., "Phosphodiesterase 4 inhibitors and the treatment of asthma - Where are we now and where do we go from here?", *Drugs*, Vol. 59, pp. 193-212, 2000.

21. Houslay, M. D., P. Schafer and K. Y. J. Zhang, "Phosphodiesterase-4 as a therapeutic target", *Drug Discovery Today*, Vol. 10, pp. 1503-1519, 2005.

22. Houslay, M. D. and D. R. Adams, "PDE4 cAMP phosphodiesterases: modular enzymes that orchestrate signalling cross-talk, desensitization and compartmentalization", *Biochemical Journal*, Vol. 370, pp. 1-18, 2003.

23. Conti, M., W. Richter, C. Mehats, G. Livera, J. Y. Park and C. Jin, "Cyclic AMP-specific PDE4 phosphodiesterases as critical components of cyclic AMP signaling", *Journal of Biological Chemistry*, Vol. 278, pp. 5493-5496, 2003.

24. Huang, Z. and J. A. Mancini, "Phosphodiesterase 4 inhibitors for the treatment of asthma and COPD", *Current Medicinal Chemistry*, Vol. 13, pp. 3253-3262, 2006.

25. Oki, N., S. I. Takahashi, H. Hidaka and M. Conti, "Short term feedback regulation of cAMP in FRTL-5 thyroid cells. Role of PDE4D3 phosphodiesterase activation", *Journal of Biological Chemistry*, Vol. 275, pp. 10831-10837, 2000.

26. Sette, C. and M. Conti, "Phosphorylation and activation of a cAMP-specific phosphodiesterase by the cAMP-dependent protein kinase. Involvement of serine 54 in the enzyme activation", *Journal of Biological Chemistry*, Vol. 271, pp. 16526-16534, 1996.

27. Manallack, D. T., R. A. Hughes and P. E. Thompson, "The next generation of phosphodiesterase inhibitors: structural clues to ligand and substrate selectivity of phosphodiesterases", *Journal of Medicinal Chemistry*, Vol. 48, pp. 3449-3462, 2005.

28. Card, G. L., B. P. England, Y. Suzuki, D. Fong, B. Powell, B. Lee, C. Luu, M. Tabrizizad, S. Gillette, P. N. Ibrahim, D. R. Artis, G. Bollag, M. V. Milburn, S. H. Kim, J. Schlessinger and K. Y. J. Zhang, "Structural basis for the activity of drugs that inhibit phosphodiesterases", *Structure*, Vol. 12, pp. 2233-2247, 2004.

29. Wang, H., M.-S. Peng, Y. Chen, J. Geng, H. Robinson, M. D. Houslay, J. Cai and H. Ke, "Structures of the four subfamilies of phosphodiesterase-4 provide insight into the selectivity of their inhibitors", *Biochemical Journal*, Vol. 408, pp. 193-201, 2007.

30. Sturton, G. and M. Fitzgerald, "Phosphodiesterase 4 inhibitors for the treatment of COPD", *Chest*, Vol. 121, pp. 192S-196S, 2002.

31. Dal Piaz, V. and M. P. Giovannoni, "Phosphodiesterase 4 inhibitors, structurally unrelated to rolipram, as promising agents for the treatment of asthma and other pathologies", *European Journal of Medicinal Chemistry*, Vol. 35, pp. 463-480, 2000.

32. Dyke, H. J. and J. G. Montana, "Update on the therapeutic potential of PDE4 inhibitors", *Expert Opinion on Investigational Drugs*, Vol. 11, pp. 1-13, 2002.

33. Schmeing, T. M. and V. Ramakrishnan, "What recent ribosome structures have revealed about the mechanism of translation", *Nature*, Vol. 461, pp. 1234-1242, 2009.

34. Rodnina, M. V. and W. Wintermeyer, "Recent mechanistic insights into eukaryotic ribosomes", *Current Opinion in Cell Biology*, Vol. 21, pp. 435-443, 2009.

35. Wallin, G., S. C. L. Kamerlin and J. Aqvist, "Energetics of activation of GTP hydrolysis on the ribosome", *Nature Communications*, Vol. 4, pp., 2013.

36. Sprang, S. R., Z. Chen and X. Du, "Structural basis of effector regulation and signal termination in heterotrimeric G alpha proteins", *Mechanisms and Pathways of Heterotrimeric G Protein Signaling*, Vol. 74, pp. 1-65, 2007.

37. Li, G. P. and X. J. C. Zhang, "GTP hydrolysis mechanism of Ras-like GTPases", *Journal of Molecular Biology*, Vol. 340, pp. 921-932, 2004.

38. Sprang, S. R., "G protein mechanisms: Insights from structural analysis", *Annual Review of Biochemistry*, Vol. 66, pp. 639-678, 1997.

39. Bourne, H. R., D. A. Sanders and F. McCormick, "The Gtpase Superfamily - A Conserved Switch For Diverse Cell Functions", *Nature*, Vol. 348, pp. 125-132, 1990.

40. Daviter, T., H. J. Wieden and M. V. Rodnina, "Essential role of histidine 84 in elongation factor Tu for the chemical step of GTP hydrolysis on the ribosome", *Journal of Molecular Biology*, Vol. 332, pp. 689-699, 2003.

