Stony Brook University

OFFICIAL COPY

Targeted Design of Dual Kinase Inhibitors for Breast Cancer

Yulin Huang

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

May 2013

Stony Brook University

The Graduate School

Yulin Huang

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Robert C. Rizzo - Dissertation Advisor
Associate Professor, Applied Mathematics and Statistics

David F. Green - Chairperson of Defense
Associate Professor, Applied Mathematics and Statistics

Erwin London
Professor, Biochemistry and Cell Biology

Yuefan Deng
Professor, Applied Mathematics and Statistics

This dissertation is accepted by the Graduate School

Charles Taber

Interim Dean of the Graduate School

Abstract of the Dissertation

Targeted Design of Dual Kinase Inhibitors for Breast Cancer

by

Yulin Huang

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

2013

In this dissertation, several projects for kinase drug design are presented. These projects employ computational techniques to study binding specificity and resistance of kinase inhibitors with breast cancer target proteins through structural and energetic analysis.

In Chapter 1, an introduction to molecular targeted therapeutics for breast cancer and computational techniques for structure-based drug discovery is described. ErbB family members and IGFIR are important targets for breast cancer. In Chapter 2, we have characterized the determinants which drive binding affinity for the FDA-approved small molecule drug lapatinib specificity with the goal of uncovering the origins of the specificity across different ErbB family members. Results have incorporated homology modeling, molecular dynamic simulations, free energy calculations, binding analysis, and hydration analysis. A key finding of our work was identification of a physically unique water-mediated H-bond network which compellingly

explains why lapatinib has high affinity for EGFR and HER2 but not the highly homologous ErbB4. Our model also helps to explain drug resistance which can arise due to disruption of the water-mediated network.

In Chapter 3, we have employed MM-GBSA method with the same protocol used in Chapter 2 to quantify the binding free energies of lapatinb two lapatinib conformations (conf1 and conf2) with EGFR, HER2 and ErbB4. To further explore the energy and the probability of two conformations, umbrella sampling with potential mean force (PMF) calculation was performed. Both MM-GBSA and PMF results show that conf1 always yield more favorable binding energy with all three proteins than conf2, indicating that conf1 with ErbB4, as seen in crystal structure 3BBT, is only a local minimum and conf1 is the global minimum. The two conformations may co-exist in an equilibrium with ErbB4, which may also help explain why lapatinib binds to ErbB4 less tightly than to EGFR and HER2.

In Chapter 4, thermodynamic integration (TI) method was used to examine the structure activity relationship (SAR) for a series of ligands with an imidazopyrazine scaffold interacting with the intermediate form of IGF-IR. Twelve different $\Delta\Delta G_{bind}$ relationships were studied as well as several "null" transformations to validate the simulation protocols. From a series of 19 simulation windows (2ns of simulation per window), we obtained a relative binding free energy of close to zero with negligible standard error of the mean for five null transformations, indicating the model construction and simulation are robust. Overall, the results of the study were mixed. While single perturbations involving aliphatic changes (i.e. N, Me, Et) yielded excellent results compared to experiment more complicated perturbations involving bulky groups (Ph) or polar groups (OH, NH_2) yielded large errors. Studies to explore the sources of these errors are ongoing.

In Chapter 5, we have presented preliminary virtual screening results targeting intermediate and active forms of IGF-IR using DOCK 6.6 to identify new drug leads. Compounds from the ZINC/ChemDiv catalog of purchasable compounds (1.2M) were flexibly docked and the single lowest-energy pose for each compound was retained. The top 100,000 molecules were then clustered based on MACCS fingerprints and the top 250 cluster heads and all families members were selected based on the four different scoring methods: (1) standard DOCK score (2) van der Waals footprint similarity score (3) electrostatic footprint similarity score and (4) the combined footprint sum. The top 20 compounds for each category will be advanced to experimental testing after visual inspection and additional analysis. In particular, compounds will be minimized and assessed in the binding pocket of EGFR, HER2 and IGF-IR to identify possible combinations for use as dual inhibitors.

In Chapter 6, we conclude with a description of ongoing projects and ideas for future directions.

Table of Contents

List of Figures

backbone oxygen at residue 791 (S2), and ligand aniline hydrogen (S3) from 10,000 evenly spaced MD snapshots (20 ns trajectories). (b) ΔH-bond and $\Delta\Delta G_b$ calcd values (mutant - wildtype) from simulations of wildtype EGFR, C775F, T790M, and T854A mutants with lapatinb (top) and a N* $\rightarrow$ C* analog (bottom). Relative H-bonds (count) computed between the 3 waters closest to N*/C* with: N*/C*, the entire ligand, residues 775, 790, 791, 854, and the overall sum excluding N*/C*. Relative binding free energies (kcal/mol) computed using 3WAT (three closest waters to N* / C*) or 1WAT protocols (one water if within 3 Å of N* / C*). 63

List of Tables

List of Abbreviations

ACF, autocorrelation function

ATP, adenosine triphosphate

BASEM, block-averaged standard errors of the mean

BCR-ABL, tyrosine kinase drug target for chronic myelogenous leukemia

EGFR, Human epidermal growth factor receptor 1

ErbB4, Human epidermal growth factor receptor 4

HER2, Human epidermal growth factor receptor 2

IGF-IR, insulin-like growth factor 1 receptor

MD, Molecular dynamics

MM-GBSA, Molecular Mechanics Generalized Born Surface Area

PDB, Protein Data Bank

RMSD, root-mean-square-deviation

PMF, potential mean force

TI, thermodynamic integration

FPS, footprint similarity score

DCE, standard docking score

ed, Euclidean distance

Acknowledgement

Foremost, I would like to express my sincere gratitude to Dr. Robert C. Rizzo for being my advisor. My graduate experience has been fruitful because of his continuous encouragement, support, patience, knowledge and guidance. I appreciate his helpful advice on my research projects and assistance writing proposals, grants, and papers.

I am grateful to Dr. David Green and Dr. Carlos Simmerling for providing me with insightful and constructive suggestions that have greatly enhanced my projects. I am also thankful for Dr. Erwin London and Dr. Steve Smith who have offered me the unique feedback of accomplished experimentalists. I also want to thank Dr. Yuefan Deng for teaching me parallel computing. My committee members' knowledge, enthusiasm, and dedication to science have been inspiring.

I also feel very thankful for my labmates. I especially acknowledge Dr. Trent Balius whose work provided the foundation for my kinase research. Trent is a highly creative and dedicated researcher from whom I have learned a lot. Dr. Sudipto Mukherjee, a smart and enthusiastic researcher, has provided me with great help for computer related questions and docking projects. I also want to acknowledge postdoctoral fellow Dr. William Joseph Allen, in our lab, for his patient, conscientious and warmhearted guidance. Lingling Jiang has helped me set up and interpret thermodynamic integration experiments and Brian Fochtman who has helped me make figures and improve my writing. Past lab members who have aided my research include Brian McGillick, Rashi Goyal and Patrick Holden who helped established the protocols I used in the virtual screening project. I also want to thank my friends from other labs, Dr. Yukiji, AJ Campbell, Lin Fu, and Yi Shang for useful discussions.

Finally, I want to thank my parents for providing me with spiritual support during all these years that I have been abroad alone. Their dedication to me, has helped me pursue my career and my path through life. I really cherished the days that my parents and I traveled around the USA in 2008.

Chapter 1. Introduction

This chapter provides background and introduction to the theory and computational methods used in this dissertation.

1.1 Molecular Targeted Therapeutics for Breast Cancer

Breast cancer is the most common malignant tumor for women accounting for over 25% cancers (excluding skin cancers) diagnosed among females in the United States.[1] American Cancer Society statistics show that one in eight women (12.5%) are expected to develop some form of breast cancer in their lifetime. An estimated 226,870 new cases of breast cancer were diagnosed among women in 2012.[1] Despite the important successes in development of therapies for treatment of breast cancer, drug resistance and harmful side effects still pose a significant problem which highlights the need for continued development of novel, more potent, and effective drugs to prevent cancer progression and metastasis. Notably, development of breast and other anti-cancer agents in the promising new small molecule drug class known as "molecular targeted therapeutics", was in large measure enabled through structure-based design that included computer-aided modeling of inhibitors with their targets at the atomic level.[2-4] Computer modeling provides a powerful platform for deciphering which interactions are most crucial for drug activity. And, computational techniques such as virtual screening can be used in the search for new drug leads. Work presented in this thesis makes use of a variety of modeling

techniques to predict and characterize protein-ligand binding for breast cancer targets at the atomic level.

Molecular targeted therapeutics represent promising breast cancer treatment alternatives over traditional cytotoxic therapies which can have significant undesirable side effects. Important targets in this class include members of the ErbB family of receptor tyrosine kinases (mainly EGFR and HER2)[5-9] and the related protein IGF-IR.[10-15] Examples of FDA-approved drugs include erlotinib (non-small cell lung cancer), and lapatinib (breast cancer), both of which are competitive inhibitors of the native substrate ATP (Figure 1.1).[16] Unlike many compounds, lapatinib is a dual kinase inhibitor which provides important "proof-of-concept" that dual specificity could be a powerful therapeutic strategy. Although lapatinib is currently only approved for HER2+ breast cancer patients it targets both HER2 and EGFR for which heterodimer signaling has been shown to be particularly strong.[17] However, use of lapatinib can also result in activation of the alternative IGF-IR signaling pathway which leads to drug resistance.[18] We hypothesize that inhibiting proteins from both families could be an effective strategy for development of new breast cancer drugs.

The primary objective of work presented in this thesis is to construct all-atom computational models and elucidate determinants which drive molecular recognition for inhibitors with breast cancer relevant targets. Concurrently we will apply computational tools for virtual screening in the search for new breast cancer drug leads. The rational for performing the research is that development of accurate models for characterization of inhibitor binding will enable development of new therapeutics for breast cancer having enhanced potency and improved resistance profiles.

Compound	Structure	PDB code	Kinase target
AMP-PNP (ATP analog)		2ITX	ALL
Erlotinib		1M17	EGFR
Lapatinib		1XKK	EGFR HER2 ErbB4
PQIP		3D94	IGF-IR
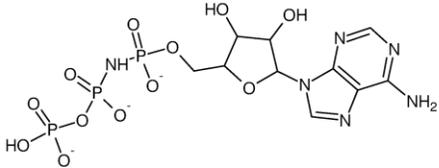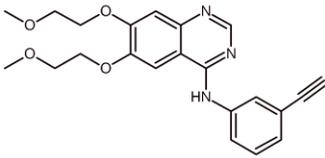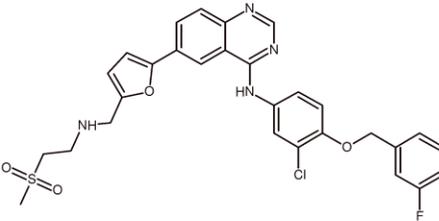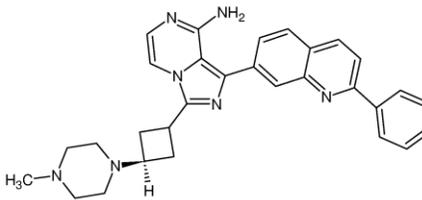
Figure 1.1. 2D structures of representative kinases inhibitors, crystallographic PDB codes, and targets.

Receptor tyrosine kinases are important targets for development of anti-cancer therapies due to their key roles in cellular signaling and growth.[19] ErbB family members and IGF-IR have both been shown to be overexpressed in breast cancer and are validated breast cancer targets. As illustrated in Figure 1.2, extracellular ligand binding induces dimerization which causes activation of intracellular domains. Both heterodimers and homodimers may form. Notably, heterodimeriztion involving HER2 has been shown to generate stronger signals compared to other combinations. Activation leads to phosphorylation which results in activation of a series of downstream cascades including PI3K and MAPK pathways. Aberrant signaling can lead to uncontrolled growth and cancer. To suppress kinase function and aberrant signaling, there are

two primary methods: (1) monoclonal antibodies which bind the extracellular domain (see Figure 1.2 and Figure 1.3) thereby blocking native ligand binding and (2) tyrosine kinase inhibitors (TKIs) which compete for ATP in the catalytic site in the intracellular kinase domain (highlighted in Figure 1.3) which stops the transfer of the γ-phosphate from ATP to various substrate proteins.
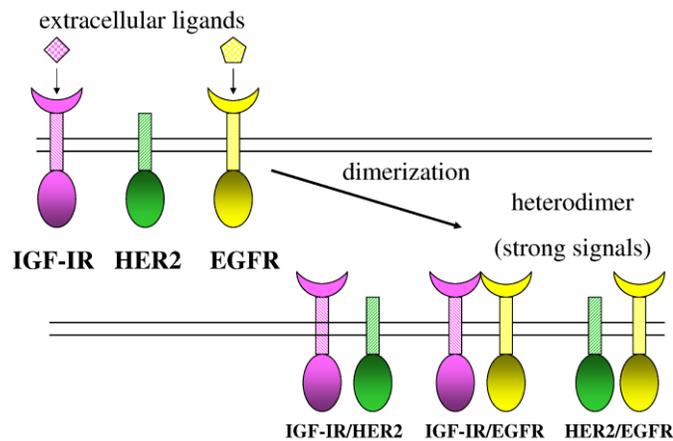
Figure 1.2. Cross-talk between the ErbB family and IGF-IR. Figure adapted from Jin et al. *J. Mammary Gland. Biol. Neoplasia* 2008, *13*, 485-498.
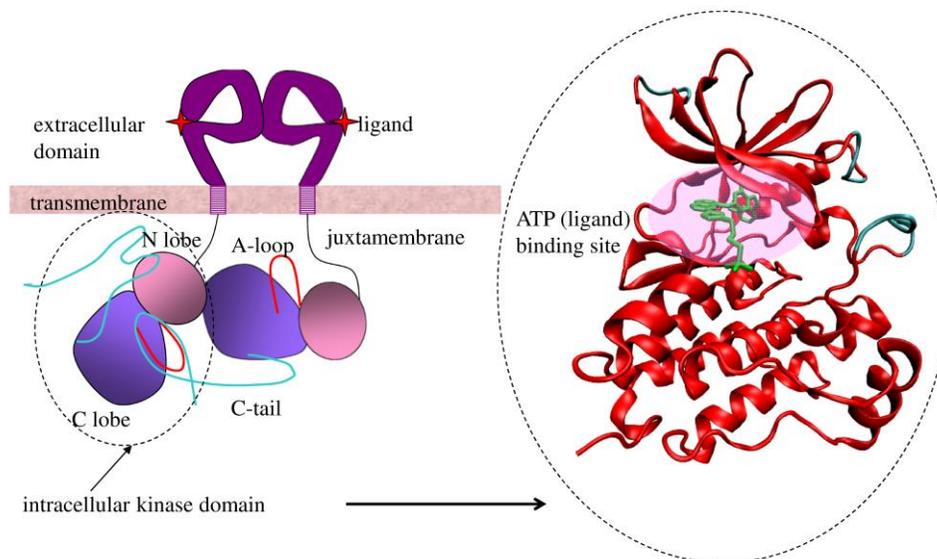
Figure 1.3. A schematic representation of the kinase structure (left) and the intracellular domain (right) with the binding site highlighted. Figure adapted from Zhang et al. *Cell* **2006,** 125, 1137-1149.

