COMPUTATIONAL MODELING OF PROTEIN KINASES: MOLECULAR BASIS FOR INHIBITION AND CATALYSIS

Yingting Liu

A DISSERTATION

in

Bioengineering

Presented to the Faculties of the University of Pennsylvania

in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2010

Supervisor of Dissertation	Graduate Group Chair
_____	_____
Ravi Radhakrishnan, Ph.D.	Susan Margulies, Ph.D.
Associate Professor of Bioengineering	Professor of Bioengineering

Dissertation Committee:

Scott L. Diamond, Ph.D., Professor of Chemical and Biomolecular Engineering and Bioengineering

Daniel A. Hammer, Ph.D., Professor of Bioengineering and Chemical Engineering, Committee Chair

Mark A. Lemmon, Ph.D., Professor of Biochemistry and Biophysics

Kim A. Sharp, Ph.D., Associate Professor of Biochemistry and Biophysics

Acknowledgment

I would like to first give my hearty acknowledgment to my advisor, Dr. Ravi Radhakrishnan, for his constantly encouragement and strongly support over the past five years. I also want to thank my collaborator, Dr. Mark Lemmon, for his support and valuable insight on my work on Epidermal Growth Factor Receptor over the years. My sincere thanks also go to Dr. Eric Meggers, Dr. Scott Diamond, Dr. Kim Sharp and Dr. Daniel Hammer for their help and encouragement. Last by not the lease, I thank my fellow colleagues Andrew Shih, Jeremy Purvis, Neeraj Agrawal, Shannon Telesco, Ravindra Venkatramani, Sung Hee Choi, Fuming Shi, Jin Park, Ryan Bradley, Peter Huwe and Jin Liu.

ABSTRACT

COMPUTATIONAL MODELING OF PROTEIN KINASES:

MOLECULAR BASIS FOR INHIBITION AND CATALYSIS

Yingting Liu

Ravi Radhakrishnan

Protein kinases catalyze protein phosphorylation reactions, i.e. the transfer of the γ-phosphoryl group of ATP to tyrosine, serine and threonine residues of protein substrates. This phosphorylation plays an important role in regulating various cellular processes. Deregulation of many kinases is directly linked to cancer development and the protein kinase family is one of the most important targets in current cancer therapy regimens. This relevance to disease has stimulated intensive efforts in the biomedical research community to understand their catalytic mechanisms, discern their cellular functions, and discover inhibitors. With the advantage of being able to simultaneously define structural as well as dynamic properties for complex systems, computational studies at the atomic level has been recognized as a powerful complement to experimental studies. In this work, we employed a suite of computational and molecular simulation methods to (1) explore the catalytic mechanism of a particular protein kinase, namely, epidermal growth factor receptor (EGFR); (2) study the interaction between EGFR and one of its inhibitors, namely erlotinib (Tarceva); (3) discern the effects of molecular alterations (somatic mutations) of EGFR to differential downstream signaling response; and (4) model the interactions of a novel class of kinase inhibitors with a common ruthenium based

organometallic scaffold with different protein kinases. Our simulations established some important molecular rules in operation in the contexts of inhibitor-binding, substrate-recognition, catalytic landscapes, and signaling in the EGFR tyrosine kinase. Our results also shed insights on the mechanisms of inhibition and phosphorylation commonly employed by many kinases.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Overview

Most protein kinases reported have been found to share a very similar structure fold, implicating an overall similarity in their mechanism of enzymatic action. However, as protein kinases control many different signaling pathways, the details of how they are regulated and how they recognize and subsequently regulate their substrates are found to be different. In this work, we focused on a biologically and pharmaceutically important protein kinase, namely the epidermal growth factor receptor (EGFR) tyrosine kinase domain (TKD). We employed a suite of molecular and multiscale simulation methods to investigate mechanisms of substrate recognition and catalysis in EGFR (Chapter 2), the interaction between EGFR and a small molecule inhibitor (Chapter 3), and the EGFR signaling network (Chapter 4). We investigated the unique properties of this kinase and also identified additional conserved properties applicable to homologous kinases. In chapter 5, we developed and employed a computational protocol to predict the binding of a novel ruthenium based inhibitors to different protein kinases to understand the selectivity profile of the new inhibitor scaffold.

1.2 Biological background on protein kinases

Protein kinases, one of the most important targets in the current cancer therapy, are the largest enzyme family involved in cell signal transduction [24, 124]. They are encoded by

approximately 2% of eukaryotic genes and more than 500 protein kinases are identified based on the human genome sequencing [168] and biochemical studies. Protein kinases catalyze the transfer of the γ-phosphate group from an ATP molecule to tyrosine, serine or threonine residues in proteins. This process plays an essential role in regulating many fundamental cellular processes [138]. Constitutive or inappropriate activation of protein kinases are seen in a variety of cancers and small molecule inhibitors are designed to target/inhibit kinase signaling in such scenarios [61]. Understanding the inhibition and phosphorylation of protein kinases are significant to guide cancer therapy. Kinases studied in this work, namely EGFR (our main focus) and other members in the ErbB family, GSK3-β, CDK2, and PIM1 are all well-established drug targets. The following sections will describe the background on protein kinases using the EGFR protein kinase as a template.

1.2.1 Protein kinases play key roles in cellular signaling

Protein kinases play important roles on cellular signaling. A typical example is the ErbB-mediated signaling network.

The ErbB family of receptors have received much attention, given their strong association with malignant proliferation [125]. The family has four members — the epidermal growth factor receptor (EGFR or ErbB1 or HER1), ErbB2 (or HER2), ErbB3, and ErbB4. They activate a multi-layered network mediating crucial pathways, including Ras-Raf-MAP-kinase, PI3K-Akt, and STAT, leading to cell proliferation, differentiation, migration and metabolism [63, 120, 218, 281] (Figure 1.1). Over-expression or mutational modification of the ErbB family are often implicated in a large percentage of

clinical cancers of various types [173, 233]. They, especially EGFR and ErbB2, have already been established as targets in cancer therapy. Small molecule inhibitors for EGFR are of significant interest as cancer therapeutic drugs [61]. A better understanding of these receptors is critical to discern the cancer mechanism and will help to design efficacious drugs.

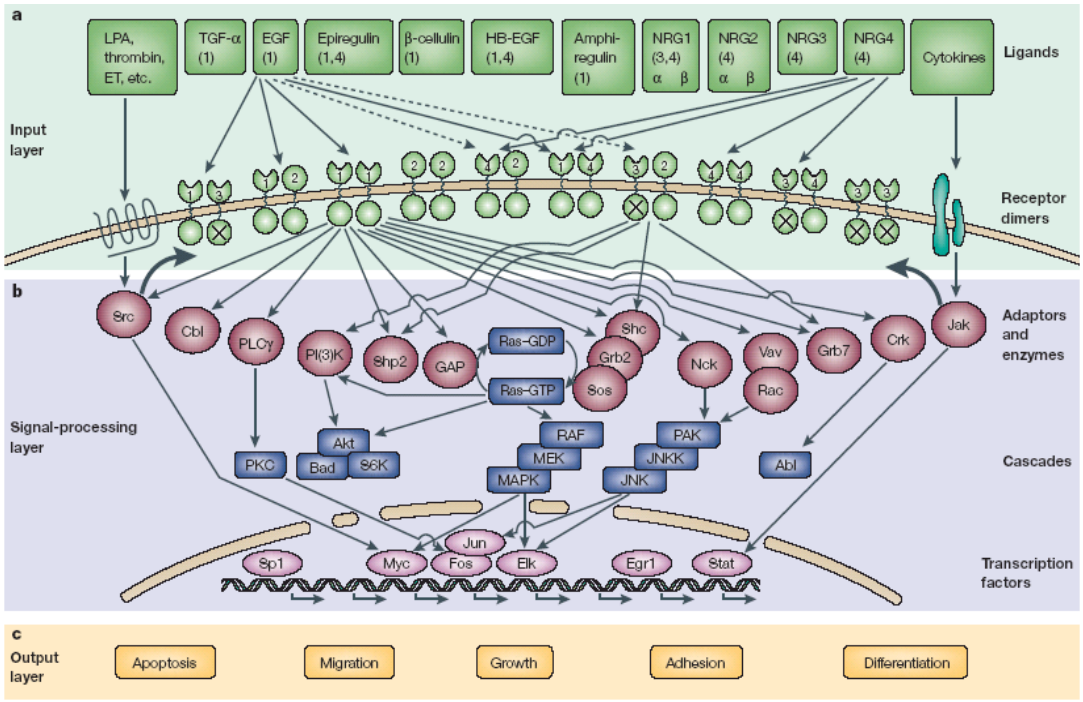

Figure 1.1 ErbB family signaling network [281].

The ErbB receptors share a common protein structure characterized by an extracellular region (serves as the ligand binding domain), a transmembrane domain, and an intracellular tyrosine kinase domain (the focus of this work) flanked by a juxtamembrane and a C-terminal regulatory region [63, 158]. In the presence of activating ligands, namely, the epidermal growth factor (EGF), transforming growth factor-α (TGF-α), and several other peptide growth factors [63], the family members associate to form

homo/hetero dimers, activating their intracellular tyrosine kinases using an unique allosteric mechanism [289], followed by phosphorylation of several tyrosine residues in the C-terminal tails. The phosphorylated tyrosines in the C-tails serve as docking sites for SH2 (Src homology 2) and PTB (phosphotyrosine binding) domains of signaling proteins, leading to activation of proteins involved downstream in the signaling cascades.

1.2.2 Active and inactive structures of protein kinases

A significant amount of work has been dedicated to characterizing the function of protein kinases. A fast expanding structural database is continuing to provide valuable insights into the molecular details of kinase function and regulation. It has been well established that most protein kinases are highly conversed in both sequence and structures [5]. Owing to the highly conserved nature of the core structure, a description of EGFR tyrosine kinase domain (TKD), a typical protein kinase, in this thesis will encompasses the main features of all members of the kinase family.

Similar to most protein kinases, EGFR TKD shares a two-lobed structural fold, an N-terminal lobe (N-lobe) and a larger C-terminal lobe (C-lobe) (Figure 1.2). The two domains, connected by a small linker region, form a binding pocket for ATP and the substrate. Several conserved components near the site of phosphoryl transfer play important catalytic functions, including the glycine-rich loop (G-loop) and αC-helix in N-lobe, as well as the activation loop (A-loop) and catalytic loop (C-loop) in the C-lobe.

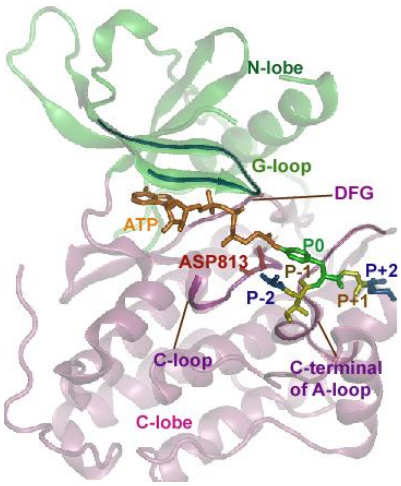

Figure 1.2 Key components of EGFR TK (PDBID: 2GS6 [288]).

Like many other kinases, two EGFR TKD conformations, namely active and inactive, have been captured by crystallography [248, 276, 284, 288]. The active EGFR TKD was solved both as an apo-kinase (i.e. in isolation) and in the presence of inhibitors, including erlotinib and gefitinib [248, 284], while the inactive EGFR TKD was crystallized by mutating a residue critical destabilizing the activating dimer interface or in complex with a bulky inhibitor, lapatinib. The most notable structural difference between active and inactive structure of the EGFR TKD is the outward rotation of the αC-helix and the conformational change of activation loop (Figure 1.3). In a fully active state, the A-loop adopts an open conformation so that it is positioned away from the catalytic center, providing a platform for substrate binding [248] (shown in Figure 1.2). The side chain of the aspartate (Asp831) in the DFG motif is directed into the ATP binding site, which is required to coordinate Mg^{2+} and to help orient the γ -phosphate group of ATP for efficient phosphoryl transfer [248, 289]. The αC-helix is positioned toward the catalytic core, as a result, the catalytic glutamate residue (Glu738) located nearly at the center of

αC-helix bridges a lysine residue (Lys721), conserved across several kinases, which is important in positioning the phosphates of ATP. In contrast, the inactive state is characterized by an outward shifting of the αC-helix and a closed conformation of the activation loop. The shift of αC-helix in the inactive EGFR TKD results in the loss of the Glu-Lys salt bridge and makes the pocket between the two lobes larger, which accommodate the extra fluorobenzyl-oxy moiety of lapatinib (Figure 1.3).

1.2.3 Protein kinase substrate recognition

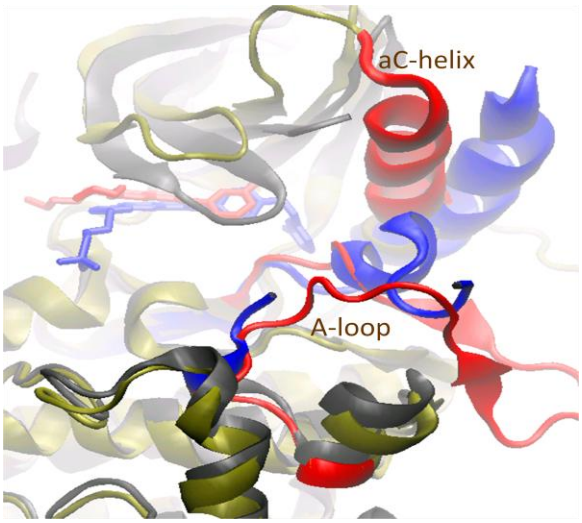

Figure 1.3 Active and inactive form of EGFR TKD. Blue: inactive and red: active.

Protein kinases regulate several essential signaling pathways by recognizing [225, 255], recruiting and phosphorylating a variety of substrates. Protein kinases poorly phosphorylate free amino acids and it is believed that the affinity and specificity of signaling cascades are in-part determined by the sequence identity of phospho-tyrosines on substrates [226]. That is, the specificity in-part depends on how the kinase recognizes the local sequence of substrate proteins at the P-site (shown in Figure 1.2), i.e., the amino

acids situated immediately proximal to the P-site [142, 261]; the nomenclature for these resides are P-1, P-2, P-3, etc., for those N-terminal to the P-site, and P+1, P+2, P+3, etc., for those C-terminal to the P-site. Experience has suggested that different kinases show preference for substrates with particular amino acid at the positions close to the P-site [28, 40, 43, 104]. EGFR TKD, for example, shows preference for substrates with glutamic acid at the P-1 position and large hydrophobic amino acids at the P+1 position [118, 123, 149, 245]. The precise relationship of such a sequence identity to substrate specificity and the catalytic mechanism is still not understood.

Besides the local sequences flanking the P-site, there are many factors influence substrate specificity of particular protein kinases. For example, CDK2 uses a hydrophobic patch, RXL, to help recruit its substrate p107, which locates about 40Å away from the site of phosphoryl transfer [225].

1.2.4 Protein kinase catalytic mechanism

Protein kinases deliver a single phosphoryl group from the gamma position of ATP to the hydroxyls of serine, threonine, and tyrosine in protein substrates. Extensive efforts have been made to understand the details of this phosphoryl transfer reaction. The cAMP-dependent protein kinase (PKA) is by far the best understood protein kinase from this point of view [167] and substantial progress has also been made with other kinases [16, 17, 30, 41]. Although several details are yet to be resolved, a common mechanism believed to be in operation among kinases can be defined:

1) Mg^{2+} ions are involved in the reaction. One Mg^{2+} ion is believed to be essential for catalysis [5]. This metal ion is chelated Mg^{2+} by the phosphates of ATP and one conserved aspartic acid (as observed in the crystallographic structure of PKA, PDBID:1ATP) [167, 256, 292]. A second metal ion is also observed in the active site of some enzymes when the crystal is soaked with high concentration of Mg^{2+} ions. The coordination of this metal varies among the different structures available [5].

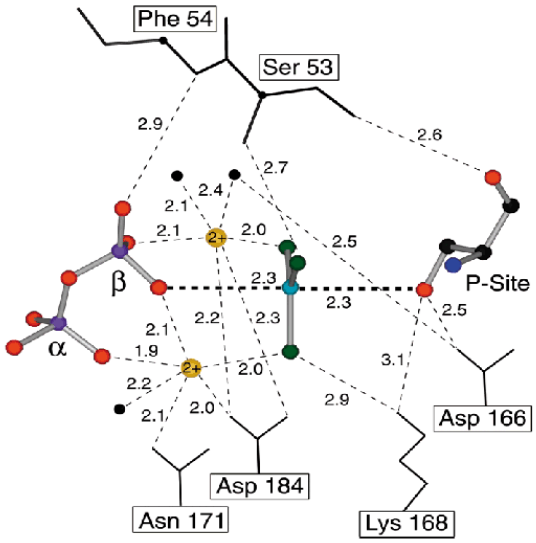

Figure 1.4 Stereo view of Mg2ADP, AIF3, SP20 (substrate serine), and catalytically important residues in the catalytic subunit of cAMP-dependent protein kinase.

2) In phosphoryl-transfer by kinases and polymerases, nucleophilic attack on the target phosphate proceeds via a conformation that resembles a trigonal-bipyramidal transition state [154, 249], shown in Figure 1.4.

3) The phosphoryl-transfer can occur through either an associative or a dissociative mechanism [64] (showed in Figure 1.5). In the dissociative mechanism, the cleavage of the bond between the phosphorus center and the leaving group oxygen precedes

the bond formation with the attacking nucleophile. On the contrary, in associative mechanism, the bond formation between the attacking nucleophile and the phosphorus center precedes the bond cleavage. Although it is well accepted that the dissociative mechanism is the preferred mechanism for reaction in solution for monoester [29, 145, 175, 220, 292] which are particularly relevant for the reaction catalyzed by protein kinases, discerning the transition state of protein kinase catalyzed reaction is complex, resulting in controversial experimental interpretations [145, 167, 294].

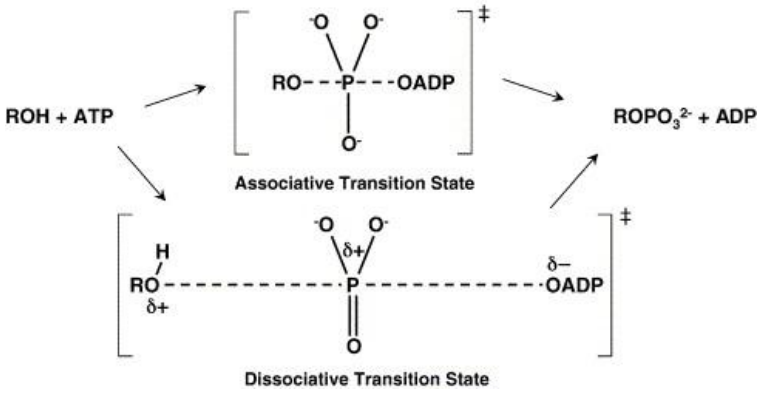

Figure 1.5 Dissociative vs. associative mechanisms for phosphoryl transfer [5].

4) The potential roles of specific residues around catalytic site are still not clearly understood. There is a debate over the role of the conserved aspartic residue (D813 in EGFR or D166 in PKA) in the catalytic loop [160]. This conserved aspartate is clearly important for efficient catalysis. Replacement of this residue with alanine in PKA and PhK results in decrease in k_{cat} by 2-4 orders of magnitude [100, 241]. However, the potential function of this conserved residue is ardently debated. The facts that in PKA as well as in other kinases, the aspartate (D166 in PKA) forms a

hydrogen bond with the substrate peptide and that the second-order rate constant, k_{cat}/K_m, for the phosphorylation of peptides is pH sensitive, support a general-base mechanism [282]. However, it is also shown that high negative charge on the hydroxyl could inhibit the reaction if charge repulsion with the γ phosphate is considered [145], which limits the advantage of a general base reaction mechanism. Another proposal is that the carboxyl group may position the hydroxyl for productive attack of the γ phosphate of the ATP and the proton would not be transferred to the aspartate. Computational modeling methods may serve to further clear this debate.

1.2.5 Pharmacophore model of kinases and ATP-competitive inhibitors

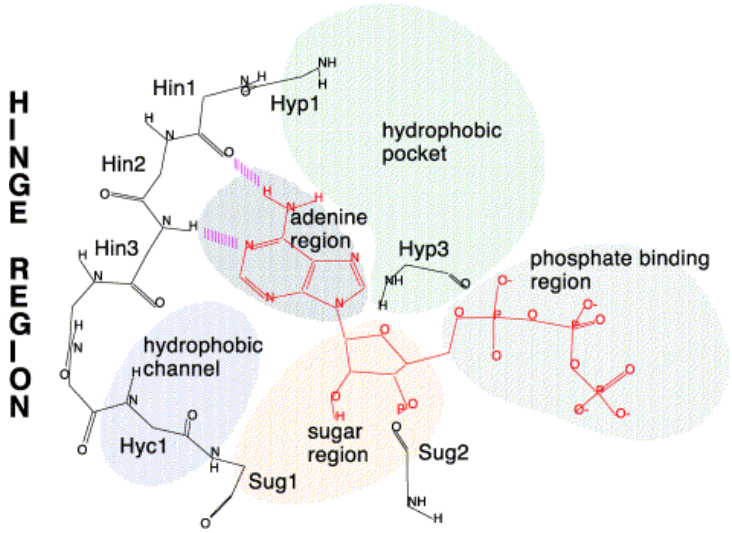

Figure 1.6 Pharmacophore model of the ATP-binding site of protein kinases. ATP is in red. Sug1, Hyp1, and Hyc1, Hin are residues lining the sugar region (Sug), hydrophobic pocket (Hyp), hydrophobic channel (Huc) and the hinge region respectively [82].

Protein kinases have been identified as one of the most important targets in the current cancer therapy. One of ways to inhibit abnormal kinase activity is to design small

molecule inhibitors to bind into the ATP binding cleft, block ATP binding and therefore inhibit the phosphorylation. The success of small-molecule ATP-competitive inhibitors such as imatinib (Gleevec) for the treatment of chronic myeloid leukemia (CML) and gastrointestinal stromal tumors (GIST) confirmed that this strategy is indeed effective [42]. In spite of the success of Gleevec, the design of ATP-competitive inhibitors that are selective (specific) for a particular kinase appears to be quite challenging due to the conserved nature of protein structures.

X-ray crystal structures for a range of kinases are available in the Protein Data Bank (PDB), which provide structural basis for understanding kinase inhibition and facilitate structure-guided design of kinase-specific inhibitors. The catalytic clefts of protein kinases usually include two hydrophobic pockets (Hyp1 and Hyc1), adenine region, 'sugar' pocket and the phosphate groove, shown in Figure 1.6. The adenine region contains the two key hydrogen bonds formed by the interaction of the N-1 and N-6 amino groups of the adenine ring with the backbone NH and carbonyl groups of the adenine anchoring hinge region (Met769 in EGFR) of the protein kinase. Many potent inhibitors use at least one of these hydrogen bonds. The hydrophobic pocket is not used by ATP, but is exploited in the design of most kinase inhibitors. It plays an important role for inhibitor selectivity, and its size is different in active and inactive kinase states. The hydrophobic channel opens to solvent and is not used by ATP and hence, it can be exploited in the design of inhibitors. The phosphate binding region offers little opportunity in terms of inhibitor binding affinity due to high solvent exposure. However, it can be utilized in improving selectivity.

1.3 Computational methods and their applications in biological studies

Besides the chemical approaches, computational methods (described below) are also widely used to understand the inhibition and phosphorylation mechanism of protein kinases.

1.3.1 Molecular dynamics (MD) simulations

Empirical force field-based molecular dynamics (MD) simulations [19] are widely applied to refine molecular structures, obtain information on the time evolution of conformations that is otherwise elusive. Along with a range of complementary computational approaches, MD simulations have the ability to analyze the relationship of structures to energies, which experimental approaches can only access indirectly, and thus have become invaluable for structure/function analyses in proteins.

To perform MD simulations, the starting point is defining the initial coordinates and initial velocities of the atoms characterizing the model system, for example, the desired biomolecule plus the biologically relevant environment; i.e. water molecules or other solvent and/or membranes. The coordinates of the desired biomolecule can usually be found as structural data (X-ray or NMR) deposited into the protein data bank (PDB); otherwise it is possible to derive initial geometry and coordinate data from model building techniques, including homology methods. This step also typically includes the placement and positioning of the environment of the molecules (solvation, ionic strength

etc.). The initial velocities are typically derived from the Maxwell-Boltzmann distributions at the desired temperature of the simulation. The potential of interactions of each of the atoms are calculated using a force field, which parameterizes the non-bonded and bonded interaction terms of each atom depending on its constituent atom connectivity: bond terms, angle terms, dihedral terms, improper dihedral terms, non-bonded Lennard-Jones terms, and electrostatic terms. The potential interactions are summed across all the atoms contained in the system, to compute an overall potential energy function for the system [103]:

$$U(\vec{R}) = \sum_{\text{bonds}} \mathrm{K_b}(\mathrm{b}-\mathrm{b_0})^2 + \sum_{\text{angles}} \mathrm{K_\theta}(\theta-\theta_0)^2 + \sum_{\text{dihedrals}} \mathrm{K_\chi}(\chi-\chi_0)^2 + \sum_{\text{nonbond}} \varepsilon_{ij}\left[\left(\frac{\mathrm{R_{ij}^{min}}}{\mathrm{r_{ij}}}\right)^{12} - \left(\frac{\mathrm{R_{ij}^{min}}}{\mathrm{r_{ij}}}\right)^{6}\right] + \sum_{\text{nonbond}} \frac{\mathrm{q_i q_j}}{\mathrm{Dr_{ij}}}$$

Taking the derivative of the potential energy function yields the force, and from Newton's second law, this is equal to mass times acceleration. Although, the process seems simple, the derivative function results in a set of 3N-coupled 2nd order ordinary differential equations that must be solved numerically. The solution consists of a numerical recipe to advance the positions and the velocities by one time step. This process is repeated over and over again to generate MD trajectories of constant energy. Constant temperature dynamics are derived by coupling the system to a thermostat using well established formulations such as the Langevin dynamics or the Nose-Hoover methodologies [37]. Application of MD simulations to biomolecules is facilitated by several popular choices of force fields such as CHARMM27 [271] (www.charmm.org), AMBER [227] (www.ambermd.org), and GROMOS[192] (www.gromacs.org), as well as

dynamic simulations packages and visualization/analysis tools such as NAMD [122] (www.ks.uiuc.edu/Research/namd/) and VMD [206] (www.ks.uiuc.edu/Research/vmd/).

With analysis of molecular dynamics trajectories, it is possible to calculate statistical properties under a variety of initial and other external conditions [162] such as hydrogen bond analyses for hydrophilic interactions and solvent accessible surface area for hydrophobic interactions. For example, by analyzing the relative positions of the hydrogen bond donors to the hydrogen bond acceptors with a preset cutoff angle and a bond length, the hydrogen bonds present in the majority of a given trajectory can be identified to record permanent stabilizing interactions and differentiate them from transient interactions [169, 170]. Similarly, using the solvent accessible surface area analysis, i.e. by mapping a surface area created using a probe sphere of 1.4 Å and accumulating the statistical data, provides a quantitative metric of hydrophobic stabilization effects.

The utility of MD simulations generally depend on the accuracy of the underlying force-field. Since force-fields are created using empirical energy functions, they are parameterized and tuned to specific class of molecules; this introduces a constraint on their transferability to model non-native systems or environments and results must often be compared to experimental results, not only to verify their accuracy, but also to identify where methodological improvements can be made. Thus, there is continued development of the basic force-field as well as the simulation methodology.

Another important consideration is the ability to perform sufficient sampling of the combinatorially large number of conformations available to even the simplest of

biomolecules [81]. In this respect, a potential disadvantage of molecular dynamics calculations is that there is an inherent limitation upon the maximum time step used for the simulation ($\leq$ 2 fs). Solvated systems of protein monomers typically consist of 40000 atoms, and those of higher order complexes such as dimers or membrane-bound proteins can be as large as 200,000 to 500,000 atoms. For such system sizes, with current hardware and software, simulation times extending into the microsecond regime and beyond is an exceedingly difficult and labor intensive endeavor which requires a combination of algorithmic enhancements as well as the utilization of high-performance computing hardware infrastructure. For example, cutoff distances reduce the number of interactions to be computed without loss of accuracy for short-range interactions but not for long-range (electrostatic) interactions; to help maintain accuracy, long-range corrections such as the particle mesh Ewald algorithm [105, 252] along with periodic boundary conditions are typically implemented. Parallelization techniques enable the execution of the simulations on supercomputing resources such as 4096 processors of a networked Linux cluster. Although a cluster of this size is a big investment, its accessibility is feasible through the US National Science Foundations's TeraGrid Initiative (founded in 2001) for academic researchers. TeraGrid resources (www.teragrid.org) currently include more than a petaflop of computing capability and more than 30 petabytes of online and archival data storage, with rapid access and retrieval over high-performance networks. Capitalizing on advances in hardware architecture, another approach is the creation of custom hardware for MD simulations, and offers one-two orders of magnitude enhancement in performance; examples include MDGRAPE-3 [230, 232] and ANTON [95, 251]. Recently, graphical processing unit

(GPU) accelerated computation has come into the forefront to enable massive speed enhancements for easily parallelizable tasks with early data indicating that GPU accelerated computing may allow for the power of a supercomputing cluster in a desktop, see examples [36, 109].

1.3.2 Molecular docking

Molecular docking can be defined as the prediction of the "correct" bound structure of the complex involving a receptor (typically a protein) and a ligand (other proteins, nucleic acids or small molecule inhibitors) [181]. The idea behind docking is to generate a comprehensive set of conformations of the receptor-ligand complex and then to rank them according to their stability. One essential component for the success of a docking program is an algorithm for conformational space searching. Ideally, conformational searching during docking for both receptor and ligands are necessary. However, to enhance computational efficiency, ligand flexibility is considered by most existing softwares, while receptors (or proteins) are usually treated as rigid during docking. However, the frequently observed changes in receptor structure upon ligand binding indicate the importance of receptor flexibility [274] and different versions of receptor flexible models with varying degrees of sophistication have been proposed. Among them, using MD simulations to pre-generate protein conformers for docking [178] will be used in this thesis. Another essential part of a docking program is a fast yet accurate scoring function. Many docking programs, for reasons of computational efficiency, usually approximate the binding free energy based on the evaluation of a single structure. These

scoring functions usually assume that the binding free energy can be estimated by a linear combination of pairwise terms [204]:

$$\Delta G = \Delta G_{vdW} + \Delta G_{hbond} + \Delta G_{elec} + \Delta G_{conform} + \Delta G_{tor} + \Delta G_{sol}$$

where, the first four terms are molecular mechanics terms, namely, dispersion/repulsion, hydrogen bonding, electrostatics and deviations from the covalent geometry; the fifth term models the restriction of internal rotors and global rotation and translation; and the last term accounts desolvation upon binding and the hydrophobic effect. This class of scoring functions is of practical use for molecular docking and is computationally tractable. On the other hand, they certainly have limitations in terms of the accuracy with which they represent the free energy of binding due to the various simplifying assumptions.

1.3.3 Free energy perturbation (FEP)

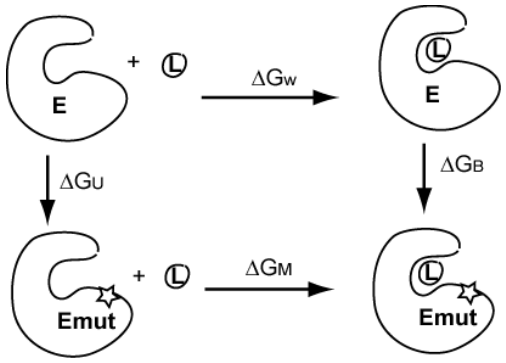

Figure 1.7 Thermodynamic cycle for the calculation of binding affinity change of ligand/substrate due to a mutation.

In this thesis, we designed a thermodynamic cycle as shown in Figure 1.7 and employed the free energy perturbation (FEP) method to calculate the effect of how mutation will affect small molecule binding. To calculate the binding free energy change $\Delta\Delta G_{W\to M}$, we can calculate the difference between the free energy changes caused by the particular mutation in the bound state (ΔG_B) and the unbound state (ΔG_U). The binding affinity change due to the mutation can be expressed as:

$$\Delta\Delta G_{W\to M} = \Delta G_M - \Delta G_W = \Delta G_B - \Delta G_U$$

The free energy changes caused by the particular mutation in the bound state (ΔG_B) and the unbound state (ΔG_U) were calculate by free energy perturbation (FEP) method. In general, the free energy difference between two related system A and B, represented by the Hamiltonians of H_A and H_B, can be calculated by

$$\Delta G = G_B - G_A = -RT \ln\langle e^{-(H_B - H_A)/RT}\rangle_A$$

where R is Boltzmann's constant, T is the temperature, and $\langle \quad \rangle_A$ refers to an ensemble average over a simulation run for state A.