41. Limmer, S., C. O. A. Reiser, N. K. Schirmer, N. W. Grillenbeck and M. Sprinzl, "Nucleotide Binding And Gtp Hydrolysis By Elongation-Factor Tu From Thermus-Thermophilus As Monitored By Proton NMR", *Biochemistry*, Vol. 31, pp. 2970-2977, 1992.

42. Harmark, K., P. H. Anborgh, M. Merola, B. F. C. Clark and A. Parmeggiani, "Substitution Of Aspartic Acid-80, A Residue Involved In Coordination Of Magnesium, Weakens The GTP Binding And Strongly Enhances The GTPase Of The G-Domain Of Elongation Factor-Tu", *Biochemistry*, Vol. 31, pp. 7367-7372, 1992.

43. Grigorenko, B. L., M. S. Shadrina, I. A. Topol, J. R. Collins and A. V. Nemukhin, "Mechanism of the chemical step for the guanosine triphosphate (GTP) hydrolysis catalyzed by elongation factor Tu", *Biochimica Et Biophysica Acta-Proteins and Proteomics*, Vol. 1784, pp. 1908-1917, 2008.

44. Schuette, J.-C., F. V. Murphy, A. C. Kelley, J. R. Weir, J. Giesebrecht, S. R. Connell, J. Loerke, T. Mielke, W. Zhang, P. A. Penczek, V. Ramakrishnan and C. M. T. Spahn, "GTPase activation of elongation factor EF-Tu by the ribosome during decoding", *Embo Journal*, Vol. 28, pp. 755-765, 2009.

45. Zeidler, W., C. Egle, S. Ribeiro, A. Wagner, V. Katunin, R. Kreutzer, M. Rodnina, W. Wintermeyer and M. Sprinzl, "Site-Directed Mutagenesis Of Thermus-Thermophilus Elongation-Factor Tu - Replacement Of His85, Asp81 And Arg300", *European Journal of Biochemistry*, Vol. 229, pp. 596-604, 1995.

46. Scarano, G., I. M. Krab, V. Bocchini and A. Parmeggiani, "Relevance Of Histidine-84 In The Elongation-Factor Tu GTPase Activity And In Poly(Phe) Synthesis - Its Substitution By Glutamine And Alanine", *Febs Letters*, Vol. 365, pp. 214-218, 1995.

47. Moazed, D., J. M. Robertson and H. F. Noller, "Interaction Of Elongation-Factors EF-G And EF-Tu With A Conserved Loop In 23S RNA", *Nature*, Vol. 334, pp. 362-364, 1988.

48. Hausner, T. P., J. Atmadja and K. H. Nierhaus, "Evidence That The G2661 Region Of 23S Ribosomal-RNA Is Located At The Ribosomal-Binding Sites Of Both Elongation-Factors", *Biochimie*, Vol. 69, pp. 911-923, 1987.

49. Sedlak, E., M. Sprinzl, N. Grillenbeck and M. Antalik, "Microcalorimetric study of elongation factor Tu from Thermus thermophilus in nucleotide-free, GDP and GTP forms and in the presence of elongation factor Ts", *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology*, Vol. 1596, pp. 357-365, 2002.

50. Bourne, H. R., D. A. Sanders and F. McCormick, "The GTPase Superfamily - Conserved Structure And Molecular Mechanism", *Nature*, Vol. 349, pp. 117-127, 1991.

51. Krab, I. M. and A. Parmeggiani, "Mechanisms of EF-Tu, a pioneer GTPase", *Progress in Nucleic Acid Research and Molecular Biology, Vol 71*, Vol. 71, pp. 513-551, 2002.

52. Krab, I. M. and A. Parmeggiani, "EF-Tu, a GTPase odyssey", *Biochimica Et Biophysica Acta-Gene Structure and Expression*, Vol. 1443, pp. 1-22, 1998.

53. Voorhees, R. M., T. M. Schmeing, A. C. Kelley and V. Ramakrishnan, "The Mechanism for Activation of GTP Hydrolysis on the Ribosome", *Science*, Vol. 330, pp. 835-838, 2010.

54. Aleksandrov, A. and M. Field, "Mechanism of activation of elongation factor Tu by ribosome: Catalytic histidine activates GTP by protonation", *Rna-a Publication of the Rna Society*, Vol. 19, pp. 1218-1225, 2013.

55. Aqvist, J. and S. C. L. Kamerlin, "The Conformation of a Catalytic Loop Is Central to GTPase Activity on the Ribosome", *Biochemistry*, Vol. 54, pp. 546-556, 2015.

56. Prasad, B. R., N. V. Plotnikov, J. Lameira and A. Warshel, "Quantitative exploration of the molecular origin of the activation of GTPase", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 110, pp. 20509-20514, 2013.

57. Adamczyk, A. J. and A. Warshel, "Converting structural information into an allosteric-energy-based picture for elongation factor Tu activation by the ribosome", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 108, pp. 9827-9832, 2011.

58. Mannhold, R., H. Kubinyi, G. Folkers and C. Sotriffer, *Virtual screening: principles, challenges, and practical guidelines*, John Wiley & Sons, 2011, Vol. 48.