Focusing on the class of reversible inhibitors which target the ATP binding site in the intracellular domain, several compounds have been designed which target the ErbB family, and more recently IGF-IR (Figure 1.1). In 2007, lapatinib was approved by the FDA to be used with capecitabine for patients with advanced or metastatic breast cancers overexpressing HER2, or who have already been treated with other drugs (anthracycline, a taxane, or trastuzumab) and cyototoxic chemotherapy.[16] Clinical trials indicated that lapatinib should be limited to the treatment of HER2-overexpressing breast cancer patients in which only 25-30% show response.[18,20] The drug shows little efficacy against tumors driven mainly by EGFR. The compound erlotinib (Tarceva, OSI Pharmaceuticals) is approved for non-small cell lung (NSCL), and pancreatic cancer, and is in Phase I clinical trials for primary breast cancer although it does not have clinical activity against HER2-positive or triple-negative breast cancers.[21] In addition, both erlotinib and lapatinib do not have the capability to inhibit EGFR kinase activity of EGFRvIII variant which is frequently expressed in breast cancer.[19,35] Lastly, the compound called PQIP in Figure 1.1 is an IGF-IR inhibitor,[22,23] for which a crystal structure has been reported, and is the parent compound for an analog termed OSI-906. Studies have shown that OSI-906 reduced growth in cell lines representative of breast and several other cancers including NSCL and pancreatic.[24]

Additionally, there is strong evidence that ErbB family members and IGF-IR cross-talk in breast cancer cells[25-28] resulting in drug-resistance through formation of HER2/IGF-IR or EGFR/IGF-IR heterodimers.[29,30] This cross-talk is associated with trastuzumab resistance among breast cancer patients[29,31-33] and resistance to EGFR inhibitors.[25,34] Thus, development of compounds which co-target IGF-IR and ErbB members could lead to drugs with improved resistance profiles and should be pursued. Several clinical studies strongly support this

concept.[27,35-37]   The combination of gefitinib and the IGF-IR inhibitor ABDP inhibits activation of IGF-IR and significantly suppresses phosphorylation of EGFR as well as cell growth compared with taking a single agent in tamoxifen-resistant cells.[27,35]   In a similar way, a dual combination of the IGF-IR inhibitor AG1024 with gefitinib (which primarily targets EGFR) increases anti-tumor effects on breast cancer cell growth.[36]   Additionally, an antibody which co-targets both IGF-IR and EGFR has achieved satisfactory anti-tumor effects in combating colorectal and pancreatic carcinoma xenografts.[37]   Overall, when complexed with relevant kinases, the inhibitors in Figure 1.1 provide a wealth of structural information for development of the computer simulations and models outlined in this thesis to co-target ErbB and IGF-IR for the development of new drugs.

1.2 Computational Techniques for Structure-Based Drug Discovery

Historically, new drugs were discovered from plants and other natural products through trial and error. Today, increasing number of 3-D structures of relevant biological targets and the success of computational structure-based approaches has paved the way for more rational drug design. As shown in Figure 1.4, modern drug discovery includes numerous stages such as pre-discovery (target identification, validation), hit identification (lead generation, profiling and optimization), preclinical trials, clinical trials, FDA review, and large-scale manufacturing.[38] This process can take 10-15 years at an estimated cost of $800 million to $1 billion.[39] Introducing accurate and efficient computational methods for structure-based drug design in the early stages of drug discovery can save significant time and money compared to other approaches, for example, synthesizing and experimental testing thousands of compounds.[2,40,41]
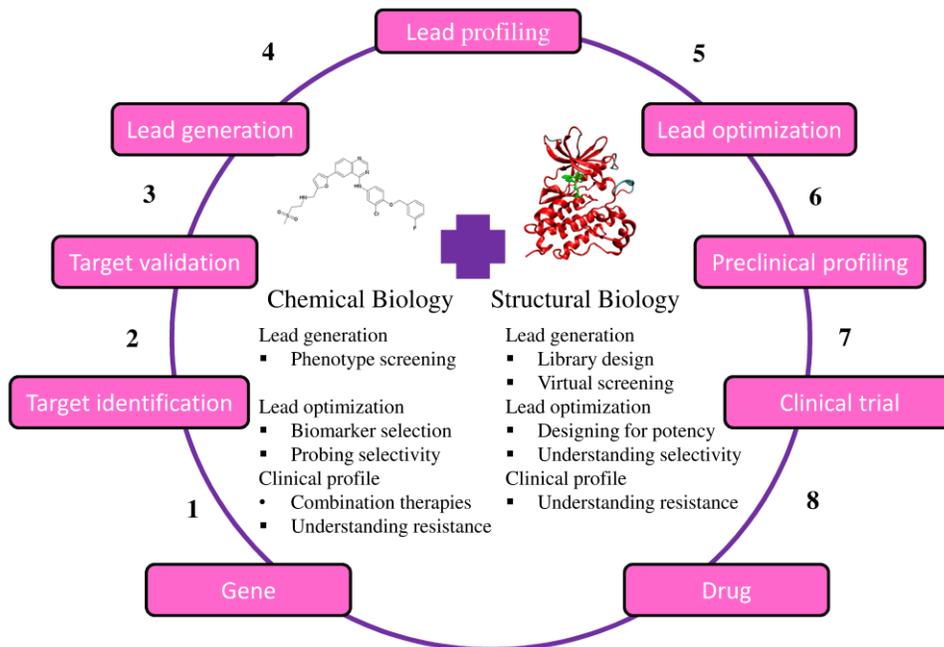
Figure 1.4. Schematic representation of the important phases of drug discovery. Figure adapted from Collins, I.; Workman, P. *Nat Chem Biol* **2006**, *2*, 689-700.

In the target identification phase, the targets to be advanced for further study should be closely linked to the cause of a human disease, and, ideally bind a small molecule or peptide. Drug targets are usually proteins having well-defined binding pockets or RNA targets with defined secondary structures. A structure of the target is critical for computational lead identification and X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR) and, in some instances computational methods such as homology modeling, are the primary methods for structure determination. The Protein Data Bank (PDB)[42,43] makes publically available thousands of structures (2013 statistics indicate >88,000 structures) from X-ray crystallography and NMR which can be used for atomistic computer simulations. If a query protein has not been deposited but a template protein with high sequence homology is available, the structure of the

query protein can be modeled using homology modeling. Therefore, many more protein structures can be studied when using related protein structures in the PDB as templates.

After a target structure and binding site in the target are identified, a variety of simulation methods can be used to develop and refine leads. Focusing on atomistic computational methods, two of the most widely used approaches are: (1) Molecular docking and virtual screening (lead development),[44,45] and (2) molecular dynamics and free energy calculation (lead refinement). Importantly, computational methods allow for molecules to be studied in detail at the atomic level and can be used to investigate properties that can be difficult to assess experimentally such as the energetic strength between specific residues or atoms, water-mediated hydrogen bonds, desolvation penalties and structure-activity relationships. In addition, unlike the static structures provided by X-crystallography, molecular dynamics simulations can be used to sample different conformations of the target proteins and their ligands which can provide insights into allosteric binding.

In terms of lead discovery, molecular docking is the computational process of searching for small molecule drug-leads from large ligand databases that are able to fit both geometrically and energetically into the binding site of a protein target. In this process, both the binding geometry (termed pose) and the interaction energy of the small molecule are predicted. The UCSF ZINC database[46] represents perhaps the world's largest source of purchasable compounds (curated by commercial vendor) and DOCK,[47-49] AUTODOCK,[50] FlexX,[51] GLIDE,[52,53] and GOLD[54] represent some widely used software programs for virtual screening. Generally, a ligand library on the order of 1 million or less is screened to the target, rank-ordered based on different scoring functions, and the top scoring leads are subjected to visual inspection so that a subset (on the order of 100 compounds) can be purchased for evaluation by experimentalists.

In terms of lead refinement, more detailed calculations based on molecular dynamics in conjunction with free energy perturbation (FEP),[45,55] thermodynamic integration (TI),[56] or more recently MM-GBSA/MM-PBSA[57-59] theories can be used. With FEP and TI methods, relative binding free energies are typically estimated through a process which involves perturbation of ligand A to ligand B in both bound and unbound states. This requires multiple simulations or "windows" and is quite time consuming. Further, the calculations are quite tedious to setup for large numbers of ligands and obtaining well-converged absolute free energies is difficult. With the MM-GBSA method however, typically only a single simulation of the ligand in the bound state needs to be performed to estimate an "absolute" free energy and the method is easily implemented. Despite the approximations (such as neglect of all entropic contributions to binding), in favorable cases, MM-GBSA can achieve similar accuracy to FEP and TI but with less computational expense. Based on calculations using either approach, atomistic detail on binding is available and leads (or related analogs) designed through additional simulations can be optimized to increase binding. Those with the most favorable free energies can be developed further.

The sections which follow provide more detailed descriptions of the primary computational tools/methods used in this thesis which include: (1) homology modeling, (2) molecular mechanics, (3) molecular docking, (4) molecular dynamics, and (5) free energy calculations.

1.3 Homology Modeling

Molecular modeling typically starts using protein structure coordinates deposited in the PDB which have been solved experimentally. But for some proteins, there may not be a suitable

crystal or NMR structure available. An important computational category of protein structure prediction involves comparative/homology modeling.[60-63] Other approaches such as *de novo* prediction can also be used. Focusing on homology modeling, which was used in this thesis, the goal is to derive a three dimensional (3D) structure for a query protein based on the sequence similarity to a protein of known structure (template) as shown schematically in Figure 1.5.[60-62] Two prerequisites for building useful models must be met. Firstly, the target sequence and the template structure must share some extent of similarity. Secondly, a significant amount of alignment between the target sequence and the template structure must be calculated correctly. The procedure of homology modeling consists of four steps, including template selection, template-target alignment, model construction, and model evaluation. If the final model is not satisfactory, the four steps are repeated until a satisfactory one is constructed.[64]
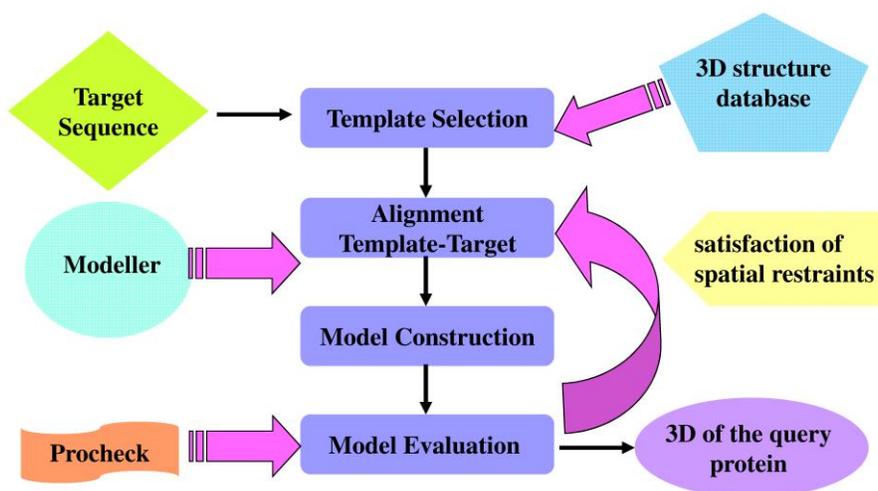
Figure 1.5. The process of constructing homology models.

The very first step of homology modeling is to select appropriate templates. Searching structure databases such as PDB database,[42,65] SCOP,[66] DALI,[67] and CATH[68] can be helpful in identifying templates for the target sequence. Typically, three methods termed pairwise

sequence-sequence comparison (FASTA, BLAST),[69,70] multiple sequence comparisons (PSI-BLAST),[71] and threading[72,73] can be used to identify similar folds. The template environment (solvent, pH, ligands, etc) as well as the quality of the experimental template structure should be considered in the selection process. If ligand binding is being studied, a template structure containing the same or similar ligands should be chosen.

In terms of target-template alignment, for closely related protein sequences with identity over 40%, the alignment is almost always correct. The alignment becomes difficult in the "twilight zone" of less than 30% sequence identity. For the difficult cases, multiple structure and sequence information will be more reliable than the single one.[74,75] CLUSTAL[76] is one of the most widely used programs for multiple sequence alignment which is based on dynamic programming techniques. For the difficult cases where the sequence similarity is lower than 30% or with large number of gaps, multiple templates will reduce the alignment errors and provide more reliable information.[77,78]

With regard to model construction, one method called satisfaction of spatial restraints[79] is employed in the most widely used software Modeller.[80,81] The 3-D models generated by Modeller are obtained by optimally satisfying spatial restraints under the guide of alignment to the templates. The restraints consist of homology-derived spatial restraints[82] (main-chain, side-chain distances and dihedral angles), non-homology-derived restraints[83-85] (dihedral angels not aligned with template residues and distances between all nonbonded atom pairs) and molecular mechanics restraints[86] that enforce proper stereochemistry. These restraints are expressed as probability density functions (pdf). Working models are obtained by optimization of pdfs to minimize/reduce the number of restraint violations. An advantage of modeling by satisfaction of spatial restraints is that the restraints can be obtained from experimental data (NMR, cross-

linking, fluorescence spectroscopy experiments, etc) which helps enforce that the models are consistent with available experimental data.

The final step is model evaluation. The program Modeller incorporates a scoring function called Discrete Optimized Protein Energy (DOPE)[87] to gauge structure reliability which was derived using 1472 crystal structures. Another program called PROCHECK[88,89] can be used to evaluate the models through checking Ramachandran plots, torsion angles, and main-chain and side-chain properties, and residue properties. Once a model is deemed reasonable, molecular modeling using the tools of molecular mechanics can be used as described in the next section.