In our simulation, systems A and B are defined by the wildtype and mutation states of bound or unbound system. In order to improve accuracy, the two thermodynamic states are broken into windows, each one involving a small enough interval to allow the incremental free energy change to be calculated accurately. A Hamiltonian $H(\lambda)$ is defined as

$$H(\lambda) = \lambda H_B + (1-\lambda)H_A$$

Where λ can vary from 0 ($H = H_A$) to 1 ($H = H_B$). The accumulated free energy change is calculated by:

$$\Delta G = G_B - G_A = -RT \sum_{\lambda=0}^{1} \ln\langle e^{-(H_{\lambda+d\lambda} - H_\lambda)/RT} \rangle_\lambda \ .$$

1.3.4 MMPBSA

While approximate binding free energies are suitable for high-throughput docking, stringent rankings or accurate free energies are often needed for lead optimization and more detailed examination of binding at the molecular level. Thermodynamic integration (TI) and free energy perturbation (FEP) [150, 253] are among the most rigorous free energy methods available currently. However, they are too computationally expensive to be widely employed. MD-based methods such as Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) [153] can achieve similar accuracy at a smaller computational cost, therefore commonly used for virtual screening of compounds. It has produced reasonable binding energies for several systems, and also been employed as a post-docking filter in the virtual screening of compounds [136, 188, 254]. *In this thesis, we use MMPBSA to re-rank the top predicted bound conformation generated by molecular docking.*

In a MMPBSA method, the binding affinity between ligand (L) and the protein receptor (R) is estimated from the free energies of three species,

$$\Delta G_{bind} = G(RL) - G(R) - G(L) \ ,$$

where all species are assumed to be in water solution. The free energy of each of the species is estimated as a sum of four terms:

$$G = \langle E_{MM} \rangle + \langle G_{solv} \rangle + \langle G_{np} \rangle - T\langle S_{MM} \rangle$$

where G_{solv} is the polar solvation energy of the molecule, estimated by the solution of the Poisson-Boltzmann (PB) equation, G_{np} is the nonpolar solvation energy, estimated form the solvent-accessible surface area of the molecule, T is the temperature, S_{MM} is the entropy of the molecule, estimated from a normal-mode analysis of harmonic frequencies calculated at the molecular mechanics (MM) level, and E_{MM} is the MM energy of the molecule, i.e., the sum of the internal energy of the molecule (i.e., bonded terms), the electrostatics and van der Waals interactions. All the terms in the above equation are averages of energies obtained from a number of snapshots taken from MD simulations. As the computationally most intensive step of the MMPBSA method is the generation of snapshots (geometries) employed in the energy calculations through MD simulations, in order to minimize computing resources, the same geometry is employed for all three species (complex, ligand, and receptor), i.e., only the complex is simulated by MD. Thereby, the bonded term in MM energy cancels out in the calculation.

The MMPBSA model, like other implicit models, is derived from the potential of mean force based on the assumption that electrostatic and nonpolar contributions to the free energy can be treated separately in a additive fashion (this represents a mean-field assumption). In the MMPBSA approach, we calculate the salvation terms based on a pre-defined model for atomic radii and solute dielectric constant, which are simplifying

assumptions. The method also makes simplifying assumptions on conformational sampling in order to gain efficiency.

1.3.5 Electronic Structure Methods

Molecular simulations using the empirical force-field approach are rooted in the validity of classical mechanics. Electronic structure methods [71, 89, 96, 99, 196, 207, 209, 228, 240, 268, 269, 290] that relieve this fundamental assumption are often necessary in force-field development as well as in understanding chemical reaction pathways and catalytic mechanisms. While electronic structure methods are computationally rather demanding, recent advances in mixed quantum mechanics molecular mechanics (QM/MM) methods also enable us to combine high level electronic structure methods with molecular mechanics [22, 44, 46, 68, 72, 76, 78, 87, 88, 106, 127, 128, 146, 157, 178, 190, 191, 221, 243, 279, 280, 283, 285, 286, 296]. This multiscale description of force-fields provides a route to extend the electronic structure methods to the nanometer scale to enable the study of biomolecular systems. Compared to their classical counter-parts [268, 269], the quantum approaches (such as the empirical valence bond [13, 67, 92, 99, 108, 291]) and mixed QM/MM approaches are more challenging to apply in the biochemical context, though several successful demonstrations are available [3, 166]. *In this thesis, we employed electronic structure methods in force-field development as well as in the study of catalytic reaction mechanisms.*

Electronic Structure Methods in Force-field Development: In order to obtain an accurate description of molecular parameters (force-field), we require specification of geometric parameters such as bond, angle, dihedral angles etc., energetic terms such as partial

charges, bond/angle/dihedral stiffness, and contributions of solvation-desolvation. Consistent with existing biomolecular force-fields, it is thus essential to parameterize any new compound/molecule/chemical modification in the model and extend the force field for accurate representation of the model system.This procedure usually involves a series of steps [254] described below. The geometry of new compounds is optimized using an electronic structure method such as Hartree-Fock (HF) [254] and by using a set of basis functions to represent the electronic wave function. Such calculations are facilitated by the aid of the quantum chemistry software such as GAUSSIAN (www.gaussian.com [91]). The procedure of geometry optimization of the molecule returns the equilibrium geometry of the molecule, the spatial electron-density at 0 K, from which the partial (Mulliken) charges on each atom are inferred. Moreover, the frequency spectrum of the molecule to characterize its vibration spectrum in the ground state is also computed using the analysis of normal modes. In order to represent delocalized (low-frequency) motions such as torsional oscillations, a torsional energy scan is also performed.

In determining empirical force field parameters for new compounds, consistency and compatibility with the parent force-field is an important criterion. For example, the CHARMM27 force-field is commonly used as an all-atom empirical force field for biomolecules; it is therefore necessary to develop force-field parameters for new compounds consistent with the rest of the CHARMM27 force-field.

To obtain these parameters [33], partial atomic charges are calculated using CHarges from ELectrostatic Potentials using a Grid based method (CHELPG) [178] in GAUSSIAN and further adjusted to reproduce interaction energy of H-bond

donor/acceptor atoms in the inhibitor compound with water molecules. The intramolecular parameters such as molecule geometry and force constants are determined by using structural and vibrational data derived from electronic structure calculations. Intramolecular force constants are initially assigned based on existing parameters of compounds with similar structure and further refined following an automatic procedure using a genetic algorithm (GA) to optimize a merit function [263], which matches both eigenvalues and eigenvectors estimated using the electronic structure calculations (i.e. using GAUSSIAN) to those resulting from a normal mode analysis calculated from classical molecular mechanics calculations, i.e. using CHARMM [228, 268].

Quantum Mechanics Molecular Mechanics (QM/MM) Simulations: In the QM/MM simulations, the system is sub-divided into two sub-regions, the quantum mechanical sub-region (QM region) where the reactive events take place, and the molecular mechanical sub-region (which provides the complete environment around the reactive chemistry) [188]. Since electronic structure methods are limited by the number of atoms they can handle (typically 50-500), the QM sub-region is restricted to a small number of atoms of the total system. For example, in an enzymatic system, the quantum region can consist of Mg^{2+} ions, water molecules within 3 Å of the Mg^{2+} ions, parts of the substrate molecules and the catalytic amino acid residues (such as aspartic acids). The remaining protein and solvent molecules are treated classically using the regular classical force-field (such as CHARMM27).

In QM/MM simulations, wave function optimizations are typically performed in the quantum (or QM) sub-region of the system using an electronic structure method such as

density functional theory (DFT) [180, 228]. In this step, the electrostatic coupling between the QM and the MM sub-regions is accounted for: i.e., the charges in the MM sub-region are allowed to polarize the electronic wave functions in the QM sub-region. The forces in the quantum sub-region are calculated using DFT, on-the-fly, assuming that the system moves on the Born-Oppenheimer surface [89]. That is, we assume a clear timescale of separation between the electronic and nuclear degrees of freedom and the electronic degrees of freedom are in their ground state around the instantaneous configurations of the nuclei. The forces on the classical region are calculated using a classical force-field. In addition, a mixed Hamiltonian (energy function) accounts for the interaction of the classical and the quantum sub-regions. For example, since the QM/MM boundary often cuts across covalent bonds one can use a link atom procedure [71] to satisfy the valences of broken bonds in the QM sub-region. Also, bonded terms and electrostatic terms between the atoms of the QM region and those of the classical region are typically included [222].

From a practitioner's stand-point, QM/MM methods are implemented based on existing interfaces between the electronic structure and the molecular dynamics programs, one implementation is between GAMESS-UK (www.cfs.dl.ac.uk [37]) (an *ab-initio* electronic structure prediction package) and CHARMM [71]. The model system can then be subjected to the usual energy minimization and constant temperature equilibration runs at the desired temperature using the regular integration procedures in operation for pure MM systems; it is customary to carry out QM/MM dynamics runs (typically limited to 10-100 ps because of the computationally intensive electronic structure calculations) using a standard 1 fs time step of integration. The main advantage of the QM/MM

simulations is that one can follow reactive events and dissect reaction mechanisms in the active site, while considering the explicit coupling to the extended region. In practice, sufficient experience and care is needed in the choices of the QM sub-region and the many alternative choices of system sizes, as well as the link-atom schemes need to be compared to ensure convergence and accuracy of results [290]. The shorter length of the dynamics runs in the QM/MM simulations (ps) relative to the MM MD simulations (ns) implies that sufficiently high resolution structures are usually necessary for setting up such runs as the simulations only explore a limited conformational space available to the system. Another challenge is an accurate and reliable representation of the mixed QM/MM interaction terms [238, 295]. These challenges are currently being overcome by the suitable design of next generation methods for electronic structure and molecular mechanics simulations [52, 193].

Chapter 2 Computational Delineation of Tyrosyl-Substrate Recognition and Catalytic Landscapes in the Epidermal Growth Factor Receptor Tyrosine Kinase Domain

2.1 Introduction

The recent availability of several crystal structures of the EGFR kinase domain [15, 248, 284, 289] together with the identification of several kinase-activating mutations [48, 164, 186] clinically related to enhanced drug response have brought to focus the importance of molecular details in the structure-function relationship, including mechanisms for substrate recognition and phosphoryl transfer.

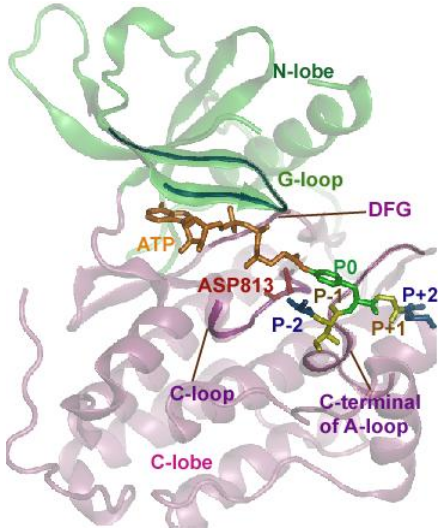

Figure 2.1 Structure of EGFR TKD co-crystallized with the bi-substrate analogs.

EGFR tyrosine kinase domain (TKD) is a typical tyrosine protein kinase, which catalyzes protein phosphorylation reactions by transferring the γ-phosphate group from an ATP molecule to the hydroxyl group of tyrosine residues in protein substrates [5]. It shares bilobe-fold catalytic domain common to several kinases [5, 30, 110] implying a conserved catalytic mechanism [5, 30, 110] in which ATP binds into the cleft between the N- and the C- lobes. In addition, two divalent metal (magnesium) ions are thought to be involved in the catalytic reaction [5, 110, 145, 256] along with a number of highly conserved regions in the catalytic domain, including the glycine-rich loop, the DFG motif of the activation loop as well as the catalytic loop serving to position the phosphate of ATP and the two Mg^{2+} ions. Then, by recognizing [225, 255], recruiting and phosphorylating a variety of peptide substrates, the kinase regulates several essential signaling pathways. Even though a complete elucidation of specificity in tyrosine phosphorylation by TKDs continues to be elusive[177], it is believed that the specificity of signaling cascades are in-part determined by the sequence identity of phosphor-tyrosines on substrates [226]. That is, the specificity in-part depends on how the kinase recognizes the local sequence of substrate proteins at the P-site (shown in Figure 2.1), i.e., the amino acids situated immediately proximal to the P-site [142, 261], denoted by P-1, P-2, P-3, etc. on the N-terminal side and P+1, P+2, P+3, etc. on the C-terminal side. It has been suggested that receptor tyrosine kinases including EGFR TKD show preference for substrates with glutamic acid at the P-1 position and large hydrophobic amino acids at the P+1 position [118, 123, 149, 245]. The precise relationship of such a sequence identity to substrate specificity and the catalytic mechanism is still not understood.

Experimental and theoretical studies have demonstrated that, in phosphoryl-transfer by kinases and polymerases, nucleophilic attack on the target phosphate proceeds via a conformation that resembles a trigonal-bipyramidal transition state [154, 249]. However, the phosphoryl-transfer can occur through either an associative or a dissociative mechanism (39, 41-44). In the dissociative mechanism, the cleavage of the bond between the phosphorus center and the leaving group oxygen precedes the bond formation with the attacking nucleophile. On the contrary, in associative mechanism, the bond formation between the attacking nucleophile and the phosphorus center precedes the bond cleavage. Although it is well accepted that the dissociative mechanism is the preferred mechanism for reaction in solution for monoester [64] which are particularly relevant for the reaction catalyzed by TKDs, discerning the transition state of protein kinase catalyzed reaction is complex, resulting in controversial experimental interpretations [29, 145, 175, 220, 292]. Besides, it is also important to understand the catalytic roles of specific residues around catalytic site of a protein kinase, in particular, the role of the conserved aspartic residue (D813 in EGFR) in the catalytic loop [145, 167, 294]. To address this question in other well-characterized kinase systems such as PKA, CDK2, insulin kinase, and MAPK, computational methods have been successfully used to investigate the phosphoryl transfer reaction [52, 53, 73, 74, 97, 260, 264, 293]. The challenge in extending such studies to EGFR kinase is that, while the structure for EGFR crystallized with a bi-substrate (ATP-peptide analog) is available [289] providing molecular resolution to the ternary complex of the kinase/substrate/ATP system, the covalent constraint between the ATP analog and the substrate and the fact that the peptide sequence in the bi-substrate ligand is a c-Src [159, 288] rather than a native EGFR substrate introduce an important challenge in

resolving the structural details of the catalytically productive ternary complex [28]; this brings to focus a related but equally important question on how EGFR kinase recognizes its native peptide substrates.

In this chapter, we employ methods of molecular modeling and simulation to (1) investigate how the EGFR TKD recognizes the local amino acid sequence of substrates proximal to the P-site as, and (2) characterize the EGFR TKD-catalyzed mechanism of phosphoryl transfer. Specifically, to address (1) we employ a computational protocol combining molecular dynamics simulations, ensemble molecular docking [161, 274], and Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) [102, 117, 150, 153, 231, 253, 266] to investigate the bound conformations of peptide substrates derived from two of autophosphorylation sites of the cytoplasmic tail (C-terminal) region of EGFR TKD [123], namely Y1068 (VPEYINQ) and Y1173 (NAEYLRV). We find that in our predicted bound conformations, the substrates bound into the same pocket as the bi-substrate peptide and that the P-1 glutamic acid residue and the hydrophobic amino acid at the P+1 position situate to interact with a highly conserved motif in the protein kinase family, namely Val852-Pro853-Ile854-Lys855-Trp856 in EGFR. This motif creates a binding surface to form hydrophobic contacts as well as hydrogen bonds to residues close to P-site (P0) tyrosine in the substrate and therefore helps to orient the tyrosine residue towards the active site. Using our predicted binary complex of EGFR TKD/substrate, we modeled the ternary complex structure of EGFR TKD/substrate/ATP and two Mg^{2+} ions, and address (2) by investigating the phosphoryl transfer of EGFR TKD using mixed quantum mechanics molecular mechanics (QM/MM) simulations. Our QM/MM simulations show that the EGFR protein kinase favors the dissociative mechanism,

although an alternative channel through the formation of an associative transition state is also possible. Our potential energy scans suggest that the kinase directs the reaction through the dissociative channel primarily using a hydrogen bond between the conserved aspartic acid residue in the catalytic loop and the hydroxyl group of the substrate. The alternative channel can sustain weak activity, but can become significant in non-cognate environments where the catalytic aspartic acid is mutated or the enzyme is locked in a constitutively inactive conformation. Collectively, our simulations establish key molecular rules in operation for the substrate-binding and catalytic landscapes in the EGFR TKD.

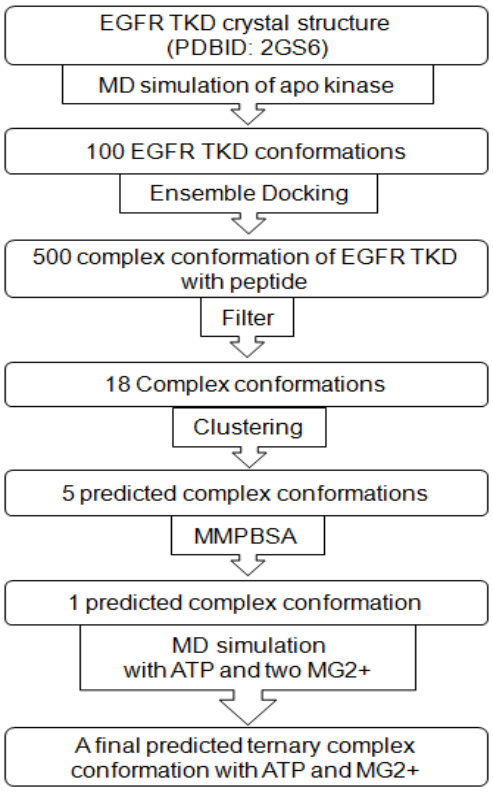

Figure 2.2 Docking protocol to predict peptide bound conformation to EGFR TKD

2.2 Methods and Materials

We employ a suite of computational methods including molecular dynamics and molecular docking (see Figure 2.2), umbrella sampling, and QM/MM simulations to investigate molecular mechanisms of substrate recognition (i.e. bound conformation of a tyrosyl-peptide) by EGFR TKD and the catalytic mechanism associated with the subsequent phosphoryl transfer reaction.

2.2.1 Molecular dynamics

All the molecular dynamics simulations in this thesis started with preparing the protein or protein complex structure by adding hydrogen atoms and missing residues using the CHARMM [37] biomolecular simulation package. The protonation states of histidine (HIS) residues were chosen according to recommendations from the WHATIF web interface (http://swift.cmbi.kun.nl/WIWWWI). Protonation states for all titratable groups distal to the catalytic site were chosen based on their individual pKa values in aqueous solution at a pH 7.0 [200]. The systems were then placed in a water (TIP3 model) box with 15Å padding with ions added randomly to neutralize the systems using the Solvate and Autoionize Plug-ins in VMD package [121]. The systems were minimized for 50000 steps with heavy atoms fixed and relaxed for another 50000 steps without any restraints. The systems were then gradually heated from 0K to 300K followed by 500ps equilibrating runs at constant pressure and temperature to allow the system to reach the proper density. A productive trajectory was then generated at 300K and 1atm. The simulations were performed using NAMD with CHARMM force field [166] with periodic boundaries enforced and long-range electrostatics taken into consideration.

2.2.2 Molecular docking

All the docking simulated in this thesis were performed with AUTODOCK 3.05 [178]. Grid maps were constructed as 126×126×126 points with a grid spacing of 0.184 Å to cover the pre-defined binding pocket. The Lamarckian Genetic Algorithm (LGA)[178] was applied to explore the conformational space of the substrate. In each docking run, the initial population was set to 100 individuals; the maximum number of energy evaluations is set to 10^8; and the generation of GA run is set to 30,000.

2.2.3 Docking of substrate tyrosyl-peptides Y1068 and Y1173 to EGFR TKD

The active conformation of the wildtype EGFR TKD domain was modeled based on the 2GS6 structure from the Protein Data Bank [289]. A 10 ns dynamics trajectory was generated by molecular dynamics simulation and an ensemble of 100 protein conformations was derived by sampling at uniform intervals from the trajectory**.** Two 7-residue peptides derived from the c-terminal region of EGFR (corresponding to Y1068: namely, VPEYINQ and to Y1173: namely, NAEYLRV) were docked to each snapshot in the 100-member ensemble by molecular docking. Five possible bound conformations were predicted for each protein conformation, which in total yielded 500 bound conformations of tyrosyl-peptide/EGFR TKD for each peptide.

Since no constraints were imposed in the docking simulation, we employed a screening filter to identify those conformations in which the tyrosine-hydroxyl oxygen Y(P0):Oη was situated <6 Å from the $O\delta_2$ oxygen of the conserved catalytic aspartic acid residue (D813). The choice of 6 Å as the filter-cutoff was made to accommodate direct as well as water-mediated hydrogen bonding between the respective groups. The resulting

structures were aligned (based on the protein backbone) to the crystal structure of EGFR TKD in complex with the bi-substrate ligand, namely 2GS6; conformations in which the peptide overlaps with ATP conformation in the bi-substrate structure were further discarded. Hence, out of the 500 predicted conformations, only 18 of them satisfied the constraints of our filter and were organized into 5 clusters based on RMSD (based on the backbone atoms and a cluster cutoff value of 3.0Å). In each cluster, we identified the conformation with the lowest binding energy score as the representative for further analysis.

In order to further sort the binding conformations generated by the ensemble docking protocol, we employed MMPBSA [102, 117, 150, 153, 231, 253, 266] to calculate the binding free energy. For each of the predicted bound conformations, a 1 ns trajectory was generated using molecular dynamics. The MMPBSA energy of each modeled complex was calculated as the average of the single point MMPBSA scores of 500 snapshots derived from the 1 ns simulation; the waters molecules and mobile ions were removed before the MMPBSA score calculation. In the MMPBSA calculation, the molecular mechanics energy, U_{MM}, was evaluated using an infinite cutoff for non-bonded interactions. The electrostatic contribution to the solvation free energy, W_{PB}, was calcuated with Poisson-Boltzmann Solver in CHARMM [126]. The reference system was assigned a solvent dielectric of 1 and 0 M salt concentration and the solvated system was assigned a dielectric of 80.0 and 100 mM salt concentration. The nonpolar contribution to the solvation free energy, W_{SA}, was approximated with the suface area model ΔW_{SA}=[0.00542 kcal/mol/Å^2]×SASA +0.92 kcal/mol [215], where the molecular surface area (SA) was estimated with a 1.4 Å slovent probe radius. The translational and rotional

entropy loss upon binding was estimated based on the change of freedom of center of mass as well as orientations (Eular Angles) by following the procedure described by Swanson et.al. [253] . Other (vibrational) entropy changes were neglected in this thesis.

The top-ranked conformation of peptide Y1068 with EGFR TKD identified based on the MMPBSA score was chosen as the initial bound structure for further analysis of the catalytic mechanism. The ATP was placed in homology with the bi-substrate structure (PDB ID: 2GS6) [289] and two Mg^{2+} ions were placed in homology with those resolved in the crystal structure of PKA (PDBID:1ATP) [52]. A 10 ns molecular dynamics simulation was performed on the ternary complex of ATP/substrate-Y/EGFR TKD.

2.2.4 Sampling the Pre-Catalytic Free Energy Landscape

Since the positions and orientations of the ions and ATP relative to the substrate tyrosines in our initial modeled structures were determined either from a bisubstrate ligand complex or by homology with protein kinase A (see above), prior to QM/MM simulations, we explored the classical free energy landscape along an a priori-chosen reaction coordinate χ_1 (the tyrosyl O-ATP P_γ nucleophilic attack distance). For computing the free energy landscape, the probability distribution $P(\chi_1)$ is calculated by dividing χ_1 into several (6-7) windows. The histograms for each window were collected by harvesting classical molecular dynamics trajectories of length 200 ps per window using a harmonic restraint $0.5\times K\times(\chi_1-\chi_1^{i,0})^2$, $\chi_1^{i,0}$ being the pre-set of the reaction coordinate characterizing the i^{th} sampling window, and K=20 kcal/mol/Å^2 being the force constant for the restraint along χ_1. The sampling data from the different windows were then processed using the weighted histogram analysis method (WHAM) [213].

2.2.5 Quantum Mechanics Molecular Mechanics (QMMM) Simulations

The QMMM approach we adopted was based on an existing interface between GAMESS-UK [222] and CHARMM [37]. For the quantum region, we chose the two Mg^{2+} ions, water molecules within 5 Å of the Mg^{2+} ions, segments of the ATP, the tyrosine residue of the substrate, and two catalytic residues (D813 and D831), yielding a total of 60 atoms . In order to reliably handle the phosphorous chemistry, the quantum region was treated using density functional theory with a B3LYP exchange correlation functional and a 6-31G* basis set [254], and the remaining protein and solvent molecules were treated classically using the CHARMM27 force field. Since the QMMM boundary cuts across covalent bonds, we applied the single link atom procedure [89] to satisfy the valences of broken bonds in the QM region. The electrostatic interactions between the quantum region and the classical atoms belonging to groups at the QMMM boundary were discarded as this choice has been shown to enhance the accuracy in the resulting energies and geometries [71]. We have extensively explored this choice of the QMMM region (including sensitivity to the size of the QM region, and the functional/basis set combination) in our prior studies of closely related systems [203, 201, 202]; others have validated the choice of the link atom [71]. Non-bonded van der Waals (VDW) interactions were treated by applying a switching function at 10 Å and truncating the VDW potential energy smoothly at a cutoff distance of 12 Å. The forces in the quantum region were calculated on-the-fly, assuming that the system moves on the Born-Oppenheimer surface defined by the QMMM Hamiltonian. The system was first subjected to 1200 steps of the adopted basis Newton-Raphson (ABNR) minimization and subsequently QMMM MD simulations were performed during which the systems were

first heated to 300 K and then subjected to constant temperature of 300 K dynamics using a Langevin thermostat for 10 ps and a standard 1 fs time step of integration.

2.2.6 Catalytic Reaction Pathways

The exact reaction coordinate is complex and multi (high) dimensional and a quantitative estimate of the free energy landscape can only be obtained using efficient long-time sampling approaches [26, 38, 56, 79, 208], which remain prohibitively expensive for QMMM applications due to large computational cost. Although we believe that the true reaction coordinate is complex, in the spirit of the a recent study [212], we described the reaction pathways in terms of several simple reaction coordinates, χ. For mechanisms of phosphoryl transfer through associative mechanism, χ includes the reactive distance tyrosine O- ATP P_γ, tyrosine O- ATP O2/3β, and ATP P_γ- ATP O2/3β distances. For mechanisms of phosphoryl transfer through dissociative mechanism, in addition to the three distances, χ includes the coordinate for proton abstraction, namely tyrosine OH-D830:$O\delta_2$. In order to drive the system along a reaction coordinate, we performed restrained minimization as well as restrained sampling simulations using the QMMM Hamiltonian along χ by augmenting the Hamiltonian with a potential bias term $(1/2)\times K\times(|\chi|-\chi_0)^2$; here K is the effective spring constant (ranging between 10 and 30 kcal/mol/Å^2) and χ_0 are reference values for sampling in different windows of χ. Complete pathways were assembled by varying the offset distances χ_0 in increments of ±0.05 Å, where the positive and negative increments represent the directionality of the energy scan. The reaction paths were computed by energy-minimizations in the presence of restraints and recalculating the single-point energies in the absence of the constraints.

For our choice of the increments of χ_0, calculations performed in forward and reverse directions satisfied path reversibility in configurations as well as energy values. Following the energy scans, constant temperature restrained dynamics runs of 300 fs-1ps for each choice of χ_0 was also performed to ensure that the reaction path was robust to thermal fluctuations.

2.3 Results

2.3.1 Tyrosyl-Peptide Substrate Recognition by EGFR TKD

Following the computational protocols described above, 500 bound conformations were predicted using ensemble docking protocol for binding of EGFR TKD to a seven-residue native EGFR peptide with P0 corresponding to autophosphorylation site Y1068 (VPEYINQ). Since a large number of the resulting structures showed peptide conformations in which the tyrosine residue was far from the active site, we constructed a histogram of D: the distance between the tyrosyl-oxygen and the $O\delta_2$ oxygen of the conserved catalytic aspartic acid residue (D813), see Figure 2.3A, and only considered those conformations with D < 6 Å; this choice for the cut-off is intended to cover situations where the hydroxyl group of the peptide forms a hydrogen bond with $O\delta_2$ of D813, either directly, or through a water mediated interaction. In addition, conformations in which the peptide substrate overlaps with the position of ATP (quantified using the bi-substrate conformation in the PDBID:2GS6 structure) were further discarded. Thus, among the 500 predicted conformations, 18 structures of the TKD-substrate complex satisfied our constraints, which were categorized into 5 clusters based on root-mean-squared deviations (RMSD) of the backbone heavy atoms.

Table 2.1 Rescoring of the top five predicted peptide conformations predicted by ensemble docking (showed in Figure 2.3B) using MMPBSA based on 1-ns MD simulation. The RMSD (Å) is calculated between the backbone atoms of the central five residue of the modeled peptide to the backbone atoms of the five residue peptide portion of the bi-substrate in structure 2GS6.The unit of energy is Kcal/mol.

Conf.	$\Delta\bar{U}_{MM}$	$\Delta\bar{W}_{PB}$	$\Delta\bar{W}_{SA}$	$\Delta\bar{G}_{trans}$ $(-T\Delta\bar{S}_{trans})$	$\Delta\bar{G}_{rot}$ $(-T\Delta\bar{S}_{rot})$	$\Delta\bar{G}_{total}$	RMSD
1	-157.4	153.3	-7.7	5.2	4.2	-2.4	5.45
2	-131.4	121.9	-7.5	5.5	4.3	-7.2	5.09
3	-134.4	109.4	-6.5	4.9	4.3	-22.2	2.72
4	-206.3	197.4	-7.9	5.4	4.2	-7.3	5.97
5	-107.5	103.7	-7.3	5.2	4.3	-1.6	3.92

The five conformations with the lowest score in each cluster are depicted in Figure 2.3B, along with the bi-substrate conformation from the PDBID:2GS6 structure. The computed MMPBSA binding free energies ΔGs of the five conformations are provided in Table 2.1; as expected, the binding free energy is dominated by the difference between the unfavorable polar solvation and the favorable molecular mechanics terms. Interestingly, the conformation 3, which corresponds to the lowest MMPBSA score, also aligns most closely to the bi-substrate conformation as shown in Figure 2.3B and Table 2.1.

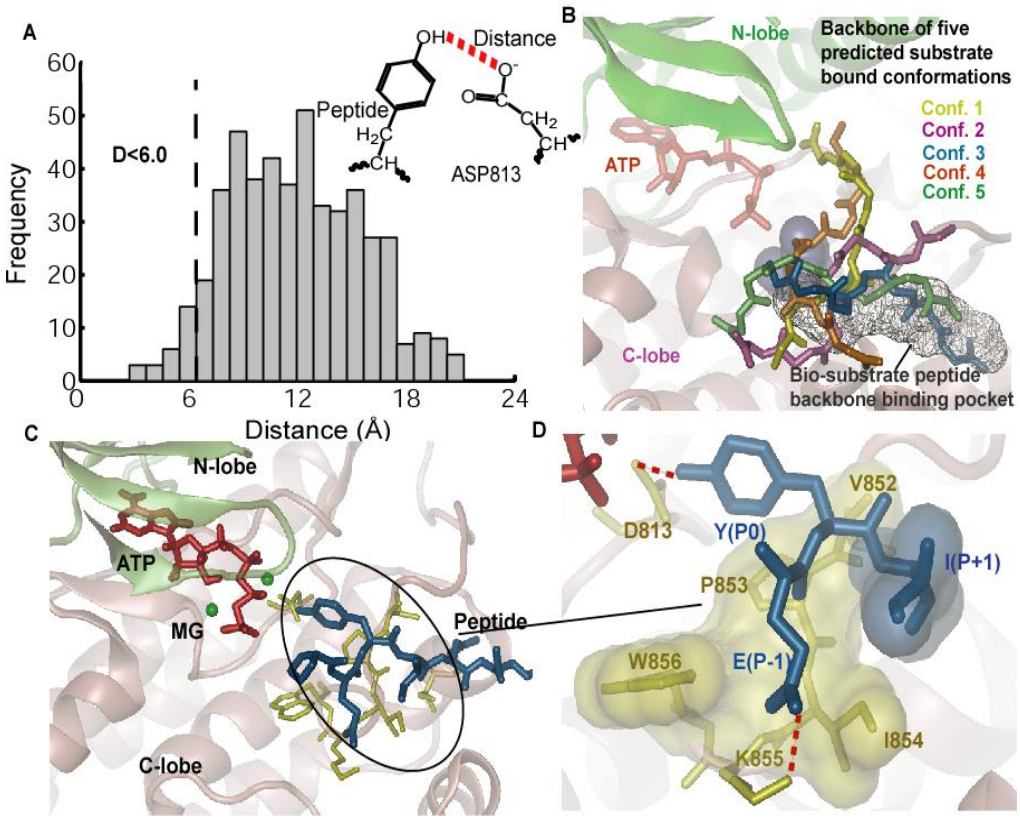

Figure 2.3 Results of docking peptide Y1068 to EGFR TKD. A) Histogram of the distance between the oxygen in hydroxyl group of the tyrosine and the oxygen atom in the carboxylate group of the D813 of all the 500 predicted conformations by ensemble docking. Conformations with this distance less than 6 Å were clustered into 5 clusters based on their backbone RMSD. B) Five predicted peptide clusters of Y1068 bound conformation generated by ensemble docking which satisfied the distance criteria in A. The bi-substrate peptide binding pocked are shown as the back surface area for comparison. All the five conformations were re-ranked using MMPBSA, shown in Table 2.1. C) The bound conformation of peptide Y1068 to EGFRTK in the presence of ATP and two MG2+ ions after 10 ns MD equilibrium. ATP are depicted in red, MG2+ in green, peptide in blue and residues in EGFR TKD which are interacted with the peptide in yellow. D) A close look of main interactions between the peptide and kinase: Y(P0) hydrogen-bonded with D813; E(P-1) formed a salt bridge interaction with residue K855; hydrophobic residue I(P+1) interacted with a hydrophobic surface in EGFR TKD consisted of residue V852-P853-I854-W856. All the hydrophobic residues are depicted as transparent surface to show the interaction.