59. Bacilieri, M. and S. Moro, "Ligand-based drug design methodologies in drug discovery process: an overview", *Current drug discovery technologies*, Vol. 3, pp. 155-165, 2006.

60. Young, D. C., *Computational drug design: a guide for computational and medicinal chemists*, John Wiley & Sons, 2009.

61. Meng, X. Y., H. X. Zhang, M. Mezei and M. Cui, "Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery", *Current Computer-Aided Drug Design*, Vol. 7, pp. 146-157, 2011.

62. Totrov, M. and R. Abagyan, "Flexible protein–ligand docking by global energy optimization in internal coordinates", *Proteins: Structure, Function, and Bioinformatics*, Vol. 29, pp. 215-220, 1997.

63. Friesner, R. A., J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley and J. K. Perry, "Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy", *Journal of Medicinal Chemistry*, Vol. 47, pp. 1739-1749, 2004.

64. Jones, G., P. Willett, R. C. Glen, A. R. Leach and R. Taylor, "Development and validation of a genetic algorithm for flexible docking", *Journal of Molecular Biology*, Vol. 267, pp. 727-748, 1997.

65. Morris, G. M., D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function", *Journal of Computational Chemistry*, Vol. 19, pp. 1639-1662, 1998.

66. Zsoldos, Z., D. Reid, A. Simon, S. B. Sadjad and A. P. Johnson, "eHiTS: a new fast, exhaustive flexible ligand docking system", *Journal of Molecular Graphics and Modelling*, Vol. 26, pp. 198-212, 2007.

67. McGann, M., "FRED and HYBRID docking performance on standardized datasets", *Journal of Computer-Aided Molecular Design*, Vol. 26, pp. 897-906, 2012.

68. Jain, A. N., "Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine", *Journal of Medicinal Chemistry*, Vol. 46, pp. 499-511, 2003.

69. Baxter, C. A., C. W. Murray, D. E. Clark, D. R. Westhead and M. D. Eldridge, "Flexible docking using Tabu search and an empirical estimate of binding affinity", *Proteins: Structure, Function, and Bioinformatics*, Vol. 33, pp. 367-382, 1998.

70. Ewing, T. J., S. Makino, A. G. Skillman and I. D. Kuntz, "DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases", *Journal of Computer-Aided Molecular Design*, Vol. 15, pp. 411-428, 2001.

71. Grosdidier, A., V. Zoete and O. Michielin, "EADock: docking of small molecules into protein active sites with a multiobjective evolutionary optimization", *Proteins: Structure, Function, and Bioinformatics*, Vol. 67, pp. 1010-1025, 2007.

72. Pang, Y. P., E. Perola, K. Xu and F. G. Prendergast, "EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases", *Journal of Computational Chemistry*, Vol. 22, pp. 1750-1771, 2001.

73. Venkatachalam, C. M., X. Jiang, T. Oldfield and M. Waldman, "LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites", *Journal of Molecular Graphics and Modelling*, Vol. 21, pp. 289-307, 2003.

74. Rarey, M., B. Kramer, T. Lengauer and G. Klebe, "A fast flexible docking method using an incremental construction algorithm", *Journal of Molecular Biology*, Vol. 261, pp. 470-489, 1996.

75. Sochacka, J., "Docking of thiopurine derivatives to human serum albumin and binding site analysis with Molegro Virtual Docker", *Acta Pol. Pharm*, Vol. 71, pp. 343-349, 2014.

76. Welch, W., J. Ruppert and A. N. Jain, "Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites", *Chemistry & biology*, Vol. 3, pp. 449-462, 1996.

77. Wu, G., D. H. Robertson, C. L. Brooks and M. Vieth, "Detailed analysis of grid-based molecular docking: A case study of CDOCKER - A CHARMm-based MD docking algorithm", *Journal of Computational Chemistry*, Vol. 24, pp. 1549-1562, 2003.

78. Jain, A. N., "Scoring functions for protein-ligand docking", *Current Protein and Peptide Science*, Vol. 7, pp. 407-420, 2006.

79. Ferreira, L. G., R. N. dos Santos, G. Oliva and A. D. Andricopulo, "Molecular docking and structure-based drug design strategies", *Molecules*, Vol. 20, pp. 13384-13421, 2015.

80. Kitchen, D. B., H. Decornez, J. R. Furr and J. Bajorath, "Docking and scoring in virtual screening for drug discovery: methods and applications", *Nature Reviews Drug Discovery*, Vol. 3, pp. 935-949, 2004.

81. Huang, S. Y. and X. Zou, "An iterative knowledge-based scoring function to predict protein–ligand interactions: I. Derivation of interaction potentials", *Journal of Computational Chemistry*, Vol. 27, pp. 1866-1875, 2006.

82. Gohlke, H., M. Hendlich and G. Klebe, "Knowledge-based scoring function to predict protein-ligand interactions", *Journal of Molecular Biology*, Vol. 295, pp. 337-356, 2000.

83. Barnum, D., J. Greene, A. Smellie and P. Sprague, "Identification of common functional configurations among molecules", *Journal of Chemical Information and Computer Sciences*, Vol. 36, pp. 563-571, 1996.

84. Wolber, G. and T. Langer, "LigandScout: 3-d pharmacophores derived from protein-bound Ligands and their use as virtual screening filters", *Journal of Chemical Information and Modeling*, Vol. 45, pp. 160-169, 2005.