1.4 Molecular Mechanics

Many biological problems are too large to employ quantum mechanics, therefore forcefield methods (also known as molecular mechanics) are employed. Molecular mechanics uses classical mechanics to compute the energy of a molecule as a function of its conformation and chemical makeup given a force field equation and a set of force field parameters

It is important to notice that a force field is specified not only by its functional form but also its parameters which may be derived based on both experimental X-ray, thermodynamic, spectroscopic[90-92] and quantum-mechanical data (high level ab initio calculations). Figure 1.6 schematically shows the component of a standard molecular mechanics equations consisting of bond stretching, angle bending and torsional terms and non-bonded interactions that include electrostatic and van der Waals interactions. The total energy is defined as in eq (1.1). The bonded terms are shown in eq (1.2) and non-bonded terms[92] are shown in eq (1.3).
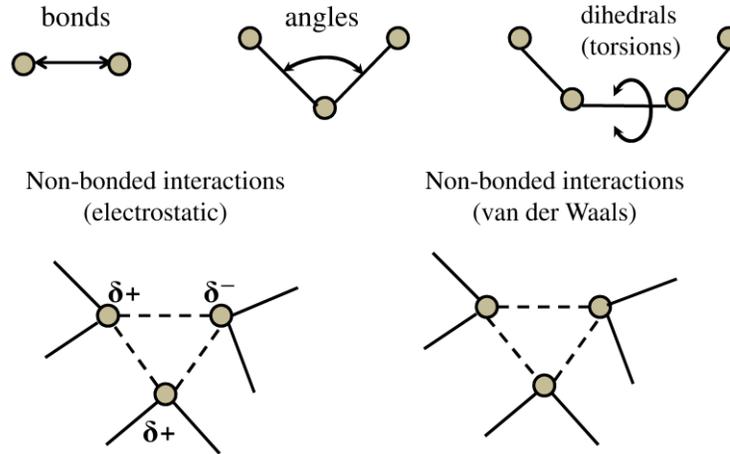
Figure 1.6. Schematic representation of the key contributors to a molecular mechanics force field estimation.

$$E_{total} = E_{bonded} + E_{nonbonded} \qquad (1.1)$$

$$E_{bonded} = \sum_{i \in bonds} k_r (r_i - r_0)^2 + \sum_{i \in angles} k_\theta (\theta_i - \theta_0)^2 + \sum_{i \in dihed} k_\chi (1 + \cos(n\chi_i - \delta)) \qquad (1.2)$$

$$E_{nonbonded} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j e^2}{D r_{ij}} \right) \qquad (1.3)$$

For the bonded interactions in eq (1.2), bond r is defined as the distance between the nuclei of two bonded atoms in a molecule (e.g. C-N, C-O, C-H, etc.); angle θ is calculated between three connected atoms (e.g. C-C-C, C-O-C, C-C-H, etc.); dihedral angle (torsion) χ is estimated between three consecutive bonds (e.g. C-C-C-C, C-O-C-N, H-C-C-H, etc). Unique parameters are assigned to each term based on environment including atom types. Hooke's law defines the reference values for bonds (r_0), and angles, (θ_0). The first term in eq (1.2) models the forces of chemical bonds between pairs of bonded atoms with the k_r parameter determining the stiffness of the spring. This equation estimates the energy increase with vibration from the

equilibrium bond length by using a harmonic potential. The second term of the same functional

form contains a spring constant used to compute the energy associated with deviation about the

equilibrium bond angle (θ). The third term describes the potential derived from dihedral angle

(χ) energy changes due to bond rotation. The k_χ parameter defines the vertical range of the

curve, the n parameter determines its periodic time length, and δ specifies the offset.

The non-bonded terms (through-space interactions) estimate energies in a pair-wise

manner being summed over all interacting non-bonded atoms i and j. In eq (1.3), energy is

computed using a Lennard-Jones potential for the van der Waals interactions and a Coulomb

potential term for the electrostatic interactions. The Lennard-Jones potential employs σ

(distance) and ε (well depth) parameters to estimate energy as a function of interatomic distance

containing a repulsive (r^{-12}) and attraction (r^{-6}) regimes that approach zero as the distance

between particles goes to infinity. In eq (1.3), q_i, q_j represents the partial charges for atom i and

j; r_{ij} defines the distance between atom i and j. For the AMBER function form used in this

work,[92] 1-2 (atoms one bond away), and 1-3 (atoms two bonds away) VDW and Coul

interactions are neglected while 1-4 interactions are scaled.

1.5 Molecular Docking (Sampling and Scoring)

Molecular docking serves as the most widely used computational tool for screening large

ligand databases to select promising drug-like leads by predicting their binding geometry[93]

(pose) using a scoring function. DOCK was the first software that aims at identifying the

binding mode between small organic molecules and a macromolecule. In terms of scoring,

DOCK is based on the AMBER molecular mechanics force field[95,102] and ligands are ranked

based on a very crude and approximate binding energy consisting of only the electrostatic and

van der Waals terms. Despite the approximations, this simple function is remarkably good for pose identification. In general, DOCK[94] calculations consist of two key parts: a search algorithm used to sample poses and a scoring function for ranking these poses. Docking is usually performed with the receptors kept rigid, although flexible receptor [95-97] docking is an active area of interest. For the ligands, torsions about single bonds are usually sampled. Docking programs are evaluated primarily based on their ability to accurately reproduce known crystallographic poses across large databases such as SB2010 (N=780 systems) as well as enrich for known active ligands from a large number of decoy compounds using databases such as DUD.[98-100]

Sampling. With regard to pose orientation, rigid ligand docking involves sampling the six degrees of rotational and translational freedom while flexible ligand docking also includes the sampling of each flexible torsion. DOCK uses the incremental construction algorithm, termed anchor-and-grow[47] for ligand sampling. Firstly, all ligand rotatable bonds are identified and the largest rigid segment (termed anchor) is isolated. The anchor is then rigidly placed in the active site, and multiple poses of the anchor are generated, minimized, and clustered. Next, each segment in turn is added to the anchor and at each stage of growth, poses are energy minimized, ranked, clustered, and pruned to maintain a diverse number of conformations with favorable energies. Additional segments are sequentially added to the growing molecule until all layers have been added. A successful docking experiment will generate the correct ligand conformation and correctly orient the conformation into the binding pocket.

Scoring. In order to identify the most promising leads, poses are evaluated and ranked based on a scoring function. The scoring function in DOCK uses the AMBER molecular force field consisting of only the intermolecular ver der Waals (VDW) and electrostatic energies shown in eq (1.4). Additionally, to prevent ligand internal clashes, DOCK6.4 includes the

repulsive van der Waals term as part of the internal energy.[99] Here, *D* represents the dielectric

constant of the solvent surrounding the receptor and the ligand and in DOCK this effect is

typically modeled as a distance-dependent dielectric constant ($D = 4r$).[101] Importantly, the

potential energy derived from the receptor can be calculated beforehand and stored on a grid[102]

to speed up the calculations. Work in this thesis also employs a new scoring function recently

incorporated into DOCK6 termed footprint similarity score (FPS)[100] which has been shown to be

useful for pose identification and ligand ranking when a known reference compound is available.

FPS utilizes residue-based decomposition of the electrostatic, VDW, and in some cases hydrogen

bonding interactions to compute the overlap between a reference ligand (usually an approved

drug or ligand in the pose seen a crystal structure) and a query compound. In this way, the

ligands that make similar interaction signature to a reference can be identified. References used

in the present work employed the known IGF-IR inhibitor called PQIP.

$$E_{nonbonded} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j e^2}{D r_{ij}} \right)$$ **(1.4)**

1.6 Molecular Dynamics

With a suitable structure and a molecular mechanics force field assigned, the technique of

molecular dynamics (MD), first developed in late 1970s, can be used to explore the dynamics of

the system. The potential energy determined by the force fields can be used in MD to determine

the forces through solving Newton's equation of motion (F=ma). A typical molecular dynamics

include 4 steps: (1) An initial model is prepared from the crystal structure, NMR structure or

homology modeling; (2) The forces acting on each atom in the system are calculated based on

Newton second law; (3) The positions of the atoms are determined using Newton's law. (4) The

positions and velocities of the atoms are updated at each timestep.

The way to implement molecular dynamics is to determine the position of all atoms as a function of time. For a given atom i in the system, Newton's equation of motion can be expressed as $\dfrac{\vec{F_i(x)}}{m_i} = a_i = \dfrac{d^2 x}{dt^2}$. This equation defines the motion of one atom with mass m_i under the force F_i along x axis. The position of the atom i in Cartesian space includes coordinates in x, y, and z directions (x_i, y_i, z_i) and can be described as $\vec{r_i} = (x_i, y_i, z_i)$. For a system with N atoms, $\vec{r} = (\vec{r_1}, \vec{r_2}, \cdots, \vec{r_N})$, $\vec{F} = (\vec{F_1}, \vec{F_2}, ..., \vec{F_N})$, and $\vec{a} = (\vec{a_1}, \vec{a_2}, ..., \vec{a_N})$, the motions on all the atoms can be calculated in the same way simultaneously. Using eq (1.5) and (1.6) we can obtain the forces and accelerations (data from y and z coordinates can be calculated in the same way):

$$v_{x,i}(t) = \int_0^t a_{x,i}(t')\, dt' \tag{1.5}$$

$$x_i(t) = \int_0^t v_{x,i}(t')\, dt' \tag{1.6}$$

In a real world of intermolecular interactions, the forces on each atom will vary corresponding to changing position of the atom itself or the interaction partners. Since it is hard to analytically solve the motions of all the particles that are coupled together, a finite difference method is used to break down the integration into each stage by using a fixed time Δt. Determined by the force, the accelerations, the positions, and velocities can be calculated at time t, t+Δ, t+2Δ and so on which can be approximated as a Taylor series expansions. The leap-frog algorithm which is derived from Verlet algorithm is used in the AMBER software package:

$$v(t + \Delta t) = r(t) + \Delta t v(t + \frac{1}{2}\Delta t) \tag{1.7}$$

$$v(t + \frac{1}{2}\Delta t) = v(t - \frac{1}{2}\Delta t) + \Delta t a(t) \tag{1.8}$$

17

Typically the simulation time is advanced every 1-2 fs which is equivalent to the vibration period of a bond. Due to high computing demand, molecular simulations are usually performed on computer clusters or supercomputers using hundreds of processors in parallel. The most widely used simulation software package includes AMBER,[103] CHARMM,[104] GROMOS[105] and NAMD[106] employ Message Passing Interface (MPI) which allow a calculation to be spread out among hundreds of processors.

A single static conformation from NMR, crystallography, or homology modeling provides little information about dynamic processes of molecular recognition and drug binding. Molecular dynamics simulations can overcome such limitations by providing ensembles of receptor structures. However, MD simulations still face two challenges, and it may be still limited by two primary challenges,[107] (i) the inaccuracy of force fields and (ii) sampling challenges than may require a prohibitive computational time. In terms of the force fields problem, many attempts have been made, including introducing quantum-mechanical calculations into classic MD force fields,[108] and development of polarizable force fields.[109] In terms of sampling, accelerated MD,[110] umbrella sampling, and more recently development that employ accelerated hardware graphics-processing-units (GPUs)[110,111] are approaches that can be used to explore additional phase space that may be intractable with standard MD.

1.7 Free Energy Calculation

Although challenging, the prediction of free energies of binding (ΔG_{bind}) are becoming more and more routine as codes become better and molecular mechanic force fields have improved. Ligand binding takes place in a solvent, therefore hydration/solvation plays an exceptionally important role in binding energy calculation.[45,55] By performing solvated MD

time-averaged energies can be obtained that include such effects. The next section describes two methods, MM-GBSA and thermodynamic integration, used in this thesis to estimate ΔG_{bind}.

1.7.1 MM-GBSA

One of the most widely used methods for calculating free energy of binding that includes the effects of solvent in the calculation is the Molecular Mechanics Generalized Born Solvent Accessible Surface Area (MM-GBSA) Method.[57,112] Although it is an approximate approach compared to other methods, it can provide reasonable accuracy within a reasonable timeframe and is easily implemented. Figure 1.7 shows a schematic representation of a binding event and an accompanying thermodynamic cycle used to estimate the binding energy.

Here, the calculated free energy of binding is estimated as the sum of nonbonded gas-phase ($\Delta G_{gas} = \Delta E_{vdw} + \Delta E_{coul}$) interactions and the change in hydration free energy ($\Delta\Delta G_{hyd} = \Delta\Delta G_{polar} + \Delta\Delta G_{nonpolar}$) for ligand complexation. The gas phase energy is computed using the standard MM force field and consists of nonbonded VDW and Coulombic interaction terms. The hydration free energy consists of both polar and nonpolar terms. The polar term ($\Delta\Delta G_{polar}$) due to the electrostatic interaction between the solute and solvent can be computed using Generalized Born (GB) model[113] described in eq (1.10). Here, q_i and q_j are the partial charges associated with each atom and α_i represents the Born radii for which several functional forms have been proposed. The current work employs the form Hawkins-Cramer-Truhlar (HCT)[112] form implemented in the program AMBER. The nonpolar term consists of van der Waals interaction between the solute and solvent, as well as the free energy required to form a cavity in the solvent which can be estimated with solvent accessible surface area (SASA) using the linear equation (1.11) $G_{nonpolar} = \gamma SASA + \beta$ with $\gamma = 0.00542$ kcal/mol Å^2 and $\beta = 0.92$ kcal/mol).[114]
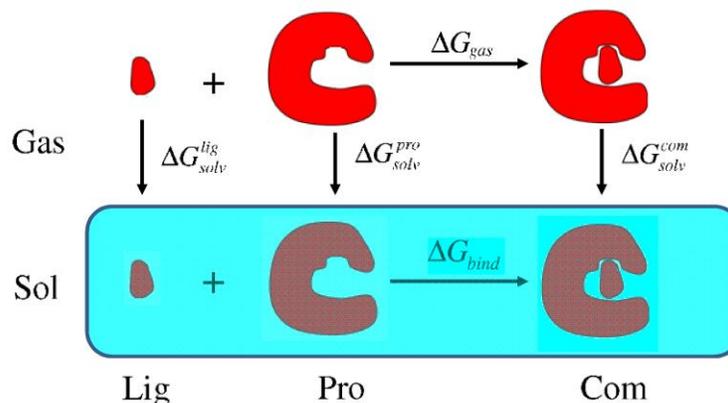
Figure 1.7. Schematic representation of the thermodynamic cycle used to calculate free energies of binding (ΔGb calcd).

$$\Delta G_b \text{ exptl} \approx \Delta G_b \text{ calcd} = \Delta E_{gas} + \Delta \Delta G_{hydration} = (\Delta E_{vdw} + \Delta E_{coul}) + (\Delta \Delta G_{polar} + \Delta \Delta G_{nonpolar}) \qquad \textbf{(1.9)}$$

$$G_{polar} = -166(1 - \frac{1}{\varepsilon})\sum_{i=1}^{N}\sum_{j=1, j \neq i}^{N}\frac{q_i q_j}{f_{GB}} - 166(1 - \frac{1}{\varepsilon})\sum_{i=1}^{N}\frac{q_i^2}{\alpha_i} \qquad \textbf{(1.10)}$$

$$f_{GB} = \left\{ r_{ij}^2 + \alpha_{ij}^2 \exp[-r_{ij}^2/(4\alpha_{ij}^2)] \right\}^{0.5} \qquad \textbf{(1.11)}$$

It is important to note that GB is a continuum model and instead of including every single solvent molecule in the system the solvent acts as "bulk medium" which, in favorable instances, provides a sufficiently accurate description of the screening effects of water. And, although the GB models are thought to capture bulk entropic effects associated with the "hydrophobic effect", changes in solute entropy are not explicitly accounted for. Procedures including normal mode or quasi-harmonic analysis[115] have been employed in an attempt to include solute entropic changes however such calculations are in general quite be difficult to perform, noisy, and have poor convergence. For most studies (including the work done in this thesis), in which energy changes are being computed/compared for a related series of ligands (or receptors), solute entropy changes may reasonably be assumed to cancel and extra terms are ignored.