A 9-ns MD simulation performed using conformation 3 in the presence of ATP and two Mg2+ ions reached steady state after 3 ns (Figure 2.4). The peptide backbone is stable

during the entire simulation as monitored by the backbone RMSD relative to the initial conformation (Figure 2.4). Therefore our predicted substrate conformation emerges to be in close alignment with the bi-substrate structure in PDBID:2GS6 structure as shown in Figure 2.3B. The side-chain residues, however, adjust conformations to result in stable hydrogen bonds and hydrophobic interactions in the final predicted bound structure of the ternary complex (EGFR TKD/tyrosyl-substrate/ATP) as shown in Figure 2.3C&D**.**

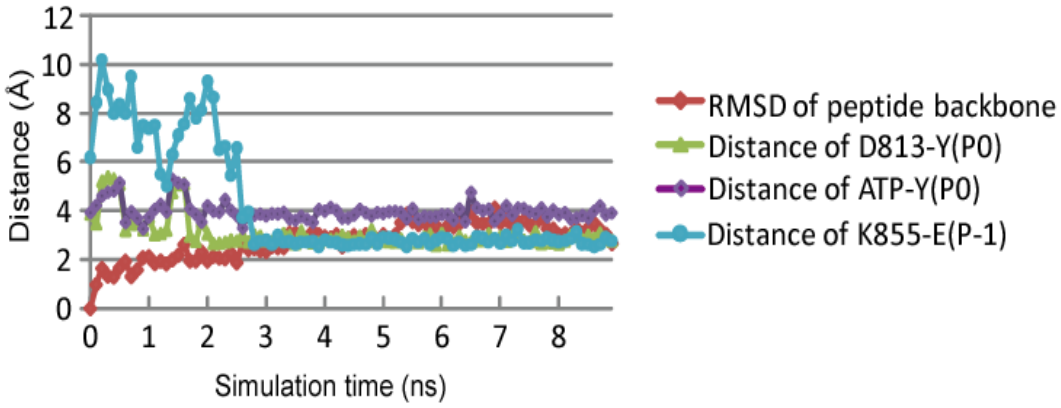

Figure 2.4 Evolution of RMSD and key distances for 10 ns MD simulation. The color schemes are: the RMSD calculated based the backbone atoms of the central 5 residue of the peptide conformation to the initial conformation (red); the distance between D813:Oδ_2 and Y(P0):Oη (green); the distance between Y(P0):Oη and ATP:P$_\gamma$ (purple); and the salt bridge distance between K855 and E(p-1) (light blue).

The predicted conformation (Figure 2.3C&D) provides new insights into the substrate recognition mechanism of EGFR TKD: (1) with respect to the P0-tyrosine, the C-terminal region of the peptide is anchored into a binding pocket formed by residues in the C-terminal region of EGFR activation loop, and the N-terminal of the peptide extends to the front of the C-terminal lobe of EGFR TKD (Figure 2.3C); (2) The hydroxyl oxygen of the P0 tyrosine is 2.9 Å away from the Oδ_2 oxygen atom of residue D813, within a hydrogen bond distance (Figure 2.3D); (3) the P-1 glutamic acid residue forms a salt-

bridge with a highly conserved lysine residue (K855); and (4) the isoleucine in P+1 position situates in a hydrophobic pocket formed by Val852, Pro853, Ile854 and Trp856 (Figure 2.3D).

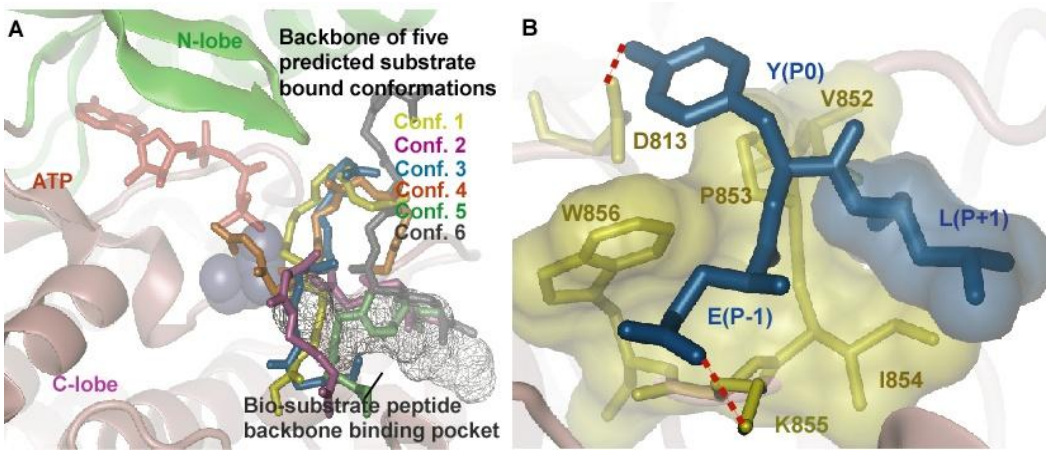

Figure 2.5 Results of docking peptide Y1173 to EGFR TKD. A) Six predicted peptide clusters of Y1173 bound conformation generated by ensemble docking which satisfied the distance criteria in Figure 2.3A. The bi-substrate peptide binding pocked are shown as the back surface area for comparison. All the 6 conformations were re-ranked using MMPBSA, shown in Table 2.2. B) A close look of main interactions between the peptide and kinase: Y(P0) hydrogen-bonded with D813; E(P-1) formed a salt bridge interaction with residue K855; hydrophobic residue L(P+1) interacted with a hydrophobic surface in EGFR TKD consisted of residue V852-P853-I854-W856. All the hydrophobic residues are depicted as transparent surface to show the interaction.

To further validate the robustness of this predicted mode of tyrosyl-substrate interaction, we also predicted the bound conformation of a different EGFR-native peptide sequence (NAEYLRV) derived from the auto-phosphorylation site Y1173 of EGFR (shown in Figure 2.5 and Table 2.2). The conformation predicted for Y1173 is very similar to that for Y1068 (Figure 2.5B). In particular, the hydrogen bond between the hydroxyl of tyrosine and D813, the salt-bridge between E(P-1) and K855, and the hydrophobic contact between L(P+1) and VAL852-Pro853-Ile854 are all preserved.

Table 2.2 Rescoring of the top six predicted peptide conformations predicted by ensemble docking (Figure 2.5) using MMPBSA based on 1-ns MD simulation. The RMSD (Å) is calculated between the backbone atoms of the central five residues of the modeled peptide to the backbone atoms of the five residues in the peptide portion of the bi-substrate in structure 2GS6. The unit of energy terms is Kcal/mol.

Conf.	$\Delta\bar{U}_{MM}$	$\Delta\bar{W}_{PB}$	$\Delta\bar{W}_{SA}$	$\Delta\bar{G}_{trans}$ $(-T\Delta\bar{S}_{trans})$	$\Delta\bar{G}_{rot}$ $(-T\Delta\bar{S}_{rot})$	$\Delta\bar{G}_{total}$	RMSD
1	-436.5	394.6	-8.0	5.9	4.5	-39.4	9.6
2	-238.0	217.1	-6.9	4.7	4.3	-18.8	10.0
3	-243.8	226.2	-8.1	5.4	4.3	-16.1	7.9
4	-114.2	111.8	-7.4	5.1	3.9	-0.7	10.6
5	-297.0	241.3	-7.1	5.3	4.7	-52.8	3.6
6	-250.1	228.5	-8.0	5.4	4.3	-19.9	9.7

2.3.2 Pre-Catalytic Conformational Landscape Orchestrates Multiple Reaction Channels

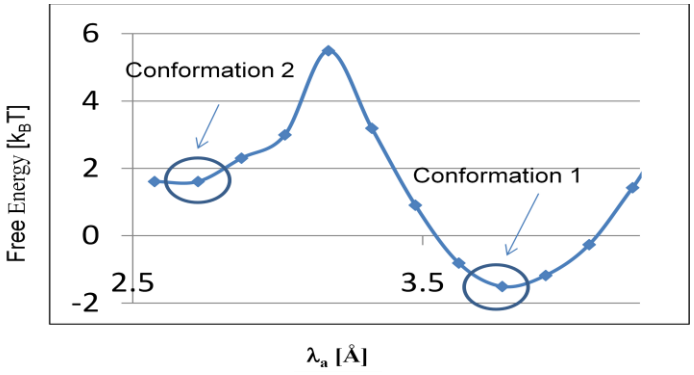

Figure 2.6 Pre-Catalytic Conformational Landscape along the the nucleophilic attack distance λa.

During the course of the MD simulation of the ternary complex, the conformational equilibration of ATP and the two Mg^{2+} ions relative to the substrate tyrosine resulted in small rearrangements of the phosphate tail of ATP within the active site. Hence, we chose to explore the free energy landscape for the pre-organization of the nucleophilic attack

distance λ_a (Y(P0):Oη - ATP:P_γ distance) as this distance is most relevant to catalysis and most sensitive to the phosphate tail conformation (Figure 2.6). In conformation 1, obtained from energy minimization of our model structure, $\lambda_a \approx 4$ Å. However, in the course of umbrella sampling, the system settles into another metastable state with λ_a <3 Å (conformation 2), surpassing a nominal free energy barrier of 3.5 k_BT (2 Kcal/mol). This subtle pre-catalytic conformational transition corresponds to a conformational rearrangement of γ-phosphate of ATP (compare snapshots). An analysis of conformations within the window corresponding to conformation 2 revealed that the proton of the tyrosine-hydroxyl (Hη) is primed for transfer to two possible positions owing to two mutually exclusive hydrogen bonds: (1) hydrogen bonded to $O\delta_2$ oxygen of D813 (indicated by a small value of δ_I), and (2) hydrogen bonded to the O1γ oxygen of ATP (indicated by a small value of δ_{II}).

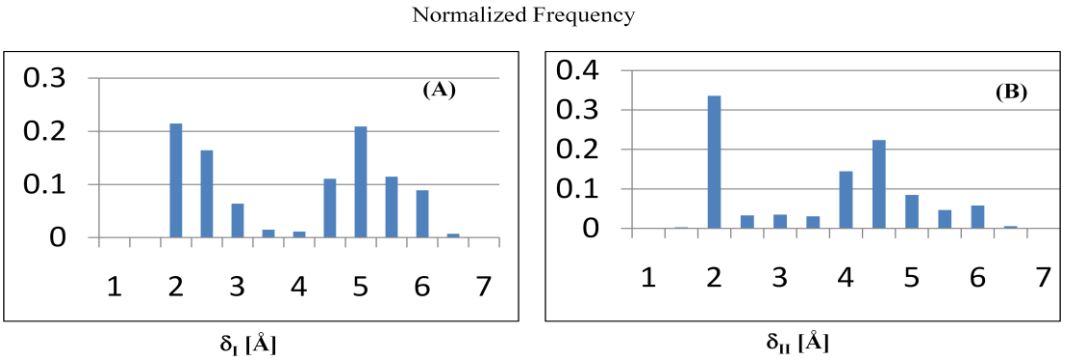

Figure 2.7 Normalized histograms representing the distributions of values for the distances between the proton of the substrate tyrosine hydroxyl and (A) the Oδ2 oxygen of D813 in EGFR or equivalent in ErbB3 (δI) or (B) the O1γ oxygen of ATP (δII) recorded in our simulations of the ground state (reactant) dynamics.

We hypothesize that the bimodal distributions of δ_I and δ_{II} values represented in the normalized histograms in Figure 2.7 serve to select either of two competing catalytic routes, labeled as pathways I and II in Figure 2.8 for proton transfer and concomitant phosphoryl transfer. We explore this hypothesis further by investigating the possible mechanisms of phosphoryl transfer in EGFR TKD following the protocols for QM/MM minimization and QM/MM MD simulations. Our simulations (see Figure 2.9) reveal that the phosphoryl transfer via pathway I in which the proton is abstracted through the catalytic aspartic acid residue can occur via both the dissociative and associative mechanisms, while pathway II supports only the associative mechanism. For pathway I, (highlighted red in Figure 2.8), the phosphoryl transfer step occurs through either a dissociative mechanism (Figure 2.9A), or an associative mechanism (see Figure 2.9B), and proceeds concomitant with migration of the proton from the hydroxyl group of the tyrosine substrate to the $O\delta_2$ oxygen of D813. In the dissociative reaction channel, the sum of the nucleophilic attack distance λ_a (between Y-OH and ATP Pγ) and the bond-cleavage distance λ_c (between ATP Pγ and ATP O2/3β) in the transition state is greater than the corresponding sum in the reactant and product states; this scenario is reflected in Figure 2.9A where the $\lambda_a+\lambda_c$ values for transition state (TS) are greater than those for Rectant (R) or Product (P) states, i.e., the TS region lies above the line joining the R and P states in the λ_a versus λ_c plot. In contrast, a nucleophilic attack through an associative mechanism is evident in Figure 2.9B where the $\lambda_a+\lambda_c$ values for TS are smaller than those for R or P, i.e., TS lies below the line joining the R and P in the λa versus λc plot. Pathway II (highlighted blue in Figure 2.8) is characterized by the migration of the proton

from the hydroxyl group of the tyrosine substrate to the O1γ oxygen of ATP and the subsequent transfer of the proton to the O3β atom of ATP. The phosphoryl transfer takes place concomitant with the proton transfer steps but only through an associative mechanism, see Figure 2.9C; interestingly, for this channel, our attempts to drive the system through a dissociative mechanism yielded an associative mechanism with the O2β atom of ATP playing the surrogate (as the leaving group) to the O3β atom of ATP.

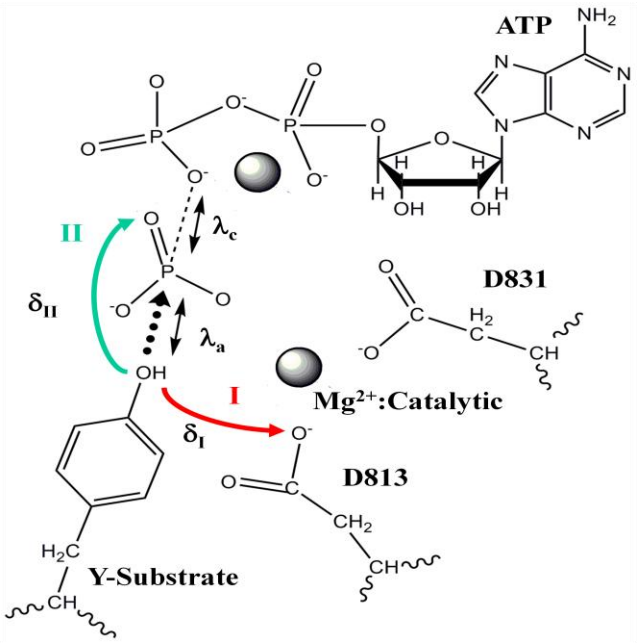

Figure 2.8 Proposed schematic of pathway for phosphoryl-transfer in EGFR-TKD. Mg2+ ions and the catalytic aspartates (D831) are marked. Two potential pathways for proton migration (concomitant with phosphoryl-transfer) are shown. Pathway I (red) involves proton abstraction from the substrate tyrosine –OH group by D813 of EGFR, and pathway II (green) involves proton migration to the O1γ oxygen of ATP. δI and δII denote distances between the proton of the substrate tyrosine hydroxyl and either the Oδ2 oxygen of D813 (δI) or the O1γ oxygen of ATP (δII). The nucleophilic attack distance, λa (distance between the tyrosine oxygen and the ATP Pγ) and the bond cleavage distance, λc (distance between the ATP Pγ and ATP O2/3β) are also noted.

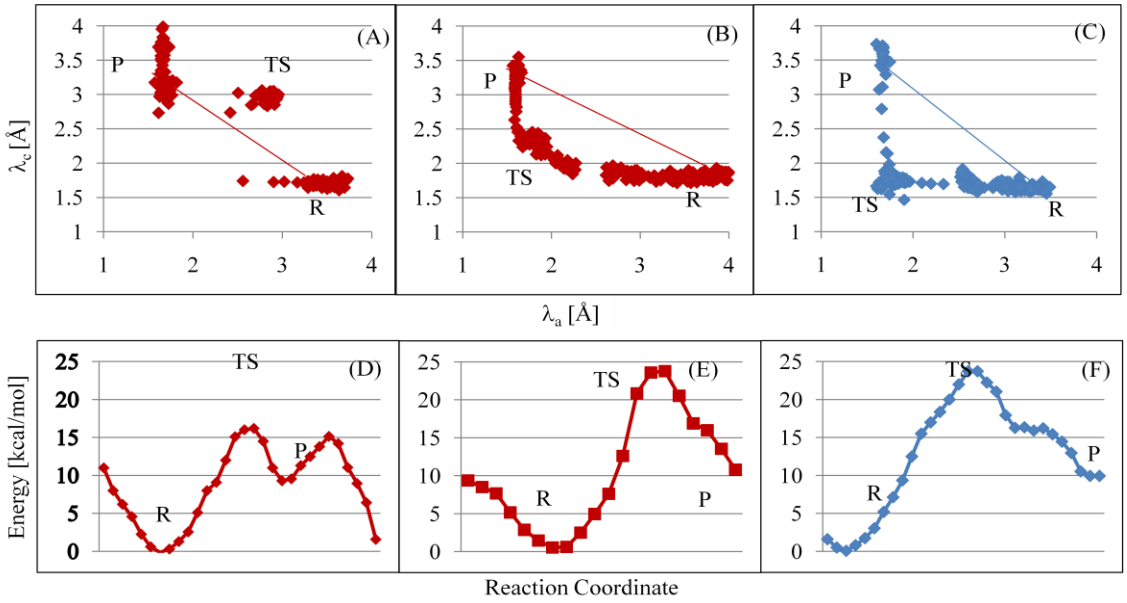

Figure 2.9 QM/MM results for proposed pathways. (A)&(D): Pathway I and dissociative; (B)&(E): Pathway I and associative; and (C)&(F): Pathway II and associative. R = reactant; P = product; and TS = transition state characterized by a trigonal bipyramidal geometry about P_γ. (A-C) Distributions of values for the nucleophilic attack distance, λ_a (distance between the tyrosine oxygen and the ATP P_γ) and the bond cleavage distance, λ_c (distance between the ATP P_γ and ATP $O2/3_\beta$) along the reaction pathways. *(D-F)* Energy changes along the reaction pathways in QM/MM simulations for the noted mechanisms.

In Figure 2.9(D-F), we depict the computed energy changes along the reaction pathways in active EGFR TKD, which suggest that for EGFR, the dissociative mechanism of phosphoryl transfer through pathway I corresponds to the lowest activation energy, E_a=16 kcal/mol (Figure 2.9D), while energy profiles for the associative mechanisms through pathways I and II yield barriers of 24 kcal/mol (Figure 2.9E) and 26 kcal/mol (Figure 2.9F), respectively.

2.4 Discussion

Crystal structures of multiple TKDs have already been solved, which has led to important insights in understanding the kinase function. However, most of these crystal structures

are solved in the presence of ATP analogs and small molecule inhibitors. Due to the relatively low binding affinity of the interaction between TKD and peptide substrates, crystallization of TKD-peptide substrate complexes is still a challenge using experimental techniques. In this chapter, we suggest that a rich landscape exists for the binding of tyrosyl-substrate and subsequent pre-catalytic and catalytic events in the ternary complex of EGFR TKD/ATP/tyrosyl-substrate, from which we outline new molecular rules, which may, in-part, contribute to specificity in phosphorylation reactions.

Our ensemble molecular docking protocol is successful in predicting bound conformations of a 7-residue peptide sequence derived from the C-terminal tail region of EGFR to EGFR TKD, with predicted bound conformation very similar to that found in a crystal structure of a bi-substrate analog co-crystalized with EGFR TKD [289] (PDBID: 2GS6). In our predicted complex, the aspartic residue D813 forms a hydrogen bond with the hydroxyl oxygen the P0 tyrosine of the peptide substrate. The aspartic residue has been shown to hydrogen-bond with the hydroxyl group of serine/threonine/tyrosine in other kinase systems [119, 235, 292]. In the only existing EGFR TKD/bi-substrate analog structure (i.e. PDBID: 2GS6), the distance between the aniline nitrogen and the $O\delta_{1/2}$ oxygen of D813 is 6.37 Å, which makes it difficult to ascribe a role for the aspartic residue. The conformation in our modeled ternary complex clearly places the hydroxyl oxygen of the P0 tyrosine 2.9 Å away from the $O\delta_2$ oxygen atom D813, i.e., within hydrogen bonding distance, which is consistent with the conceived role of D813 as a facilitator of proton abstraction.

Moreover, it has been suggested that protein tyrosine kinase (including EGFR) select substrates with glutamic acid at the P-1 position and large hydrophobic amino acids at the P+1 position [85, 86, 118, 123, 149, 245]. Our predicted peptide bound ternary complex suggests that a highly conserved region, corresponding to Val852-Pro853-Ile854-Lys855-Trp856 in EGFR TKD is essential for substrate binding: Lys855 forms a salt bridge with the highly preferred glutamic acid residue at the P-1 site, and the hydrophobic pocket formed by Val852, Pro853, Ile854 and Trp856 serves to anchor a large hydrophobic residue at P+1 position. This observed binding mode for our top-ranked configuration is consistent with the hypothesis that the substrate binding surface formed through hydrophobic contacts and hydrogen bonds help orient the tyrosine residue for catalysis [149]. Collectively, this extended region, i.e. Val852-Pro853-Ile854-Lys855-Trp856 in EGFR, has been identified as a highly conserved motif in the protein tyrosine kinase family [28], which further adds significance to our predicted conformation for the ternary complex. This allows us to conclude that a conserved binding surface on the kinase sandwiching the P0-tyrosine and making hydrophobic contact to the P+1 and hydrogen bond with P-1 residues, respectively, helps orient the tyrosine residue for catalysis [149]. Interestingly, the corresponding motif in serine/theronine kinase, with a sequence of Gly200-Thr201-Pro202-Glu203-Tyr204 in PKA, has been named as P+1 binding pocket and shown to be critical for substrate specificity [35, 104, 149]. In this case, the glutamic acid residue (E203) in this motif has been proposed to interact with preferred positively charged residues of the N-terminal site in PKA substrates [149], and residues Gly200, Pro 202, Thr201 and Tyr204 favor a hydrophobic residue in the P+1 position [261].

Two of the main points of contention in kinase-catalyzed phosphoryl transfer reactions are whether the phosphoryl-transfer reaction occurs through an associative or a dissociative mechanism (39, 41-44) and whether the conserved aspartic (D813 in EGFR) functions as an acceptor for proton transfer from the hydroxyl group of the substrate. With the predicted peptide bound conformation, we were able to build the ternary structure of EGFR TKD and peptide with ATP and Mg^{2+} ions, and investigate the mechanistic details of EGFR TKD catalytic reaction mechanism using QM/MM simulations. Based on a pre-catalytic energy landscape of γ-phosphate of ATP we identify two alternative preferred proton positions of the tyrosine hydroxyl based on which we hypothesize two possible pathways: Pathway I involves proton abstraction from the substrate tyrosine hydroxyl group by D813 of EGFR, and pathway II involves proton migration to the $O1_{\gamma}$ oxygen of ATP. Indeed, our QM/MM simulations support both pathways, with the phosphoryl transfer via pathway I sustained via both dissociative and associative mechanisms, while that via pathway II sustained by just the associative mechanism. The dissociative mechanism through pathway I has the most favorable energy barrier of 16 kcal/mol, while energy profiles for the associative mechanisms through pathways I and II each yield barriers of 24 kcal/mol. Combined with the pre-catalytic reaction energy barrier, our estimate for the net energy barrier $\Delta E_{barrier}$ for activation through the most favorable reaction channel (i.e. for pathway I through dissociative mechanism) in the active EGFR TKD is 16+2=18 kcal/mol. This value compares favorably with the experimental value for the free energy barrier for EGF activated receptor system in the cellular context for which $\Delta G_{barrier}=k_BTln[k_BT/(k_{cat}h)]=18.5$ kcal/mol was estimated from the reported k_{cat} of 17.4

min^{-1} by Gibbes-Johnson et. al. [84]. Our estimate for the net energy barrier $\Delta E_{barrier}$ for activation through pathway II for active EGFR TKD is 24+2=26 kcal/mol. This translates to a k_{cat} of $\leq 10^{-6}$ s^{-1} or a reduction by at least a factor 10^{-6} relative to the experimental k_{cat} of EGF activated receptor or a factor of 10^{-5} relative to experimental k_{cat} in the absence of EGF [85, 185]. The reduction in k_{cat} manifests itself as little or very weak residual activity through pathway II, which is consistent with recent experiments [147, 244].

Given that the protonation states of titratable side-chains in the active site can have considerable effect on the stability of the ground state as well as on the energies of the reaction landscape, we explored the different protonation states of the conserved catalytic residue D813 in EGFR. Multiple protonation states of the conserved catalytic residue (D831 of the DFG motif) was not considered because the location is relatively remote from the atoms involved in the catalytic reaction pathway. Although recent studies of the Abl kinase have reported that the protonation state of this residue is important in the flipping of the DFG motif toward the active conformation [257]. Such a flip for EGFR is already facilitated, even in the unprotonated state for D831, by kinase dimerization [162]. For the D813-protonated system, the phosphoryl transfer step occurs through an associative mechanism and proceeds with the simultaneous migration of the proton from the hydroxyl group of the tyrosine substrate to the $O\delta_1$ oxygen of D813 and the migration of $H\delta_2$ of the protonated $O\delta_2$ oxygen of D813 to O3γ atom of ATP (data not shown). Our attempts to drive the system through a dissociative mechanism yielded an associative mechanism with the O2β atom of ATP playing the surrogate (as the leaving group) to the O3β atom of ATP.

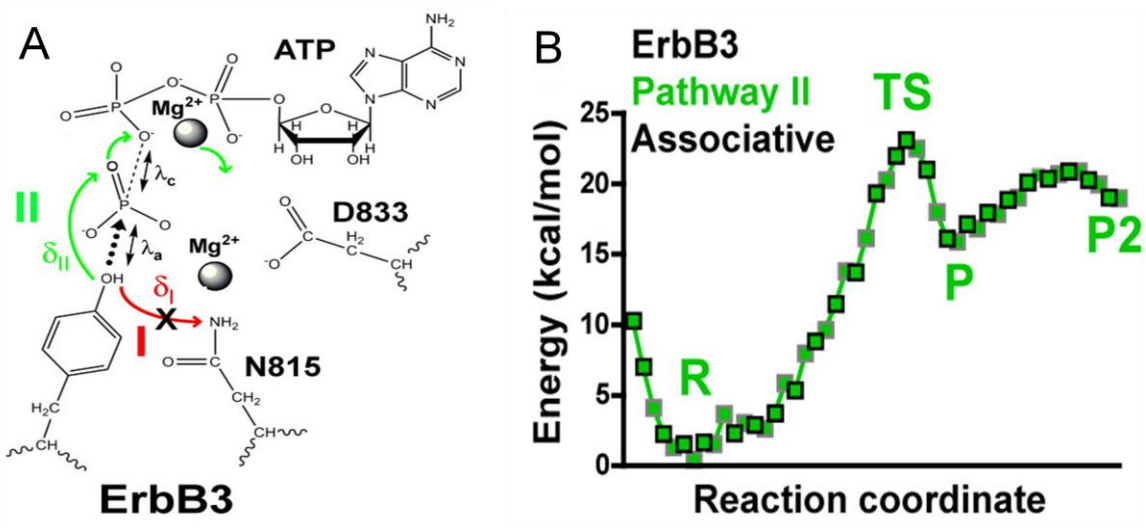

Figure 2.10 Possible catalytic reaction pathway for ErbB3 modeled by QM/MM simulation.

ErbB3/HER3 is another member of the human epidermal growth factor receptor (EGFR/HER) or ErbB receptor tyrosine kinase family. ErbB3 is considered to be an inactive pseudokinase because it lacks several key conserved (and catalytically important) kinase domain residues – including the catalytic base aspartate. Very recently, our collaborator, Dr. Mark Lemmon's laboratory, reported that despite this residue alteration, the kinase domain of ErbB3 retains phosphoryl-transferase activity, as evidenced by *trans*-autophosphorylation of the purified ErbB3 intracellular region [237]. Kinase activity appears to be ~1000-fold weaker than that seen for EGFR. Moreover, they showed that the ErbB3 kinase domain binds ATP with a K_D of approximately 1.1 μM, similar to that seen for known active kinases. A crystal structure showed that the overall structure of ErbB3-TKD$^{665\text{-}1001}$ closely resembles the inactive conformations of the EGFR, ErbB4, and Src kinases [198, 287]. The structure also clearly showed the electron density for both AMP-PNP and Mg^{2+}, which are similarly positioned as in the EGFR TKD structures.

To address this observation, we hypothesized that the weak kinase activity of ErbB3 might utilize a mechanism distinct from that seen in other kinases, and might be carried out by the 'inactive-like' state. We therefore undertook quantum mechanics molecular mechanics (QM/MM) simulations to delineate the energy landscape of phosphoryl-transfer from ATP to a substrate tyrosine catalyzed by the ErbB3 structure, and compared the outcomes with those seen in parallel computational studies of EGFR. Due to the presence of an asparagine (rather than aspartate) at N815, pathway I proposed for EGFR is not sustainable in ErbB3. However, like active EGFR TKD, the proton migrates to the $O1_{\gamma}$ oxygen of ATP, through pathway II in Figure 2.10. Phosphoryl-transfer concomitant with this pathway can be catalyzed by ErbB3 in the inactive-like conformation, although it is predicted to be several orders of magnitude slower (estimated E_a is 24 kcal/mol) than the most favorable reaction channel in EGFR (*i.e.* for pathway I through a dissociative mechanism) – consistent with experimental observations.

Indeed, a D813A-mutated variant of EGFR variant has been reported to retain its ability to promote EGF-dependent DNA synthesis and MAP kinase activation despite exhibiting greatly reduced receptor autophosphorylation [65]. A low level of autophosphorylation in D813A-mutated EGFR, similar to that shown for ErbB3, may therefore be sufficient to mediate certain key aspects of its signaling.

ErbB3 signaling is important in development of resistance of NSCLC to the EGFR inhibitor gefitinib through amplification of the gene for Met [80], the hepatocyte growth factor receptor. ErbB3 also plays a key role in the ability of ErbB2-overexpressing cells to escape growth inhibition by the EGFR/ErbB2 dual-specific kinase inhibitor lapatinib

[229]. Our data suggest that evaluating the signaling importance of ErbB3's weak kinase activity is also worthwhile, and could represent a valuable therapeutic target. Indeed, ErbB3 autophosphorylation *in vitro* is not inhibited by lapatinib or by the EGFR inhibitors gefitinib or erlotinib, leaving open the possibility that ErbB3 kinase activity promotes resistance to growth inhibition by EGFR/ErbB2-targeted drugs. If this is the case, selective inhibition of the ErbB3 kinase could be an additional effective therapeutic strategy.

Chapter 3 Interaction of Erlotinib to Different Conformational States of the Epidermal Growth Factor Receptor Tyrosine Kinase Domain (EGFR TKD)

3.1 Introduction

EGFR and other ErbB-family receptor tyrosine kinases have been associated with the development, progression and aggressiveness of a number of malignancies [173, 233]. The EGFR kinase has long been recognized as a potential drug target. Small molecule receptor tyrosine kinase inhibitors (TKI) for EGFR TKD are of significant interest as cancer therapeutic drugs [61]. Erlotinib and gefitinib are two EGFR TKD inhibitors which have been approved for the treatment of advanced Non-small cell lung cancer (NSCLC) [8, 242]. These two drugs caused a dramatic response in a sub-group of patients in clinical trials. A class of mutations, including the point mutation in the activation loop (L834R), have been reported correlate highly with favorable clinical responses of patients to gefitinib and erlotinib [48, 164, 186]. The discovery of these mutations and their correlation with the drug response has fueled an intense interest in studies of the oncogenic effects and drug binding properties of these mutations. Although a lot effort has been made, a mechanistic basis for such a correlation remains unclear.

While many factors may contribute to drug response [80, 101, 244], a basic question to be answered is whether the mutations will directly affect the drug binding. Available crystal structures have shown that L834R EGFR TKD shares a very similar fold with the

wildtype kinase and gefitinib was shown to be bound very similarly to wildtype and L834R mutatant EGFR TKD [77, 284]. Erlotinib complexed with wildtype EGFR TKD [248, 284] shows a very similar binding mode to that of the gefinitb EGFR TKD. Despite the similarity in binding modes, several groups have reported varying results for the binding affinities of gefitinib and erlotinib to EGFR TKD (shown in Table 3.1), thereby creating some controversy over the notion that the binding mechanism is a common one. Although two separate studies [83, 101] reported no overall effect of the mutations on drug binding, other studies have claimed an increase in binding affinity for the drug in the L834R mutant (relative to wildtype) [48, 284].

Table 3.1 Binding affinities reported for Erlotinib and Gefitinib binding to EGFR TKD.

EGFR	K_i (nM)				
	Erlotinib		Gefitinib		
WT	0.67[83]	17.5[48]	0.9[83]	0.19[101]	53.5[284]
L834R	0.97[83]	6.26[48]	1.3[83]	0.20[101]	2.6[284]

Like many other kinases, two EGFR TKD conformations, namely active and inactive, have been captured in crystal structures [248, 276, 284, 288]. Both erlotinib and gefitinib were solved in complex with active EGFR TKD [248, 284], while another inhibitor, lapatinib, was crystallized in an inactive EGFR TKD conformation. Such structural studies have lead to a prevailing conclusion that erlotinib/gefitinib are specific to active EGFR TKD form whereas lapatinib preferentially binds to the inactive conformation. L834R kinase structure has been found to adopt a similar conformation to the active

wildtype structure. As the L834 residue lies in the helical turn of the activation loop in the inactive state of the kinase, it has been proposed that mutation L834R will destabilize the inactive structure by disrupting the hydrophobic interaction which stabilizes the relatively displaced αC-helix in the inactive state, thereby providing an activating stimulus. As a result of the assumed preference of active form of EGFR TKD in L834R system, it has been proposed that a tighter binding of erlotinib and gefitinib would result in the mutant TKD. Moreover, very recently, a new experimental study has shown that the relative selectivity of erlotinib to the active-state EGFR TKD is only 5-10 fold higher than that to inactive-state EGFR TKD[199], which raises the possibility of crossover binding of these inhibitors to inactive TKD conformations. To test the hypothesis that L834R affects inhibitor binding by altering the equilibrium between active and inactive states of EGFR TKD, a detailed study of the binding mechanism of erlotinib/gefitinib to wildtype and L834R TKD is necessary.

In this chapter , we computationally docked the erlotinib to L834R mutated EGFR TKD using an induced fit docking protocol, and then compared the binding of erlotinib with wildtype and L834R mutated TKD using molecular dynamics. We also employed free energy perturbation to further elucidate the free energy difference between wildtype and L834R systems. Our results suggested that erlotinib binds in a similar fashion to wildtype and mutated kinase systems in the active form. We further compared the erlotinib binding with both active and inactive EGFR TKD and proposed that erlotinib is capable of binding to both forms with similar binding efficacy. Our simulation results illuminate the complexity of inhibitor binding in EGFR TKD and highlight the importance of targeting erlotinib and other inhibitors to different conformations of the kinase.