85. Dixon, S. L., A. M. Smondyrev, E. H. Knoll, S. N. Rao, D. E. Shaw and R. A. Friesner, "PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results", *Journal of Computer-Aided Molecular Design*, Vol. 20, pp. 647-671, 2006.

86. Chemical Computing Group, I., Montreal, Quebec, Canada, Molecular and Operating Environment (MOE).

87. Triballeau, N., F. Acher, I. Brabet, J. P. Pin and H. O. Bertrand, "Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4", *Journal of Medicinal Chemistry*, Vol. 48, pp. 2534-2547, 2005.

88. Mannhold, R., H. Kubinyi, G. Folkers, T. Langer and R. D. Hoffmann, *Pharmacophores and pharmacophore searches*, John Wiley & Sons, 2006, Vol. 32.

89. Karplus, M. and J. A. McCammon, "Molecular dynamics simulations of biomolecules", *Nature Structural & Molecular Biology*, Vol. 9, pp. 646-652, 2002.

90. Alder, B. and T. Wainwright, "Phase transition for a hard sphere system", *The Journal of chemical physics*, Vol. 27, pp. 1208-1209, 1957.

91. Adcock, S. A. and J. A. McCammon, "Molecular dynamics: survey of methods for simulating the activity of proteins", *Chemical reviews*, Vol. 106, pp. 1589-1615, 2006.

92. Genheden, S. and U. Ryde, "How to Obtain Statistically Converged MM/GBSA Results", *Journal of Computational Chemistry*, Vol. 31, pp. 837-846, 2010.

93. Gohlke, H. and G. Klebe, "Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors", *Angewandte Chemie-International Edition*, Vol. 41, pp. 2645-2676, 2002.

94. Zwanzig, R. W., "High-Temperature Equation of State by a Perturbation Method .1. Nonpolar Gases", *Journal of Chemical Physics*, Vol. 22, pp. 1420-1426, 1954.

95. Kongsted, J. and U. Ryde, "An improved method to predict the entropy term with the MM/PBSA approach", *Journal of Computer-Aided Molecular Design*, Vol. 23, pp. 63-71, 2009.

96. Aqvist, J., C. Medina and J. E. Samuelsson, "New Method For Predicting Binding-Affinity in Computer-Aided Drug Design", *Protein Engineering*, Vol. 7, pp. 385-391, 1994.

97. Wong, S., R. E. Amaro and J. A. McCammon, "MM-PBSA Captures Key Role of Intercalating Water Molecules at a Protein-Protein Interface", *Journal of Chemical Theory and Computation*, Vol. 5, pp. 422-429, 2009.

98. Rastelli, G., G. Degliesposti, A. Del Rio and M. Sgobba, "Binding Estimation after Refinement, a New Automated Procedure for the Refinement and Rescoring of Docked Ligands in Virtual Screening", *Chemical Biology & Drug Design*, Vol. 73, pp. 283-286, 2009.

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

100. Onufriev, A. In *Annual Reports in Computational Chemistry, Vol 4*; Wheeler, R. A., Spellmeyer, D. C., Eds. 2010; Vol. 4, p 125-137.

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

102. Weis, A., K. Katebzadeh, P. Soderhjelm, I. Nilsson and U. Ryde, "Ligand affinities predicted with the MM/PBSA method: Dependence on the simulation method and the force field", *Journal of Medicinal Chemistry*, Vol. 49, pp. 6596-6606, 2006.

103. Jeffery, P., "Phosphodiesterase 4-selective inhibition: novel therapy for the inflammation of COPD", *Pulmonary Pharmacology & Therapeutics*, Vol. 18, pp. 9-17, 2005.

104. Castro, A., M. J. Jerez, C. Gil and A. Martinez, "Cyclic nucleotide phosphodiesterases and their role in immunomodulatory responses: Advances in the development of specific phosphodiesterase inhibitors", *Medicinal Research Reviews*, Vol. 25, pp. 229-244, 2005.

105. Spina, D., "The potential of PDE4 inhibitors in respiratory disease", *Curr Drug Targets Inflamm Allergy*, Vol. 3, pp. 231-236, 2004.

106. O'Donnell, J. M. and H. T. Zhang, "Antidepressant effects of inhibitors of cAMP phosphodiesterase (PDE4)", *Trends in Pharmacological Sciences*, Vol. 25, pp. 158-163, 2004.

107. Draheim, R., U. Egerland and C. Rundfeldt, "Anti-inflammatory potential of the selective phosphodiesterase 4 inhibitor N-(3,5-dichloro-pyrid-4-yl)- 1-(4-fluorobenzyl)-5-hydroxy-indole-3-yl -g lyoxylic acid amide (AWD 12-281), in human cell preparations", *Journal of Pharmacology and Experimental Therapeutics*, Vol. 308, pp. 555-563, 2004.

108. Banner, K. H. and M. A. Trevethick, "PDE4 inhibition: a novel approach for the treatment of inflammatory bowel disease", *Trends in Pharmacological Sciences*, Vol. 25, pp. 430-436, 2004.