20

The MM-GBSA simulations in this work involved explicit solvent MD dynamics simulations used to compute absolute binding energies (ΔG_{bind}).   A single explicit solvent simulation for the complex was performed and three gas phase trajectories for the complex, the receptor, and the ligand were obtained by stripping the water from the fully solvated trajectories. The final absolute binding energy can be computed based on $\Delta G_{bind} = \Delta G_{com} - (\Delta G_{rec} + \Delta G_{lig})$. using the simulation output (e.g. $\Delta G_{com} = \Delta E_{vdw} + \Delta E_{coul} + \Delta G_{polar} + \Delta G_{nonpolar}$) for each species of the complex, the receptor, and the ligand.

1.7.2 Thermodynamic Integration

Distinctly different from the MM-GBSA method described above, an important category of free energy calculation methods include "computational alchemy" approaches such as thermodynamic integration (TI) which involves molecular transformations along alchemical (or non-physical) paths.   A potential advantage over MM-GBSA is that all entropic effects, including changes in solute entropy, are included. By computing the free energy change between ligands in both the bound and unbound states, a "relative" free energy of binding ($\Delta\Delta G_b$ calcd) can be obtained and compared with experiment ($\Delta\Delta G_b$ exptl). Figure 1.8 shows a representative thermodynamic cycle for a TI calculation involving bound and unbound transformations of ligand A into B during an explicit solvent MD simulation. The relationship between the two computational and two experimental legs of the thermodynamic cycle is given by eq (1.13).
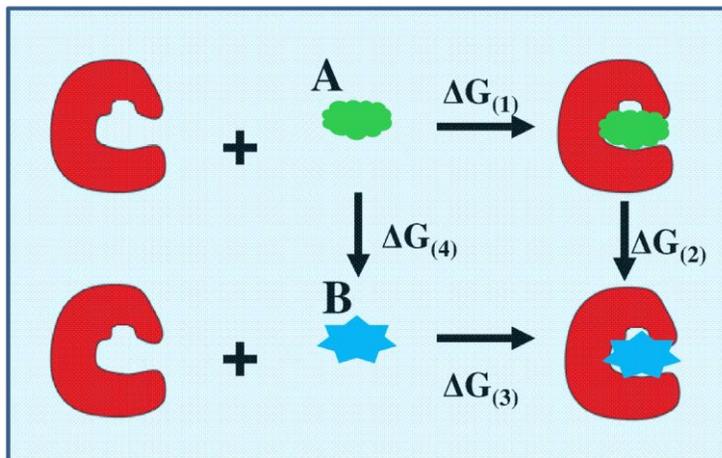
Figure 1.8. Schematic representation of the TI thermodynamic cycle used to compare computed relative free energies of binding ($\Delta\Delta G_b$ calcd = $\Delta G_2 - \Delta G_4$) with experimental relative free energies of binding ($\Delta\Delta G_b$ exptl = $\Delta G_3 - \Delta G_1$).

$$\Delta\Delta G_b \text{ calcd} = \Delta G_2 - \Delta G_4 = \Delta\Delta G_b \text{ exptl} = \Delta G_3 - \Delta G_1 \qquad \textbf{(1.12)}$$

It is important to note that to promote convergence, the transformation from A to B is done through a coupling parameter λ which varies from $\lambda = 0$ (initial state) to $\lambda = 1$ (final state) with the mixing of states given by eq (1.13). The total free energy difference can be estimated by the Hamiltonian as a function of λ in eq (1.14) where angular brackets denote a Boltzmann-weighted ensemble average, using V(λ) as the potential. In practice, the free energy of binding can be obtained by performing simulation at discrete λ values and carrying out numerical integration. The integration can be approximated by using trapezoidal rule instead of being solved analytically.

$$V(\lambda) = \lambda V_1 + [1 - \lambda]V_0 \qquad \textbf{(1.13)}$$

$$\Delta G = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \qquad \textbf{(1.14)}$$

In the most recent version of the AMBER program, TI calculations can be performed using softcore potentials.[116] This procedure is much more straightforward than developing a perturbation that requires use of dummy atoms, and appearing and disappearing functional groups (e.g. groups in state A but not state B) can be present at the same time. In principle this allows for changes involving both Lennard Jones and Coulombic energies to be computed at the same time. The Lennard Jones for softcore potentials calculations are shown in eq (1.15) and (1.16) which smoothly switch on or off non-bonded interactions unique to either state (A or B). Here, ε and σ are standard LJ parameters, r is the interatomic distance and α is an adjustable constant. Analogous equations for the Coulomb softcore potentials are shown in eq (1.17) and eq (1.18). Here, q_i and q_j are the atomic partial charges, r is the interatomic distance, and β is an adjustable constant.

$$V_{vdw,A} = 4\varepsilon(1-\lambda)\left(\frac{1}{\left[\alpha\lambda + (\frac{r}{\sigma})^6\right]^2} - \frac{1}{[\alpha\lambda + (\frac{r}{\sigma})^6]} \right) \tag{1.15}$$

$$V_{vdw,B} = 4\varepsilon\lambda\left(\frac{1}{\left[\alpha(1-\lambda) + (\frac{r}{\sigma})^6\right]^2} - \frac{1}{[\alpha(1-\lambda) + (\frac{r}{\sigma})^6]} \right) \tag{1.16}$$

$$V_{Coulomb,A} = (1-\lambda)\frac{q_i q_j}{4\pi\varepsilon_0\sqrt{\beta\lambda + r^2}} \tag{1.17}$$

$$V_{Coulomb,B} = \lambda\frac{q_i q_j}{4\pi\varepsilon_0\sqrt{\beta(1-\lambda) + r^2}} \tag{1.18}$$

The TI simulations in this work involved explicit solvent MD dynamics simulations in which one functional group was perturbed to another (i.e. Ethyl to Methyl) for several series of structurally related ligands. The calculations employed 19 windows (λ from 0.05 to 0.95 with

the interval of 0.05) and were done for both the bound and unbound states. The relative binding free energy is estimated by the trapezoid method through integration of DV/Dλ versus λ across all windows.

1.8 Research Projects

This dissertation describes several research projects involving application of computational techniques to simulate kinase inhibitors implicated in for breast cancer. The project in Chapter 2 describes simulations of the FDA-approved drug lapatinib binding with highly homologous ErbB family member kinases EGFR, HER2 and ErbB4. The aim is to understand lapatinib binding specificity and resistance. Chapter 3 describes use of umbrella sampling to explore the energetic change between two lapatinib conformations seen in different crystal structures and to determine which the most favorable conformation for lapatinib with ErbB4. Chapter 4 describes ongoing work to estimate free energy of binding using the thermodynamic integration method to specify IGF-IR binding with a series of related analogs. Chapter 5 describes preliminary virtual screening for residues targeting active, semi-active and inactive forms of IGF-IR with the goal of identifying dual kinase inhibitors for subsequent experimental evaluation.

Chapter 2. A Water-based Mechanism of Specificity and Resistance for Lapatinib with ErbB Family Kinases

Author contributions. YH and RCR designed research; YH performed research, analyzed data, and wrote initial draft; YH and RCR wrote the paper.

2.1 Abstract

The dual kinase inhibitor lapatinib has high affinity for EGFR and HER2, but weak affinity for ErbB4, although the factors driving specificity for these highly homologous members of the ErbB family of receptor tyrosine kinases are not well understood. In this report, homology modeling, molecular dynamics simulations, and free energy calculations are employed with the goal of uncovering the energetic and structural molecular basis of lapatinib specificity and resistance. The results reveal a distinct network of three binding site water molecules which yield strikingly similar hydration patterns for EGFR and HER2 in contrast to ErbB4 which shows a different pattern with reduced occupancy at one of the positions. The primary cause was traced to a single amino acid change in the binding site (EGFR position 775), involving a swap from C or S (EGFR, HER2) to V (ErbB4) for which the side chain is bulkier, hydrophobic, and lacks H-bond capability with water. Notably, excellent quantitative agreement with experimental activities are obtained across the series EGFR > HER2 > ErbB4 when key waters

are included in the calculations. Quantitatively, Coulombic interactions and H-bond counts between network waters and species involved in the network are less favorable in ErbB4 by ca. 40% relative to EGFR or HER2. Additional simulations with clinically relevant EGFR (C775F, T854A, and T790M) and HER2 (T790I) mutants demonstrate that resistance can also be understood in terms of changes that occur in the binding site water network. Overall, the results of this study have yielded a physically reasonable water-based mechanism for describing lapatinib specificity and resistance.

2.2 Introduction

Excluding skin cancer, breast cancer is the most common type of cancer for women in the United States and one in eight women are expected to develop breast cancer in their lifetime.[1] Members of the highly homologous ErbB family of receptor tyrosine kinases (EGFR, HER2, HER3, and ErbB4) play key roles in breast cancer;[17] and overexpression of EGFR[117] and HER2[118] are important prognostic markers. The design of agents targeting EGFR or HER2 alone, or in combination, has been a major therapeutic focus.[17,19] Successful examples include the HER2 neutralizing antibody trastuzumab (Herceptin) and the small molecule inhibitor lapatinib (Tykerb) (Table 2.1), which targets EGFR and HER2.[119] Unfortunately, less than one third of patients with HER2 amplification benefit from trastuzumab.[19] Importantly, lapatinib has shown activity against trastuzumab-treated breast cancer cell lines[120] and in patients previously treated with trastuzumab.[121] However, acquired resistance to lapatinib involving alternative signaling pathways[122,123,124-126] or potentially deleterious point mutations[127,128] will likely hamper long-term clinical utility. Uncovering mechanisms associated with sensitivity and resistance to ErbB family inhibitors is an important long-term therapeutic goal. The focus of the present work is

improved understanding of the atomic and molecular-level details driving lapatinib dual specificity for EGFR and HER2 (strong) vs ErbB4 (weak)[129,130] (Table 2.1) and what leads to a loss of binding with specific point mutations.[127,128]

Table 2.1. Experimental activities for lapatinib with ErbB family members.

	Activity Type	EGFR	HER2	ErbB4
	IC_{50}[a]	10.8±0.53	9.2±0.75	367.0±4.2
	K_i[b]	3.0±0.2	13±1	347.0±16
	ΔG_{bind}[c]	-11.62	-10.75	-8.81
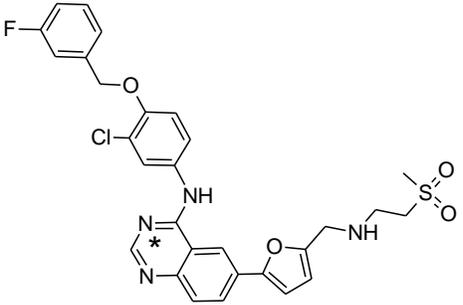
[a]IC_{50} values (nM) from Rusnak et al [129] [b]EGFR and HER2 (K_i) and Erb4 (calculated K_i) values (nM) from Wood et al.[130] [c]ΔG_{bind} exptl ≈ RT ln (K_i) at 298.15 K. Note that N* indicates a key ligand atom involved in water mediated H-bonding (see text for discussion).

Members of the ErbB family share a similar overall structural architecture comprising: (i) extracellular ligand binding domain, (ii) transmembrane domain, (iii) intracellular juxtamembrane domain, (iv) intracellular tyrosine kinase domain, and (v) C-terminal regulatory region where phosphorylation occurs.[131] Interestingly, no exogenous ligands have been identified which bind to HER2 and HER3 lacking kinase activity.[17,19] Figure 2.1 highlights the similarities in sequence (EGFR numbering), including the lapatinib binding site, across the kinase domain of the proteins studied in this report (EGFR, HER2, and ErbB4). Ligand binding to the extracellular domain of EGFR, HER3, or ErbB4 can promote homo or heterodimerization with another ErbB family member resulting in activation of the intracellular kinase domain.[17,19] Activation (inactive → active form) is mediated by a structural reorganization which primarily involves conformational changes in the C-helix and activation loop regions.[131] Active homo/heterodimeric complexes can, in turn, cause activation of a series of downstream signaling

cascades.   Aberrant signal transduction can promote cell proliferation, differentiation, and

eventually lead to malignant transformation.[17,19]

```
                           ▼        • • •    •    •                     • • •    •
EGFR        702  ALLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILD
HER2        702  AQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILD
ErbB4       702  AQLRILKETELKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNETTGPKANVEFMD
consensus   702  *..*******....*************.*.*.**.********.*.*.*.****.*..*

                          • • • • •              • • • •    •• •• •    •
EGFR        762  EAYVMASVDNPHVCRLLGICLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQI
HER2        762  EAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQDLLNWCMQI
ErbB4       762  EALIMASMDHPHLVRLLGVCLSPTIQLVTQLMPHGCLLEYVHEHKDNIGSQLLLNWCVQI
consensus   762  **..**...  *.. ****.**...*.** *****.****...*.*.....** *****.**

                                •     •              • • • • •
EGFR        822  AKGMNYLEDRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWM
HER2        822  AKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWM
ErbB4       822  AKGMMYLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLEGDEKEYNADGGKMPIKWM
consensus   822  **** ***..************.* .********** .**...*.**.*.*.***.*****

                                                 •
EGFR        882  ALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTI
HER2        882  ALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTI
ErbB4       882  ALECIHYRKFTHQSDVWSYGVTIWELMTFGGKPYDGIPTREIPDLLEKGERLPQPPICTI
consensus   882  ***.*.  *..***********.*******.******..**...**************

                                    ▼                    HOMOLOGY
EGFR        942  DVYMIMVKCWMIDADSRPKFRELIIEFSKMARDPQRYLVIQ      EGFR-HER2    (81%)
HER2        942  DVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQRFVVIQ      EGFR-ErbB4   (78%)
ErbB4       942  DVYMVMVKCWMIDADSRPKFKELAAEFSRMARDPQRYLVIQ      HER2-ErbB4   (77%)
consensus   942  ****.*********...**.*.**. ***.*******..***
```

● Binding site residues   ● Binding site sequence changes   ▼ kinase domain range

Figure 2.1.   Sequence comparisons between EGFR, HER2, and ErbB4 kinase domains.   Numbering corresponds to EGFR (PDB entry 1XKK).