3.2 Methods and Materials

3.2.1 System preparation and molecular docking

Protein structures were taken from the Protein Data Bank: the active conformation of the wildtype EGFR TKD was modeled on 1M17 [248] and 2ITX [284] and the L834R model was built using the 2ITV structure [284]. The inactive EGFR TKD was modeled based on the 2GS7 [288] and 1XKK [276] structures. Initial erlotinib conformation was also taken from the 1M17 structure. Protein and ligand conformations were prepared using the protein preparation wizard and the "LigPrep" tools available within the Schrödinger computational chemistry package[224]. All simulations using Schrodinger were performed with the OPLS force field[137] . In particular, the docking simulations were performed using Schrödinger's Induced Fit Docking (IFD) package [236]. With the induced fit docking protocol, the ligand was first docked to a rigid protein conformation using Glide XP [98]; for the resulting top 20 complex conformations, the protein side chains residues within 5.0 Å of the ligand poses were subject to conformational search and minimization using the prime structure prediction method [129, 130]; finally the ligand was re-docked to the 20 new receptor conformations. Therefore, this docking protocol includes possible binding modes of ligands as well as the associated conformational reconfiguration within the receptor active site.

3.2.2 Parameterization of erlotinib

The resulting conformations from induced fit docking were analyzed using molecular dynamics simulations and the CHARMM27 forcefield [93], which consists of the following terms:

$$U(\vec{R}) = \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\chi(\chi - \chi_0)^2$$

$$+ \sum_{\text{nonbond}} \varepsilon_{ij} \left[\left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{12} - \left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{6} \right] + \sum_{\text{nonbond}} \frac{q_i q_j}{D r_{ij}}$$

, where, K_b , K_θ , K_χ are the bond, angle and dihedral force constants; the b_0, θ_0, χ_0represent the equilibrium value of bond, angle and dihedral; $R_{ij}^{\min}$and ε_{ij} are, respectively, the distance between atoms i and j at which the LJ potential is zero and the depth of the LJ potential well for the same pair of atoms; D is the effective dielectric constant and q_i is the partial atomic charge on atom. Since the CHARMM27 force-field did not contain the parameters for erlotinib, all these parameters had to be defined for each atom type of erlotinib (see Figure 3.1).

The atom types describing the erlotinib molecule are shown in Figure 3.1. 9 new atom types were added to the CHARMM topology file. Initial partial atomic charges (q_i) were calculated using CHarges from ELectrostatic Potentials using a Grid based method (CHELPG) [33] in the ab-initio electronic structure package GAUSSIAN, by fitting the molecular mechanics-derived electrostatic potential to that obtained quantum mechanically. The van der Waals constants ($R_{ij}^{\min}$and ε_{ij}) were transferred from existing CHARMM parameters and were not modified during refinement as their values depend mostly on atomic properties and are assumed to be transferable to the molecular environment. The equilibrium constants (b_0 , θ_0) were obtained from optimized structures of the small molecules based on the crystal conformation [248] using geometry

optimization in ab-initio calculations, and were also not changed during refinement. An initial guess, based on an analogy with other existing CHARMM parameters is made for all missing intermolecular force field constants K_b, K_θ, K_χ, and the equilibrium constants for the dihedral terms, namely, n, and δ. In particular, for the carbon-carbon triple bond, we first performed ab-initio potential energy surface scan along the bond length, then use parabolic potential function to fit the potential surface and get initial estimation of this bond constant.

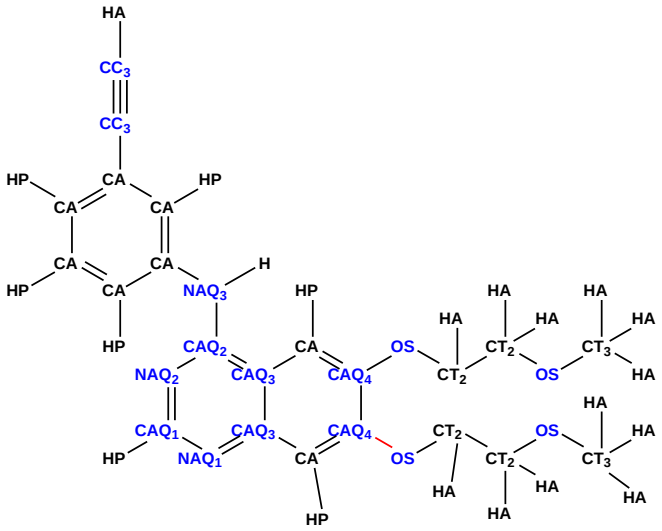

Figure 3.1 Atom type defined for erlotinib. New atom types are shown in blue.

Partial atomic charges (q_i) were adjusted to reproduce interaction energy and geometry of H-bond donor/acceptor atoms in the inhibitor compound with water molecules (Figure 3.2). The hydrogen bonded complexes shown in the figure were individually optimized via ab-initio calculations using the 6-31G* basis set with fixed monomer geometries. The partial atomic charges in CHARMM are manually adjusted to reproduce the ab-initio results for the geometry and energy. Following the strategy for generating the CHARMM

force field [93, 166, 189] , the ab-initio interaction energies are scaled by 1.16 and the distances are offset by -0.2 Å.

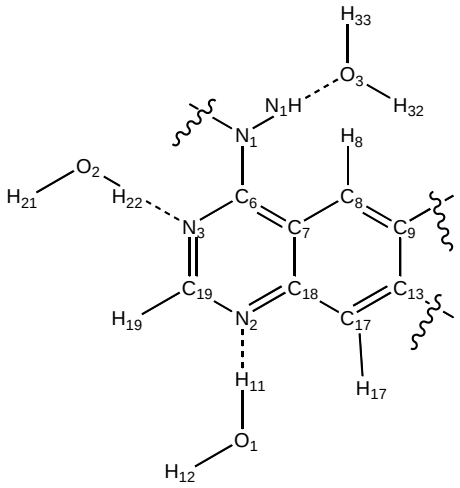

Figure 3.2 Water interactions used to refine the force field.

The force constants K_b , K_θ and, K_χ were refined by reproducing the vibrational eigenvalues and eigenvectors from ab-initio calculations following the procedure employed by Viana et. al. [263], which ensured that both vibrational frequencies and vibrational modes (defined by eigenvectors) calculated from CHARMM and those from ab-initio methods match closely. In the algorithm, the current parameter set is used for energy minimization and the calculation of normal modes υ_i^C and $\overline{\chi}_i^c$ (eigenvalues and eigenvectors) with CHARMM. Each of the modes is projected to the eigenvector sets $\overline{\chi}_j^G$ (the corresponding quantity calculated from Gaussian) and a best projection, (mode i in CHARMM projecting to mode $j\max$ in GAUSSIAN), is obtained according to two criteria: (1) there is a one-to-one correspondence of the two-sets of eigenvectors, and (2) $\prod_i \frac{1}{\max_j(\overline{\chi}_i^c \cdot \overline{\chi}_j^G)}$ is a minimum. Then, the penalty function value is calculated by:

$\sigma = \sqrt{\frac{\sum_{i=1}^{3N-6} \omega_i (\upsilon_i^C - \upsilon_{j\max}^G)^2}{3N-6}}$, $\omega_i = \frac{1}{\max_j (\bar{\chi}_i^c \cdot \bar{\chi}_j^G)}$. In the ideal case, $\upsilon_i^C = \upsilon_{j\max}^G$ $\bar{\chi}_i^c \cdot \bar{\chi}_j^G = \delta_{ij}$, which implies that the CHARMM parameter set can perfectly reproduce the frequency spectrum from GAUSSIAN.

For the dihedral potential surface fitting, relaxed potential surface scan for key dihedral angle (C4-C4-N1-C6) is performed by both CHARMM and GAUSSIAN and defined as $\bar{D}^C, \bar{D}^G$ and the penalty function is defined as $\sigma = \sqrt{\sum_{i=1}^{NGRID} (D_i^C - D_i^G)^2}$.

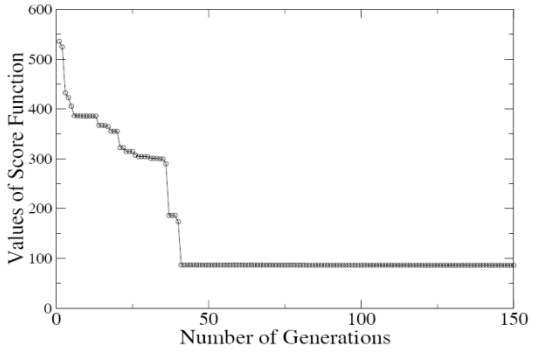

Figure 3.3 Optimization of the merit function based on a genetic algorithm.

An automated procedure using a genetic algorithm (GA) [178] is developed for the refinement of the intermolecular force field constants and the dihedral potential energy surfaces. The efficiency of the GA algorithm is showed in Figure 3.3 . We iteratively repeat the procedure to refine 1) partial charges; 2) frequencies and eigenvectors; and 3) the dihedral surface scan until reaching a force field with which, the target data – the water interaction, the dihedral potential energy surface, and the vibrational normal modes

– calculated by CHARMM matches well with the corresponding values calculated by Gaussian.

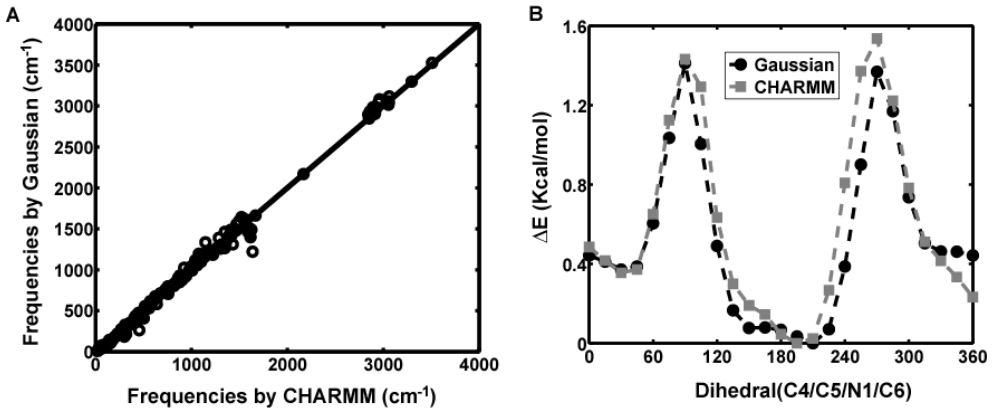

Figure 3.4 Target data matching between electronic structure (Gaussian) and molecular mechanics calculations (CHARMM). A: one-to-one correspondence of frequencies. The frequencies are matched based on eigenvectors. B: dihedral scan.

An empirical parameter set for erlotinib optimized to be consistent with the CHARMM27 force field set (see Table 3.2 and Figure 3.4) was developed by following the procedure described above. This parameter set was employed in molecular dynamics simulations.

3.2.3 Molecular dynamics simulations

The conformations resulting from induced fit docking were minimized and subsequently equilibrated by performing molecular dynamics simulations. Each system was subjected to a constant temperature and constant pressure molecular dynamics run at 300 K and 1 atm followed by a 10 ns constant temperature equilibration at 300K with periodic boundaries enforced and long-range electrostatics taken into consideration.

Table 3.2 Water-erlotinib interactions and erlotinib dipole moment calculated by ab-initio (Gaussian) and CHARMM. The ab-initio interaction energies in the table have been scaled by 1.16 (See text) and the distances are scaled by -0.2 Å.

Hydrogen bond	Interaction Energies (Kcal/mol)		Distances (Å)	
	GAUSSIAN	CHARMM	GAUSSIAN	CHARMM
N2…HOH	-6.69	-6.61	1.93	1.91
N3…HOH_2	-5.33	-5.30	2.12	2.01
N1H…OHH_2	-6.52	-6.52	2.44	2.63
Dipole moment (Debye)	GAUSSIAN		CHARMM	
	4.87		5.07	

3.2.4 MMPBSA calculation

We employed MMPBSA [102, 117, 150, 153, 231, 253, 266] to calculate the absolute binding free energy and compared it to the glide score. MMPBSA Energies of each modeled complex conformation were calculated as the average of single point MMPBSA energy of 500 snapshots taken from the 10 ns of the simulation. Waters and ions were removed from the trajectory before calculation. The molecular mechanics energy, U_{MM}, was evaluated by averaging energies over all structures using an infinite cutoff for non-bonded interactions. The electrostatic contribution to the solvation free energy, W_{PB}, was calculated with Poisson-Boltzmann Solver in CHARMM [126]. The reference system had a solvent dielectric of 1 and 0 M salt concentration and the solvated system had a dielectric of 80.0 and 100 mM salt concentration. Non-polar contribution to solvation free energy, W_{SA}, was approximated with the suface area model ΔW_{SA}=[0.00542 kcal/mol/Å^2]×SASA +0.92 kcal/mol [215], where the molecular surface area (SA) was estimated with a 1.4 Å slovent probe radius. Vibrational entropy changes were neglected.

3.2.5 Free energy perturbation (FEP)

We designed a thermodynamic cycle shown in Figure 3.5 to directly calculate binding free energy changes using FEP. To calculate the binding free energy change $\Delta\Delta G_{W\to M}$, we calculated the difference between the free energy changes caused by mutation in the bound state (ΔG_B) and the unbound state (ΔG_U). The binding free energy change due to the mutation is obtained as $\Delta\Delta G_{W\to M} = \Delta G_M - \Delta G_W = \Delta G_B - \Delta G_U$.

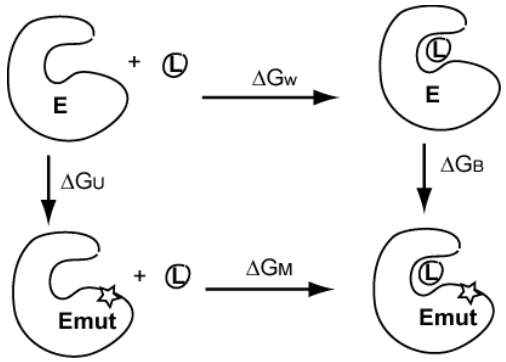

Figure 3.5 Thermodynamic cycle for the calculation of binding free energy change of ligand/substrate due to mutation

The free energy changes caused by the particular mutation in the bound state (ΔG_B) and the unbound state (ΔG_U) were calculate by free energy perturbation (FEP) method [171, 204, 205] . The FEP procedure was performed using the alchemical method in NAMD 2.7b2 using the dual-topology paradigm. For each state (bound or unbound), we performed FEP calculation in both the forward and backward direction to ensure convergence and define an error bar. For each direction, the perturbation was divided into 72 windows (λ= 0, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 0.005, 0.01, 0.015, 0.02, 0.03-0.1 at an interval of 0.01, 0.1-0.9 at an interval of 0.02, 0.9- 0.98 at an interval of 0.01, 0.098, 0.0985, 0.099, 0.995, 0.999, 0.9999, 0.99999, 0.999999, 1). In each window, the system

was equilibrated for 20 ps and run for another 100 ps for data collection. Larger window sizes and longer simulations have also been tested and the current setup yields reasonable convergence in the final binding affinity. To avoid the "end-point catastrophes", a soft-core potential was used to gradually scale the unbounded interaction potential. The scaling of van der Waals (vdW) and electrostatic interactions were handled separately. For appearing atoms, vdW interactions are linearly coupled to the simulation from λ=0 (fully decoupled) to λ=1 (fully coupled) and electrostatic interactions are coupled over the range of λ=0.5 to λ=1. For vanishing atoms, the vdW interactions are linearly decoupled from the simulation over the value range of 0 to 1, and the electrostatic interactions are decreased gradually from λ=0 to λ=0.5.

3.3 Results

3.3.1 Erlotinib binds to wildtype and L834R with similar modes and affinities.

The crystal structure of erlotinib bound to wildtype EGFR TKD was solved in 2002 [248]. As a control, we first docked erlotinib to two wildtype EGFR TKD conformations using Glide: one from the exact crystal complex (1M17) [248] and another from a crystallized EGFR TKD with the presence of ATP analog (2ITX) [284]. In both cases, we predicted the compound in a very similar orientation compared to the crystal conformation (Figure 3.6). The quinazoline moiety forms one hydrogen bond with residue Met769 of the kinase and the distances between N2 of the quinazoline and the amide nitrogen of Met769 are 2.9 Å in both the predicted structures, close to that in crystal structure (2.7 Å). The acetylene moiety extends to the back of the ATP binding

pocket, surrounded by the side chains of the Leu764, Lys721 and Thr766. Although the "tails" of the compound in our predicted conformation do not overlapped with those in the crystal structure, the differences are not significant, as the tails project into the solvent and do not make significant contacts with the protein residues.

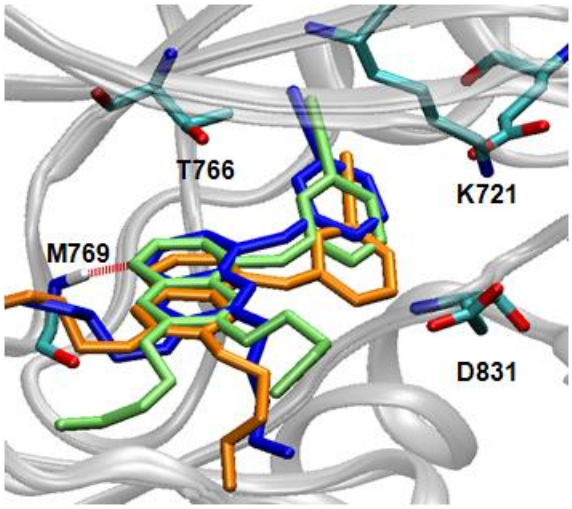

Figure 3.6 The docked conformation of erlotinib bound to wildtype (blue) and L834R mutant (orange) EGFR TKD compared with the crystallized complex of erlotinib and wildtype EGFR TKD (green).

We then prepared the complex structure of erlotinib with L834R system by docking the inhibitor to L834R EGFR TKD (PDBID: 2ITZ). The best predicted conformation is shown to align very closely to that of the wildtype (Figure 3.6) with the 'tails' projecting into solvent and the aniline moiety binding in the pocket formed by Leu764, Lys721 and Thr766. The distance between N1 of the quinazoline and the amide nitrogen of Met769 is 3.15 Å, which is within the hydrogen bonding distance.

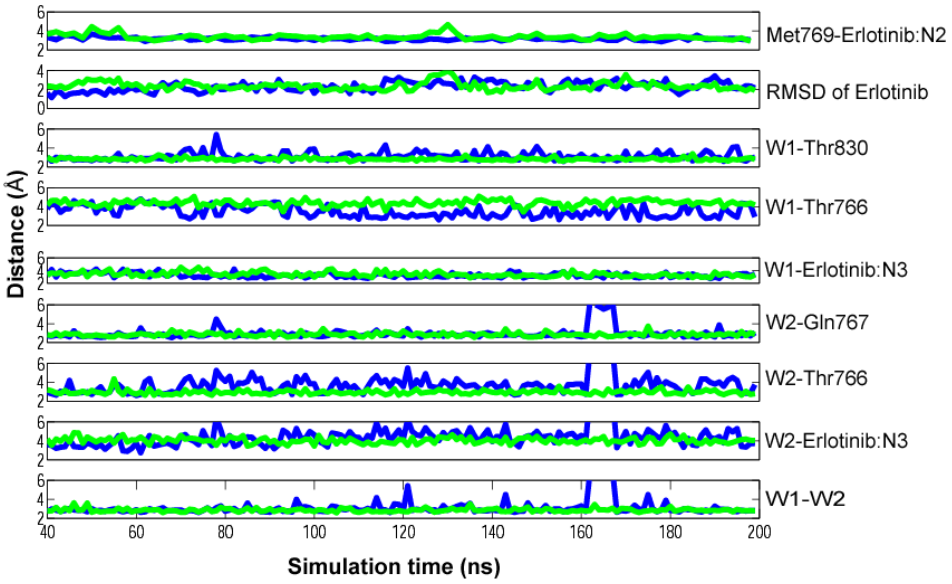

Figure 3.7 Monitoring of root mean squared deviation (RMSD) of erlotinib and key distances between EGFRTKD, erlotinib as well as water molecules in a 10 ns molecular dynamics simulation. Data from the first 2 ns (pre-equilibration) are not shown.

In both wildtype and L834R mutant systems, the erlotinib conformation is stable in the predicted position during the entire course of the 10 ns simulation. The hydrogen bond between quinazoline and Met769 in the wildtype and mutant system were found to fluctuate around an average value of 3.17 Å for wildtype and 3.13 Å for L834R mutant (Figure 3.7). The aniline moiety remains in the pocket of the Leu764, Lys721 and thr766 during the dynamics simulation while the two tails of the erlotinib are seen to be flexible in both systems.

Table 3.3 Distances between donor and acceptor atoms in the water mediated hydrogen bond network (all distances have a units of Å)

Distances	W1-Erlotinib	W1-T830	W1-T766	W1-W2	W2-T766	W2-Q767	W2-Erlotinib
Wildtype	3.25	3.12	3.50	2.96	3.56	2.90	4.20
L834R	3.40	2.83	4.38	2.86	2.95	2.85	4.03

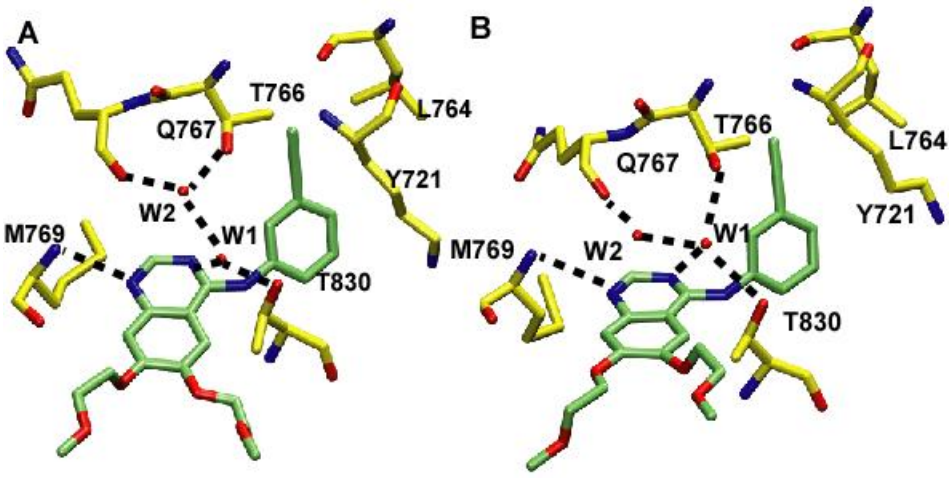

Figure 3.8 Two typical water mediated networks in the erlotinib / EGFR TKD complex.

In the crystal structure (PDBID: 1M17), one bound water molecule is resolved, which is described as bridging the gap between Thr766 and the quinazoline nitrogen atom (N3). In our MD simulations, two water molecules were found to enter the (back) pocket for ATP binding and formed a stable hydrogen bond network between erlotinib and residues Thr830, Thr766 and Gln766 after the initial 2 ns of equilibration. The average distances between donors and acceptors in the networks for both wildtype and L834R are described in Table 3.3 and the fluctuations are depicted in Figure 3.7. We also depict two typical hydrogen networks in Figure 3.8A&B. In both networks, the first water molecule (W1) bridges the N3 atom of erlotinib, Thr830 and the second water molecule (W2) interacts with Gln767. In network A, W2 interacts with Thr766 while in network B, W1 interacts with Thr766. The two networks were found in both wildtype and L834R systems. While they were seen to be equally favored in wildtype system, the L834R system showed a preference for network A, which can explain the differences in the average values of W1-Thr766 and W2-Thr766 distances in Table 3.3. A similar conclusion of two-water-

mediated interactions between erlotinib and EGFR TKD has also been reported in an earlier study involving a different force field [12].

Table 3.4 Binding affinities calculated by Glide, MMPBSA and FEP for wildtype and L834R system.

	Glide Score (kcal/mol)	MMPBSA Score (kcal/mol)	FEP (kcal/mol)
Wildtype	-9.34	-28.2	
L834R	-9.35	-30.7	
ΔΔG	-0.01	-2.5	0.831±1.158

To compare the binding affinities of erlotinib to wildtype and mutant EGFR TKD, we provide the absolute binding energy calculated using Glide and MMPBSA in Table 3.4. Both the Glide docking scores as well as the MMPBSA energies are very similar in wildtype and L834R. We also report the difference in relative binding free energy computed based on the thermodynamic cycle in Figure 3.5 and FEP calculations. In the FEP calculation, the free energy difference between the wildtype and the L834R system for the bound and unbound state were calculated in both forward and backward directions of FEP integration in order to obtain a convergence measure. The cumulative free energy differences are shown in Figure 3.9 and the resulting $\Delta\Delta G_{W\rightarrow M}$ values are 0.68 kcal/mol and 0.98 kcal/mol in the forward and backward directions, respectively; thus, the FEP calculations yield an average of 0.83±1.16 kcal/mol in the free energy difference between the wildtype and the L834R systems.

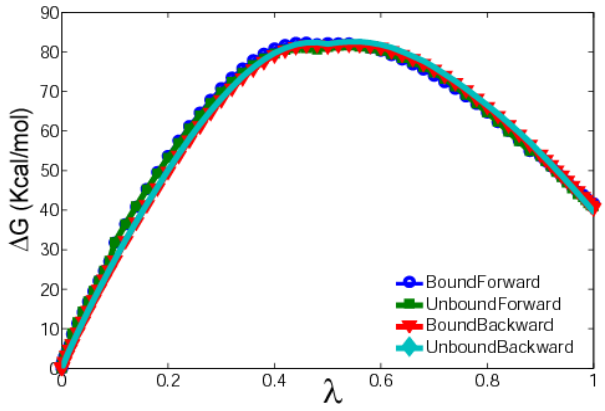

Figure 3.9 Cumulative free energy (kcal/mol) calculated for bound and unbound systems in both forward and backward directions. .

3.3.2 Erlotinib interacts with the inactive EGFR TKD

It has been established that specific kinase inhibitors can target different forms of the same enzyme [151, 217]. In EGFR TKD, erlotinib has been crystallized with active wildtype EGFR TKD, based on which it is assumed widely that erlotinib is more effective in inhibiting the activated EGFR kinase. To challenge this notionHere, we explore the possibility for erlotinib to bind with the inactive EGFR TKD.

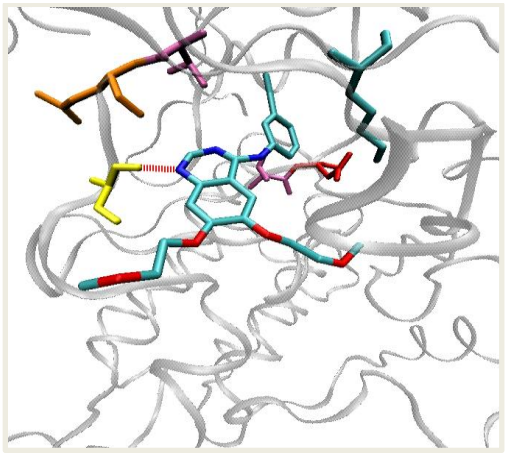

Figure 3.10 The docked erlotinib conformation to wildtype inactive EGFR TKD.

The docked conformation of erlotinib to wildtype inactive EGFR TKD is shown in Figure 3.10. The conformation is very similar to that presented in the active EGFR TKD, with the same orientation and the critical hydrogen bond between N2 of erlotinib and Met769. This conformation has also been verified to be stable in 10 ns MD simulations. Besides the two water molecules present in the active kinase, there is a third water molecule entering the back pocket to fill the void formed by Met742, Leu753, Thr766, Thr830, Phe832 and Leu834, where the 3-fluobobenzyloxy group of the lapatinib overlays in the crystal structure (PDBID: 1XKK). The erlotinib conformations (after 10 ns MD simulations) in both active and inactive wildtype EGFR TKD were aligned based on EGFR TKD backbone, see Figure 3.11. Except for the additional water molecule, erlotinib has a very similar binding mode in both kinase conformations. Morever, the glide score for erlotinib binding to inactive TKD is 9.72 kcal/mol, which is very close to that for the the active form of EGFR TKD, namely 9.34 kcal/mol.

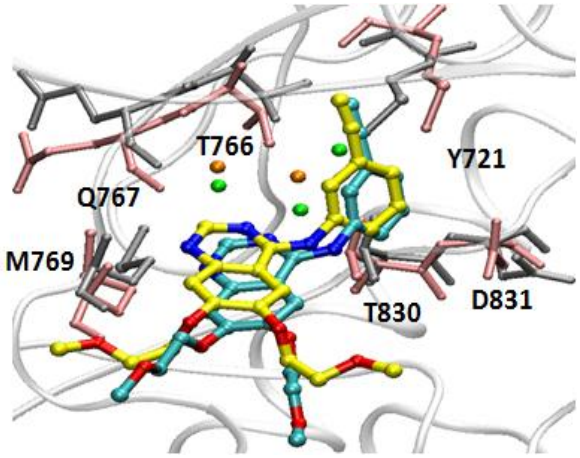

Figure 3.11 Comparison of erlotinib conformations in active and inactive EGFR TKD.

3.4 Discussion

Mutations in the EGFR TKD, including the point mutation L834R, have been identified in approximately 10% of human non-small cell lung cancer cases which confer the patient response to be very sensitive to treatment with erlotinib. Several factors contribute to the effects of mutations on the kinase inhibitor efficacy, at the core of which is changes in the inhibitor binding. To address this question, the K_i values of erlotinib binding to wildtype and mutant EGFR TKD have been reported by several groups. Some groups have reported similar K_i values for erlotinib (or Gefitinib, an inhibitor with very similar clinical and biochemical properties to erlotinib) forin both wildtype and mutant systems [83, 101]. However, others have reported higher binding affinities for the L834R EGFR TKD relative to Wildtype [284], even though the crystal structures of gefinitib with both wildtype and L834R mutation EGFR TKD show no conformational differences in inhibitor binding.

The structure of the wild-type EGFR TKD has been determined in both active and inactive conformations. The crystal structure in complex with erlotinib [248] reveals an active conformation of the kinase, while the structure of wild-type EGFR kinase in complex with lapatinib [276] shows the kinase adopting an inactive conformation. Hence, it is widely believed that erlotinib is more effective at inhibiting the activated EGFR kinase and that lapatinib is more effect at inhibiting the inactive EGFR TKD.

Considering the different conformational preferences of the kinase inhibitors, the dynamic equilibrium between the active and inactive conformations of the ErbB kinase domains may have an important role in determining their sensitivity to these inhibitors. It has been proposed that Leu834 plays an important role in stabilizing the inactive

conformation of the EGFR kinase domain and mutation L834R likely destabilizes the inactive conformation and shifts the equilibrium toward the active conformation, hence leading to possible kinase activation as well as higher sensitivity to erlotinib [284]. Therefore, one possible explanation is that this increase in affinity for gefitinib results from the stabilization of the active-like conformation of the EGFR TKD due to mutation. Hence, a detailed mechanistic picture of how erlotinib binds to different EGFR kinase conformations and how the mutation will affect the binding in each conformation is necessary.

Most of the experimental data have also dealt with the total population of EGFR kinase and not the relative fractions of active and inactive states. Computational analysis and simulations are therefore a vital complement to experiments in defining the equilibrium between the conformational states. Our results confirmed that the L834R mutation does not significantly change the inhibitor binding free energy to EGFR TKD in the active form. Moreover, we have shown that it is very likely that erlotinib binds to the inactive EGFR TKD with similar affinity as that to the active EGFR TKD. This finding highlights the importance for future investigations to address the effects of different kinase conformations on the inhibitor binding as well as inhibitor sensitivity.

Chapter 4 A Multiscale Computational Approach to Dissect Early Events in EGFR Mediated Differential Signaling and its Relevance to Oncogenic Transformations

4.1 Introduction

It is well established that alteration or over expression of EGFR is correlated with a variety of clinical cancers [173, 233]. Hence, small molecule receptor tyrosine kinase inhibitors for EGFR TKD are also of significant interest as cancer therapeutic drugs. Small molecule tyrosine kinase inhibitors (e.g., anilinoquinazoline compounds such as gefitinib and erlotinib), which are ATP analogues [9, 10, 134, 179, 210], have been shown to exhibit growth inhibitory effects against a wide variety of human tumors [57, 58, 60, 59, 61]. The RTK inhibition approach has shown promise in some clinical trials, but results have been quite mixed [66]. In particular, despite the promise in a small (demographic) sub-population with significantly advanced disease, clinical responses to gefitinib and erlotinib varied among population samples. Recently, somatic mutations L834R, L837Q, G685S, del L723-P729 ins S (there is an alternate clinical numbering scheme used in the literature with an offset of 24 residues according to which these mutations are denoted by L858R, L861Q, G719S, del L747-P753 ins S) in the EGFR TKD domain were reported to correlate with the clinical response of gefitinib and erlotinib [48, 49, 165, 186]. In vitro, these EGFR mutants demonstrated enhanced

tyrosine kinase activity in comparison to wildtype EGFR and increased sensitivity to inhibition [10, 55, 69, 70, 165, 186, 187, 247].

Although, the underlying biochemical basis involving the drug sensitivity to the mutations is still not clear, the remarkable sensitivity of cell lines carrying gefitinib sensitizing mutants appears not to be centered around inhibitor or ATP binding affinities, but rather on other biochemical mechanisms regulating the EGFR TKD activity and EGF-mediated signaling [107]. In a recent review, the ERbB signaling network is likened to a bow-tie-configured, evolvable network, displaying modularity, redundancy, and control circuitry [63]. Under this framework, identifying the role and significance of drug sensitizing mutations of ErbB receptors would require a systems level understanding of the signaling network [131-133]. This would prove to be of enormous value in effectively designing kinase inhibitors in development for cancer therapy to target many points along the signaling pathway simultaneously [83, 135]. Evidence emerging from biochemical studies points to differential phosphorylation rates associated with different tyrosine phosphorylation sites in the receptor [7, 139, 247, 273]. Moreover, recent structural studies have indicated that the clinically identified mutations are able to alter the regulatory elements within the receptor such as bypassing the dimer-mediated activation step leading to constitutive activity. Therefore, implications of the effect of such mutations on preferential signaling can be crucial in shaping the signaling response in altered cell lines derived from cancer patients.