109. Calverley, P. M. A., F. Sanchez-Torill, A. McIvor, P. Teichmann, D. Bredenbroeker and L. M. Fabbri, "Effect of 1-year treatment with roflumilast in severe chronic obstructive pulmonary disease", *American Journal of Respiratory and Critical Care Medicine*, Vol. 176, pp. 154-161, 2007.

110. Rabe, K. F., E. D. Bateman, D. O'Donnell, S. Witte, D. Bredenbroker and T. D. Bethke, "Roflumilast - an oral anti-inflammatory treatment for chronic obstructive pulmonary disease: a randomised controlled trial", *Lancet*, Vol. 366, pp. 563-571, 2005.

111. Jeon, Y. H., Y. S. Heo, C. M. Kim, Y. L. Hyun, T. G. Lee, S. Ro and J. M. Cho, "Phosphodiesterase: overview of protein structures, potential therapeutic applications and recent progress in drug development", *Cellular and Molecular Life Sciences*, Vol. 62, pp. 1198-1220, 2005.

112. Skoumbourdis, A. P., R. Huang, N. Southall, W. Leister, V. Guo, M. H. Cho, J. Inglese, M. Nirenberg, C. P. Austin, M. Xia and C. J. Thomas, "Identification of a potent new chemotype for the selective inhibition of PDE4", *Bioorg Med Chem Lett*, Vol. 18, pp. 1297-1303, 2008.

113. Trott, O. and A. J. Olson, "Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading", *Journal of Computational Chemistry*, Vol. 31, pp. 455-461, 2010.

114. Lagorce, D., O. Sperandio, H. Galons, M. A. Miteva and B. O. Villoutreix, "FAF-Drugs2: Free ADME/tox filtering tool to assist drug discovery and chemical biology projects", *Bmc Bioinformatics*, Vol. 9, pp., 2008.

115. Kirchmair, J., G. Wolber, C. Laggner and T. Langer, "Comparative performance assessment of the conformational model generators omega and catalyst: a large-scale survey on the retrieval of protein-bound ligand conformations", *Journal of Chemical Information and Modeling*, Vol. 46, pp. 1848-1861, 2006.

116. Chen, Z., G. Tian, Z. Wang, H. Jiang, J. Shen and W. Zhu, "Multiple Pharmacophore Models Combined with Molecular Docking: A Reliable Way for Efficiently Identifying Novel PDE4 Inhibitors with High Structural Diversity", *Journal of Chemical Information and Modeling*, Vol. 50, pp. 615-625, 2010.

117. Dym, O., I. Xenarios, H. Ke and J. Colicelli, "Molecular docking of competitive phosphodiesterase inhibitors", *Mol Pharmacol*, Vol. 61, pp. 20-25, 2002.

118. Garrett M. Morris, R. H., Arthur J. Olson, "UNIT 8.14 Using AutoDock for Ligand-Receptor Docking", Vol., pp., 2008.

119. Baell, J. B. and G. A. Holloway, "New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays", *Journal of Medicinal Chemistry*, Vol. 53, pp. 2719-2740, 2010.

120. Callahan, S. M., N. W. Cornell and P. V. Dunlap, "Purification and Properties of Periplasmic 3'/5'-Cyclic Nucleotide Phocphodiesterase- A Novel Zinc-Containing Enzyme From The Marine Symbiotic Bacterium Vibrio-Fischeria", *Journal of Biological Chemistry*, Vol. 270, pp. 17627-17632, 1995.

121. Ke, H. and H. Wang, "Crystal structures of phosphodiesterases and implications on substrate specificity and inhibitor selectivity", *Current Topics in Medicinal Chemistry*, Vol. 7, pp. 391-403, 2007.

122. Kang, N.-S., S.-J. Hong, C.-H. Chae and S.-E. Yoo, "Comparative molecular field analysis (CoMFA) for phosphodiesterase (PDE) IV inhibitors", *Journal of Molecular Structure-Theochem*, Vol. 820, pp. 58-64, 2007.

123. Rastelli, G., A. Del Rio, G. Degliesposti and M. Sgobba, "Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA", *Journal of Computational Chemistry*, Vol. 31, pp. 797-810, 2010.

124. Genheden, S., T. Luchko, S. Gusarov, A. Kovalenko and U. Ryde, "An MM/3D-RISM Approach for Ligand Binding Affinities", *Journal of Physical Chemistry B*, Vol. 114, pp. 8505-8516, 2010.

125. Alexander, R. P., G. J. Warrellow, M. A. Eaton, E. C. Boyd, J. C. Head, J. R. Porter, J. A. Brown, J. T. Reuberson, B. Hutchinson, P. Turner, B. Boyce, D. Barnes, B. Mason, A. Cannell, R. J. Taylor, A. Zomaya, A. Millican, J. Leonard, R. Morphy, M. Wales, M. Perry, R. A. Allen, N. Gozzard, B. Hughes and G. Higgs, "CDP840. A prototype of a novel class of orally active anti-inflammatory phosphodiesterase 4 inhibitors", *Bioorg Med Chem Lett*, Vol. 12, pp. 1451-1456, 2002.