From a therapeutic standpoint, small molecule kinase inhibitors have been developed

which preferentially target either the inactive (lapatanib, Table 2.1)[29] or active (erlotinib,[132]

gefitinib[133]) kinase forms.   These compounds bind competitively at the ATP-binding site,

blocking phosphorylation and subsequent downstream signaling. Lapatinib is approved to treat

HER2-overexpressing breast cancer while erlotinib and gefitinib are approved to treat non-small

cell lung cancer.[119]   Unlike erlotinib and gefitinib which primarily target EGFR, lapatinib is a

dual kinase inhibitor[134] targeting both EGFR and HER2. Interestingly, despite the high sequence

homology (Figure 2.1), lapatinib only weakly inhibits ErbB4 (Table 2.1).   This fact may be

28

clinically relevant as it has been reported that ErbB4 may lead to a potentially beneficial antiproliferative response in human breast cancer cells[135] and overexpression of ErbB4 correlates with reduced recurrence of breast cancer.[136] Prior studies, including crystallographic reports of lapatinib in complex with inactive-form EGFR[130] or ErbB4,[137] have not offered clear reasons why lapatinib preferentially binds EGFR and HER2 but not ErbB4.

There is growing concern that drug resistance, as has been observed with the related kinase inhibitors erlotinib and gefitinib, may occur upon long term treatment with lapatinib.[127,128,138] Known EGFR mutations associated with acquired resistance to erlotinib or gefitnib include T790M and T854A.[139,140] Here, the so called gate keeper mutation at position 790 (EGFR or HER2) is similar to the T315I mutation arising from treatment with imatinib (Gleevec) in the BCR-ABL kinase system.[141] Avizienyte et al[127] and Trowe et al[128] have reported that several ErbB family mutations, including EGFR (C775F, T854A, T790M), and HER2 (T790I) also negatively affect lapatinib. Proposed resistance mechanisms include: (i) a steric clash for EGFR C775F,[127] EGFR T790M,[138] or HER2 T790I,[128] (ii) increased affinity for ATP for EGFR T790M,[142] and (iii) loss of contact for EGFR T854A.[140] Improved understanding of what contributes to lapatinib binding at the structural level will be important to explain the impact of observed mutations as well as the design of next generation EGFR/HER2 inhibitors with improved resistance profiles.

Examples of prior computational work addressing ligand binding in ErbB systems have employed homology and molecular modeling,[143-145] comparative molecular field analysis,[146,147] docking/virtual screening,[146-148] molecular dynamics,[123,143,144,146,147,149,150] and Monte Carlo simulations.[151] Surprisingly, few studies have addressed binding across multiple members of the ErbB family. Exceptions include Kamath et al[144] who explored ATP selectivity between EGFR

and HER2 using molecular dynamics and Scaltriti et al[145] who explored binding of lapatinib with EGFR, HER2 and HER3 using homology modeling in combination with energy minimization. Prior calculations addressing drug resistance include reports by Liu et al[149] for gefitinib in complex with EGFR (T790M, L858R&T790M), and Balius et al[123] for erlotinib, gefitinib, and AEE788 with EGFR (wildtype, L858R, L858R&T790M). Focusing on the latter study, Balius et al[123] compellingly showed that drug resistance in EGFR, in contrast to proposed mechanisms primarily involving a steric clash[139,152] or altered affinity for ATP,[142] more likely involve disruption of favorable interactions including an important water-mediated H-bond network. Free energy calculations for inhibitors with EGFR, reported by Michel et al,[151] similarly revealed a local water network which mediates binding. In general, hydration is known to be an important factor in drug design.[151,153-156] In particular, specific waters can help to mediate protein-ligand binding.[154] Examples of other FDA-approved drugs[119] which make water-mediated interactions include erlotinib,[132] zanamivir,[157] and nevirapine.[158] For the lapatinib simulations presented here, bridging waters also appear to play a key role in modulating both specificity and resistance among ErbB family tyrosine kinases.

Our long-term goal is development of improved anti-cancer small molecule drugs. Goals of the specific study are threefold: (i) construct robust all-atom computational models to quantify lapatinib binding with inactive forms of wildtype EGFR, HER2 and ErbB4 and deleterious mutations including EGFR C775F, EGFR T854A, EGFR T790M, and HER2 T790I, (ii) determine how receptor sequence and associated structural changes lead to variation in computed lapatinib activities for comparison with experimental results, and (iii) elucidate the primary factors controlling lapatinib specificity and resistance. The design of inhibitors with tailored

30

selectivity and resistance profiles will ultimately be enabled through use of well-tested models for quantifying and predicting molecular recognition at the atomic level.

2.3 Methods

2.3.1 Model Construction

Simulation-ready models for three different ErbB systems (EGFR, HER2, ErbB4) were constructed using one of two crystallographic structures termed here template #1 = 1XKK containing EGFR with lapatinib,[130] and template #2 = 2R4B containing ErbB4 with a lapatinib-like covalent inhibitor.[159] 1XKK was chosen as a template because it was the only inactive structure available for EGFR. 2R4B was chosen as a template for inactive ErbB4 because there was only a short section missing (6 residues) in comparison to an alternative structure 3BBT[137] which was missing 14 residues in the activation loop region. No inactive structures were available for HER2 although an active form structure was recently published.[160] Structural changes in the binding site reveal lapatinib is compatible only with inactive conformations. The program Modeller9v6[79] was employed to construct homology models using both inactive templates. Figure 2.2 outlines the overall workflow schematically.
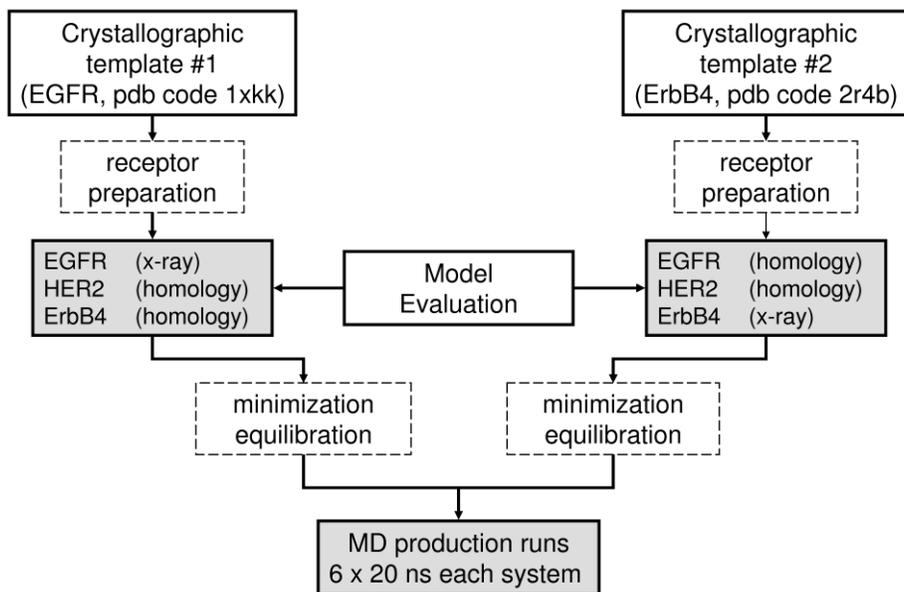
Figure 2.2. Overall workflow schematic showing model construction for EGFR, HER2, and ErbB4 using two crystallographic templates.

For lapatinib, the ligand conformation observed in template 1XKK was used as the initial set of coordinates for all simulations. Template 2R4B contained a lapatinib-like covalent inhibitor and there was incomplete ligand electron density in 3BBT. Interestingly, 3BBT contained an alternative rotamer amounting to a ca. 180° ring flip of the fluorophenyl moiety relative to 1XKK or 2R4B. However, average ligand B-factors in 3BBT were more than double (100.2) compared to 1XKK (41.8) which suggests substantial uncertainty. In addition, preliminary studies showed an energetic preference for the 1XKK conformation and some simulations starting from the 3BBT conformation flipped to 1XKK but never vice-versa. Additional studies to investigate the discrepancy observed in the two crystallographic structures are in progress.

Based on sequence alignments using the ClustalW program,[161] the kinase receptor domains share very high sequence identity: EGFR & HER2 (81%), EGFR & ErbB4 (78%), and

32

HER2 & ErbB4 (77%) thus homology models for the receptors can be constructed with high confidence. To assess the effects of homology modeling, two complete sets of three systems each were constructed (EGFR, HER2, ErbB4) using both templates (1XKK, 2R4B). Missing regions in 1XKK (residues 734-737, 750-753 and 868-875) were completed based on PDB entry 2GS7[131] containing inactive EGFR with ligand AMP-PNP while the missing region in 2R4B (residues 755-760) was completed using Modeller. Otherwise, conformations for identical residues in the alignments were kept the same as those in the relevant template.

For each complex, 10 homology models were generated and the model with the lowest Discrete Optimized Protein Energy (DOPE)[162] score was selected as the initial structure for molecular dynamic simulations. Models were evaluated using the program PROCHECK[88] which showed no residues in disallowed regions and ca. 90% of residues in the most favored regions. These metrics help confirm the structures are of overall good quality. In general, the six completed receptors, based on template 1XKK (EGFR, HER2, ErbB4) or template 2R4B (EGFR, HER2, ErbB4), were structurally similar except for some differences in the activation loop, the C-helix, and N-lobe regions (Figure 2.3 arrows).

To probe how point mutations would affect ligand binding, additional setups for mutant forms of EGFR (C775F, T854A, T790M), HER2 (T790I), and ErbB4 (V775C) were constructed through modification of the relevant wildtype structures using the program MOE.[163] Starting rotamers for mutant sidechains were chosen to overlap well with the wildtype sidechains subject to visual inspection to ensure there were no steric clashes. The same equilibration and production protocols described below for wildtype systems were employed for the mutants.
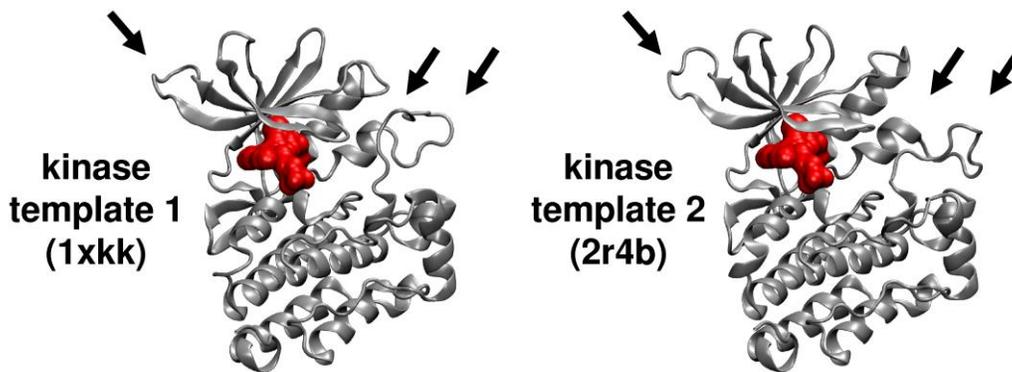
Figure 2.3. Homology modeling templates derived from crystallographic coordinates of inactive EGFR (entry 1XKK) or inactive ErbB4 (entry 2R4B). Arrows indicate regions with structural differences between the templates.

2.3.2 Simulation Setups and Molecular Dynamics Details.

Molecular dynamic simulations were performed with NAMD2.6[106] using input files prepared using the AMBER8[164] suite of programs which was used to assign force-field parameters and assemble and solvate each complex in a periodic box containing ca. 45000 waters (~ 80 Å3). No crystallographic waters were employed. Force-field parameters used were FF99SB[165] for the protein, GAFF[166] for the ligand, and TIP3P[167] for water. Ligand charges were obtained at the HF/6-31G* level of theory using the ChelpG[168] method as implemented in Gaussian98.[169] Based on environment, residue HIS 803 in the ErbB4 binding site was modeled as protonated. Lapatinib was modeled as neutral. Prior to the production runs, a nine-step equilibration procedure was used to relax the protein and solvent in a sequential way and consisted of three minimization steps (1000 steps of steepest decent) and six molecular dynamics runs (50 ps each). The production runs employed weak restraints only on short stretches of the N-termini (10 a.a.) and C-termini (14 a.a.). Slightly different equilibration protocols were used for crystallographic vs homology models. Table 2.2 summarizes the various equilibration steps.

34

Table 2.2. Equilibration protocol for crystallographic and homology models.

Step[a]	Crystallographic restraints[b] (weight)[c]	Homology restraints (weight)
1 min	wat O (5), pro all (5), lig all (5)	wat O (5), pro homo all (5), pro nonhomo hev (5), lig all (5)
2 min	pro all (5), lig all (5)	pro hom all (5), pro nonhomo back (5), pro nonhomo side (1), lig all (5)
3 min	pro hev (5), lig hev (5)	pro homo hev (5), pro nonhomo back (5), lig hev (5)
4 md	pro hev (5), lig hev (5)	pro homo hev (5), pro nonhomo back (5), lig hev (5)
5 md	pro hev (1), lig hev (1)	pro homo hev (1), pro nonhomo back (1), lig hev (1)
6 md	pro Cα (1), pro hev (0.1), lig hev (0.1)	pro Cα (1), pro homo hev (0.1), pro nonhomo back (0.1), lig hev (0.1)
7 md	pro Cα (1)	pro Cα (1)
8 md	pro Cα (0.5)	pro Cα (0.5)
9 md	pro Cα (0.1) 702-711, 969-982	pro Cα (0.1) 702-711, 969-982

[a]min = energy minimization, md = molecular dyanmics, [b]wat = water, pro = protein, lig = ligand, hev = heavy atoms, homo = residues homologous to template, nonhomo = residues non-homologous to template, back = backbone Cα, C, N, O, side = sidechain. [c]Restraint weights in parenthesis in kcal/mol Å^2.

Molecular dynamic simulations were performed in the NPT ensemble using Langevin dynamics[170] at a specified constant temperature of 298.15K and pressure of 1.01325 bar. MD equilibrations employed a 1 fs time step while production runs employed a 2 fs time step which required use of SHAKE.[171] Additional key input parameters include use of the Particle Mesh Ewald[172] to compute long range electrostatics (1.0 Å grid spacing) and a 12 Å direct space cutoff (10 Å smoothing switch) for non-bonded interactions. MD snapshots were saved every 1 ps during the productions runs for subsequent analysis. For each wildtype system six 20 ns production simulations were performed (different random seeds) while for each mutant system one 20 ns production simulation was performed.