In this chapter, we adopt a multiscale modeling approach to study EGFR mediated differential signaling. Through these modeling approaches, we are able to extend the

prior modeling of EGF-mediated signal transduction by considering specific EGFR TKD docking interactions mediated by differential binding and phosphorylation of different C-terminal peptide tyrosines on the RTK tail. By modeling signal flows through branching pathways of the EGFR TKD resolved on a molecular basis, we are able to transcribe the effects of molecular alterations in the receptor (e.g., mutant forms of the receptor) to differing kinetic behavior and downstream signaling response. We then extended this approach to study the effects of EGFR inhibition through TKIs. By employing molecular docking in combination with network modeling, we are able to quantify changes in the EC_{50} of receptor phosphorylation (i.e. 50% inhibition in the cellular context), and EC_{50} for the inhibition of downstream markers (ERK and Akt) upon treatment with the TKI erlotinib, in cell-lines carrying both wildtype (WT) and mutant forms of the receptor. Based upon our results, we conclude that the remarkable efficacy of the inhibitor erlotinib in the L834R mutant cell line can be attributed to relative gain in efficiency over the WT in inhibiting the Akt response.

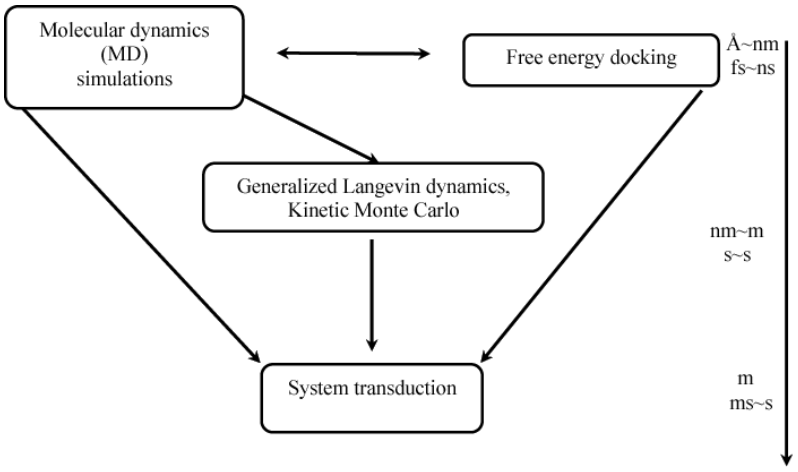

Figure 4.1 Scheme for hierarchical multiscale simulations.

4.2 Methods

4.2.1 Overall method description

A hierarchical multiscale computational approach (showed in Figure 4.1) based on molecular dynamics simulations, free energy based molecular docking simulations, and deterministic network-based kinetic modeling is utilized to study EGFR mediated differential signaling network.

4.2.2 Signal transduction

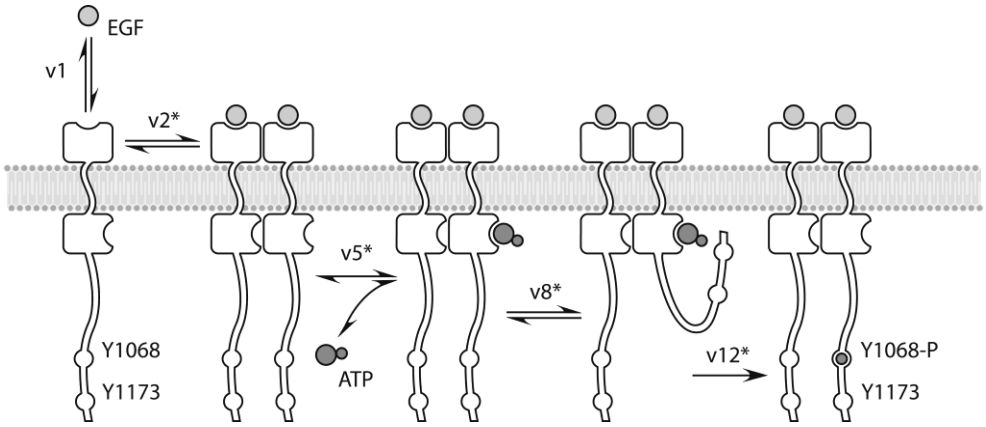

Figure 4.2 Phosphorylation model for EGFR. (v1) EGF binds reversibly to EGFR, which increases the receptor's affinity for other ligand-bound receptors (v2). Basal levels of spontaneous dimerization also occur (v3, v4; not shown). ATP binds reversibly to the catalytic site of the receptor (v5). The transient ATP-bound EGFR must then encounter a free unphosphorylated peptide on either cytoplasmic tail in a diffusion-limited step (v8 – v11). (v12) The receptor catalyzes autophosphorylation at either Y1068 or Y1173 (for clarity, only the Y1068 site is shown). Reactions with an asterisk denote more than one reaction involving similar kinetics but differing substrates (see Table 4.1).

Table 4.1 Reaction mechanisms and parameterization for EGFR phosphorylation reactions

Reaction	Mechanism	Parameter values		References
v1	[EGFR] + [EGF] $\rightleftharpoons$ [EGFR:EGF]	k_f = 3 • 10^{-7} M^{-1} s^{-1}	k_r = 3.8 • 10^{-3} s^{-1}	[143]
v2	[EGFR:EGF] + [EGFR:EGF] $\rightleftharpoons$ [$EGFR_2$]	k_f = 1 • 10^{-7} M^{-1} s^{-1}	k_r = 0.1 s^{-1}	[143]
v3 [a]	[EGFR] + [EGFR] $\rightleftharpoons$ [$EGFR_2$]	k_f = 1 • 10^{-7} M^{-1} s^{-1}	k_r = 1 • 10^{3} s^{-1}	[143] Modeled
v5 [b]	[$EGFR_2$] + [ATP] $\rightleftharpoons$ [$EGFR_2$:ATP]	K_m = 2.4 • 10^{-8} M^{-1}		[34]
v8 [c]	[$EGFR_2$:ATP] $\rightleftharpoons$ [$EGFR_2$:ATP:Y1068]	k_f = 6.51 • 10^5 s^{-1}	k_r = 769 s^{-1}	[85], Modeled
v10 [d]	[$EGFR_2$:ATP] $\rightleftharpoons$ [$EGFR_2$:ATP:Y1173]	k_f = 1.98 • 10^5 s^{-1}	k_r = 48.1 s^{-1}	[85], Modeled
v12 [e]	[$EGFR_2$:ATP:Y1068] $\rightarrow$ [$EGFR_2^{Y1068\text{-}P}$]	k_{cat} = 0.29 s^{-1}		[85]
v14 [f]	[$EGFR_2$:ATP:Y1173] $\rightarrow$ [$EGFR_2^{Y1173\text{-}P}$]	k_{cat} = 0.25 s^{-1}		[85]
v16 [g]	[$EGFR_2^{Y1068\text{-}P}$] $\rightleftharpoons$ [$EGFR_2$]	V_{max} = 4.5 • 10^{-7} M s^{-1}	K_m = 5.0 • 10^{-8} M	[143]

[a] An identical rate equation was employed for spontaneous dimerization of EGFR and EGFR:EGF (v4)

[b] Identical rate equations were employed for ATP binding to $EGFR_2^{Y1068}$ (v6) and $EGFR_2^{Y1173}$ (v7)

[c] An identical rate equation was employed for Y1068 binding to $EGFR_2^{Y1173\text{-}P}$:ATP (v9)

[d] An identical rate equation was employed for Y1173 binding to $EGFR_2^{Y1068\text{-}P}$:ATP (v11)

[e] An identical rate equation was employed for phosphorylation of $EGFR_2^{Y1173\text{-}P}$:ATP:Y1068 (v13)

[f] An identical rate equation was employed for phosphorylation of $EGFR_2^{Y1068\text{-}P}$:ATP:Y1173 (v15)

[g] Identical rate equations were employed for dephosphorylation of $EGFR_2^{Y1173\text{-}P}$ (v17) and $EGFR_2^{Y1173\text{-}P}$ (v18)

EGF stimulation in a cell results in the simultaneous activation of multiple pathways that are functionally interlinked [139, 218, 219]. The major pathways we included in this

chapter are EGF-ERK route and PI3K-Akt pathway. In the EGF-ERK pathway, the activation of Ras initiates a multistep phosphorylation cascade that leads to the activation of MAPKs, ERK1, and ERK2 [234]. ERK1 and ERK2 regulate transcription of molecules that are linked to cell proliferation, survival, and transformation. The PI3K-Akt pathway results in the activation of the downstream protein-serine/threonine kinase Akt. Prior studies have established that some growth hormone-stimulated membrane tyrosine kinase receptors interact with Shc adapter protein and phosphatidylinositol3'-kinase (PI3K), and consequently PI3K-activated Akt inhibits Raf-1 and the following ERK activity [39, 47, 111, 211]. Akt transduces signals that trigger a cascade of responses from cell growth and proliferation to survival and motility [174, 197, 214, 234]. The pathways we do not include in our model are phosphoinositol metabolism and signaling [20, 21, 114, 262, 272], activation and nuclear translocation of STATs [23, 25, 148, 277], clathrin-mediated endocytosis [75, 265] and molecular cross-talk with other receptor systems [27, 45, 90, 144, 163, 194, 195].

In the kinetic model employed here, signaling through the EGFR is modeled by combining three published models and augmented by our own set of reactions and calculations (see Figure 4.2 and Table 4.1). Phosphorylation and docking reactions are modeled according to Ref. [143]; the MAP kinase pathway reactions are modeled after Ref. [223]; Akt and PI3K activation are incorporated into the model as described in Ref. [39]. The similar parameterization and topology in these models allows us to construct a consistent, stable, and comprehensive system with results in good agreement with published experimental data [216]. Altogether, our model comprises of 74 reactions and 67 species. 17 of these reactions are novel and represent enhanced molecular resolution

and detail in EGFR activation, phosphorylation, and docking reactions (Figure 4.2 and Table 4.1). We note that we restrict our models to the early signaling phase (before receptor internalization) and thus our model does not take into consideration receptor signaling from endosomal compartments [115, 112, 113].

Receptor Dimerization. Ligand-bound receptors associate into dimers in a diffusion-limited step [143] with a second-order forward rate constant of 0.1 M^{-1} s^{-1} and a reverse rate of 0.1 s^{-1} (v2 in Table 4.1). To account for the observed basal levels of autophosphorylation in the absence of an agonist [54], we also implemented reactions in which unactivated EGFR can form homo- or heterodimers with other inactivated or activated receptors (v3, v4 in Table 4.1). Assuming a dissociation constant of 0.1 mM, we estimate the reverse rate for these reactions ($k_{r,3,4}$ = 1000 s^{-1}) to be 10,000-fold greater than that of two activated receptors.

Receptor Phosphorylation. In our model, the phosphorylation of the receptor occurs in three steps: (1) ATP binds reversibly to the active site of EGFR (v5-v7 in Table 4.1). The wildtype K_m = 24 nM reported by [34] was used for this reaction. (2) An unphosphorylated tyrosine peptide on the cytoplasmic tail diffuses and binds transiently to the active site (v8-v11 in Table 4.1). Peptide dissociation rates were parameterized according to values reported in the literature [85]. The reverse rate for these reactions was estimated by setting $k_{off} = (k_{on}) \bullet e^{+\Delta G/kT}$, where ΔG was calculated from the K_m for each peptide. (3) In the third step, a phosphate group from ATP is transferred to either Y1068 or Y1173 (v12-v15). k_{cat} values of 0.29 s^{-1} and 0.25 s^{-1} were used for catalysis of Y1068 and Y1173, respectively [85]. By undergoing two rounds of phosphorylation, a single

receptor may be phosphorylated at both sites. Phosphorylation reactions and parameters are listed in Table 4.1.

Docking interactions. Differential signaling is modeled by allowing only a subset of signaling proteins to bind to each form of the phosphorylated receptor (Figure 4.2). Using the most comprehensive model of network interactions to date [184], we partition signaling proteins into two groups. Grb2 and PI3K bind only to Y1068; PLCγ and Shc bind only to Y1173 [139, 247]. Higher order docking reactions are modeled as described previously [39, 143, 223].

Inhibition reactions. To model competitive inhibition by an ATP analog, we added the reaction EGFR + inhibitor $\rightleftharpoons$ EGFR:inhibitor in our model. The inhibition constant (K_I) of erlotinib for WT (17.5 nM) and L834R (6.25 nM) receptors are obtained by obtained from published experimental data [48].

The complete model is assembled and analyzed using the Systems Biology Toolbox for MATLAB (http://sbtoolbox.org/). Simulations are carried out using the CVODE solver from Sundials (http://www.llnl.gov/CASC/sundials) along with a compiled version of the model.

4.2.3 Molecular dynamics (MD)

An active conformation of the wildtype EGFR TKD domain is modeled using the 1M17 structure [248, 288] by adding hydrogen atoms and missing residues using the CHARMM [37] biomolecular simulation package. The resulting conformations are first energy minimized and subsequently equilibrated by performing molecular dynamics. A

L834R mutant active structure was also generated on the basis of the wildtype structure (mutation performed using the INSIGHT II package [4]) and subsequently energy minimized using CHARMM. The proteins are then solvated and neutralized by placing ions; the solvated models are minimized and equilibrated using the NAMD program [182] with periodic boundaries enforced and long-range electrostatics taken into consideration. Each system is subjected to a constant temperature and constant pressure molecular dynamics runs at 300 K and 1 atm followed by constant temperature equilibrium at 300K for 10 ns.

4.2.4 Molecular docking

Molecular docking is used to predict ligand binding in the absence of a ligand-bound crystal structure and functional affinity data. We employ AutoDock, an automated docking tool designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3-dimensional structure [178]. We perform a global conformational search using a multiple conformation docking strategy, in which the protein flexibility is taken into account implicitly by using multiple conformations, i.e., by sampling from several (100) frames of the 10 ns molecular dynamics trajectory. The lead docked conformations (based on lowest binding free energy) are extracted and clustered on the basis of their root-mean-squared deviation or RMSD values.

As control simulations, we validated the predicted results of our docking protocol by applying our method to study ligand binding to enzymes for which a solved crystal structure of the bound complex was available. In all three cases, namely ATP binding to cAMP, ATP binding to insulin receptor kinase, and erlotinib binding to EGFR TKD, the

predictions of the location and conformation of the bound ligand matched the crystallographic data very closely (data not shown). This multiple conformation docking protocol is employed in calculating the ATP, inhibitor, and Y1068/Y1173 peptide substrate binding affinities.

4.3 Results

4.3.1 Substrate binding affinities for EGFR TKD.

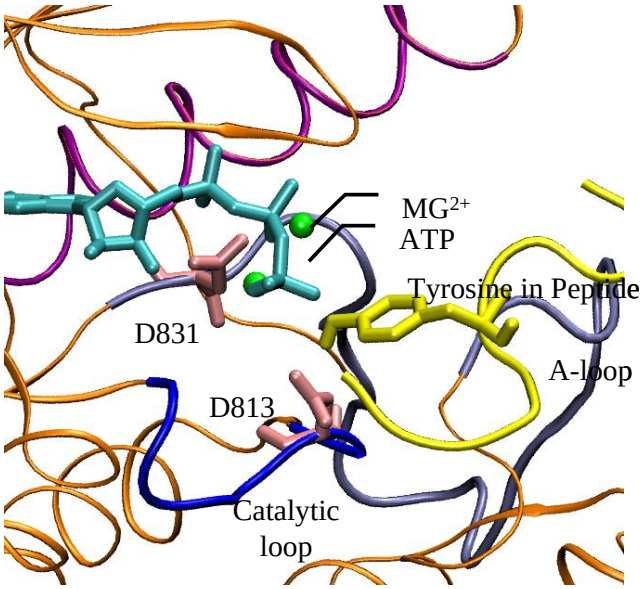

Figure 4.3 Binding modes for ATP (cyan) and the optimal peptide sequence (yellow) in the EGFR TKD domain. The geometry of the active site is consistent with that prescribed for an ideal two-metal-ion catalyzed mechanism and reveals a state that is poised for the phosphoryl transfer reaction. The alignment of the two divalent Mg2+ ions as well as the catalytic aspartates D813 and D831 are also shown.

In order to explore the C-terminal peptide-substrate-bound pre-catalytic conformation, we dock EGFR TKD substrate peptides to the catalytic site. First, we focus on an optimal peptide sequence EEEEYFELV as a control. This optimal sequence is selected for a maximal value in k_{cat}/K_M among EGFR TKD substrate peptides [246]. Based on 500-

1000 different genetic algorithm runs for docking of this peptide to EGFR TKD, we select the best binding conformation for binding as one that yields the lowest binding free energy simultaneously with the correct orientation for the tyrosine binding into the active site. The latter criterion is essential in ensuring that k_{cat} is not significantly compromised in obtaining a stronger binding. We depict in Figure 4.3, the peptide conformation bound to the active site, which is seen to be ideally poised for a two-metal-ion catalyzed phosphoryl transfer from ATP to the bound tyrosine [176, 250]: i.e., the relative positions of the ATP, Mg^{2+} ions, and the catalytic aspartates are in good alignment. Encouraged by this result, we performed peptide docking simulations for two substrate peptides (each being seven amino acids long) corresponding to the Y1068 (VPEYINQ) and Y1173 (NAEYLRV) sites of the C-terminal tail for both wildtype and L834R mutant structures. Based on similar criteria of lower binding free energy and optimal catalytic geometry, we calculated the peptide binding affinity corresponding to sites Y1068 and 1173 for wildtype and L834R mutant systems. Upon mutation, the calculated K_M for the Y1068 peptide decreased 20-fold (calculated $\Delta\Delta G$[wildtype$\rightarrow$L834R]= -2.2 kcal/mol), while the calculated K_M for the Y1173 peptide increased 4-fold (calculated $\Delta\Delta G$[wildtype$\rightarrow$L834R]= +0.9 kcal/mol). Analysis of the binding modes reveals that the relative alignment of the tyrosine with respect to the catalytic site is ideal (2.2 Å in the observed distance ξ between Tyr OH oxygen and Asp 813 $O_{\delta 2}$) for the two peptide tyrosines binding to wildtype, with the two mutant systems showing a 0-1.5 Å deviation. The differences can translate into differences in k_{cat} values associated with the tyrosine phosphorylation. Assuming a harmonic energy profile along the reaction coordinate ξ with a force constant K_ξ, the work done in reducing this distance is given by $1/2K_\xi\xi^2$. The

value of K_ξ is obtained by recording the fluctuations of ξ in the MD simulations ($K_\xi = k_B T/\sigma_\xi^2$=0.77 kcal/mol/Å^2 for the wildtype and 1.01 kcal/mol/Å^2 for L834R), where σ_ξ is the standard deviation associated with the fluctuations in ξ in the trajectories). The normalized values of k_{cat}/K_M can then be compared for the four systems at the same value of ξ=2.2 Å); the relative increase in k_{cat}/K_M: 1068 for the L834R in comparison to the wildtype is 20-fold (increase by 1-order of magnitude), while the relative decrease for the 1073 is 200-fold (decrease by 2-orders of magnitude). This differential sensitivity translates into differences in phosphorylation of the associated tyrosine sites, which ultimately leads to differential signaling and downstream activation in the EGFR network: the Y1173-phosphorylation is associated Shc mediated ERK activation and Y1068-phosphorylation is associated with Akt activation [139, 247].

4.3.2 Differential signaling through EGFR TKD

By employing the deterministic model we have calculated transient evolution of the EGF-mediated signaling network under a variety of initial conditions and parameter values. In particular, we have focused on two pools of normal ([EGFR]=100 nM or 33,000 receptors per cell) and over expressed ([EGFR]=2000nM or 660,000 receptors per cell) EGFR concentrations, each in the absence (serum starved, [EGF]=0 nM) and presence (serum cultured, [EGF]=8 nM) of the EGF ligand. Since our molecularly resolved model is currently focused on differential signaling through Y1068 and Y1173 phosphorylation sites of EGFR TKD, we track the integrated response of ERK and Akt phosphorylation levels in our simulations. Each output state was quantified according to the peak level of phosphorylation over the simulated time of 1000 s. Specifically, we perform a two-

dimensional scan over k_{cat}/K_M values associated with Y1068 and Y1173 phosphorylation in which the respective k_{cat}/K_M values are allowed to deviate from their default (wildtype) value over a logarithmic range of five log units. The result is a two-dimensional matrix in which each element represents output from a single simulation involving a unique pair of parameters (see Figure 4.4).

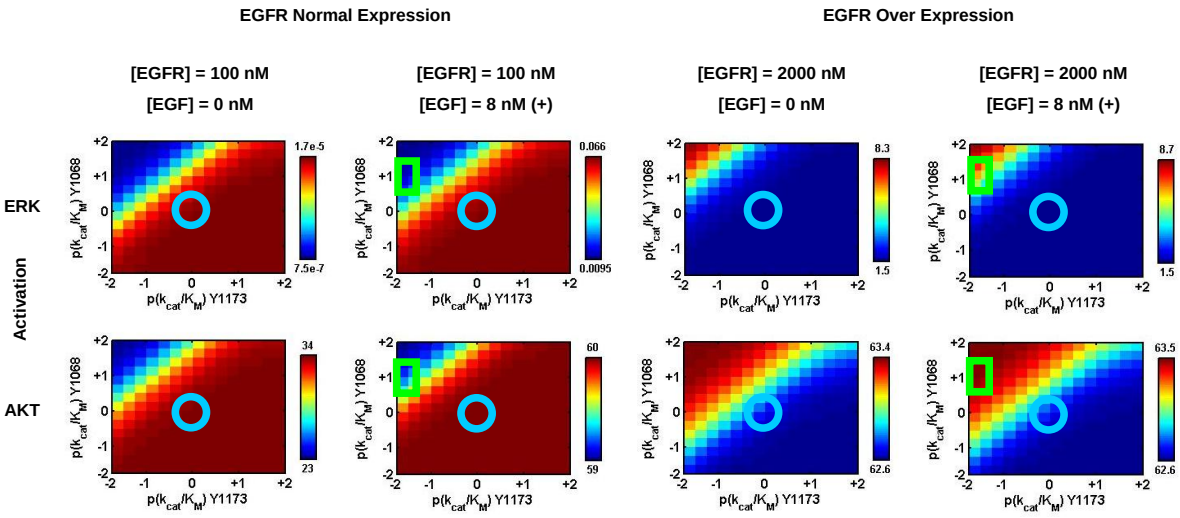

Figure 4.4 Calculated ERK and Akt phosphorylation levels in units of nM (peak-levels over the time course of 1000 s) under serum starved (EGF-) and serum cultured (EGF+) conditions for cell types with normal EGFR expression and EGFR over expression. In each panel, the blue circle denotes the wildtype levels and the green square denotes those relevant to the L834R mutant cell lines.

4.3.3 Inhibition of EGFR phosphorylation

Next, we examined the sensitivity of WT and mutant systems to inhibition in the cellular context by calculating receptor phosphorylation over a range of erlotinib concentrations. Simulations were performed for both 'normal' ([EGFR] = 100 nM) and 'over-expressed' ([EGFR] = 1000 nM) systems. All simulations were performed with (50 ng/ml or 8 nM) or without ligand (EGF) stimulation.

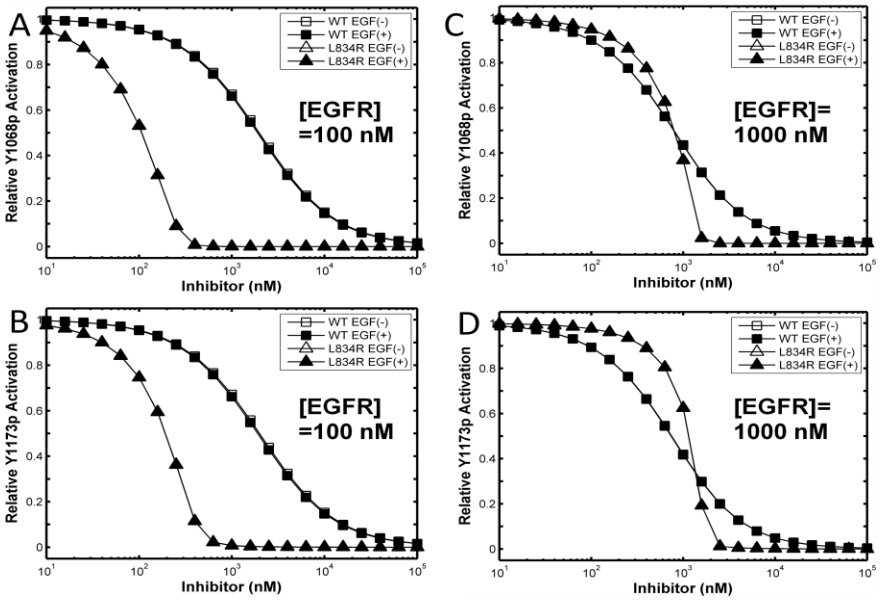

Figure 4.5 Relative inhibition of receptor phosphorylation. Tyr phosphorylation levels relative to that without inhibitor are plotted.

The EC_{50} (inhibitor concentration at which 50% of the activity is suppressed in the cellular context; this is different from IC_{50} because the non-linear and temporal effects due to signal-transduction is accounted for) for L834R was 50-fold lower (see Figure 4.5: A&B) than that of the WT (100 nM vs. 5000 nM) in the presence as well as absence of ligand. There were also no prominent differences between the inhibition at 1068 and 1173 sites. In over-expressed systems, there was no significant difference in EC_{50} among the four groups (Figure 4.5: C&D). All had an EC_{50} near 1000 nM.

4.3.4 Inhibition of downstream activation

The sensitivity of downstream signaling molecules ERK and Akt to inhibition for a range of erlotinib concentrations is shown in Figure 4.6.

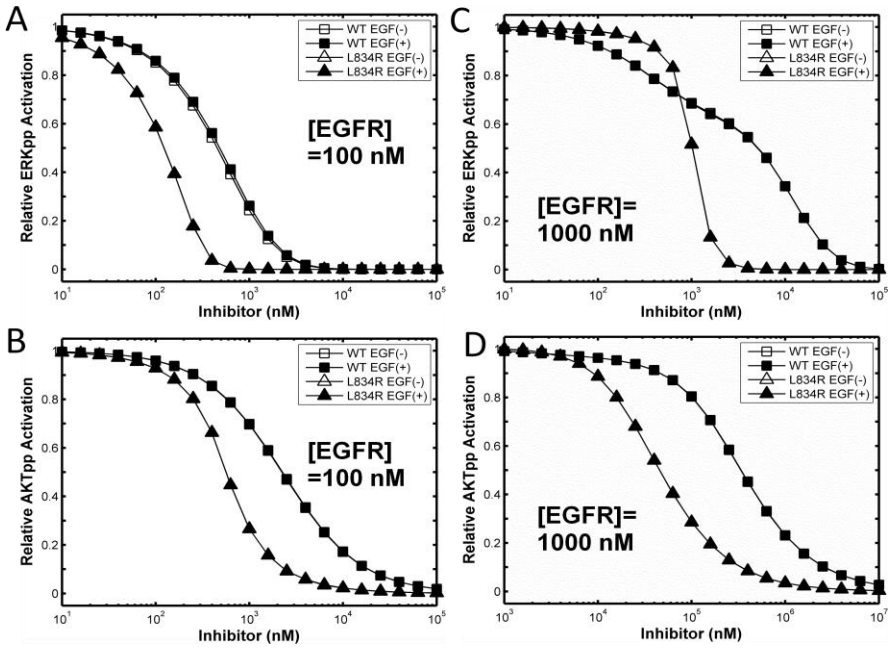

Figure 4.6 Inhibition of ERK and Akt activation in systems with normal receptor expression. Responses plotted relative to that without inhibition under the same conditions.

In normal expression systems (Figure 4.6 A&B), there was a nearly 7-fold increase in the efficiency of pERK inhibition for L834R (EC_{50} = 100 nM) compared to WT (EC_{50} = 700 nM) with and without ligand present. With respect to Akt activation, there was a 4-fold decrease in EC_{50} for L834R compared to WT (300 nM vs. 1200 nM) with and without ligand present. For the case of over-expressed receptors (Figure 4.6 C&D), inhibition of pERK is 4-fold more efficient for L834R compared to WT, while pAkt inhibition is almost 10-fold more efficient for the L834R mutant, although significantly more erlotinib is required to achieve inhibition (EC_{50}=40 μM). This result is consistent with the dramatically elevated Akt levels expected in systems that bear both the L834R mutation and a higher receptor count.

4.4 Discussion and conclusion

Based on an integration of a large body of knowledge accumulated over the years, the modular nature of EGFR-mediated signaling has become apparent [63, 131]. Through a hierarchical multiscale modeling scheme, we have realized a molecular resolution to this system-level view, which can potentially augment our predictive as well as comprehensive capabilities and rationalize the changes in early signaling characteristics of altered and mutant cell lines. Our overall findings are summarized below.

Our results from the free energy docking simulations suggest enhanced binding of the inhibitor erlotinib to the L834R mutant system in comparison to wildtype binding. Our results also reveal preferential binding and phosphorylation of certain C-terminal tyrosine substrates (but not others) to the L834R mutant in comparison to the wildtype: that is, the calculated change in the turn over rate (k_{cat}/K_M) upon the L834R mutation is markedly different for the Y1173 peptide in comparison to the Y1068 peptide. This results in a differential sensitivity at the receptor level. Specifically, the estimated increase in k_{cat}/K_M for the Y1068 peptide is ~10-fold in comparison to a decrease of ~100-fold for the Y1173 peptide. Thus, we expect to see a net increase in phosphorylation of Y1068 and a net decrease in phosphorylation of Y1173 for the L834R mutant, in comparison to wildtype levels. The biochemical results of Sordella et al. [247] and Chen et al. [51] corroborate this view qualitatively; in these studies, the immunoblots measuring the degree of phosphorylated receptor at the Y1173 site show a modest decrease for the L834R mutant in contrast to a strong increase for the Y1068 site, both in comparison to their respective wildtype measurements.

Our systems level simulation illustrates how the altered affinities at the receptor level for the wildtype and L834R translates into changes in the downstream response. In Figure 4.4, we explored how the variation in Y1173 and Y1068 phosphorylation affects the peak responses of ERK and Akt for two different EGFR concentrations, each in the absence or presence of EGF stimulation. The calculated responses for the wildtype values are highlighted by the blue circles at the origin of each of the panels. The responses for L834R are identified by the green square centered at a $(k_{cat}/K_M)/(k_{cat}/K_M)_{wildtype}$ value of 1/100 (-2 in $\log_{10}$ units) for Y1073 and 10 (+1 in $\log_{10}$ units) for Y1068. Owing to the constitutively active characteristic of the L834R mutant, the signals relevant for the L834R system (even in the absence of EGF) are analogous to those in the panels with EGF stimulation. As indicated by the color maps in these two scans, increasing the degree of phosphorylation of the Y1173 site (positive x-direction) leads to an overall increase in ERK response. This finding is consistent with our model which is based on Y1173 being an important docking site for the Shc and Grb-2 adaptors that lead to Ras activation and subsequent Erk phosphorylation. For normal EGFR expression, the calculated ERK response in the absence of EGF is low for the L834R mutant (0.01 nM) but negligible (zero) for the wildtype. In the presence of EGF the ERK response for L834R is 7-fold lower than that for the wildtype. For over expressed EGFR, the ERK activation is 7-fold higher in L834R compared to wildtype both in the presence as well as in the absence of EGF stimulation. Thus, Erk stimulation is higher in EGFR over-expressed cells. The Akt response relative to the wildtype is 2-fold higher for the mutant cells under no EGF stimulation and about the same under EGF stimulation for normal

EGFR expression. The Akt response does not seem to be affected by mutation for high EGFR expression either in the presence or absence of EGF.

Our calculated responses for ERK and Akt signaling for normal EGFR expression agree with the qualitative experimental observations of Sordella et al. [247] and Tracy et al. [259]: namely, that the Erk stimulation levels decrease and the Akt levels are maintained for the mutant cells with normal EGFR expression levels. This preferential activation of Akt in L834R mutant cell lines observed in these experiments is consistent with the systems level predictions from our simulations. Moreover, our calculations predict that the preferential Akt activation is not featured in EGFR over-expressed cells. Through multiscale modeling, we have also shown how this framework can be utilized to predict the efficacy of the inhibition in cell-lines, especially those harboring mutations in the receptor. We find that the mutant cell line is more susceptible to inhibition by TKIs both at the level of curbing the receptor phosphorylation as well as that of downstream (ERK and Akt) activation. Considering that the absolute pAkt levels are 5-fold higher than those for pERK in the WT and 100-fold higher in the mutant, the remarkable effect of the drug in non-small-cell lung cancer cell lines carrying the mutation can be attributed to the gain in efficacy with respect to Akt inhibition.

In light of the overall agreement between simulations and experiment, we can begin to explore the origins of inhibitor sensitivity in L834R mutant cell lines. The enhanced sensitivity of erlotinib to L834R mutant arises from a combination of several effects: (1) the mutant system has a higher binding affinity to the inhibitor; (2) the constitutive activity of L834R deems the mutant to be a more potent target for therapeutic inhibition;

(3) The differences in the predicted downstream response (between the wildtype and the mutant) accounted for by early signaling events can potentially lead to a divergent long-term behavior in the two systems causing a differential sensitivity to inhibition.