126. Kuang, R., H. J. Shue, D. J. Blythin, N. Y. Shih, D. Gu, X. Chen, J. Schwerdt, L. Lin, P. C. Ting, X. Zhu, R. Aslanian, J. J. Piwinski, L. Xiao, D. Prelusky, P. Wu, J. Zhang, X. Zhang, C. S. Celly, M. Minnicozzi, M. Billah and P. Wang, "Discovery of a highly potent series of oxazole-based phosphodiesterase 4 inhibitors", *Bioorg Med Chem Lett*, Vol. 17, pp. 5150-5154, 2007.

127. Guay, D., L. Boulet, R. W. Friesen, M. Girard, P. Hamel, Z. Huang, F. Laliberte, S. Laliberte, J. A. Mancini, E. Muise, D. Pon and A. Styhler, "Optimization and structure-

activity relationship of a series of 1-phenyl-1,8-naphthyridin-4-one-3-carboxamides: identification of MK-0873, a potent and effective PDE4 inhibitor", *Bioorg Med Chem Lett*, Vol. 18, pp. 5554-5558, 2008.

128. Xu, R. X., W. J. Rocque, M. H. Lambert, D. E. Vanderwall, M. A. Luther and R. T. Nolte, "Crystal structures of the catalytic domain of phosphodiesterase 4B complexed with AMP, 8-Br-AMP, and rolipram", *Journal of Molecular Biology*, Vol. 337, pp. 355-365, 2004.

129. Salter, E. A. and A. Wierzbicki, "The mechanism of cyclic nucleotide hydrolysis in the phosphodiesterase catalytic site", *Journal of Physical Chemistry B*, Vol. 111, pp. 4547-4552, 2007.

130. Chen, X., X. Y. Zhao, Y. Xiong, J. J. Liu and C. G. Zhan, "Fundamental Reaction Pathway and Free Energy Profile for Hydrolysis of Intracellular Second Messenger Adenosine 3 ',5 '-Cyclic Monophosphate (cAMP) Catalyzed by Phosphodiesterase-4", *Journal of Physical Chemistry B*, Vol. 115, pp. 12208-12219, 2011.

131. Oliveira, F. G., C. M. R. Sant'Anna, E. R. Caffarena, L. E. Dardenne and E. J. Barreiro, "Molecular docking study and development of an empirical binding free energy model for phosphodiesterase 4 inhibitors", *Bioorganic & Medicinal Chemistry*, Vol. 14, pp. 6001-6011, 2006.

132. Niu, M., F. Dong, S. Tang, G. Fida, J. Qin, J. Qiu, K. Liu, W. Gao and Y. Gu, "Pharmacophore Modeling and Virtual Screening for the Discovery of New type 4 cAMP Phosphodiesterase (PDE4) Inhibitors", *PLoS ONE*, Vol. 8, pp., 2013.

133. Voet, D. and J. G. Voet, *Biochemistry*, Wiley, New York, 1990.

134. Lybrand, T. P., J. A. McCammon and G. Wipff, "Theoretical Calculation of Relative Binding-Affinity in Host Guest Systems", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 83, pp. 833-835, 1986.

135. Wang, J. M., P. Morin, W. Wang and P. A. Kollman, "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", *Journal of the American Chemical Society*, Vol. 123, pp. 5221-5230, 2001.

136. Kollman, P. A., I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, "Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models", *Accounts of Chemical Research*, Vol. 33, pp. 889-897, 2000.

137. Beveridge, D. L. and F. M. Dicapua, "Free-Energy Via Molecular Simulation - Applications to Chemical And Biomolecular Systems", *Annual Review of Biophysics and Biophysical Chemistry*, Vol. 18, pp. 431-492, 1989.

138. Chipot, C., X. Rozanska and S. B. Dixit, "Can free energy calculations be fast and accurate at the same time? Binding of low-affinity, non-peptide inhibitors to the SH2 domain of the src protein", *Journal of Computer-Aided Molecular Design*, Vol. 19, pp. 765-770, 2005.

139. Liu, T., Y. Lin, X. Wen, R. N. Jorissen and M. K. Gilson, "BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities", *Nucleic Acids Research*, Vol. 35, pp. D198-D201, 2007.

140. Dal Piaz, V., M. P. Giovannoni, C. Castellana, J. M. Palacios, J. Beleta, T. Domenech and V. Segarra, "Heterocyclic-fused 3(2H)-pyridazinones as potent and selective PDE IV

inhibitors: Further structure-activity relationships and molecular modelling studies", *European Journal of Medicinal Chemistry*, Vol. 33, pp. 789-797, 1998.

141. Hersperger, R., K. Bray-French, L. Mazzoni and T. Muller, "Palladium-catalyzed cross-coupling reactions for the synthesis of 6,8-disubstituted 1,7-naphthyridines: A novel class of potent and selective phosphodiesterase type 4D inhibitors", *Journal of Medicinal Chemistry*, Vol. 43, pp. 675-682, 2000.

142. Morris, G. M., R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, "AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility", *Journal of Computational Chemistry*, Vol. 30, pp. 2785-2791, 2009.

143. Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. M. Wang and P. Kollman, "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations", *Journal of Computational Chemistry*, Vol. 24, pp. 1999-2012, 2003.