2.3.3 Calculation of Binding Free Energies

Free energies of binding (ΔG_b calcd) were estimated using the well described MM-GBSA method[57,173] with AMBER which our laboratory has employed to successfully characterize a number of systems[174-177] including active-form EGFR.[123] For this work, free energies were estimated using a four term equation [ΔG_b calcd = ΔE_{vdw} + ΔE_{coul} + $\Delta\Delta G_{polar}$ + $\Delta\Delta G_{nonpolar}$] consisting of intermolecular van der Waals (ΔE_{vdw}) energy, intermolecular Coulombic (ΔE_{coul}) energy, changes in polar hydration energy ($\Delta\Delta G_{polar}$), and changes in nonpolar hydration energy ($\Delta\Delta G_{nonpolar}$). No additional entropic terms were included. Polar hydration energies were obtained using the GB model described by Onufriev et al[113] (igb = 5) with mbondi2 radii and interior and exterior dielectric constants of 1 and 78.5. Nonpolar hydration energies were obtained from solvent accessible surface areas through the relationship $\Delta G_{nonpolar} = \gamma$ SASA + β with γ = 0.00542 kcal/mol Å^2 and β = 0.92 kcal/mol.[114] The AMBER distribution file src/sander/mdread.f was modified to include radii of fluorine (1.50 Å) and chlorine (1.70 Å)

required by GBSA calculations for lapatinib. Free energies of binding were computed using three different protocols termed 0WAT, 3WAT, or 1WAT depending on the number of explicit waters retained from the fully solvated MD trajectories. 0WAT includes no waters and is the default MM-GBSA protocol, 3WAT includes the three closest waters to N* (labeled in Table 2.1), and 1WAT includes one water if within 3 Å of N*.

2.4 Results and Discussion

2.4.1 System Behavior and Convergence

2.4.1.1 Structural Stability (crystallographic vs homology starting coordinates).

To assess the behavior of the simulations and, in particular, to assess stability of the homology models, root-mean-square-deviations (RMSD) were examined as a function of time for each wildtype setup. To more accurately estimate binding free energies, six simulations using different random seeds were performed. Figure 2.4 plots results initiated from the crystallographic structure of EGFR (1XKK) and the two homology models of HER2 and ErbB4 constructed using 1XKK as the template. Here, light gray points show instantaneous RMSDs with colored lines representing the running average of the previous 100 snapshots for receptor Cα, C, N, and O backbone atoms (green line), lapatinib heavy atoms (red lines), or lapatinib quinazoline plus aniline rings defined as the scaffold (blue lines). RMSDs were computed after each MD production snapshot was fit to the initial model coordinates using receptor Cα backbone atoms as the match criteria. Importantly, for each of the six replicas across all systems, the receptor backbone RMSDs are within 2 Å which is an indication that both the crystallographic, as well as the homology models have structural integrity. While ligand total RMSDs, in some cases, show fluctuations > 2 Å, the ligand scaffold RMSDs always remain low

37

(1~1.5 Å) which suggest it is only the solvent-exposed methyl sulfone tail (see Table 2.1) which fluctuates significantly. Similar ligand RMSD results were obtained in an earlier study of inhibitors which target the active form of EGFR.[123]
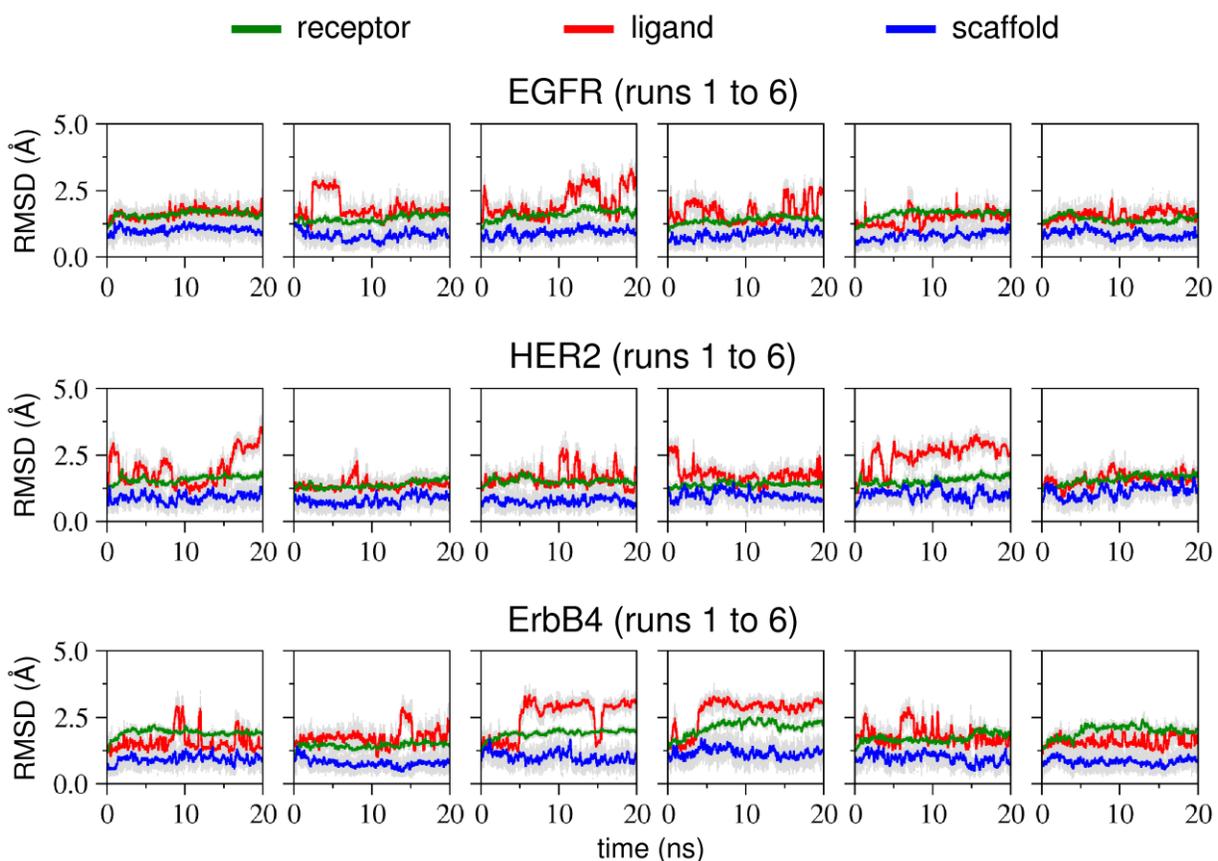
Figure 2.4. Root-mean-square-deviation (RMSD) vs time (ns) for lapatinib with EGFR (top), HER2 (middle), and ErbB4 (bottom) for six MD runs based on template 1XKK. RMSDs are shown for the receptor backbone (green line), ligand (red line), and ligand scaffold quinazoline plus aniline rings (blue line).

2.4.1.2 Binding Free Energy (crystallographic vs homology starting coordinates)

To help gauge convergence of the computed free energies of binding (ΔG_b calcd), autocorrelation functions (ACFs) and block-averaged standard errors of the mean (BASEM)

analysis were also performed.[178,179] For a given time series, as demonstrated in Figure 2.5 for the group of six simulations based on template 1XKK, ACFs provide a means to assess to what extent data are correlated while BASEM plots can be used to quantify statistical noise. The ΔG_b calcd data in Figure 2.5 were obtained using a three water protocol (3WAT), as described further below, however two other protocols (0WAT, 1WAT) lead to similarly good convergence. Focusing on the EGFR results, Figure 2.5 reveals the ACFs drop quickly and by ca. 1 ns lag time the data is largely uncorrelated (80-90%). And, the accompanying BASEM results show the expected monotonical increase as block averaging size increases (from 1 ps to 10 ns) which begins to reach a plateau indicating the error estimates are converging. The same general trend observed for EGFR is maintained for the other systems including the simulations based on the alternative template 2R4B or other protocols. Interestingly, one HER2 simulation yielded ACF and BASEM results which showed poor convergence behavior in comparison to others simulations. A simulation generated using an alternative random seed which was better behaved was substituted for the unconverged run. For comparison, Figure 2.5 quantifies ACF and BASEM values computed at four different lag time and block size values (1 ps, 100 ps, 1000 ps, and 2000 ps). Overall, by using a 1 ns block size, uncertainties in ΔG_b calcd values on the order of 0.5 kcal/mol may be considered to be a reasonable estimate. ACF data at this interval is relatively uncorrelated (~90%) and a sufficient number of blocks (N = 20) can be employed for BASEM calculations.
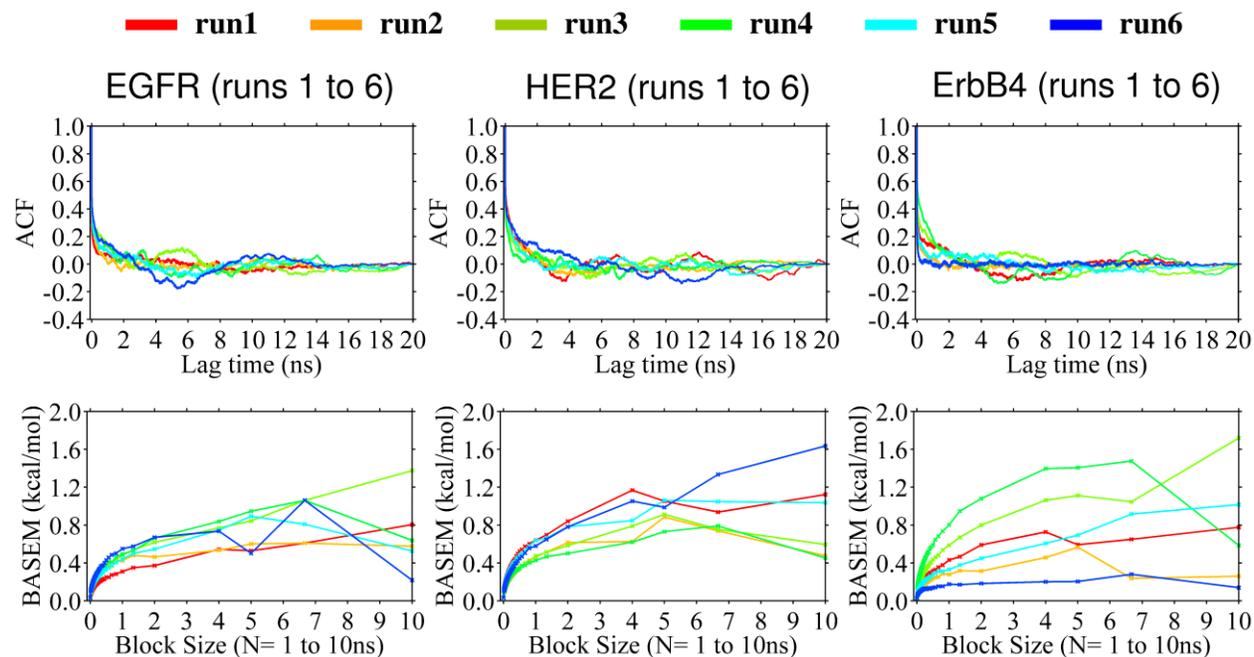
Figure 2.5. Top panels: Autocorrelation functions (ACFs) for free energies of binding of lapatinib with EGFR (top), HER2 (middle), and ErbB4 (bottom) computed from six independent 20 ns MD simulations each. Bottom panels: Block-averaged standard errors of the mean (BASEM) in kcal/mol as a function of block size. Free energies of binding calculated with 3 binding site waters (3WAT protocol).

2.4.1.3 Influence of Homology Model Templates on Interaction Energy Variability

To further explore how use of different templates might affect the energetic results, protein-ligand van der Waals and Coulombic interaction energies, on a per-residue basis, were plotted against each other as shown in Figure 2.6 for the range -7 to +3 kcal/mol. Notably, both homology templates yield very similar energetics. In all cases, per-residue van der Waals interactions are essentially identical (Figure 2.6, open circles) and only a few residues (D855, N842, R841, D837, D800, K745, G719, L718) show variability in their gas-phase Coulombic energies (Figure 2.6, red squares). It is interesting that many of the residues showing Coulombic differences are charged and/or in loop regions[132] thus larger variation may not be unexpected.

40

Overall, the strong similarity in per-residue interaction starting from two different homology modeling templates is striking. We attribute the good correspondence to: (i) the homology models used are based on very high sequence homology of ca. 80%, and (ii) the EGFR, HER2, ErbB4 kinase domains are all of the same length (no insertions / deletions) which minimized the actual homology modeling required (i.e. side-chains in common were kept in their respective crystallographic template position).
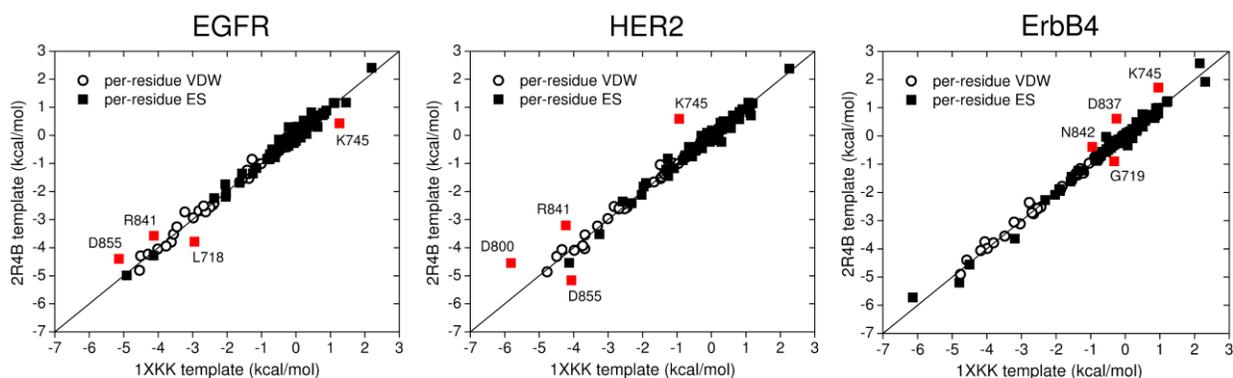
Figure 2.6. Per-residue van der Waals (VDW) and Coulombic (ES) interaction energies for lapatinib with EGFR (left), HER2 (middle), and ErbB4 (right) in which the receptors were constructed using different homology modeling templates based on PDB code 1XKK (EGFR) or 2R4B (ErbB4). Average energies computed from 120,000 MD snapshots (see text for details). For each plot, red colored squares indicate the four residues with the largest variation between the two templates. Diagonal line Y = X.