In conclusion, we have described the role of a hierarchical multiscale computational approach in resolving at a molecular level the systems-level model of EGFR-mediated signal transduction, to help rationalize the collective information emerging from biochemical, cell biology, and clinical studies. Our approach enables us to transcribe differences in molecular mechanisms as well as functional activity originating at the single molecule level into tangible differences in early signaling events. Therefore, subject to the well appreciated modeling limitations, this predictive capability is useful in differentiating signaling characteristics of mutant cell lines from the native wildtype case. This molecular-based refinement of the overall EGFR signal transduction model is a first attempt at untangling the differential nature of early signaling events triggered by receptor activation and their consequences in shaping the signal transduction in altered cell lines derived from different pathologies. A complete model, which is currently being pursued, will not only require resolving the differential characteristics of all of the tyrosine phosphorylation sites in EGFR TKD and their associated substrate recognition properties in EGFR, but also the extension to other Erb family members in the context of homo- and hetero-dimers. Transactivation of EGFR occurring through ligand-induced receptor heterodimerization [270] combined with a potential for differential signaling adds a palette of finer control elements in the ErbB-family signaling network. Indeed, recent studies have identified possible mutations in ErbB2 that may correlate with drug sensitivity in different cancers [155, 156]. The computational tools described here are

ideal for assessing the likely effect of novel EGFR and ErbB2 mutations and determining whether the drug-sensitizing mutations implicated in non-small-cell lung cancer also occur in other cancers. We believe that our model driven approach will in the long-term significantly impact the optimization of future small molecule therapeutic inhibition strategies.

Chapter 5 A Flexible-protein Molecular Docking Study of the Binding of Ruthenium Complex Compounds to PIM1, GSK-3β, and CDK2/Cyclin A Protein Kinases

5.1 Introduction

As the overall protein fold is conserved in the kinase family, one of the major challenges is to understand the molecular basis for inhibitor sensitivity and to design small-molecule compounds which are highly selective (specific) to the targeted protein kinase. Progress towards the design of selective small-molecule kinase inhibitors based on the exploitation of distinct features presented by ATP-binding site has recently been reviewed [258], [183]. Vast majority of specific enzyme inhibitors are small organic molecules that gain their specificity by a combination of weak interactions, including hydrogen bonding, electrostatic contacts, and hydrophobic interactions. Recently, a novel strategy has been introduced for the design of small-molecule enzyme inhibitors by using substitutionally inert organometallic scaffolds [31, 172], based on the hypothesis that complementing organic elements with a metal center may provide new opportunities for building three-dimensional structures with unique and defined shapes. In particular, a class of half-sandwich ruthenium complex compounds based on the scaffold shown in Figure 5.1have been reported to show high affinities and promising selectivity profiles for protein kinases and lipid kinases [11, 32, 278].

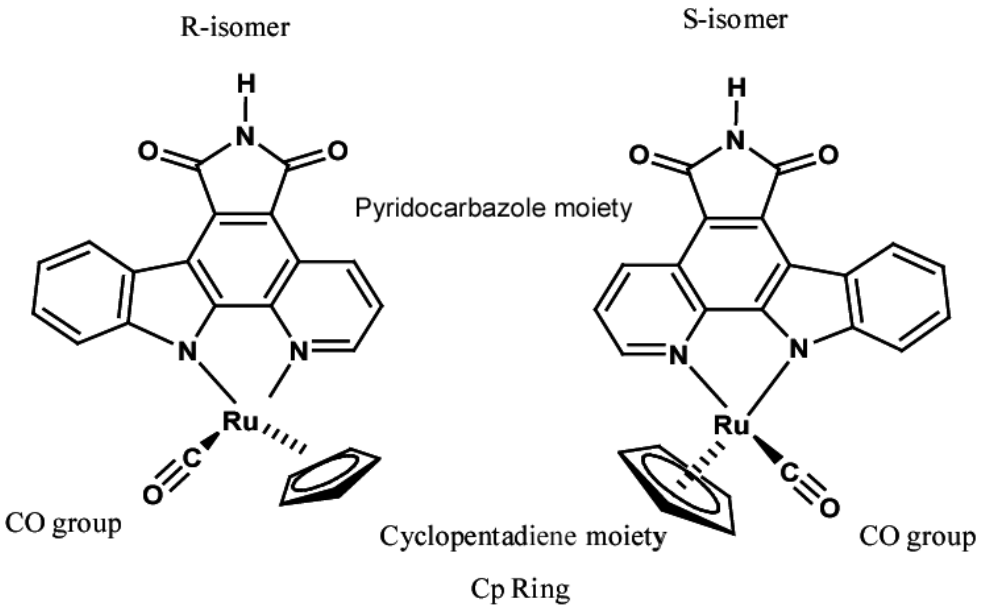

Figure 5.1 Chemical structure of ruthenium complex compound (named compound 1 in this chapter) analyzed in this chapter. The R- enantiomer is depicted on the left and the S-enantiomer on the right.

The new ruthenium complex compounds are designed to mimic the shape of staurosporine, a well-known protein kinase inhibitor, by replacing the indolocarbazole alkaloid scaffold with metal complexes in which the structural features of the indolocarbazole heterocycle is retained. The ruthenium metal center plays a structural role by organizing the organic ligands in the three-dimensional space. As shown in Figure 5.1, the coordination geometry around the ruthenium is pseudo-octahedral, formed by the pyridocarbazole ligand, the CO group oriented perpendicular to the pyridocarbazole plane, and the cyclopentadiene (Cp) moiety. The compound was designed to bind with the kinase by forming hydrogen bonds to the backbone residues at the hinge region of the kinases. However, unlike staurosporine, which is a nonspecific nanomolar inhibitor for most protein kinases, these ruthenium half-sandwich compounds show remarkable selectivity profiles. In particular, profiling the racemic mixture of (*R/S*)-**1** against more

than 50 protein kinases in vitro shows high selectivity of this class of compounds for PIM1 (IC_{50}= 3nM, 100 μm ATP) and GSK-3β (IC_{50}= 50nM, 100 μm ATP)[11]. Interestingly, the phylogenetically and structurally closely related cyclin-dependent kinases (CDKs) are not significantly inhibited, with IC_{50}= 3μM (100 μm ATP) for the CDK2/cyclin A complex [11].

Table 5.1 Sequence alignment of residues around active site of the three protein kinases, PIM1, GSK3, and CDK2. Residues which are significantly different are colored.

Kinases	N-terminal lobe							Linker region					C-terminal lobe			
PIM-1	L44	G45	S46	F49	V52	A65	I104	L120	E121	R122	-	P123	E171	L174	I185	D186
GSK-3β	I62	G63	N64	F67	V70	A83	V110	L132	D133	Y134	V135	P136	Q185	L188	C199	D200
CDK2	I10	G11	E12	Y15	V18	A31	V64	F80	E81	F82	L83	H84	Q131	L134	A144	D145

Sequence alignment analysis (Table 5.1) shows a high degree of conservation among the ATP binding pocket residues of the three kinases. Superposition of the kinase active sites using the α-carbon atoms of the residues around the ATP pocket also shows that the structure of the binding site is highly similar among the three kinases (Figure 5.2). Despite these similarities, differences in specific amino acid positions are noted (colored in Table 5.1 and labeled in Figure 5.2) , which possibly have a bearing on the differences in the binding interactions, and hence the inhibitor selectivity.

Although it is of great interest to understand the interaction of ruthenium compounds with the three protein kinases in atomic details, [217, 267] only one crystal structure, i.e., PIM1 bound with (*R*)-**1**, is currently available. It is not clear how the other enantiomer, (*S*)-**1** is bound to PIM1. The bound conformations of both (*R*)-**1** and (*S*)-**1** are not known from experiments either. In this chapter, we modeled the interactions between the

ruthenium scaffold and three protein kinases, PIM1, GSK-3β and CDK2/cyclin A, targeting the selectivity profile of the compound.

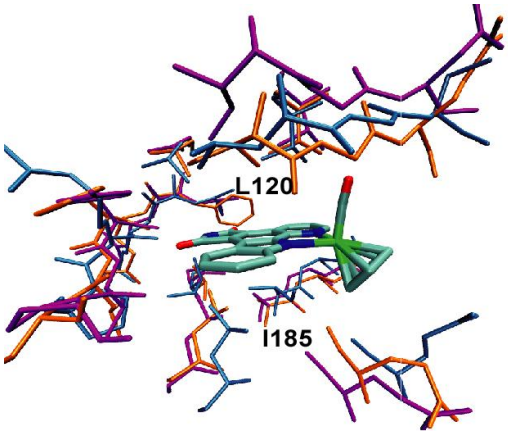

Figure 5.2 Structural alignment of the ATP binding pockets of three protein kinases, PIM1, GSK3, and CDK2/CYCLIN A.

Molecular docking is frequently used to predict ligand binding in the absence of ligand-bound crystal structures and functional affinity data [36, 109]. While ligand flexibility is accounted for in most docking programs, many treat the protein as a rigid body. However, in practical scenarios in which the receptor structure is derived either from an experimentally determined structure of the apo- kinase or from a structure where the kinase is complexed with a different ligand, the rigid-receptor docking often fails to predict proper bound conformation. Hence, efforts have been made to account for the protein flexibility [50, 275]. In this chapter, we applied the ensemble docking procedure to predict bound conformations of ruthenium compounds (*R/S*)-**1** against the three protein kinases PIM1, GSK-3β and CDK2/cyclin A. Our docking results reveal the protein-ligand interaction and suggest possible structural factors that may give rise to different binding affinities of the proteins.

5.2 Methods

The flow chart in Figure 5.3 depicts the ensemble docking protocol we have followed. To summarize, protein conformations are sampled with all-atom molecular dynamics (MD) simulations with explicit solvent. Charges and geometry of the ligands are calculated using electronic structure (ab-initio) methods. The ligands were then docked into each of the protein conformations using the docking program AutoDock3.0. The predicted bound conformations of ligand are subsequently clustered.

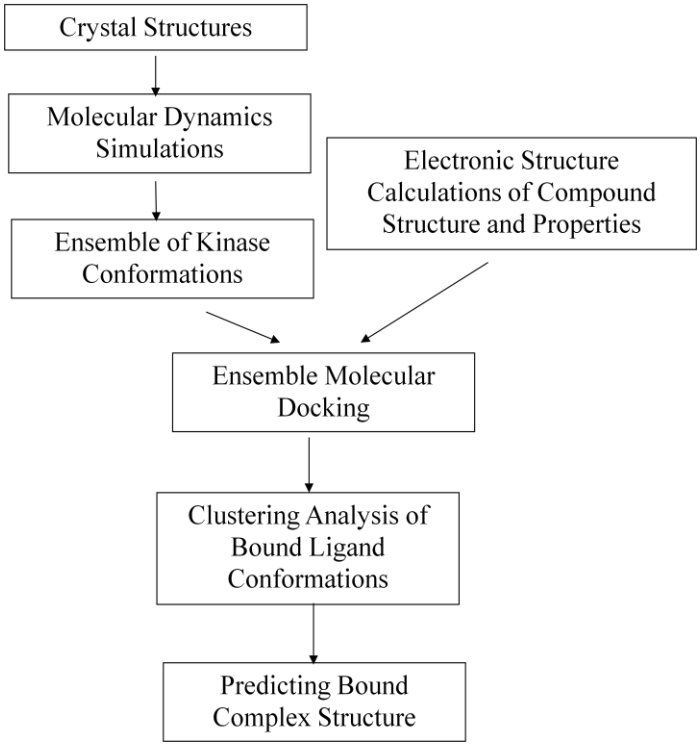

Figure 5.3 Flow chart of ensemble docking method

5.2.1 Ab-Initio electronic-structure calculations of Ru-complex compounds

The geometries of both enantiomers of compound 1 are optimized using the Gaussian 98 [254] program at the Hartree-Fock (HF) [254] level. For the hydrogen, carbon, oxygen and nitrogen, 6-31G* basis set is used while for the Ru atom, the Los Alamos ECP (effective core potential) plus DZ (double zeta) basis set (LanL2DZ [136]) is used. The geometry optimization of the molecule returns the minimum-energy geometry of the molecule and the spatial electron-density map at 0 K. Following the procedure of Aiken et.al.[6], the partial (Mulliken) charges on each atom are inferred from the ab-initio calculations which are used in determining the molecular descriptors of the inhibitor compound and are used further in the docking calculations.

5.2.2 Protein conformation preparation

The native and non-native protein conformations of PIM1 kinase are constructed based on two different crystallographic structures from the protein data bank: (1) the structure of the PIM1/(*R*)-**1** bound complex (PDB ID: 2BZH), and (2) the structure of PIM1 bound to an ATP analog (PDB ID: 1YXT). For the GSK-3β and CDK2/cyclin A systems, conformations derived from crystal complex structures with the ATP analog (1I09 for GSK-3β and 1QMZ for CDK2/cyclin A) are employed in the docking simulations. Four apo-kinase conformations are generated based on these crystal structures, by removing ligands and adding hydrogen atoms and other missing residues using CHARMM [37] biomolecular simulation package.

To generate an ensemble of conformations for each kinase, molecular dynamics (MD) simulations for each protein kinase are performed with the respective crystal structures serving as the starting conformation. The proteins are explicitly solvated using the TIP3P

model for water and neutralized by placing ions (sodium and chloride) at an ionic strength of 150 mM. The ions are placed at positions of electrostatic extrema predicted by the mean-field Debye-Huckel calculations. The solvated models are energy minimized, heated to 300 K, and equilibrated at constant temperature and constant pressure (300 K and 1 atm) using the NAMD simulation package [141] in conjunction with the CHARMM27 force-field [166]. All our simulations are performed on a fully periodic system and include long-range electrostatics using the Particle Mesh Ewald algorithm [14]. The extent of the equilibration phase is determined by tracking the plateau behavior of the RMSD (root mean square deviation) of the protein calculated with respect to its starting conformation. For each protein kinase system, 10 ns of dynamics runs are generatedFor each system, we extract 100 protein conformations at uniform intervals from the last 8 ns of the respective trajectories for use in our ensemble docking protocol. In order to make the comparison of different systems easier, all the protein conformations are aligned with respect to the PIM1 coordinates in 2BZH structure based on residues within 15 Å of the ATP binding site.

5.2.3 Docking protocol

AutoDock3.0.5 program [178] is used for docking simulation. For each protein structure, the non-polar hydrogen atoms are merged to heavy atoms and Kollman charges and solvation parameters are then assigned to the protein atoms using AutoDockTool (ADT) [1]. Grid maps for each protein conformation of dimension 126×126×126 points with a grid spacing of 0.184 Å are constructed, encompassing the entire ATP binding site. The Lamarckian Genetic Algorithm (LGA)[178] is applied to explore the conformational

space of the ligand using the scoring function. In each docking run, the initial population is set to 50 individuals, the maximum number of energy evaluations is set to 10^8, and the generation of GA run is set to 10,000. 30 GA runs are performed for each protein conformation. This choice for the set of run-parameters is adequate to achieve convergence of the docking results.

5.2.4 Ligand conformation clustering

Ligand conformations are clustered using the hierarchical clustering algorithm in Matlab [2]. For each docking run, the root mean squared deviation (RMSD) value between each pair of all docking-generated ligand conformations (30 for single conformation docking with the crystal structure and 3000 for ensemble docking) are calculated and used as the distance matrix to build a hierarchical tree by progressively merging clusters using unweighted average distance of each cluster. Then a cut-off RMSD value is used to cluster ligand conformations based on the hierarchical tree. We used the pairwise RMSD distribution of docked conformation to determine the optimal radius [152]. The distribution of pairwise RMSD of all the docked conformations are depicted in Figure 5.4. The minimum RMSD values after the first peak are chosen as the optimal clustering radius, namely, 2.0 Å for single conformation docking and 2.5 Å for ensemble docking. The conformations with the lowest docked energy in each cluster are reported as the predicted bound conformations for each system.

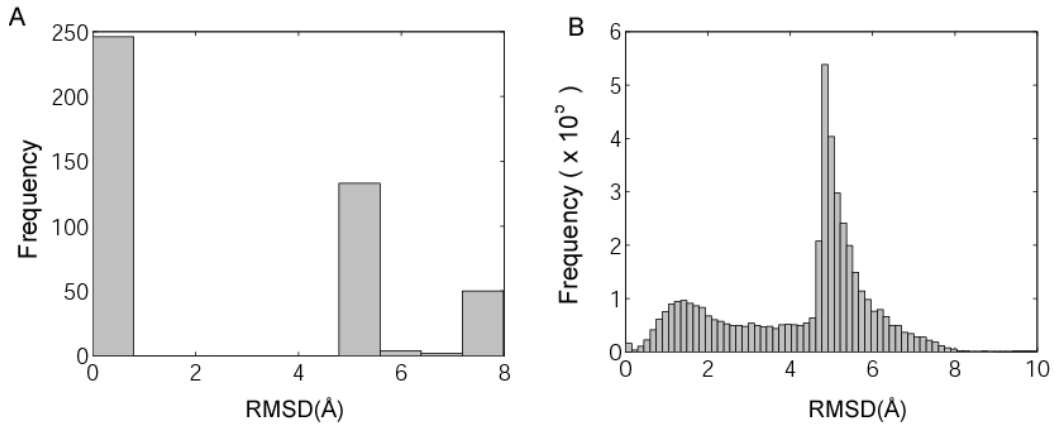

Figure 5.4 Pairwise RMSD distribution histogram of docked conformations for PIM1/(R)-1 system. A: single conformation docking, B: Ensemble docking

5.3 Results

5.3.1 Ensemble docking protocol counts the protein flexibility and improves the prediction of bound complex structure

In this chapter, we are interested in predicting and comparing the interaction between Ru-complex compound and three protein kinases, PIM1, GSK-3β and CDK2/cyclin A to provide insight on structural basis of the selectivity profile of the compound. Docking ligand to non-native protein conformation (apo- protein or protein complexed with other ligands) is still a challenging task. To account for the protein flexibility, we used the ensemble docking protocol, i.e., docking the Ru-complex compound into an ensemble of protein snapshots generated by MD simulations.

The protocol was first tested using PIM1/(*R*)-**1** system, where a crystal structure of the exact complex has been solved. To examine the reliability of different docking protocols, we performed two test cases to dock the (*R*)-**1** compound to the PIM1 protein kinase. The

first test was the docking to the PIM1 structure (PDB ID: 2BZH) in which (*R*)-**1** was co-crystallized. In the second test, the crystal structure of PIM1 (PDB ID: 1YXT) bound with an ATP analog was used instead. In each test, both the traditional docking protocol based on single protein structure and the ensemble-docking (using 100 conformations) protocol were applied to generate complex structures. The actual complex structure (2BZH) determined from crystallography serves as the "native" pose, or a reference to evaluate the docking results.

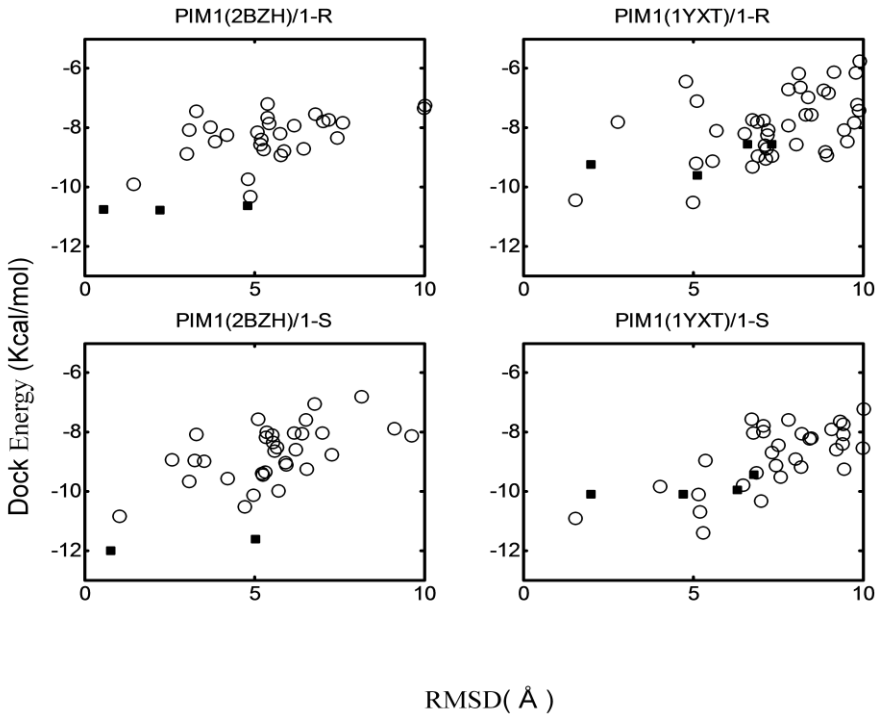

Figure 5.5 RMSD relative to the reference structure versus the docked energy score for each predicted conformation of compounds (R/S)-1 to two PIM1 structures, i.e. the native structure of PIM1 bound to compound (R)-1 (PDBID: 2BZH) and the non-native structure of PIM1 bound to an ATP analog (PDBID:1YXT). The symbols ■ corresponds to single conformation docking results and the symbols ○ correspond to ensemble docking results.

Our results in Figure 5.5 show that the predicted ligand-bound structures from single-conformation docking can be classified into several clusters with similar dock energies.

When ensemble docking was used and the flexibility of the protein was thus accounted, more complex conformations were generated and a better discrimination between the lowest (dock) energy conformations of different clusters was observed. Furthermore, conformations similar to the native pose emerge among the top clusters when ranked by dock energy.

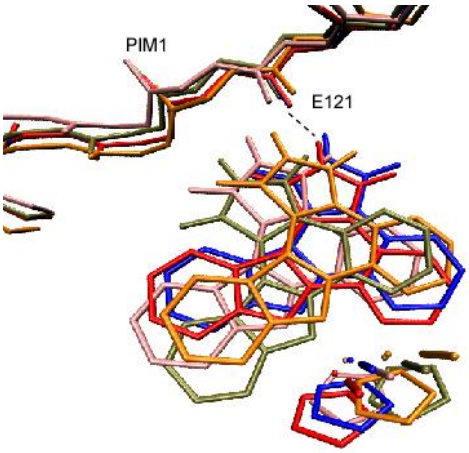

Figure 5.6 Predicted bound conformations of the inhibitor with the lowest RMSD to the reference structure from single conformation and ensemble conformation docking are aligned together with crystal structure. Blue: the single conformation docking of compound (R)-1 to the native PIM1 structure (PDBID: 2BZH). Orange: Ensemble conformation docking of compound (R)-1 to the PIM1 ensemble of structures based on native PIM1 structure. Tan: Single conformation docking of compound (R)-1 to the non-native PIM1 structure (PDBID: 1YXT). Pink: Ensemble conformation docking of compound (R)-1 to the ensemble of PIM1 structures based on non-native PIM1 structure. Red: Crystal structures (PDB ID: 2BZH).

All predicted complex structures were superimposed against the native pose (2BZH), with the closest one from each docking task shown in Figure 5.6. We also listed the parameters of the critical hydrogen bond between the NH group of the inhibitor and GLU121 of the protein in Table 5.2. When single-conformation docking was used in combination with the exact protein structure (2BZH), not surprisingly, the results (*blue* structure) closely resemble the native pose present in the same crystal structure. However,

when the approximate protein structure (1YXT) was used, the single-conformation docking failed to correctly present the critical hydrogen bond in the native pose. In contrast, when ensemble docking was applied on either the exact or the approximate protein structure, reasonable bound conformations (*orange* and *pink* structures) with the proper hydrogen bond and consistent with the native pose were among the sampled poses.

Table 5.2 Parameters of the hydrogen bond between the NH group of compound 1-R and the residue GLU121 in PIM1 for complex conformations in Figure 5.6.

Structures	Distance (Å)	Angle (Degrees)
Red (Crystal Structure)	2.0	151.02
Blue (native/single docking)	1.72	140.16
Orange (native/ensemble docking)	1.76	137.39
Tan (nonnative/single docking)	3.55	114.33
Pink (nonnative/ensemble docking)	1.65	135.03

To show how the ensemble of conformations provides a more suitable ligand docking environment, the 100 conformations of the PIM1 ensemble corresponding to non-native structure (PDBID: 1YXT) are aligned together and depicted in Figure 5.7 (A). It is evident that the snapshots in our ensemble simulations sample the kinase flexibility by exploring both backbone and side chain fluctuations around the crystallographic conformation used as the starting point for the MD simulations. In Figure 5.7 (B), the initial non-native PIM1 conformation (PDBID: 1YXT) and the conformation which binds the inhibitor with the lowest dock energy are aligned together and the key residues

around the binding pocket are depicted. The comparison reveals the shifting of key residues, in particular, GLU171, ASP128 and ASP186, from the initial non-native structure, and the rearrangement of the flexible Gly-rich loop (Gly45) to better position the CO group and the Cp ring.

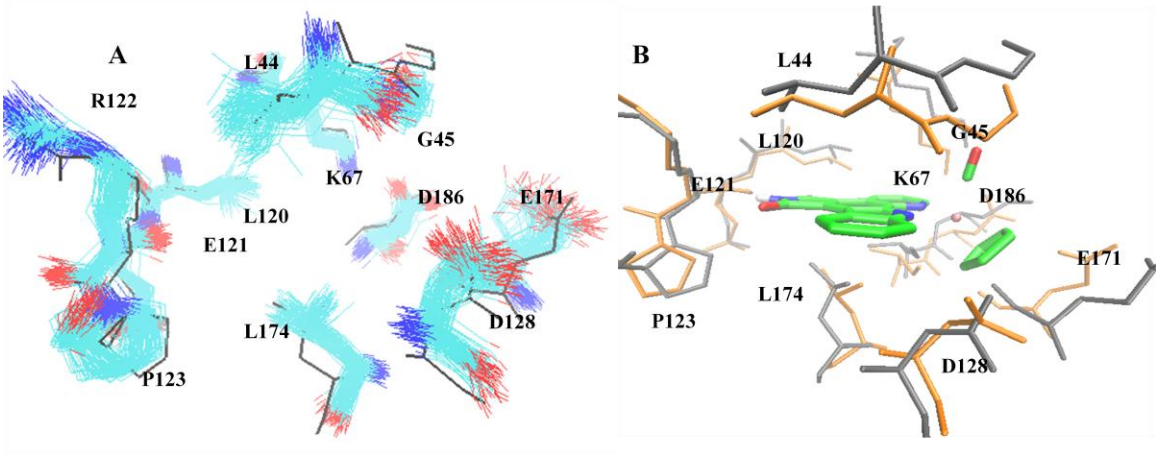

Figure 5.7 (A) Snapshots of PIM1 conformations generated from MD simulations based on the non-native PIM1 structure (PDBID: 1YXT). Snapshots are aligned by overall RMSD, and colored by atom types. The black conformation is the crystallized non-native PIM1 conformation. (B) Protein conformation comparison between the non-native crystallized structure of PIM1 (PDBID: 1YXT, gray) and the predicted protein bound conformation with the lowest RMSD to the reference structure using the ensemble conformation docking corresponding to this non-native PIM1 system (orange).

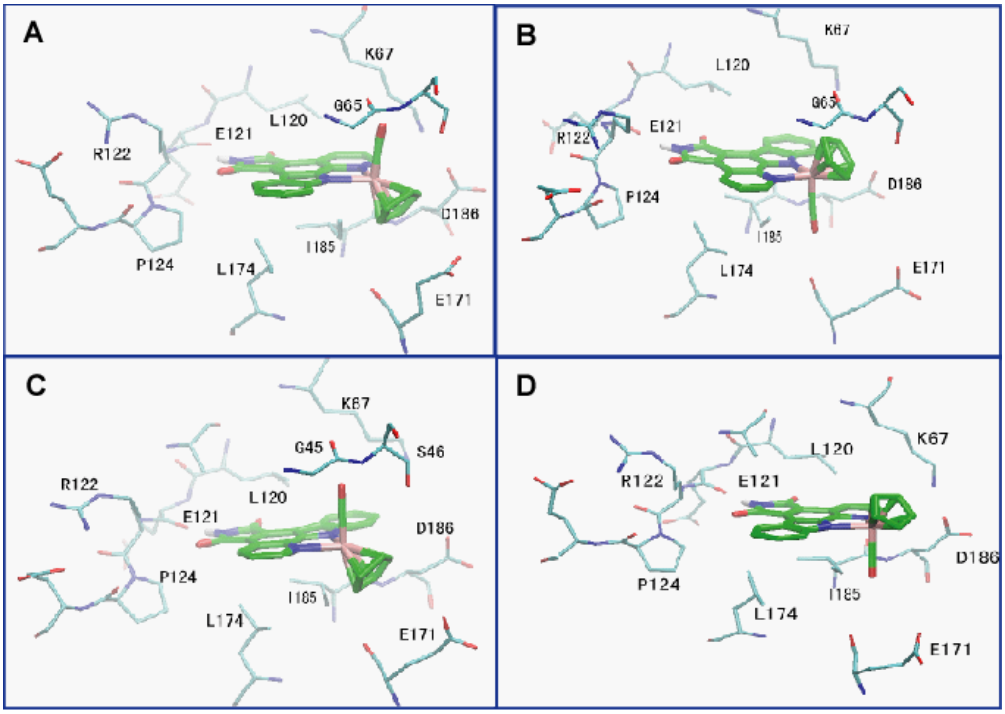

Figure 5.8 Top two ranks of conformations of compound (R/S)-1 bound to PIM1 predicted by ensemble docking. (A), (B): compound (R)-1. (C), (D): compound (S)-1. We denote the similar conformations in (A), (C) as conformation I and those in (B), (D) as conformation II.

As we noted in Figure 5.6, the non-native PIM1 structure fails to account for the conserved hydrogen bond between GLU121:O – (*R*)-**1**:H1, which is captured after the slight rearrangement of active-site residues using the ensemble docking protocol. Thus, the implicit protein flexibility in our protocol yields a better docked score for this conformation. Even though protein kinases are known to assume multiple conformational states in a rugged energy landscape [94] which are not exhaustively sampled in short MD simulations, we find that by sampling the fluctuations of protein conformation around a given initial state of our control system, our ensemble protocol demonstrates clear improvement in the prediction of the ligand-bound structure of the complex.

5.3.2 Two dominated conformations are predicted for (*R/S*)-1 to bind with GSK-3β and PIM-1

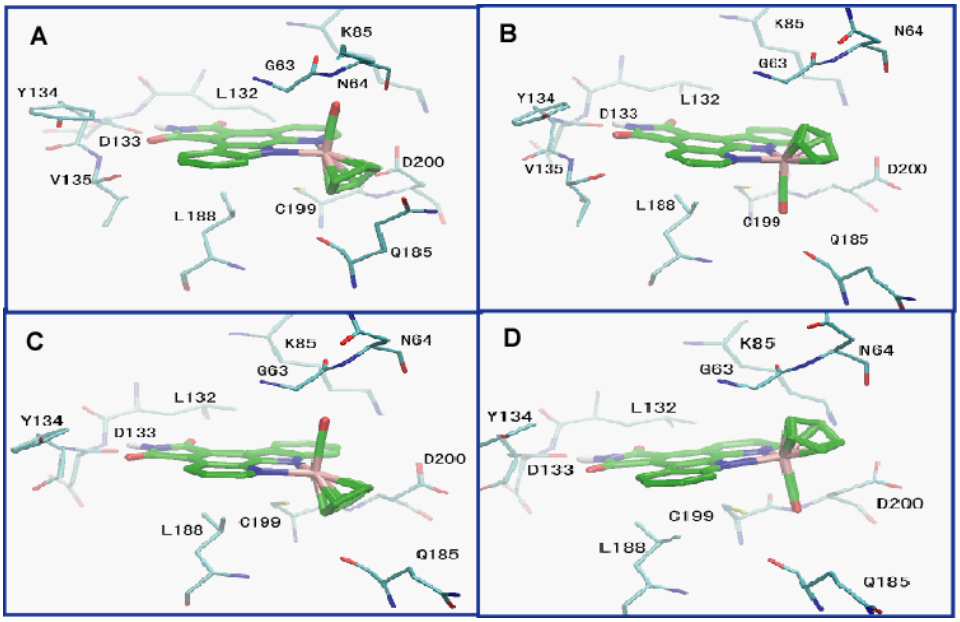

Figure 5.9 Top two ranks of conformations of compound (R/S)-1 bound to GSK3-β predicted by ensemble docking. (A), (B): compound (R)-1. (C), (D): compound (S)-1. We denote the similar conformations in (A), (C) as conformation I and those in (B), (D) as conformation II.

In Figure 5.5, we can see that for both the enantiomers, two conformations are predicted with relatively lower docking energies compared to other conformation, with RMSDs of 2 Å and 5 Å from the crystallized conformation, respectively. The two predicted structures for both enantiomers are depicted in Figure 5.8. In all of the conformations, the pyridocarbazole moiety forms a hydrogen bond between the maleimide NH group and the backbone carbonyl oxygen atom of residue within the hinge region (GLU 121) of the kinase, which mimics the hydrogen-bonding pattern of the ATP as well as staurosporine with the kinase, and is present in the PIM1/(*R*)-**1** crystal structure. One of the two conformations, which we denote as conformation I (Figure 5.8 A and C), is very similar

to the reference structure constructed from crystallography. In the second conformation, which we denote as conformation II, the CO and Cp groups occupy opposite (swapped) positions relative to conformation I. Thus the two conformations I and II are related by a 180° flip around an axis through the NH group of the pyridocarbazole moiety. The existence of conformation II as a stable structure has not been confirmed through crystallographic studies of the PIM1 kinase. However, this conformation is very close to a recent structure of compound (*S*)-**2** bound to a lipid kinase (PI3Kγ) [278]. Hence, we propose conformation II as a competing alternative bound conformation for the inhibitor bound to the protein kinase.

The same two conformations of compound (*R/S*)-**1** also turn out to be the most dominant bound conformation with GSK-3β through ensemble docking as depicted in Figure 5.8, making the binding characteristics of the (*R/S*)-**1** Ru-compound with GSK-3β similar to those with PIM1.

5.3.3 A unique bound conformation of the ruthenium compound dominates its binding to CDK2

Our results for the bound conformations (*R/S*)-**1** to CDK2/cyclin A are in stark contrast to those for PIM1 and GSK-3β. The conformation close to the two conformation for PIM1 and GSK-3β are ranked lower (4^{th} for (*R*)-**1**) or not present in our prediction ((*S*)-**1**). The disfavor of these two conformations is an evidence for unfavorable binding of this compound to CDK2/cyclin A in the native conformation observed in PIM1 and GSK-3β.