144. Case, D. A. D., T.A.; Cheatham, III, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Walker, R.C.; Zhang, W.; Merz, K.M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A.W.; Kolossváry, I.; Wong, K.F.; Paesani, F.; Vanicek, J.; Wolf, R.M.; Liu, J.; Wu, X.; Brozell, S.R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.J.; Cui, G.; Roe, D.R.; Mathews, D.H.; Seetin, M.G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P.A., "AMBER12", Vol., pp., 2012.

145. Sondergaard, C. R., M. H. M. Olsson, M. Rostkowski and J. H. Jensen, "Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK(a) Values", *Journal of Chemical Theory and Computation*, Vol. 7, pp. 2284-2295, 2011.

146. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, "Comparison of Simple Potential Functions For Simulating Liquid Water", *Journal of Chemical Physics*, Vol. 79, pp. 926-935, 1983.

147. Salomon-Ferrer, R., A. W. Goetz, D. Poole, S. Le Grand and R. C. Walker, "Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald", *Journal of Chemical Theory and Computation*, Vol. 9, pp. 3878-3888, 2013.

148. Goetz, A. W., M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, "Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born", *Journal of Chemical Theory and Computation*, Vol. 8, pp. 1542-1555, 2012.

149. Wang, J. M., R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, "Development and testing of a general amber force field (vol 25, pg 1157, 2004)", *Journal of Computational Chemistry*, Vol. 26, pp. 114-114, 2005.

150. Ryckaert, J. P., G. Ciccotti and H. J. C. Berendsen, "Numerical - Integration of Cartesian Equations of Motion of a System With Constraints - Molecular-Dynamics of N-Alkanes", *Journal of Computational Physics*, Vol. 23, pp. 327-341, 1977.

151. Ugur, I. M., A.; Aviyente, V.; Monard, G., "Why Does Asn71 Deamidate Faster Than Asn15 in the Enzyme Triosephosphate Isomerase? Answers from Microsecond Molecular Dynamics Simulation and QM/MM Free Energy Calculations", *Biochemistry*, Vol., pp., 2015.

152. Hawkins, G. D., C. J. Cramer and D. G. Truhlar, "Pairwise Solute Descreening of Solute Charges from a Dielectric Medium", *Chemical Physics Letters*, Vol. 246, pp. 122-129, 1995.

153. Tsui, V. and D. A. Case, "Theory and applications of the generalized Born solvation model in macromolecular Simulations", *Biopolymers*, Vol. 56, pp. 275-291, 2001.

154. Hawkins, G. D., C. J. Cramer and D. G. Truhlar, "Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium", *Journal of Physical Chemistry*, Vol. 100, pp. 19824-19839, 1996.

155. Onufriev, A., D. Bashford and D. A. Case, "Exploring protein native states and large-scale conformational changes with a modified generalized born model", *Proteins-Structure Function and Bioinformatics*, Vol. 55, pp. 383-394, 2004.

156. Mongan, J., C. Simmerling, J. A. McCammon, D. A. Case and A. Onufriev, "Generalized Born model with a simple, robust molecular volume correction", *Journal of Chemical Theory and Computation*, Vol. 3, pp. 156-169, 2007.

157. Nguyen, H., D. R. Roe and C. Simmerling, "Improved Generalized Born Solvent Model Parameters for Protein Simulations", *Journal of Chemical Theory and Computation*, Vol. 9, pp. 2020-2034, 2013.

158. Weiser, J., P. S. Shenkin and W. C. Still, "Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO)", *Journal of Computational Chemistry*, Vol. 20, pp. 217-230, 1999.

159. Genheden, S., O. Kuhn, P. Mikulskis, D. Hoffmann and U. Ryde, "The Normal-Mode Entropy in the MM/GBSA Method: Effect of System Truncation, Buffer Region, and Dielectric Constant", *Journal of Chemical Information and Modeling*, Vol. 52, pp. 2079-2088, 2012.

160. Genheden, S. and U. Ryde, "Comparison of the Efficiency of the LIE and MM/GBSA Methods to Calculate Ligand-Binding Energies", *Journal of Chemical Theory and Computation*, Vol. 7, pp. 3768-3778, 2011.

161. Genheden, S., "MM/GBSA and LIE estimates of host-guest affinities: dependence on charges and solvation model", *Journal of Computer-Aided Molecular Design*, Vol. 25, pp. 1085-1093, 2011.

162. Nguyen, H., J. Maier, H. Huang, V. Perrone and C. Simmerling, "Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent", *Journal of the American Chemical Society*, Vol. 136, pp. 13959-13962, 2014.

163. Andersen, G. R., P. Nissen and J. Nyborg, "Elongation factors in protein biosynthesis", *Trends Biochem Sci*, Vol. 28, pp. 434-441, 2003.

164. Spirin, A. S., "Ribosome as a molecular machine", *Febs Letters*, Vol. 514, pp. 2-10, 2002.

165. Krab, I. M. and A. Parmeggiani, "Mutagenesis of three residues, isoleucine-60, threonine-61, and aspartic acid-80, implicated in the GTPase activity of Escherichia coli elongation factor Tu", *Biochemistry*, Vol. 38, pp. 13035-13041, 1999.

166. Rodnina, M. V., T. Pape, R. Fricke and W. Wintermeyer, "Elongation factor Tu, a GTPase triggered by codon recognition on the ribosome: mechanism and GTP consumption", *Biochem Cell Biol*, Vol. 73, pp. 1221-1227, 1995.