Table 2.3. Autocorrelation function percent uncorrelated data (ACF %) and block-averaged standard errors of the mean (BASEM) for ΔGb calcd results from simulations of lapatinib with EGFR, HER2, and ErbB4 for various block lengths based on 1XKK and 2R4B templates.

	N = 1 ps		N = 100 ps		N = 1000 ps		N = 2000 ps	
template 1[a]	ACF %[b]	BASEM[c]	ACF %	BASEM	ACF %	BASEM	ACF %	BASEM
EGFR	47.55	0.03	26.79	0.18	10.11	0.45	3.68	0.56
HER2	52.68	0.03	31.78	0.21	11.24	0.54	6.43	0.68
ErbB4	44.50	0.03	23.64	0.17	8.93	0.43	4.22	0.57
average	48.24	0.03	27.40	0.19	10.09	0.47	4.78	0.60
template 2	ACF %	BASEM	ACF %	BASEM	ACF %	BASEM	ACF %	BASEM
EGFR	43.36	0.03	22.91	0.16	7.84	0.41	1.51	0.52
HER2	48.75	0.03	27.44	0.19	10.81	0.48	6.29	0.63
ErbB4	46.71	0.03	25.26	0.18	8.95	0.43	6.77	0.55
average	46.27	0.03	25.20	0.18	9.20	0.44	4.86	0.57

[a]Template 1 based on PDB entry 1XKK, template 2 based on PDB entry 2R4B. [b]ACF in % of correlated data. [c]BASEM energies in kcal/mol. Free energies of binding calculated with 3 binding site waters (3WAT protocol).

2.4.2 The Importance of Binding Site Waters

In general, water molecules are known to play important roles in molecular recognition.[151,153-156] Our laboratory's prior study of the active form of EGFR identified an important network of waters involved in mediating resistance to the anti-EGFR drugs erlotinib and gefitinib.[123] For the inactive forms studied here, examination of explicit solvent MD trajectories of EGFR, HER2, and ErbB4 with lapatinib reveals a similar network of waters. As illustrated in Figure 2.7, which plots a representative MD trajectory for each system, the analysis identified that for EGFR and HER2 there are three distinct sites with high water occupancies (defined here S1, S2, S3) in contrast to ErbB4 for which the S1 site, on average across all six runs, shows much lower occupancy. The definitions are based on waters within 3 Å of ligand atom N* (S1), kinase carbonyl oxygen at residue 791 (S2), or ligand aniline hydrogen (S3). Based on such solvation patterns, we hypothesized that ligand binding would be influenced by differences in hydration which could occur as a result of differences in primary a.a. sequence. It is important to note that these water positions are a result of MD sampling as no crystallographic waters were included in the initial setups. Comparison with available crystallographic data is interestingly mixed. For EGFR with lapatinib (1XKK) water at S1 is observed. For HER2 no inactive structures have yet been published. For ErbB4 with lapatinib (3BBT), strangely, no waters are included in the binding site although the surrounding environment contains water. For ErbB4 with a related covalently bound analog (2R4B), water at S1 is observed. In an attempt to quantify the effects of hydration at these sites, as described below binding free energies were computed using different numbers of key explicit waters.
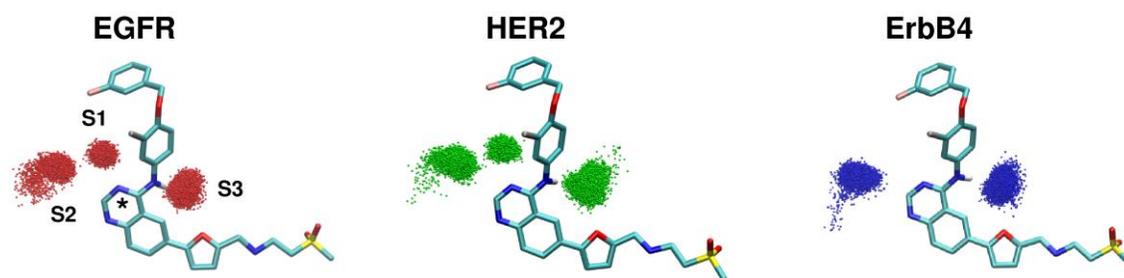
Figure 2.7. Representative MD trajectories showing primary hydration sites labeled S1, S2, and S3 from simulations of lapatinib with EGFR (left, red), HER2 (middle, green), and ErbB4 (right, blue) based on template 1XKK. Images show overlaid waters (O atoms) within 3 Å of ligand atom N* (S1), kinase carbonyl oxygen at residue 791 (S2), and ligand aniline hydrogen (S3) from 10,000 evenly spaced MD snapshots (20 ns trajectories).

2.4.3 Correlation with Experimental Binding Trends

Table 2.4 shows free energy of binding values (ΔG_b calcd) obtained using one of three calculation protocols that includes zero (0WAT), three (3WAT), or one (1WAT) explicit waters. 0WAT represents the default MM-GBSA protocol, 3WAT is designed to capture hydration effects at all sites S1-S3, and 1WAT is designed to isolate effects primarily due to S1. It is important to note that for any given simulation a single solvated trajectory is postprocessed to derive 0WAT, 3WAT, and 1WAT protocols, the underlying protein-ligand ensembles in each case are identical, and it is only the number of explicit waters retained which are different. A number of prior studies[180-182] have also included key binding site waters using related free energy calculation protocols. For each receptor setup, ΔG_b calcd results are presented for the combined average (avg) of templates 1XKK (#1) and 2R4B (#2) as well as each individual template. Each template group represents the average of six well-converged 20 ns simulations. Discussions which follow employ the template-averaged data. In terms of magnitude, the overestimation of absolute ΔG_b calcd values in Table 2.4 relative to experiment, is a well-known occurrence when explicit solute entropic terms are omitted, as was the case with the present study. Focusing on the differential energies ($\Delta\Delta G_b$), for which neglected entropic terms may be reasonably assumed

44

to cancel, the template-averaged results show remarkable agreement with the experimental trends provided that key explicit waters (3WAT or 1WAT protocols) are included in the calculations. Calculations in which no explicit waters are retained (0WAT protocol) incorrectly predict lapatinib to bind most tightly to ErbB4.

Table 2.4. Calculated vs experimental free energies of binding.

Absolute	0WAT[a] ΔG_b calcd[d]	3WAT[b] ΔG_b calcd	1WAT[c] ΔG_b calcd	ΔG_b exptl[e]
EGFR avg[f]	-61.73	-68.23	-65.33	
EGFR #1[g]	-62.19	-68.80	-65.84	-11.62
EGFR #2[h]	-61.27	-67.66	-64.81	
HER2 avg	-61.08	-67.71	-64.57	
HER2 #1	-61.39	-67.82	-64.53	-10.75
HER2 #2	-60.77	-67.60	-64.61	
ErbB4 avg	-62.96	-66.91	-63.20	
ErbB4 #1	-62.93	-67.19	-63.35	-8.81
ErbB4 #2	-62.99	-66.63	-63.05	
Differences (avg)	**0WAT** $\Delta\Delta G_b$ calcd	**3WAT** $\Delta\Delta G_b$ calcd	**1WAT** $\Delta\Delta G_b$ calcd	$\Delta\Delta G_b$ exptl
EGFR – HER2	-0.65	-0.52	-0.76	-0.87
HER2 – ErbB4	1.88	-0.80	-1.37	-1.94
EGFR – ErbB4	1.23	-1.32	-2.13	-2.81

[a]0WAT protocol computed with zero waters. [b]3WAT protocol computed with three closest waters to N* (labeled in Table 2.1). [c]1WAT protocol computed using one water if within 3 Å of N*. [d]ΔG_b calcd values in kcal/mol from six 20 ns simulations. [e]ΔG_b exptl $\approx$ RT ln (activity) at 298.15K using K_i values from Table 2.1. [f]Average of template #1 and #2. [g]Computed using template #1 (1XKK). [h]Computed using template #2 (2R4B).

As shown in Table 2.4, use of the 3WAT protocol increases the averaged absolute binding energies, relative to 0WAT, nearly identically for EGFR (-6.50 kcal/mol) and HER2 (-6.63 kcal/mol) in contrast to ErbB4 (-3.95 kcal/mol) which is lower. The smaller energetic gain for ErbB4 makes physical sense given the lack of S1 water observed in the hydration patterns (Figure 2.7). For the 1WAT protocol, again nearly identical increases in binding are observed for EGFR (-3.60 kcal/mol) and HER2 (-3.49 kcal/mol) in comparison to ErbB4 which is significantly lower (-0.24 kcal/mol). Here, the effect for ErbB4 is more pronounced given that

only the S1 water is considered. Notably, the 1WAT protocol yields remarkable numerical agreement with the experimental $\Delta\Delta G_b$ differences: (i) EGFR–HER2 = -0.76 calcd vs -0.87 exptl, (ii) HER2–ErbB4 = -1.37 calcd vs -1.94 exptl, and (iii) EGFR–ErbB4 = -2.13 calcd vs -2.81 exptl. Overall, the solvated results underscore the importance of considering binding site waters in these systems. In particular, as discussed further below, different amounts of solvent in the S1 site appear to not only drives specificity of lapatinib for the three different receptors but also plays a role in how clinically relevant point mutations affect ligand binding.

2.4.4 Component Contributions to Affinity

To gauge which of the underlying energy terms comprising ΔGb calcd change most as a result of including explicit waters, component analysis was performed as shown in Table 2.5 for the different protocols. Here, as was observed in our earlier study of active-form EGFR, affinity for lapatinib with the inactive kinases also appears to be most strongly driven by favorable intermolecular van der Waals interactions (Table 2.5). Favorable intermolecular $\Delta Ecoul$ terms are roughly half of the accompanying favorable $\Delta Evdw$ terms and the unfavorable desolvation terms (ΔG_{polar}) are significant. Taken together, the results suggest molecular association in these systems is primarily driven by steric packing. However, selectivity across the series (EGFR > HER2 > ErbB4) does not appear to correlate with variation in packing. In every case, ΔE_{vdw} terms for ErbB4 are more favorable ($\geq$ -77 kcal/mol) compared with the other two receptors (< -77kcal/mol).

Table 2.5. Absolute (Δ) binding energy components (kcal/mol).

	ΔE_{vdw}	ΔE_{coul}	ΔG_{polar}	$\Delta G_{nonpolar}$	ΔG_b calcd	ΔG_b exptl
			0WAT			
EGFR avg	-75.91	-31.40	53.17	-7.56	-61.73	-11.62
HER2 avg	-75.83	-32.89	55.22	-7.58	-61.08	-10.75
ErbB4 avg	-77.00	-35.59	57.22	-7.60	-62.96	-8.81
			3WAT			
EGFR avg	-76.86	-40.61	56.67	-7.44	-68.23	-11.62
HER2 avg	-76.54	-42.50	58.78	-7.45	-67.71	-10.75
ErbB4 avg	-77.74	-42.61	60.93	-7.49	-66.91	-8.81
			1WAT			
EGFR avg	-76.22	-35.64	54.07	-7.53	-65.33	-11.62
HER2 avg	-76.07	-36.92	55.95	-7.53	-64.57	-10.75
ErbB4 avg	-77.00	-35.94	57.33	-7.59	-63.20	-8.81

As shown in Table 2.5, addition of binding site waters (3WAT, 1WAT) affects electrostatic contributions (ΔE_{coul}, ΔG_{polar}) much more dramatically than the packing terms (ΔE_{vdw}, $\Delta G_{nonpolar}$). As water-mediated hydrogen-bonding is primarily electrostatic in nature this observation is reasonable. The largest changes in ΔE_{vdw} are less than 1 kcal/mol. In sharp contrast, for the 3WAT relative to 0WAT protocol, ΔE_{coul} favorably increases by ca. -9 kcal/mol for EGFR and HER2 vs -7 kcal/mol for ErbB4. And, for the 1WAT protocol the differences are ca. -4 kcal/mol for EGFR and HER2 vs -0.4 kcal/mol for ErbB4. The negligible ΔE_{coul} change for ErbB4 is a function of less favorable hydration at the S1 site. In support of this conclusion, all the energetic terms (ΔE_{vdw}, ΔE_{coul}, ΔG_{polar}, $\Delta G_{nonpolar}$) for ErbB4 from the 0WAT or 1WAT results are nearly identical. In fact, the smaller increases in ΔE_{coul} (favorable) and ΔG_{polar} (unfavorable) for ErbB4 vs the other two systems is the primary reason why the correct experimental ordering EGFR > HER2 > ErbB4 is obtained using solvated ΔG_b calcd protocols. Finally, it is interesting to note that of all the terms, ΔG_{polar} most closely tracks the experimental trend. This provides additional evidence that selectivity for these receptors is a function of differential solvation in the bound state.