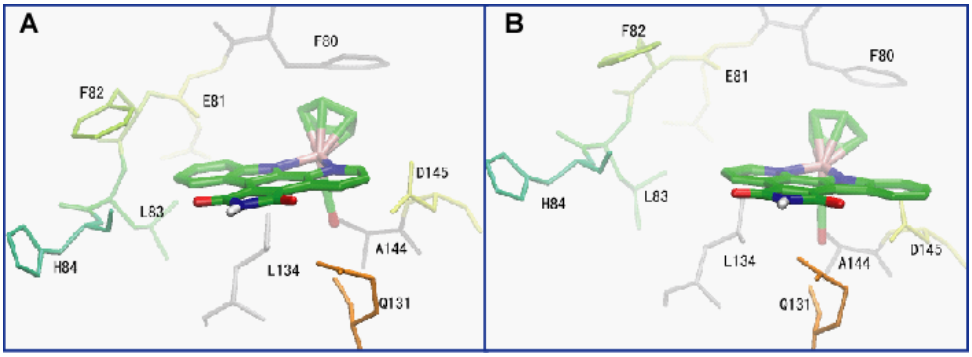

Figure 5.10 The top conformation predicted for both enantiomers bound to CDK2.

While the conformations I and II are not favorable for CDK2, we found the top-ranked conformation for both enantiomers are very similar and this conformation is unique to CDK2, not PIM1 and GSK-3β. In Figure 5.10**,** we depict the lowest-energy conformations predicted for CDK2/cyclin A. The lowest-rank conformations for both enantiomers are found to be quite similar, which strikingly, instead of forming a hydrogen bond with the hinge region residue, show that the NH group of the pyridocarbazole moiety points outside the binding pocket. Moreover, the plane of the Cp ring stacks with the plane of the aromatic PHE80, the gate-keeper residue. The CO group orients into the small binding pocket formed by ALA144, ASP145, ASN132 and GLN131. In this conformation, even though the compounds appear to fit nicely within the CDK2 binding site, there is the distinct lack of the conserved hydrogen bond, which warrants further investigation and experimental validation of the structure of the inhibitor complexed to CDK2.

5.4 Discussion

Based on in vitro protein kinases profiling results, it has been determined that the racemic mixture of ruthenium-based organometalic protein kinase inhibitor scaffold (compound

(*R/S*)-**1**) prefers to inhibit PIM1, GSK-3β over CDK2/cyclin A, even though the active sites of the three kinases show high degree of sequence and structure similarity. Here, we find that the ruthenium-based compounds bind in a similar fashion to PIM1 and GSK-3β but show a novel conformation with marked differences in binding to CDK2/cyclin A. Based on our structural analysis, we suggest the following reasons to explain this selectivity: (1) despite the extensive structure and sequence homology, CDK2 differs in the active site at the location of PHE80, which possibly stabilizes a novel bound confromation due to a stacking interaction with the Cp ring. This conformaiton also lacks the characteristic hydrogen bond between the inhibitor and linker region residue of the kinase and may cause a non-preference to the bound conformation. (2) Compared to PIM1 and GSK-3β, it is less preferrable for CDK2/cyclin A to bind the inhibitor (especially the S-isomer) in a conformation which preserves the conserved hydrogen bond. By aligning the CDK2 structure to PIM1 and GSK-3β, we suggest that residue PHE80 may play a negative role in positioning the carboxyl group in the pyridocarbazole plane (Figure 5.2). We therefore suggest to examine the selectivity profile of compound **2** shown in Figure 5.11 (C), which may help to further clarify the possible reason for the different binding of compound 1 with GSK-3β and CDK2.

Using the existing crystal structure of PIM1/(*R*)-**1** as a control test, we also have shown that by implicitly accounting for the protein flexibility, the ensemble docking protocol is able to help refine the structural features as well as discriminate the docked energies associated with bound conformations of two enantiomeric forms of the inhibitor to the three kinases. The comparison of results from single point docking and ensemble docking demonstrate that the traditional docking method based on a single protein conformation is

sensitive to the protein structure used. Although it generates good docking poses on the exact protein structure, the results become considerably worse on even a slightly different protein structure. The ensemble docking represents a significant improvement in this aspect, yielding consistent results regardless of the initial protein structure used. For the cases presented above, in which the native pose is unknown, the ensemble docking protocol is more robust and reliable in sampling the true complex structure.

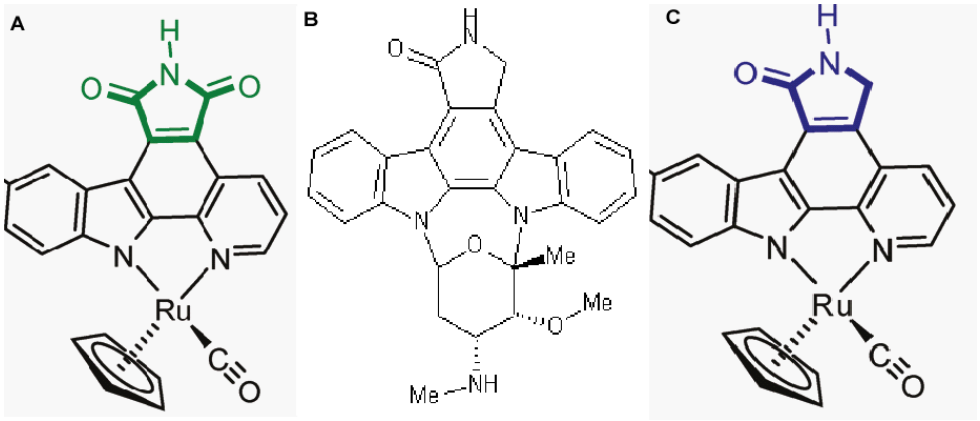

Figure 5.11 Structure of compound 1 (A), staurosporine (B) and the new compound 2 we proposed for further testing (C).

In our ensemble docking analysis, we employ the lowest dock energy conformation to characterize the clusters and show that for PIM1 and GSK-3β, the ensemble docking protocol ranks the native-like conformations as one of the top clusters. Another popular choice for scoring the conformations is based on the frequency of an observed cluster [274]. In our case, we find that many of the top-ranked clusters based on the dock energy would also be identified by the high-frequency criterion. Hence the two criteria tend to yield similar predictions for the top-ranked conformations.

In the ensemble docking protocol we have employed, we do not consider the induced effect of the inhibitor to the protein conformation, i.e., our docking protocol only accounts for protein flexibility implicitly. Our assumption is that by sampling the protein kinase in and around its unbound (apo) state, the dynamics of the protein will account for the small induced effect and the bound conformations will select for those protein conformations resulting in lower dock scores. Thus for a good scoring function (with large correlation between experimental binding affinity and the computed docking score on an extensive test set of compounds), the ensemble docking is expected to yield conformations with the lowest binding affinities. We advocate that our ensemble docking approach is a good first step in cases where we have some knowledge of the protein structure (through experimental or modeling schemes), such as the prediction of the bound complex with a new inhibitor scaffold.

We also note the organometalic inhibitor compounds we have studied are very new and fully flexible force-fields for these inhibitors are not yet available. Therefore, our focus in this chapter is on the structure prediction of the bound complex as opposed to the free energy of binding. Once the structural aspects are validated, our predictions are then logical candidates for a more rigorous evaluation using computationally intensive free energy protocols in fully atomistic systems [102, 150, 153, 204, 239].

Chapter 6 Perspectives and Future Directions

In this thesis, we developed and applied several computational tools at multiple length and time scales to model the interaction inhibitors with protein kinases and the catalytic

reaction mechanism of protein kinases. Our work on EGFR TKD, sheds insight on the unique properties of this important kinase. Moreover, our understanding of the EGFR kinase also helps in delineating common mechanisms across the whole protein kinase family.

In Chapter 2, we predicted how the EGFR TKD recognizes its substrates and performed the catalytic phosphoryl transfer reaction using a computational protocol combining molecular dynamics simulations, ensemble molecular docking [161, 274], free energy calculation[102, 117, 150, 153, 231, 253, 266] and QM/MM. We found that in our predicted bound conformations the P-1 glutamic acid residue and the hydrophobic amino acid at the P+1 position interact with a highly conserved motif in the protein kinase family, namely Val852-Pro853-Ile854-Lys855-Trp856 in EGFR. This motif creates a binding surface to form hydrophobic contacts as well as hydrogen bonds to residues close to P-site (P0) tyrosine in the substrate and therefore helps to orient the tyrosine residue towards the active site. Moreover, our QM/MM simulations show that the EGFR protein kinase favors the dissociative mechanism, although an alternative channel through the formation of an associative transition state is also possible. Collectively, our simulations establish some key molecular rules in operation for the substrate-binding and catalytic landscapes in the EGFR TKD.

A possible future direction is to extend this work to understanding how mutations affect substrate binding. The kinase-activating mutation L834R closely locates to the catalytic site. Based on our simulation results of how EGFR TKD recognize its substrate

sequences, we are able to further investigate whether mutating the hydrophobic residue LEU to a positively charged residue ARG will affect the peptide binding affinity.

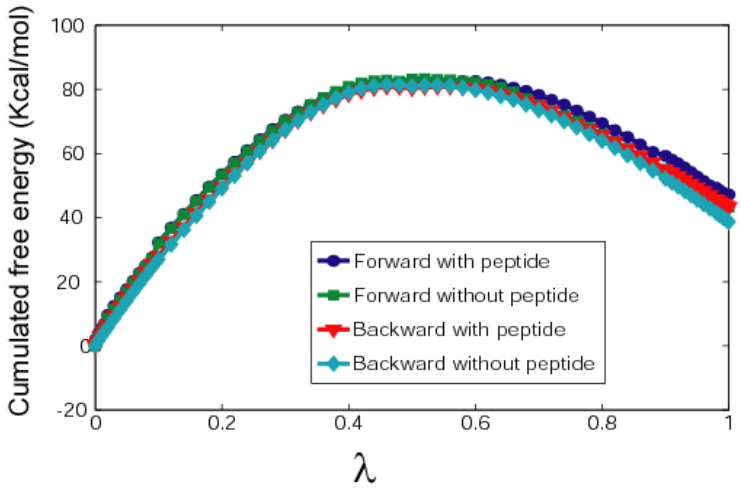

Figure 6.1 Cumulative free energy (kcal/mol) calculated for bound and unbound systems.

In our preliminary studies, we employ the free energy perturbation (FEP) method to calculate the binding free energy difference between the wildtype and mutant systems for peptide Y1068. The FEP results (Figure 6.1 and Table 6.1) showed that L834R mutation resulted in a binding free energy change ($\Delta\Delta F$) of 4.26 $\pm$ 4.2 kcal/mol. Our results imply that the mutation does not increase (but rather decreases) the substrate binding affinity of the particular peptide. Moreover, by comparing how L834R will alter the binding of different substrates (Y1068 and Y1173 for example), we can further gain insight on the effect of this point mutation to differential signaling pathways.

Another promising avenue is to explore the roles of L834R on EGFR kinase catalysis. Experimental data has shown that the value of the catalytic efficiency (k_{cat}/K_M) for the L834R mutant is ~20-fold higher than that for the wild-type kinase domain [289]. Although the mutated residue in the L834R mutant is not directly involved in the

catalytic mechanism, we proposed that the residue is very close to reaction site, three amino acids displaced from D831, which is critical to catalysis. Moreover, the phosphoryl transfer reaction is highly sensitive to the electrostatic environment. A mutation from a neutral residue to a highly positively charged residue close the reaction area may affect the reaction mechanism, which is a topic of significant interest.

Table 6.1 Cumulative free energy (kcal/mol) calculated for bound and unbound systems. The error bar of our current calculation is about 4.2kcal/mol

	ΔG		ΔΔG
	With peptide	No peptide	
Forward	47.188	43.247	3.941
Backward	43.3	38.716	4.584

Moreover, the ErbB family of receptors has four members — the epidermal growth factor receptor (EGFR or ErbB1 or HER1), ErbB2 (or HER2), ErbB3, and ErbB4. They share high sequence homology and a common protein structure. Among all the members, ErbB3/HER3 stands out among the mammalian ErbB receptors since it is generally considered to be kinase-inactive [62] because a conserved aspartate that is thought to function as a catalytic base in all protein kinases [167] – deprotonating the substrate hydroxyl group – is replaced by an asparagine in ErbB3 (N815). HER2, on the other hand, is thought to be an orphan receptor, with none of the EGF family of ligands able to activate it. ErbB2 and ErbB3 can only be activated through heterodimerization [18, 62], and neuregulin-induced ErbB2/ErbB3 heterodimers are considered to be the most potent

mitogenic signaling complexes in the ErbB receptor network [62, 116]. This presents a conundrum given current models for ErbB receptor kinase activation, in which one kinase domain allosterically activates its neighbor and becomes *trans*-autophosphorylated as a consequence [140, 287]. In such a mechanism, it is not clear how ErbB2 could be *trans*-phosphorylated within an ErbB2/ErbB3 heterodimer unless ErbB3 does have kinase activity. Our previous work, collaborating with Dr. Mark Lemmon's group on ErbB3 (see Chapter 2), proposed a mechanism for ErbB3 to catalyze the phosphryl transfer reaction. A further study on the catalysis of ErbB2 and ErbB4 could to gain more insight on functional and mechanistic inter-relationships among the family members in controlling substrate activation and signaling.

Finally, our studies on erlotinib (Chapter 3) showed that the L834R does not change the binding affinity of erlotinib to EGFR TKD in the active state. We further extended our study to understand how the erlotinib interacts with inactive EGFR TKD. Our simulation results proposed the possible binding of erlotinib to both active and inactive state of the kinase. Our collaborator, Dr. Mark Lemmon's group, is working on crystallizing the complex between erlotinib and inactive EGFR TKD. Further investigation of inhibitor-bound conformations of the EGFR kinase may illuminate the concept that erlotinib is effective at inhibiting the ligand-activated EGFR kinase and it may also help to investigate new compound to inhibit EGFR TKD. This is a promising prospect as it allows us to target inactive and alternative conformations of kinases for inhibition. Even though structures of active conformations of several kinases are structurally homologous, there seems to be a greater variation in the structures of inactive and intermediate forms.

Thus, the strategy to target conformations different from the active state opens new possibilities in designing specific kinase inhibitors for cancer therapy.

Reference

1. *AutoDockTools_ i86Linux2_1.5.2*, The Scripps Research Institute.
2. *Matlab7.0*, The MathWorks.
3. *All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution.* J. Comp. Chem., 2000. **21**: p. 105-120.
4. Accelrys, *Cerius2 Modeling Environment, Release 4.8.* 2003.
5. Adams, J.A., *Kinetic and catalytic mechanisms of protein kinases.* Chemical Reviews, 2001. **101**(8): p. 2271-2290.
6. Aikens, C.L., A. Laederach, and P.J. Reilly, *Visualizing complexes of phospholipids with streptomyces phospholipase D by automated docking.* Proteins-Structure Function and Bioinformatics, 2004. **57**(1): p. 27-35.
7. Albeck, J.G., et al., *Collecting and organizing systematic sets of protein data.* Nat Rev Mol Cell Biol, 2006. **7**(11): p. 803-12.
8. Argiris, A. and N. Mittal, *Gefitinib as first-line, compassionate use therapy in patients with advanced non-small-cell lung cancer.* Lung Cancer, 2004. **43**(3): p. 317-22.
9. Arteaga, C.L., *Epidermal growth factor receptor dependence in human tumors: more than just expression?* Oncologist, 2002. **7 Suppl 4**: p. 31-9.
10. Arteaga, C.L. and C.I. Truica, *Challenges in the development of anti-epidermal growth factor receptor therapies in breast cancer.* Seminars in Oncology, 2004. **31**(1 Suppl 3): p. 3-8.
11. Atilla-Gokcumen, G.E., et al., *Organometallic compounds with biological activity: A very selective and highly potent cellular inhibitor for glycogen synthase kinase 3.* Chembiochem, 2006. **7**(9): p. 1443-1450.
12. Balius, T.E. and R.C. Rizzo, *Quantitative Prediction of Fold Resistance for Inhibitors of EGFR.* Biochemistry, 2009. **48**(35): p. 8435-8448.
13. Banerjee, A., et al., *Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA.* Nature, 2005. **434**(7033): p. 612-618.

14. Batcho, P., D.A. Case, and T. Schlick, *Optimized Particle-Mesh Ewald / Multiple-Timestep Integration for Molecular Dynamics Simulations.* J. Chem. Phys., 2001. **115**: p. 4003-4018.

15. Beard, W.A., et al., *Enzyme-DNA interactions required for efficient nucleotide incorporation and discrimination in human DNA polymerase beta.* J. Biol. Chem., 1996. **271**: p. 12141-12144.

16. Benkovic, S.J. and S. Hammes-Schiffer, *A perspective on enzyme catalysis.* Science, 2003. **301**(5637): p. 1196-202.

17. Benkovic, S.J. and S. Hammes-Schiffer, *Biochemistry. Enzyme motions inside and out.* Science, 2006. **312**(5771): p. 208-9.

18. Berger, M.B., J.M. Mendrola, and M.A. Lemmon, *ErbB3/HER3 does not homodimerize upon neuregulin binding at the cell surface.* FEBS Lett., 2004. **569**: p. 332-336.

19. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res., 2000. **28**: p. 235-242.

20. Bhalla, U.S. and R. Iyengar, *Emergent properties of biological networks.* Science, 1999. **283**: p. 381-387.

21. Bhalla, U.S., P.T. Ram, and R. Iyengar, *MAPKinase phosphotase as a locus of flexibility in a mitogen activated protein kinase signaling network.* Science, 2002. **297**: p. 1018-1023.

22. Bindewald, E. and J. Skolnick, *A scoring function for docking ligands to low-resolution protein structures.* J. Comput. Chem., 2005. **26**: p. 374-383.

23. Biscardi, J.S., et al., *c-Src-mediated Phosphorylation of the Epidermal Growth Factor Receptor on Tyr845 and Tyr1101 Is Associated with Modulation of Receptor Function.* J. Biol. Chem., 1999. **274**(12): p. 8335-8343.

24. Blume-Jensen, P. and T. Hunter, *Oncogenic kinase signalling.* Nature, 2001. **411**(6835): p. 355-365.

25. Boerner, J.L., et al., *Phosphorylation of Y845 on the Epidermal Growth Factor Receptor Mediates Binding to the Mitochondrial Protein Cytochromec Oxidase Subunit II.* Mol. Cell Biol., 2004. **24**: p. 7059-7071.

26. Bolhuis, P.G., et al., *Transition path sampling: Throwing ropes over rough mountain passes, in the dark.* Annual Review of Physical Chemistry, 2002. **53**: p. 291-318.

27. Bookout, A.L., et al., *Targeting G$\beta \gamma$ signaling to inhibit prostate tumor formation and growth.* J. Biol. Chem., 2003. **278**: p. 37569-37573.

28. Bose, R., et al., *Protein tyrosine kinase-substrate interactions.* Current Opinion in Structural Biology, 2006. **16**(6): p. 668-675.

29. Bossemeyer, D., et al., *Phosphotransferase and substrate binding mechanism of the cAMP - dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 A structure of the complex with Mn2+ adenylyl imidodiphosphate and inhibitor peptide PKI(5-24).* Embo J., 1993. **12**: p. 849-859.

30. Bossemeyer, D., *Protein-Kinases - Structure and Function.* Febs Letters, 1995. **369**(1): p. 57-61.

31. Bregman, H., P.J. Carroll, and E. Meggers, *Rapid access to unexplored chemical space by ligand scanning around a ruthenium center: discovery of potent and selective protein kinase inhibitors.* J Am Chem Soc, 2006. **128**(3): p. 877-84.

32. Bregman, H. and E. Meggers, *Ruthenium half-sandwich complexes as protein kinase inhibitors: An N-succinimidyl ester for rapid derivatizations of the cyclopentadienyl moiety.* Organic Letters, 2006. **8**(24): p. 5465-5468.

33. Breneman, C.M. and K.B. Wiberg, *DETERMINING ATOM-CENTERED MONOPOLES FROM MOLECULAR ELECTROSTATIC POTENTIALS - THE NEED FOR HIGH SAMPLING DENSITY IN FORMAMIDE CONFORMATIONAL-ANALYSIS.* Journal of Computational Chemistry, 1990. **11**(3): p. 361-373.

34. Brignola, P.S., et al., *Comparison of the biochemical and kinetic properties of the type 1 receptor tyrosine kinase intracellular domains - Demonstration of differential sensitivity to kinase inhibitors.* Journal of Biological Chemistry, 2002. **277**(2): p. 1576-1585.

35. Brinkworth, R.I., R.A. Breinl, and B. Kobe, *Structural basis and prediction of substrate specificity in protein serine/threonine kinases.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(1): p. 74-79.

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

37. Brooks, B.R., et al., *Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations.* Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.

38. Brooks, C.L. and D.A. Case, *Theory and simulation - the control and timescale of structure and reactivity in biological systems: from peptide folding to cellular networks.* Curr. Opin. Struct. Biol., 2003. **13**: p. 143-145.

39. Brown, K.S., et al., *The statistical mechanics of complex signaling networks: nerve growth factor signaling.* Phys. Biol., 2004. **1**: p. 184-195.

40. Brown, N.R., et al., *The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases.* Nature Cell Biology, 1999. **1**(7): p. 438-443.

41. Bruice, T.C. and S.J. Benkovic, *Chemical basis for enzyme catalysis.* Biochemistry, 2000. **39**(21): p. 6267.

42. Buchdunger, E., T. O'Reilly, and J. Wood, *Pharmacology of imatinib (STI571). [Review] [58 refs]*. European Journal of Cancer, 2002. **38 Suppl 5**: p. S28-36.

43. Bullock, A.N., et al., *Structure and substrate specificity of the Pim-1 kinase.* Journal of Biological Chemistry, 2005. **280**(50): p. 41675-41682.

44. Bursulaya, B.D., et al., *Comparative study of several algorithms for flexible ligand docking.* Journal of Computer-Aided Molecular Design, 2003. **17**(11): p. 755-763.

45. Cabodi, S., et al., *Integrin regulation of epidermal growth factor (EGF) receptor and of EGF-dependent responses.* Biochem. Soc. Trans., 2004. **32**: p. 438-442.

46. Camacho, J.C., et al., *Scoring dockied conformations generated by rigid body protein protein docking*. Proteins, 2000. **40**: p. 525-537.

47. Canteley, L.C., *The PI3K pathway.* Science, 2002. **296**: p. 1655-1657.

48. Carey, K.D., et al., *Kinetic analysis of epidermal growth factor receptor somatic mutant proteins shows increased sensitivity to the epidermal growth factor receptor tyrosine kinase inhibitor, erlotinib.* Cancer Research, 2006. **66**(16): p. 8163-8171.

49. Chatterjee, A. and et al., *Binomial distribution based tau-leap accelerated stochastic simulation.* J.Chem.Phys., 2005: p. 024112.

50. Chaudhury, S. and J.J. Gray, *Conformer selection and induced fit in flexible backbone protein-protein docking using computational and NMR ensembles*. Journal of Molecular Biology, 2008. **381**(4): p. 1068-1087.

51. Chen, Y.R., et al., *Distinctive activation patterns in constitutively active and gefitinib-sensitive EGFR mutants.* Oncogene, 2006. **25**(8): p. 1205-15.

52. Cheng, Y.H., Y.K. Zhang, and J.A. McCammon, *How does the cAMP-dependent protein kinase catalyze the phosphorylation reaction: An ab initio QM/MM study.* Journal of the American Chemical Society, 2005. **127**(5): p. 1553-1562.

53. Cheng, Y.H., Y.K. Zhang, and J.A. McCammon, *How does activation loop phosphorylation modulate catalytic activity in the cAMP-dependent protein kinase: A theoretical study.* Protein Science, 2006. **15**(4): p. 672-683.

54. Choi, S.H., J.M. Mendrola, and M.A. Lemmon, *EGF-independent activation of cell-surface EGF receptors harboring mutations found in gefitinib-sensitive lung cancer.* Oncogene, 2006.

55. Chou, T.Y., et al., *Mutation in the Tyrosine Kinase Domain of Epidermal Growth Factor Receptor Is a Predictive and Prognostic Factor for Gefitinib Treatment in Patients with Non-Small Cell Lung Cancer.* Clinical Cancer Research, 2005. **11**(10): p. 3750-3757.

56. Chu, J.-W., B.L. Trout, and B.R. Brooks, *A super-linear minimization scheme for the nudged elastic band method.* The Journal of Chemical Physics, 2003. **119**(24): p. 12708-12717.

57. Ciardiello, F., *Epidermal growth factor receptor tyrosine kinase inhibitors as anticancer agents. [Review] [44 refs]*. Drugs, 2000. **60**: p. 25-32.

58. Ciardiello, F., et al., *Inhibition of growth factor production and angiogenesis in human cancer cells by ZD1839 (Iressa), a selective epidermal growth factor receptor tyrosine kinase inhibitor.* Clinical Cancer Research, 2001. **7**(5): p. 1459-65.

59. Ciardiello, F. and G. Tortora, *A novel approach in the treatment of cancer: targeting the epidermal growth factor receptor. [Review] [101 refs]*. Clinical Cancer Research, 2001. **7**(10): p. 2958-70.

60. Ciardiello, F., et al., *ZD1839 (IRESSA), an EGFR-selective tyrosine kinase inhibitor, enhances taxane activity in bcl-2 overexpressing, multidrug-resistant MCF-7 ADR human breast cancer cells.* International Journal of Cancer, 2002. **98**(3): p. 463-9.

61. Ciardiello, F., et al., *The role of EGFR inhibitors in nonsmall cell lung cancer. [Review] [54 refs]*. Current Opinion in Oncology, 2004. **16**(2): p. 130-5.

62. Citri, A., K.B. Skaria, and Y. Yarden, *The deaf and the dumb: the biology of ErbB-2 and ErbB-3.* Exp. Cell. Res., 2003. **284**(1): p. 54-65.

63. Citri, A. and Y. Yarden, *EGF-ERBB signalling: towards the systems level.* Nature Reviews Molecular Cell Biology, 2006. **7**(7): p. 505-516.

64. Cleland, W.W. and A.C. Hengge, *Mechanisms of phosphoryl and acyl transfer.* Faseb Journal, 1995. **9**(15): p. 1585-1594.

65. Coker, K.J., J.V. Staros, and C.A. Guyer, *A kinase-negative epidermal growth factor receptor that retains the capacity to stimulate DNA synthesis.* Proc. Natl. Acad. Sci. U. S. A., 1994. **91**: p. 6967-6971.

66. Couzin, J., *Cancer drugs - Smart weapons prove tough to design.* Science, 2002. **298**(5593): p. 522-+.

67. Cui, Q. and M. Karplus, *Triosephosphate isomerase: A theoretical comparison of alternative pathways.* Journal of the American Chemical Society, 2001. **123**(10): p. 2284-2290.

68. Daggett, V., *Long Timescale Simulations.* Curr. Opin. Struct. Biol., 2000. **10**: p. 160-164.

69. Dancey, J., *Epidermal growth factor receptor inhibitors in clinical development. [Review] [40 refs]*. International Journal of Radiation Oncology, Biology, Physics, 2004. **58**(3): p. 1003-7.

70. Dancey, J.E., *Predictive factors for epidermal growth factor receptor inhibitors-the bull's-eye hits the arrow. [Review] [31 refs]*. Cancer Cell, 2004. **5**(5): p. 411-5.

71. Das, D., et al., *Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method.* Journal of Chemical Physics, 2002. **117**(23): p. 10534-10547.

72. Daura, X., et al., *Reversible Peptide Folding in Solution by Molecular Dynamics Simulation.* J. Mol. Biol., 1998. **280**: p. 925-932.

73. De Vivo, M., et al., *Computational study of the phosphoryl transfer catalyzed by a cyclin-dependent kinase.* Chemistry-a European Journal, 2007. **13**(30): p. 8437-8444.

74. Diaz, N. and M.J. Field, *Insights into the phosphoryl-transfer mechanism of cAMP-dependent protein kinase from quantum chemical calculations and molecular dynamics simulations.* Journal of the American Chemical Society, 2004. **126**(2): p. 529-542.

75. Dikic, I., *Mechanisms controlling EGF receptor endocytosis and degradation.* Biochem. Soc. Trans., 2003. **31**: p. 1178-1181.

76. Duan, Y. and P.A. Kollman, *Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution.* Science, 1998. **282**: p. 740-744.

77. Eck, M.J. and C.H. Yun, *Structural and mechanistic underpinnings of the differential drug sensitivity of EGFR mutations in non-small cell lung cancer.* Biochimica Et Biophysica Acta-Proteins and Proteomics. **1804**(3): p. 559-566.

78. Elber, R., J. Meller, and R. Olender, *Stochastic path approach to compute atomically detailed trajectories: Application to the folding of C peptide.* J. Phys. Chem. B, 1999. **103**: p. 899-911.

79. Elber, R., et al., *Bridging the gap between long time trajectories and reaction pathways.* Adv. Chem. Phys, 2003. **126**: p. 93-129.

80. Engelman, J.A., et al., *MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling.* Science, 2007. **316**(5827): p. 1039-1043.

81. Essmann, U., *A smooth particle mesh Ewald method.* J. Chem. Phys., 1995. **103**: p. 8577-8593.

82. Fabbro, D., et al., *Protein kinases as targets for anticancer agents: from inhibitors to useful drugs.* Pharmacology & Therapeutics, 2002. **93**(2-3): p. 79-98.

83. Fabian, M.A., et al., *A small molecule-kinase interaction map for clinical kinase inhibitors.* Nature Biotechnology, 2005. **23**(3): p. 329-336.

84. Fan, Y.-X., et al., *Ligand Regulates Epidermal Growth Factor Receptor Kinase Specificity.* J. Biol. Chem., 2004. **279**: p. 38143-38150.

85. Fan, Y.X., et al., *Ligand regulates epidermal growth factor receptor kinase specificity - Activation increases preference for GAB1 and SHC versus autophosphorylation sites.* Journal of Biological Chemistry, 2004. **279**(37): p. 38143-38150.

86. Fan, Y.X., L.L. Wong, and G.R. Johnson, *EGFR kinase possesses a broad specificity for ErbB phosphorylation sites, and ligand increases catalytic-centre activity without affecting substrate binding affinity.* Biochemical Journal, 2005. **392**: p. 417-423.

87. Ferrara, P., J. Apostolakis, and A. Caflisch, *Computer simulations of protein folding by targeted molecular dynamics.* Proteins, 2000. **39**: p. 252-260.

88. Ferrara, P., J. Apostolakis, and A. Caflisch, *Targeted Molecular Dynamics Simulations of Protein Unfolding*. J. Phys. Chem. B., 2000. **104**: p. 4511-4518.

89. Field, M.J., P.A. Bash, and M. Karplus, *A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations.* J. Comput. Chem., 2002. **11**: p. 700-733.

90. Fischer, O.M., et al., *EGFR signal trans activation in cancer cells.* Biochem. Soc. Trans., 2003. **31**: p. 1203-1208.

91. Floppe, N. and J.A.D. MacKerell, *Intrinsic conformational properties of deoxyribonucleosides: implicated role of cytosine in the equilibrium among the A, B and Z forms of DNA.* Biophys. J., 1999. **76**: p. 3206-3218.

92. Florian, J., M.F. Goodman, and A. Warshel, *Computer simulation of the chemical catalysis of DNA polymerases: Discriminating between alternative nucleotide insertion mechanisms for T7 DNA polymerase.* Journal of the American Chemical Society, 2003. **125**(27): p. 8163-8177.

93. Foloppe, N. and A.D. MacKerell, *All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data.* Journal of Computational Chemistry, 2000. **21**(2): p. 86-104.

94. Frauenfelder, H., S.G. Sligar, and P.G. Wolynes, *The Energy Landscapes and Motions of Proteins.* Science, 1991. **254**(5038): p. 1598-1603.

95. Friedrichs, M.S., et al., *Accelerating molecular dynamic simulation on graphics processing units.* J Comput Chem, 2009. **30**(6): p. 864-72.

96. Friesner, R.A., et al., *How iron-containing proteins control dioxygen chemistry: a detailed atomic level description via accurate quantum chemical and mixed quantum mechanics/molecular mechanics calculations.* Coordination Chemistry Reviews, 2003. **238-239**: p. 267-290.

97. Friesner, R.A. and V. Guallar, *Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis.* Annual Review of Physical Chemistry, 2005. **56**: p. 389-427.

98. Friesner, R.A., et al., *Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes.* Journal of Medicinal Chemistry, 2006. **49**(21): p. 6177-6196.

99. Garcia-Viloca, M. and J. Gao, *Generalized Hybrid Orbital for the treatment of boundary atoms in combined quantum mechanical and molecular mechanical calculations using the semiempirical parameterized model 3 method.* Theoretical Chemistry Accounts, 2004. **111**: p. 280-286.

100. Gibbs, C.S. and M.J. Zoller, *Rational Scanning Mutagenesis of a Protein-Kinase Identifies Functional Regions Involved in Catalysis and Substrate Interactions.* Journal of Biological Chemistry, 1991. **266**(14): p. 8923-8931.

101. Gilmer, T.M., et al., *Impact of common epidermal growth factor receptor and HER2 variants on receptor activity and inhibition by lapatinib.* Cancer Research, 2008. **68**(2): p. 571-579.

102. Gilson, M.K. and H.X. Zhou, *Calculation of protein-ligand binding affinities.* Annual Review of Biophysics and Biomolecular Structure, 2007. **36**: p. 21-42.

103. Glenn, J.M., J.T. Douglas, and L.K. Michael, *Constant pressure molecular dynamics algorithms.* The Journal of Chemical Physics, 1994. **101**(5): p. 4177-4189.

104. Goldsmith, E.J., et al., *Substrate and docking interactions in serine/threonine protein kinases.* Chemical Reviews, 2007. **107**(11): p. 5065-5081.

105. Gota, K., et al., *Application of MDGRAPE-3, a special purpose board for molecular dynamics simulations, to periodic biomolecular systems.* Journal of Computational Chemistry, 2009. **30**(1): p. 110-118.

106. Grayson, P., E. Tajkhorshid, and K. Schulten, *Mechanisms of selectivity in channels and enzymes studied with interactive molecular dynamics.* Biophys. J., 2003. **119**: p. 36-48.

107. Grubmüller, H., et al., *A small molecular-kianse interaction map for clinical kinase inhibitors.* Nature Biotechnology, 2005. **3**: p. 329--336.

108. Guo, H., et al., *Substrate conformational transitions in the active site of chorismate mutase: Their role in the catalytic mechanism.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(16): p. 9032-9037.