167. Bilgin, N., F. Claesens, H. Pahverk and M. Ehrenberg, "Kinetic properties of Escherichia coli ribosomes with altered forms of S12", *J Mol Biol*, Vol. 224, pp. 1011-1027, 1992.

168. Pape, T., W. Wintermeyer and M. V. Rodnina, "Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E. coli ribosome", *EMBO J*, Vol. 17, pp. 7490-7497, 1998.

169. Villa, E., J. Sengupta, L. G. Trabuco, J. LeBarron, W. T. Baxter, T. R. Shaikh, R. A. Grassucci, P. Nissen, M. Ehrenberg, K. Schulten and J. Frank, "Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 106, pp. 1063-1068, 2009.

170. Kiparisov, S. V., P. V. Sergiev, A. A. Bogdanov and O. A. Dontsova, "Structural changes in the ribosome during the elongation cycle", *Molecular Biology*, Vol. 40, pp. 675-687, 2006.

171. Maracci, C., F. Peske, E. Dannies, C. Pohl and M. V. Rodnina, "Ribosome-induced tuning of GTP hydrolysis by a translational GTPase", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 111, pp. 14418-14423, 2014.

172. Kjeldgaard, M., P. Nissen, S. Thirup and J. Nyborg, "The crystal structure of elongation factor EF-Tu from Thermus aquaticus in the GTP conformation", *Structure*, Vol. 1, pp. 35-50, 1993.

173. Nissen, P., S. Thirup, M. Kjeldgaard and J. Nyborg, "The crystal structure of Cys-tRNA(Cys)-EF-Tu-GDPNP reveals general and specific features in the ternary complex and in tRNA", *Structure with Folding & Design*, Vol. 7, pp. 143-156, 1999.

174. Polekhina, G., S. Thirup, M. Kjeldgaard, P. Nissen, C. Lippmann and J. Nyborg, "Helix unwinding in the effector region of elongation factor EF-Tu-GDP", *Structure*, Vol. 4, pp. 1141-1151, 1996.

175. Coleman, D. E., A. M. Berghuis, E. Lee, M. E. Linder, A. G. Gilman and S. R. Sprang, "Structures Of Active Conformations Of G(I-Alpha-1) And The Mechanism Of GTP Hydrolysis", *Science*, Vol. 265, pp. 1405-1412, 1994.

176. Scheffzek, K., M. R. Ahmadian, W. Kabsch, L. Wiesmuller, A. Lautwein, F. Schmitz and A. Wittinghofer, "The Ras-RasGAP complex: Structural basis for GTPase activation and its loss in oncogenic Ras mutants", *Science*, Vol. 277, pp. 333-338, 1997.

177. Knudsen, C. R. and B. F. C. Clark, "Site-directed mutagenesis of Arg58 and Asp86 of elongation factor Tu from Escherichia coli: Effects on the GTPase reaction and aminoacyl-tRNA binding", *Protein Engineering*, Vol. 8, pp. 1267-1273, 1995.

178. Liang, Z. M., T. Mather and G. P. Li, "GTPase mechanism and function: new insights from systematic mutational analysis of the phosphate-binding loop residue Ala(30) of Rab5", *Biochemical Journal*, Vol. 346, pp. 501-508, 2000.

179. Eargle, J., A. A. Black, A. Sethi, L. G. Trabuco and Z. Luthey-Schulten, "Dynamics of recognition between tRNA and elongation factor Tu", *Journal of Molecular Biology*, Vol. 377, pp. 1382-1405, 2008.

180. Kulczycka, K., M. Dlugosz and J. Trylska, "Molecular dynamics of ribosomal elongation factors G and Tu", *European Biophysics Journal with Biophysics Letters*, Vol. 40, pp. 289-303, 2011.

181. Lee, M. C. and Y. Duan, "Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model", *Proteins-Structure Function and Bioinformatics*, Vol. 55, pp. 620-634, 2004.

182. D.A. Case, V. B., J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E., H. G. Duke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko,, S. L. T.S. Lee, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui,, G. S. R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. and W. a. P. A. Kollman, "AMBER 14", *University of California, San Francisco.*, Vol., pp., 2014.

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

184. Meagher, K. L., L. T. Redman and H. A. Carlson, "Development of polyphosphate parameters for use with the AMBER force field", *Journal of Computational Chemistry*, Vol. 24, pp. 1016-1025, 2003.

185. Zwanzig, R., "Nonlinear generalized Langevin equations", *Journal of Statistical Physics*, Vol. 9, pp. 215-220, 1973.

186. Kaus, J. W., L. T. Pierce, R. C. Walker and J. A. McCammon, "Improving the Efficiency of Free Energy Calculations in the Amber Molecular Dynamics Package", *Journal of Chemical Theory and Computation*, Vol. 9, pp. 4131-4139, 2013.

187. Simonson, T., J. Carlsson and D. A. Case, "Proton binding to proteins: pK(a) calculations with explicit and implicit solvent models", *Journal of the American Chemical Society*, Vol. 126, pp. 4167-4180, 2004.

188. Kuhle, B. and R. Ficner, "A monovalent cation acts as structural and catalytic cofactor in translational GTPases", *Embo Journal*, Vol. 33, pp. 2547-2563, 2014.