109. Halperin, I., et al., *Principles of docking: An overview of search algorithms and a guide to scoring functions.* Proteins-Structure Function and Genetics, 2002. **47**(4): p. 409-443.

110. Hanks, S.K., A.M. Quinn, and T. Hunter, *The Protein-Kinase Family - Conserved Features and Deduced Phylogeny of the Catalytic Domains.* Science, 1988. **241**(4861): p. 42-52.

111. Hatakeyama, M., et al., *A computational model on the modulation of MAPK and Akt pathways in Heregulin-induced ERB signaling.* Biochem. J., 2003. **373**: p. 451-463.

112. Haugh, J.M. and D.A. Lauffenburger, *Analysis of receptor internalization as a mechanism for modulating signal transduction.* J Theor Biol, 1998. **195**(2): p. 187-218.

113. Haugh, J.M., et al., *Effect of epidermal growth factor receptor internalization on regulation of the phospholipase C-gamma1 signaling pathway.* J Biol Chem, 1999. **274**(13): p. 8958-65.

114. Haugh, J.M., A. Wells, and D.A. Lauffenburger, *Mathematical modeling of epidermal growth factor receptor signaling through the phospholipase C pathway: mechanistic insights and predictions for molecular interventions.* Biotechnol Bioeng, 2000. **70**(2): p. 225-38.

115. Haugh, J.M., *Localization of receptor-mediated signal transduction pathways: the inside story.* Mol Interv, 2002. **2**(5): p. 292-307.

116. Holbrook, M.R., et al., *Epidermal growth factor receptor internalization rate is regulated by negative charges near the SH2 binding site Tyr992.* Biochemistry, 1999. **38**(29): p. 9348-9356.

117. Honig, B., K. Sharp, and A.S. Yang, *Macroscopic Models of Aqueous-Solutions - Biological and Chemical Applications.* Journal of Physical Chemistry, 1993. **97**(6): p. 1101-1109.

118. House, C., G.S. Baldwin, and B.E. Kemp, *Synthetic Peptide-Substrates for the Membrane Tyrosine Protein-Kinase Stimulated by Epidermal Growth-Factor.* European Journal of Biochemistry, 1984. **140**(2): p. 363-367.

119. Hubbard, S.R., *Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog.* Embo J, 1997. **16**(18): p. 5572-81.

120. Hubbard, S.R. and J.H. Till, *Tyrosine kinase structure and function.* Annu. Rev. Bio. Chem., 2000. **69**: p. 373-398.

121. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics, 1996. **14**(1): p. 33-&.

122. Humphrey, W., A. Dalke, and K. Schulten, *VMD - Visual Molecular Dynamics.* Journal of Molecular Graphics, 1996. **14**: p. 33--38.

123. Hunter, T. and J.A. Cooper, *Protein-Tyrosine Kinases.* Annual Review of Biochemistry, 1985. **54**: p. 897-930.

124. Hunter, T., *Signaling - 2000 and beyond.* Cell, 2000. **100**(1): p. 113-127.

125. Hynes, N.E. and H.A. Lane, *ERBB receptors and cancer: The complexity of targeted inhibitors (vol 5, pg 341, 2005).* Nature Reviews Cancer, 2005. **5**(7): p. 1.

126. Im, W., D. Beglov, and B. Roux, *Continuum Solvation Model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation.* Computer Physics Communications, 1998. **111**(1-3): p. 59-75.

127. Isralewitz, B., et al., *Steered Molecular Dynamics Investigations of Protein Function.* J. Mol. Graph. Model., 2001. **19**: p. 13-25.

128. Izrailev, S., et al., *Steered Molecular Dynamics Simulation of the Rieske Subunit Motion in the Cytochrome B Complex.* Biophys. J., 1999. **77**: p. 1753-1768.

129. Jacobson, M.P., et al., *On the role of the crystal environment in determining protein side-chain conformations.* Journal of Molecular Biology, 2002. **320**(3): p. 597-608.

130. Jacobson, M.P., et al., *Force field validation using protein side chain prediction.* Journal of Physical Chemistry B, 2002. **106**(44): p. 11673-11680.

131. Janes, K.A., et al., *A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis.* Science, 2005. **310**(5754): p. 1646-53.

132. Janes, K.A., et al., *The response of human epithelial cells to TNF involves an inducible autocrine cascade.* Cell, 2006. **124**(6): p. 1225-39.

133. Janes, K.A. and D.A. Lauffenburger, *A biological approach to computational models of proteomic networks.* Curr Opin Chem Biol, 2006. **10**(1): p. 73-80.

134. Janmaat, M.L., et al., *Response to epidermal growth factor receptor inhibitors in non-small cell lung cancer cells: limited antiproliferative effects and absence of apoptosis associated with persistent activity of extracellular signal-regulated kinase or Akt kinase pathways.* Clinical Cancer Research, 2003. **9**(6): p. 2316-26.

135. Jarvis, L.M., *Battling breast cancer.* Chemical & Engineering News, 2006. **84**(32): p. 21-27.

136. Jensen, F., *Introduction to Computational Chemistry*. 2007, Chichester: John wiley & sons.

137. Jorgensen, W.L. and J. Tiradorives, *The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin.* Journal of the American Chemical Society, 1988. **110**(6): p. 1657-1666.

138. Jorissen, R.N., et al., *Epidermal growth factor receptor: mechanisms of activation and signalling.* Experimental Cell Research, 2003. **284**(1): p. 31-53.

139. Jorrinssen, R.N., *Epidermal growth factor receptor: mechanism of activation and signaling.* Exp. Cell. Res., 2003. **284**: p. 31-53.

140. Jura, N., et al., *Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment.* Cell, 2009. **137**: p. 1293-1307.

141. Kale, L., et al., *NAMD2: Greater scalability for parallel molecular dynamics.* Journal of Computational Physics, 1999. **151**(1): p. 283-312.

142. Kennelly, P.J. and E.G. Krebs, *Consensus Sequences as Substrate-Specificity Determinants for Protein-Kinases and Protein Phosphatases.* Journal of Biological Chemistry, 1991. **266**(24): p. 15555-15558.

143. Kholodenko, B.N., et al., *Quantification of short-term signaling by EGFR.* J. Biol. Chem., 2004. **274**: p. 30169-30181.

144. Kim, J., et al., *Regulation of EGFR internalization by GPCR.* Biochemistry, 2003. **42**: p. 2887-2894.

145. Kim, K. and P.A. Cole, *Kinetic analysis of a protein tyrosine kinase reaction transition state in the forward and reverse directions.* J. Amer. Chem. Soc., 1998. **120**: p. 6851-6858.

146. Kirkpatrick, S., C. Gelatt, and M. Vecchi, *Optimization by simulated annealing.* Science, 1983. **220**: p. 671-680.

147. Kita, Y.A., et al., *NDF/heregulin stimulates the phosphorylation of Her3/erbB3.* FEBS Lett, 1994. **349**(1): p. 139-43.

148. Kloth, M.T., et al., *STAT5b, a Mediator of Synergism between c-Src and the Epidermal Growth Factor Receptor*. J. Biol. Chem., 2003. **278**: p. 1671-1679.

149. Knighton, D.R., et al., *Structural Features That Specify Tyrosine Kinase-Activity Deduced from Homology Modeling of the Epidermal Growth-Factor Receptor*. Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(11): p. 5001-5005.

150. Kollman, P.A., et al., *Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models.* Accounts of Chemical Research, 2000. **33**(12): p. 889-897.

151. Kontopidis, G., et al., *Differential binding of inhibitors to active and inactive CDK2 provides insights for drug design.* Chemistry & Biology, 2006. **13**(2): p. 201-211.

152. Kozakov, D., et al., *Optimal clustering for detecting near-native conformations in protein docking*. Biophysical Journal, 2005. **89**(2): p. 867-875.

153. Kuhn, B., et al., *Validation and use of the MM-PBSA approach for drug discovery*. Journal of Medicinal Chemistry, 2005. **48**(12): p. 4040-4048.

154. Lahiri, S.D., et al., *The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction.* Science, 2003. **299**(5615): p. 2067-2071.

155. Lee, J.W., et al., *ERBB2 kinase domain mutation in a gastric cancer metastasis.* APMIS, 2005. **113**(10): p. 683-7.

156. Lee, J.W., et al., *Somatic mutations of ERBB2 kinase domain in gastric, colorectal, and breast carcinomas.* Clinical Cancer Research, 2006. **12**(1): p. 57-61.

157. Lee, K., et al., *An efficient molecular docking using conformational space annealing.* J. Comput. Chem., 2005. **26**: p. 78-87.

158. Lemmon, M.A. and J. Schlessinger, *Cell Signaling by Receptor Tyrosine Kinases.* Cell. **141**(7): p. 1117-1134.

159. Levinson, N.M., et al., *A Src-like inactive conformation in the Abl tyrosine kinase domain.* Plos Biology, 2006. **4**(5): p. 753-767.

160. Liao, J.J.L., *Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors.* Journal of Medicinal Chemistry, 2007. **50**(3): p. 409-424.

161. Lin, J.H., et al., *The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme.* Biopolymers, 2003. **68**(1): p. 47-62.

162. Liu, Y., et al., *A Multiscale Computational Approach to Dissect Early Events in the Erb Family Receptor Mediated Activation, Differential Signaling, and Relevance to Oncogenic Transformations.* Annals of Biomedical Engineering, 2007. **35**(6): p. 1012-1025.

163. Luttrell, L.M., Y. Daaka, and R.J. Lefkowitz, *Regulation of tyrosine kinase cascades by GPCR.* Curr. Opin. Cell Biol., 1999. **11**: p. 177-183.

164. Lynch, T.J., et al., *Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib.* New England Journal of Medicine, 2004. **350**(21): p. 2129-2139.

165. Lynch, T.J., et al., *Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib.[see comment]*. New England Journal of Medicine, 2004. **350**(21): p. 2129-39.

166. MacKerell, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.

167. Madhusudan, et al., *Camp-Dependent Protein-Kinase - Crystallographic Insights into Substrate Recognition and Phosphotransfer.* Protein Science, 1994. **3**(2): p. 176-187.

168. Manning, G., et al., *The protein kinase complement of the human genome.* Science, 2002. **298**(5600): p. 1912-+.

169. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of Folded Proteins.* Nature, 1977. **267**: p. 585-590.

170. McCammon, J.A. and S.C. Harvey, *Dynamics of Proteins and Nucleic Acids*. 1987, Cambridge, MA: Cambridge University Press.

171. McCarrick, M.A. and P.A. Kollman, *Predicting relative binding affinities of non-peptide HIV protease inhibitors with free energy perturbation calculations.* Journal of Computer-Aided Molecular Design, 1999. **13**(2): p. 109-121.

172. Meggers, E., *Exploring biologically relevant chemical space with metal complexes.* Current Opinion in Chemical Biology, 2007. **11**(3): p. 287-292.

173. Mendelsohn, J. and J. Baselga, *The EGF receptor family as targets for cancer therapy.* Oncogene, 2000. **19**(56): p. 6550-6565.

174. Mendelsohn, J. and J. Baselga, *Status of EGFR antagonists in biology and treatment of cancer.* J. Clini. Onco., 2003. **21**: p. 2782-2799.

175. Mildvan, A.S. and D.C. Fry, *Nmr-Studies of the Mechanism of Enzyme Action.* Advances in Enzymology and Related Areas of Molecular Biology, 1987. **59**: p. 241-313.

176. Mildvan, A.S., *Mechanisms of signaling and related enzymes.* Proteins: Stru. Fun. Gen., 1997. **29**: p. 401-416.

177. Mok, J., et al., *Deciphering Protein Kinase Specificity Through Large-Scale Analysis of Yeast Phosphorylation Site Motifs.* Science Signaling. **3**(109).

178. Morris, G.M., et al., *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.* Journal of Computational Chemistry, 1998. **19**(14): p. 1639-1662.

179. Moulder, S.L., et al., *Epidermal growth factor receptor (HER1) tyrosine kinase inhibitor ZD1839 (Iressa) inhibits HER2/neu (erbB2)-overexpressing breast cancer cells in vitro and in vivo.* Cancer Research, 2001. **61**(24): p. 8887-95.

180. Mulholland, A.J., *Modelling enzyme reaction mechanisms, specificity and catalysis.* Drug Discovery Today, 2005. **10**(20): p. 1393-1402.

181. Murray, C.W., C.A. Baxter, and A.D. Frenkel, *The sensitivity of the results of molecular docking to induced fit effects: Application to thrombin, thermolysin and neuraminidase.* Journal of Computer-Aided Molecular Design, 1999. **13**(6): p. 547-562.

182. Nelson, M.T., et al., *NAMD: A parallel, object oriented molecular dynamics program.* International Journal of Supercomputer Applications and High Performance Computing, 1996. **10**(4): p. 251-268.

183. Noble, M.E.M., J.A. Endicott, and L.N. Johnson, *Protein kinase inhibitors: Insights into drug design from structure.* Science, 2004. **303**(5665): p. 1800-1805.

184. Oda, K., et al., *A comprehensive pathway map of epidermal growth factor receptor signaling.* Mol Syst Biol, 2005. **1**: p. E1-E17.

185. P.S Brignola and e. al., *Comparison of biochemical and kinetic properties of type 1 receptor tyrosine kinase intracellular domains.* J. Biol. Chem., 2001. **277**: p. 1576-1581.

186. Paez, J.G., et al., *EGFR mutations in lung, cancer: Correlation with clinical response to gefitinib therapy.* Science, 2004. **304**(5676): p. 1497-1500.

187. Pao, W. and M. V.A., *Epidermal Growth Factor Receptor Mutations, Small-Molecule Kinase Inhibitors, and Non Small-Cell Lung Cancer: Current Knowledge and Future Directions.* Journal of Clinical Oncology, 2005. **23**: p. 2556-2568.

188. Parr, R.G. and W. Yang, *Density-functional theory of atoms and molecules*. 1989, Oxford: Oxford University Press. ix, 333 p.

189. Pavelites, J.J., et al., *A molecular mechanics force field for NAD(+), NADH, and the pyrophosphate groups of nucleotides.* Journal of Computational Chemistry, 1997. **18**(2): p. 221-239.

190. Pearlman, D.A., *Free energy grids: A practical qualitative application of free energy perturbation to ligand design using the OWFEG method.* J. Med. Chem., 1999. **42**: p. 4313-4324.

191. Pearlman, D.A. and P.S. Charifson, *Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system.* J. Med. Chem., 2001. **44**: p. 3417-3423.

192. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* Journal of Computational Chemistry, 2005. **26**: p. 1781-1802.

193. Pichierri, F., *A quantum mechanical study on phosphotyrosyl peptide binding to the SH2 domain of p56(lck) tyrosine kinase with insights into the biochemistry of intracellular signal transduction events.* Biophysical Chemistry, 2004. **109**(2): p. 295-304.

194. Pierce, K.L., et al., *Role of endocytosis in the activation of extracellular signal-regulated kinase cascade by sequestering and nonsequestering GPCR.* Proc. Natl. Acad. Sci. USA, 2000. **97**: p. 1489-1494.

195. Prenzel, N., et al., *Tyrosine kinase signalling in breast cancer Epidermal growth factor receptor: convergence point for signal integration and diversification.* Breast Cancer Res., 2000. **2**: p. 184-190.

196. Pu, J., D.G. Truhlar, and J. Gao, *The Generalized Hybrid Orbital (GHO) method for ab initio combined QM/MM calculations.* Journal of Physical Chemistry A, 2004. **108**: p. 632-650.

197. Pulverer, B., *Nature insight: Cancer.* nature, 2003. **411**: p. 335-395.

198. Qiu, C., et al., *Mechanism of activation and inhibition of the HER4/ErbB4 kinase.* Structure, 2008. **16**(3): p. 460-467.

199. Qiu, C., et al., *In Vitro Enzymatic Characterization of Near Full Length EGFR in Activated and Inhibited States.* Biochemistry, 2009. **48**(28): p. 6624-6632.

200. Radhakrishnan, R. and T. Schlick, *Orchestration of cooperative events in DNA synthesis and repair mechanism unraveled by transition path sampling of DNA polymerase beta's closing.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(16): p. 5970-5975.

201. Radhakrishnan, R. and T. Schlick, *Fidelity discrimination in DNA polymerase beta: differing closing profiles for a mismatched (G:A) versus matched (G:C) base pair.* . J. Am. Chem. Soc., 2005. **127**: p. 13245-13253.

202. Radhakrishnan, R. and T. Schlick, *Correct and incorrect nucleotide incorporation pathways in DNA polymerase β.* Biochemical and Biophysical Research Communications, 2006. **350**: p. 521-529.

203. Radhakrishnan, R., *Coupling of Fast and Slow Modes in the Reaction Pathway of the Minimal Hammerhead Ribozyme Cleavage.* Biophys. J., 2007. **93**: p. 2391-2399.

204. Radmer, R.J. and P.A. Kollman, *Free energy calculation methods: A theoretical and empirical comparison of numerical errors and a new method for qualitative estimates of free energy changes.* Journal of Computational Chemistry, 1997. **18**(7): p. 902-919.

205. Rao, S.N., et al., *Free-Energy Perturbation Calculations on Binding and Catalysis after Mutating Asn-155 in Subtilisin.* Nature, 1987. **328**(6130): p. 551-554.

206. Redondo, A. and R. LeSar, *Modeling and Simulation of Biomaterials.* Annual Review of Materials Research, 2004. **34**(1): p. 279-314.

207. Rega, N., et al., *Hybrid ab initio empirical molecular dynamics: combining the ONIOM scheme with the atom-centered density matrix propagation (ADMP) approach.* J. Phys. Chem. B, 2004. **108**: p. 4210-4220.

208. Ren, W., et al., *Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide.* Journal of Chemical Physics, 2005. **123**(13): p. -.

209. Reuter, N., et al., *Frontier bonds in QM/MM methods: A comparison of different approaches.* Journal of Physical Chemistry A, 2000. **104**(8): p. 1720-1735.

210. Ritter, C.A. and C.L. Arteaga, *The epidermal growth factor receptor-tyrosine kinase: a promising therapeutic target in solid tumors. [Review] [69 refs]*. Seminars in Oncology, 2003. **30**(1 Suppl 1): p. 3-11.

211. Rodrigues, G.A., et al., *A Novel Positive Feedback Loop Mediated by the Docking Prote in Gab1 and Phosphatidylinositol 3-Kinase in Epidermal Growth Factor Receptor Signaling.* Mol. Cell Biol., 2000. **20**: p. 1448-1459.

212. Rosta, E., et al., *Artificial reaction coordinate "tunneling" in free-energy calculations: the catalytic reaction of RNase H.* J Comput Chem, 2009. **30**(11): p. 1634-41.

213. Roux, B., *The Calculation of the Potential of Mean Force Using Computer-Simulations.* Computer Physics Communications, 1995. **91**(1-3): p. 275-282.

214. Sachsenmaler, C., *Targeting protein kinases for tumor therapy.* Onkologie, 2001. **24**: p. 346-355.

215. Sanner, M.F., A.J. Olson, and J.C. Spehner, *Reduced surface: An efficient way to compute molecular surfaces.* Biopolymers, 1996. **38**(3): p. 305-320.

216. Saso, K., et al., *Differential inhibition of epidermal growth factor signaling pathways in rat hepatocytes by long-term ethanol treatment.* Gastroenterology, 1997. **112**(6): p. 2073-2088.

217. Schindler, T., et al., *Structural mechanism for STI-571 inhibition of abelson tyrosine kinase.* Science, 2000. **289**(5486): p. 1938-42.

218. Schlessinger, J., *Cell signaling by receptor tyrosine kinases.* Cell, 2000. **103**: p. 211-225.

219. Schlessinger, J., *Common and distinct elements in cellular signaling between EGF and FGF receptors.* Science, 2004. **306**: p. 1506-1507.

220. Schlichting, I. and J. Reinstein, *Structures of active conformations of UMP kinase drom Dictyostelium discoideum suggest phosphoryl transfer is associative.* Biochemistry, 1997. **36**: p. 9290-9296.

221. Schlick, T., *Engineering teams up with computer-simulation and visualization tools to probe biomolecular mechanisms.* Biophysical J., 2003. **85**: p. 1-4.

222. Schmidt, M.W., et al., *General Atomic and Molecular Electronic-Structure System.* Journal of Computational Chemistry, 1993. **14**(11): p. 1347-1363.

223. Schoeberl, B., et al., *Computational modeling of the dynamics of MAPKinase cascade activated by surface and internalized receptors.* Nature Biotech., 2002. **20**: p. 370-375.

224. Schrödinger, *Schrödinger software set*

225. Schulman, B.A., D.L. Lindstrom, and E. Harlow, *Substrate recruitment to cyclin-dependent kinase 2 by a multipurpose docking site on cyclin A.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(18): p. 10453-10458.

226. Schulze, W.X., L. Deng, and M. Mann, *Phosphotyrosine interactome of the ErbB-receptor kinase family.* Molecular Systems Biology, 2005.

227. Scott, W.G., *Biophysical and biochemical investigations of RNA catalysis in the hammerhead ribozyme.* Q Rev Biophys, 1999. **32**(3): p. 241-84.

228. Senn, H.M. and W. Thiel, *QM/MM methods for biological systems*, in *Atomistic Approaches in Modern Biology: from Quantum Chemistry to Molecular Simulations*. 2007, Springer-Verlag Berlin: Berlin. p. 173-290.

229. Sergina, N.V., et al., *Escape from HER-family tyrosine kinase inhibitor therapy by the kinase-inactive HER3.* Nature, 2007. **445**(7126): p. 437-441.

230. Shan, Y., et al., *A conserved protonation-dependent switch controls drug binding in the Abl kinase.* Proc Natl Acad Sci U S A, 2009. **106**(1): p. 139-44.

231. Sharp, K.A. and B. Honig, *Calculating Total Electrostatic Energies with the Nonlinear Poisson-Boltzmann Equation.* Journal of Physical Chemistry, 1990. **94**(19): p. 7684-7692.

232. Shaw, D.E., et al., *Anton, a special-purpose machine for molecular dynamics simulation.* Communications of the Acm, 2008. **51**(7): p. 91-97.

233. Shawver, L.K., D. Slamon, and A. Ullrich, *Smart drugs: Tyrosine kinase inhibitors in cancer therapy.* Cancer Cell, 2002. **1**(2): p. 117-123.

234. Shawyer, L.K., D. Slamon, and a.A. Ullrich, *Smart drugs: tyrosine kinase inhibitors in cancer therapy.* Cell, 2002. **1**: p. 117-123.

235. Shen, K., et al., *Protein kinase structure and function analysis with chemical tools.* Biochimica Et Biophysica Acta-Proteins and Proteomics, 2005. **1754**(1-2): p. 65-78.

236. Sherman, W., et al., *Novel procedure for modeling ligand/receptor induced fit effects.* Journal of Medicinal Chemistry, 2006. **49**(2): p. 534-553.

237. Shi, F.M., et al., *ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation.* Proceedings of the National Academy of Sciences of the United States of America. **107**(17): p. 7692-7697.

238. Shimojo, F., et al., *Embedded divide-and-conquer algorithm on hierarchical real-space grids: parallel molecular dynamics simulation based on linear-scaling density functional theory.* Computer Physics Communications, 2005. **167**(3): p. 151-164.

239. Shirts, M.R. and V.S. Pande, *Mathematical analysis of coupled parallel simulations.* Physical Review Letters, 2001. **86**(22): p. 4983-4987.

240. Shurki, A. and A. Warshel, *Structure/function correlations of proteins using MM, QM/MM, and related approaches: Methods, concepts, pitfalls, and current progress.* Protein Simulations, 2003. **66**: p. 249-313.

241. Skamnaki, V.T., et al., *Catalytic mechanism of phosphorylase kinase probed by mutational studies.* Biochemistry, 1999. **38**(44): p. 14718-14730.

242. Smith, J., *Erlotinib: Small-molecule targeted therapy in the treatment of non-small-cell lung cancer.* Clinical Therapeutics, 2005. **27**(10): p. 1513-1534.

243. Snow, C.D., et al., *Absolute comparison of simulated and experimental protein-folding dynamics.* Nature, 2002. **420**: p. 102-106.

244. Soltoff, S.P., et al., *ERBB3 IS INVOLVED IN ACTIVATION OF PHOSPHATIDYLINOSITOL 3-KINASE BY EPIDERMAL GROWTH-FACTOR.* Molecular and Cellular Biology, 1994. **14**(6): p. 3550-3558.

245. Songyang, Z., et al., *Catalytic Specificity of Protein-Tyrosine Kinases Is Critical for Selective Signaling.* Nature, 1995. **373**(6514): p. 536-539.

246. Songyang, Z., et al., *Catalytic specificity of protein-tyrosine kinases is critical for selective signalling.* Nature, 1995. **373**: p. 536-539.

247. Sordella, R., et al., *Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways.* Science, 2004. **305**(5687): p. 1163-7.

248. Stamos, J., M.X. Sliwkowski, and C. Eigenbrot, *Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor.* Journal of Biological Chemistry, 2002. **277**(48): p. 46265-46272.

249. Steitz, T.A., *DNA- and RNA-dependent DNA polymerases.* Curr. Opin. Struct. Biol., 1993. **3**: p. 31-38.

250. Steitz, T.A., et al., *A unified polymerase mechanism for nonhomologous DNA and RNA polymerases.* Science, 1994. **266**: p. 2022-2025.

251. Stone, J.E., et al., *Accelerating molecular modeling applications with graphics processors.* J Comput Chem, 2007. **28**(16): p. 2618-40.

252. Suenaga, A., et al., *Molecular Dynamics Simulations Reveal that Tyr-317 Phosphorylation Reduces Shc Binding Affinity for Phosphotyrosyl Residues of Epidermal Growth Factor Receptor.* 2009. **96**(6): p. 2278-2288.

253. Swanson, J.M.J., R.H. Henchman, and J.A. McCammon, *Revisiting free energy calculations: a theoretical connection between MM/PBSA and direct calculation of the association free energy.* Biophys. J., 2004. **86**: p. 67-74.

254. Szabo, A. and N.S. Ostlund, *Modern Quantum Chemistry*. 1996, Mineola, New York: Dover Publications.

255. Tanoue, T.J. and E. Nishida, *Molecular recognitions in the MAP kinase cascades.* Cellular Signalling, 2003. **15**(5): p. 455-462.

256. Taylor, S.S., et al., *Crystal-Structures of the Catalytic Subunit of Camp-Dependent Protein-Kinase Reveal General Features of the Protein-Kinase Family.* Receptor, 1993. **3**(3): p. 165-172.

257. Telesco, S.E. and R. Radhakrishnan, *Atomistic Insights into Regulatory Mechanisms of the HER2 Tyrosine Kinase Domain: A Molecular Dynamics Study.* 2009. **96**(6): p. 2321-2334.

258. Thaimattam, R., et al., *Protein kinase inhibitors: Structural insights into selectivity*. Current Pharmaceutical Design, 2007. **13**(27): p. 2751-2765.

259. Tracy, S., et al., *Gefitinib induces apoptosis in the EGFRL858R non-small-cell lung cancer cell line H3255.* Cancer Res, 2004. **64**(20): p. 7241-4.

260. Turjanski, A.G., G. Hummer, and J.S. Gutkind, *How Mitogen-Activated Protein Kinases Recognize and Phosphorylate Their Targets: A QM/MM Study.* Journal of the American Chemical Society, 2009. **131**(17): p. 6141-6148.

261. Ubersax, J.A. and J.E. Ferrell, *Mechanisms of specificity in protein phosphorylation.* Nature Reviews Molecular Cell Biology, 2007. **8**(7): p. 530-541.

262. Uyemura, T., et al., *Single-molecule analysis of epidermal growth factor signaling that leads to ultrasensitive calcium response.* Biophys J., 2005.

263. Vaiana, A.C., et al., *AFMM: A molecular mechanics force field vibrational parametrization program.* Computer Physics Communications, 2005. **167**(1): p. 34-42.

264. Valiev, M., et al., *Phosphorylation reaction in cAPK protein kinase-free energy quantum mechanical/molecular mechanics simulations.* Journal of Physical Chemistry B, 2007. **111**(47): p. 13455-13464.

265. Vieira, A.V., C. Lamaze, and S.L. Schmid, *Control of EGF Receptor signaling by clathrim-mediated endocytosis.* Science, 1996. **274**: p. 2066-2089.

266. Wang, J., et al., *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, 2001. **123**(22): p. 5221-5230.

267. Wang, Z.L., et al., *Structural basis of inhibitor selectivity in MAP kinases.* Structure with Folding & Design, 1998. **6**(9): p. 1117-1128.

268. Warshel, A., *Computer modeling of chemical reactions in enzymes and solution*. 1989, New York: John Wiley and Sons.

269. Warshel, A. and W.W. Parson, *Dynamics of Biochemical and Biophysical Reactions: Insight from Computer Simulations.* Quart. Rev. Biophys., 2001. **34**: p. 563-679.

270. Waterman, H. and Y. Yarden, *Molecular mechanisms underlying endocytosis and sorting of Erb receptor tyrosine kinases.* FEBS letters, 2001. **490**: p. 142-152.

271. Weiner, P.W. and P.A. Kollman, *AMBER: assisted model building with energy refinement.* J. Comput. Chem., 1981. **2**: p. 287-303.

272. Weng, G., U.S. Bhalla, and R. Iyengar, *Complexities in biological signaling systems.* Science, 1999. **284**: p. 92-96.

273. Wolf-Yadlin, A., et al., *Effects of HER2 overexpression on cell signaling networks governing proliferation and migration.* Mol Syst Biol, 2006. **2**: p. 54.

274. Wong, C.F., et al., *Molecular docking of balanol to dynamics snapshots of protein kinase A.* Proteins-Structure Function and Bioinformatics, 2005. **61**(4): p. 850-858.

275. Wong, C.F., *Flexible ligand-flexible protein docking in protein kinase systems.* Biochimica Et Biophysica Acta-Proteins and Proteomics, 2008. **1784**(1): p. 244-251.

276. Wood, E.R., et al., *A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): Relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells.* Cancer Research, 2004. **64**(18): p. 6652-6659.

277. Wu, W., et al., *Src-dependent phosphorylation of the epidermal growth factor receptor on tyrosine 845 is required for zinc-induced Ras activation.* Journal of Biological Chemistry, 2002. **277**(27): p. 24252-24257.

278. Xie, P., et al., *Structure-based design of an organoruthenium phosphatidyl-inositol-3-kinase inhibitor reveals a switch governing lipid kinase potency and selectivity.* Acs Chemical Biology, 2008. **3**(5): p. 305-316.

279. Yang, J., *Development and evaluation of a generic evolutionary method for protein-ligand docking.* J. Comput. Chem., 2004. **25**(6): p. 843-857.

280. Yang, L., et al., *Polymerase beta simulations suggest that Arg258 rotation is a slow step rather than large subdomain motion per se.* J. Mol. Biol., 2002. **317**: p. 651-671.

281. Yarden, Y. and M.X. Sliwkowski, *Untangling the ErbB signalling network.* Nature Reviews Molecular Cell Biology, 2001. **2**(2): p. 127-137.

282. Yoon, M.Y. and P.F. Cook, *Chemical Mechanism of the Adenosine Cyclic 3',5'-Monophosphate Dependent Protein-Kinase from Ph Studies.* Biochemistry, 1987. **26**(13): p. 4118-4125.

283. Young, M.A., et al., *Dynamic Coupling between the SH2 and SH3 Domains of c-Src and Hck Underlies Their Inactivation by C-Terminal Tyrosine Phosphorylation.* Cell, 2001. **105**: p. 115-126.

284. Yun, C.H., et al., *Structures of lung cancer-derived EGFR mutants and inhibitor complexes: Mechanism of activation and insights into differential inhibitor sensitivity.* Cancer Cell, 2007. **11**(3): p. 217-227.

285. Zagrovic, B., E.J. Sorin, and V. Pande, *beta-Hairpin Folding Simulations in Atomistic Detail Using an Implicit Solvent Model.* J. Mol. Biol., 2001. **313**: p. 151-169.

286. Zaloj, V. and R. Elber, *Parallel Computations of Molecular Dynamics Trajectories Using The Stochastic Path Approach.* Comput. Phys. Comm., 2000. **128**: p. 118-127.

287. Zhang, X., et al., *An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor.* Cell, 2006. **125**: p. 1137-1149.

288. Zhang, X., et al., *An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor.* Cell, 2006. **125**(6): p. 1137-49.

289. Zhang, X.W., et al., *An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor.* Cell, 2006. **125**(6): p. 1137-1149.

290. Zhang, Y. and W. Yang, *A pesudobond approach to combining quantum mechanical and molecular mechanical methods.* J. Chem. Phys., 1999. **110**: p. 46-54.

291. Zhang, Y., J. Kua, and J.A. McCammon, *Role of the catalytic triad and oxyanion hole in acetylcholinesterase catalysis: an ab initio QM/MM study.* J. Amer. Chem. Soc., 2002. **124**: p. 10572-10577.

292. Zheng, J.H., et al., *2.2-Angstrom Refined Crystal-Structure of the Catalytic Subunit of Camp-Dependent Protein-Kinase Complexed with Mnatp and a Peptide Inhibitor.* Acta Crystallographica Section D-Biological Crystallography, 1993. **49**: p. 362-365.

293. Zhou, B.J. and C.F. Wong, *A Computational Study of the Phosphorylation Mechanism of the Insulin Receptor Tyrosine Kinase.* Journal of Physical Chemistry A, 2009. **113**(17): p. 5144-5150.

294. Zhou, J. and J.A. Adams, *Is there a catalytic base in the active site of cAMP-dependent protein kinase?* Biochemistry, 1997. **36**(10): p. 2977-2984.

295. Zhou, R. and B.J. Berne, *A New Molecular Dynamics Method Combining the Reference System Propagator Algorithm with a Fast Multipole Method for Simulating Proteins and Other Complex Systems.* J. Chem. Phys., 1995. **103**: p. 9444-9459.

296. Zhou, R., B.J. Berne, and R. Germain, *The free energy landscape for β hairpin folding in explicit water.* Proc. Natl. Acad. Sci. USA, 2001. **98**: p. 14931-14936.