2.4.5 Footprint Contributions to Affinity

Although structure activity relationships can often be explained using per-residue energetic decompositions (molecular footprints), an examination here suggests lapatinib specificity is not governed directly by changes in residue-specific intermolecular interactions. As shown in Figure 2.8, which plots results based on template 1XKK simulations, the remarkably high overlap in the three van der Waals (ΔE_{vdw}) footprints derived from the EGFR,

HER2, and ErbB4 results reveals no specific residue(s) which might play a role. While not a particularly useful indicator of ErbB family specificity, the tight ΔE_{vdw} overlap in does provide additional evidence the homology model results are robust and well-converged. In contrast, Coulombic energy (ΔE_{coul}) footprints do show variability at ca. 5 positions (KKK 745, CSV 775, DDE 800, RRH 803, DDD 855) and these sites were examined in greater detail. Nomenclature here refers to the a.a. present at any Figure 2.8 given position following the receptor order EGFR-HER2-ErbB4. At positions KKK 745, CSV 775, and RRH 803 the interactions between lapatinib and the accompanying residue in ErbB4 are in fact less favorable than those in EGFR, HER2, or both, thus these residues could possibly play a role in reduced binding. However, these ErbB4 losses are not enough to compensate for gains in favorable ΔE_{coul} energies at positions DDE 800 and DDD 855. While a localized increase at position 800 with ErbB4 may make physical sense, in terms of a DDE change contributing to a more favorable global ΔE_{coul} term (see 0WAT results Table 2.5) the net effect is in the wrong direction if ligand-protein intermolecular Coulombic interactions alone were driving specificity. Similarly, the favorable increase at residue 855 with ErbB4 is counterintuitive, although it is interesting that this position involves the DFG motif (855-857) which can adopt an inactive (DFG-out) or active (DFG-in) conformation.[183] Although the footprints do not provide a direct route for understanding specificity, indirectly, the relatively small energetic change at position 775 involving a swap from a polar (C or S) residue in EGFR or HER2 to a nonpolar (V) residue in ErbB4 has very large consequences in terms of changing ligand hydration.
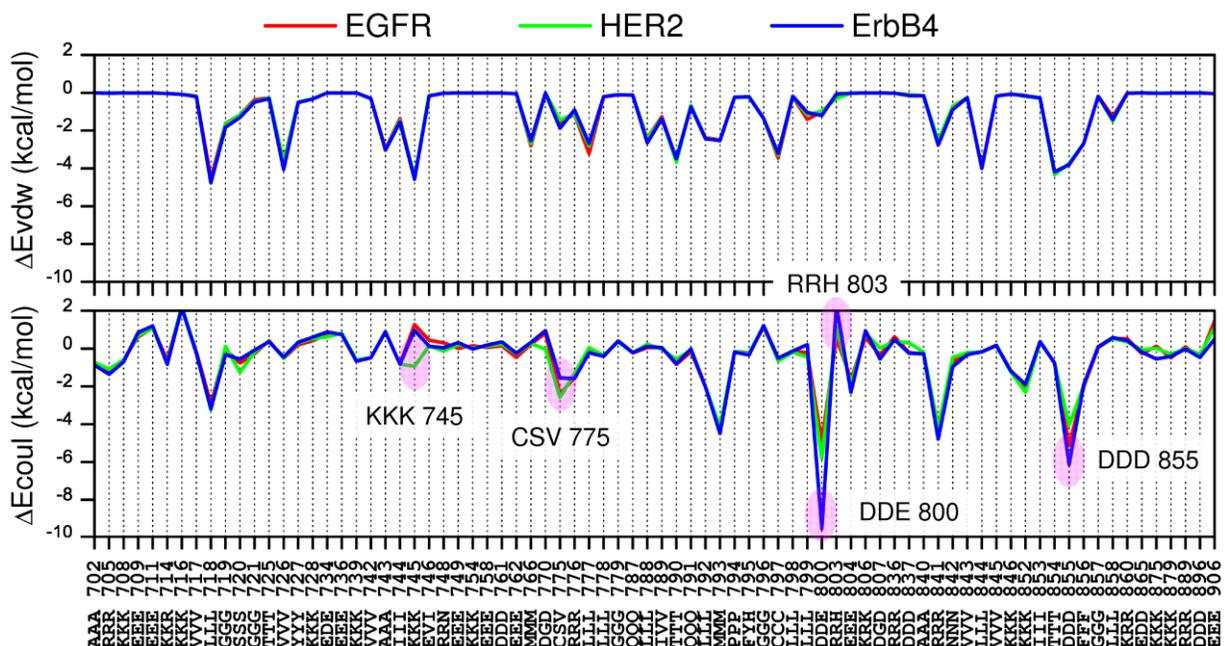
Figure 2.8. Average per-residue van der Waals (ΔE_{vdw}) and Coulombic (ΔE_{coul}) footprints for lapatinib with EGFR, HER2, and ErbB4 from six 20 ns simulations derived from the 1XKK template. Single letter a.a. codes in the x-axis legend indicate, at any given position, the side chain present for EGFR (bottom), HER2 (middle), or ErbB4 (top).

2.4.6 Primary Sequence Differences Alter Solvation which Drives Specificity

Examination of MD trajectories with explicit solvent reveals hydration at the S1 and S2 sites to be mediated primarily by four residues close to the ligand: CSV 775, TTT 790, QQQ 791, and TTT 854 (Figure 2.9). Of these positions, only 775 involve an amino acid change. The representative snapshots shown in Figure 2.9a highlight the quadrifurcated H-bonding network involving S1 and S2 sites while Figure 2.9b quantifies differences in terms of H-bond counts and Coulombic energy for the 3 waters closest to N* with nearby functionalities (ligand N* atom, all ligand heavy atoms, CSV 775, TTT 790, QQQ 791, or TTT 854 residues). H-bond counts employed a 3Å acceptor-donor pair cutoff and an angle between 120-180° for X_D-H_D----X_A. Values represent averaging over 120,000 MD snapshots using the structures derived from

50

template 1XKK. Importantly, the results show high occupancy waters of > 70 % with all four

residues being examined, which indicate long-lived interactions, with the notable exception that

for simulations of lapatinib with ErbB4 the H-bound counts and Coulombic energies are

dramatically lower.

	Population (count)			ΔE_{coul} energy (kcal/mol)		
Residues	**EGFR**	**HER2**	**ErbB4**	**EGFR**	**HER2**	**ErbB4**
ligand atom N*	0.88	0.75	0.11			
ligand	2.30	2.15	1.65	-9.26	-9.19	-7.21
CSV 775	0.71	0.85	0	-1.63	-4.47	-0.90
TTT 790	0.76	0.90	0.77	-3.87	-4.55	-3.36
QQQ 791	1.37	1.13	0.98	-6.33	-6.77	-5.63
TTT 854	0.89	0.76	0.11	-8.19	-6.90	-1.81
sum (excluding N*)	6.03	5.79	3.51	-29.28	-31.88	-18.91
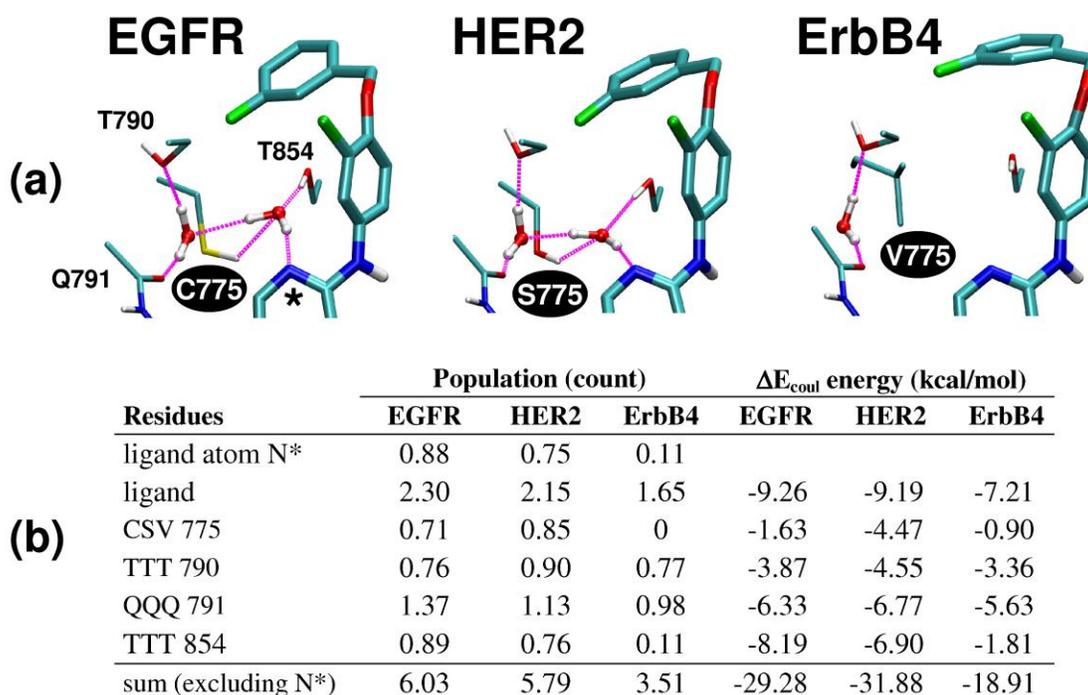
Figure 2.9. Water-mediated H-bonds involving lapatinib and nearby residues in EGFR, HER2, and ErbB4. **(a)** Representative snapshots showing quadrifurcated H-bonding network involving S1 and S2 waters. **(b)** Averaged H-bond counts and pair-wise Coulombic interaction energies (N=120,000 frames each) for the 3 waters closest to N* with: N*, the entire ligand, residues CSV 775, TTT 790, QQQ 791, or TTT 854. Results based on template 1XKK.

The binding site graphic in Figure 2.9a visually highlights why a bulkier side chain at

position 775 in ErbB4 (V), without the same H-bonding capability as in EGFR (C) or HER2 (S),

is the most likely cause of reduced water occupancy. Quantitatively (Figure 2.9b), water H-

bonding at this position is zero in ErbB4 compared to 0.71 and 0.85, respectively, in EGFR and

HER2. And, the change to valine reduces counts involving the N* atom of lapatinib with ErbB4

(0.11) compared with EGFR (0.88) and HER2 (0.75). Counts for the total ligand show a similar

trend (ErbB4 = 1.65 vs EGFR = 2.30 vs HER2 = 2.15). It is probable these differences are the primary contributor of the larger desolvation penalties in ErbB4 (higher ΔG_{polar} terms in Table 2.5). Interestingly, water H-bonds involving residue TTT 790 do not appear to be reduced in simulations of ErbB4 relative to EGFR, however, at QQQ 791, and in particular for TTT 854, there are reductions (Figure 2.9b). In the latter case for T854, the number of interactions ErbB4 (0.11) vs EGFR (0.89) or HER2 (0.76) are essentially the same as those obtained from calculations involving only the ligand N* atom. The correspondence here confirms a highly coupled water site. Interestingly, counts with the ligand track with the experimental ordering: EGFR (2.30) > HER2 (2.15) > ErbB4 (1.65). Overall, the summed H-bond counts for waters with the ligand plus the four nearby residues are ca. 40% lower in ErbB4 (3.51) compared with EGFR (6.03) or HER2 (5.79).

Examination of Coulombic energies (ΔE_{coul}) reveals a similar trend with averaged water-residue interactions being significantly less favorable in the ErbB4 binding site than in EGFR or HER2 (Figure 2.9b). The fact the summed ΔE_{coul} interactions (ligand plus four nearby residues) are again reduced by roughly 40% for ErbB4 (-18.91 kcal/mol) vs EGFR (-29.28 kcal/mol) or HER2 (-31.88 kcal/mol) indicates that in these systems Coulombic losses scale linearly with changes in hydrogen bonding. As before, the most significant losses occur at CSV 775 and TTT 854. And, similar to the H-bond results, variation in ΔE_{coul} for waters with the ligand also tracks with experiment: EGFR (-9.26 kcal/mol) > HER2 (-9.19 kcal/mol) > ErbB4 (-7.21 kcal/mol). Interestingly, the ca. 2 kcal/mol lower ΔE_{coul} for ErbB4 versus the other receptors is roughly similar in magnitude to $\Delta\Delta G_b$ exptl (ErbB4–EGFR = 2.81 kcal/mol and ErbB4–HER2 = 1.94 kcal/mol).

2.4.7 The Effects of Mutation

At least two mechanisms of resistance impact anti-cancer drugs which target receptor tyrosine kinases: (i) redundant survival pathways activated as the consequence of inhibition and (ii) point mutations which directly alter ligand binding. Focusing on the latter mechanism, although point mutations in EGFR or HER2 are not commonly associated with lapatinib resistance as a result of clinical use to treat breast cancer, there is a concern that mutation sites observed as a result of treatment with related inhibitors for other cancers could eventually impact lapatinib.[127,128,138] Using in vitro screening, Avizienyte et al[127] and Trowe et al[128] both reported numerous ErbB family mutations which negatively affect lapatinib activity. Compellingly, several mutations identified (C775F, T854A, T790M, S775P, T790I) map here to sites involved in water-mediated ligand binding. Analogous to the EGFR > HER2 > ErbB4 selectivity arguments discussed above, it is reasonable to propose that changes affecting water-mediated binding would also lead to drug-resistance. To test this hypothesis additional MD simulations and analysis (Figure 2.10, Table 2.6) for lapatinib complexed with the following mutant kinases were performed: EGFR (C775F, T854A, T790M), HER2 (T790I), and ErbB4 (V775C). Simulations were based on template 1XKK and for each mutant system a single 20 ns MD run was performed.
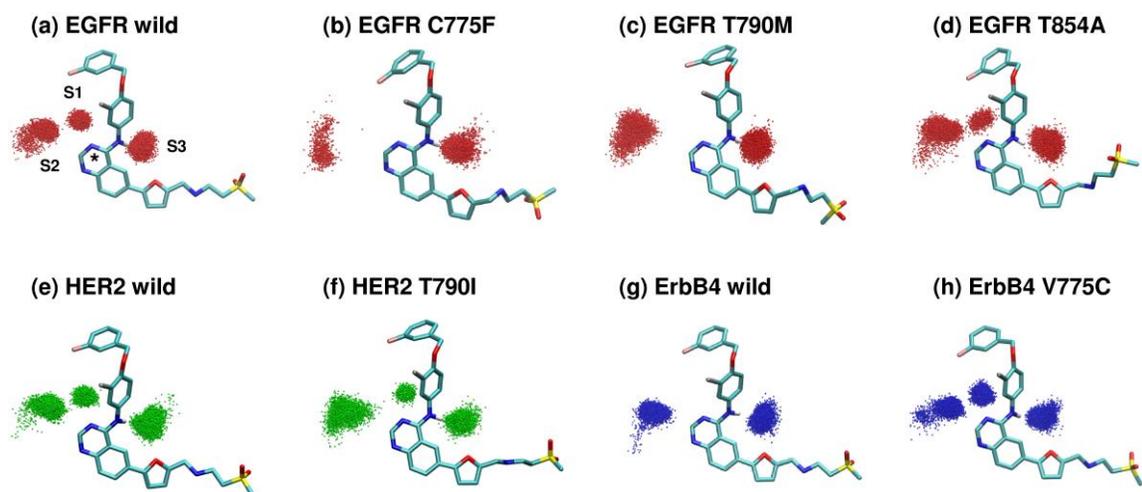
Figure 2.10. Water-mediated H-bonding patterns (sites S1-S3) from lapatinib with wildtype and mutant forms of EGFR (red), HER2 (green), and ErbB4 (blue). Images show overlaid waters (O atoms) within 3 Å of ligand atom N* (S1), kinase backbone oxygen at residue 791 (S2), and ligand aniline hydrogen (S3) from 10,000 evenly spaced MD snapshots (20 ns trajectories)

Table 2.6. Average number of water-mediated H-bonds and binding free energies for lapatinib with wildtype and mutant kinases.

Residue (wildtype)	(a) EGFR wild[a]	(b) EGFR C775F[b]	(c) EGFR T790M[b]	(d) EGFR T854A[b]	(e) HER2 wild[a]	(f) HER2 T790I[b]	(g) ErbB4 wild[a]	(h) ErbB4 V775C[b]
	H-bonds with water (count)[c]							
ligand (N*)	0.88	0	0	0.22	0.75	0.54	0.11	0.95
ligand	2.30	1.37	1.40	1.76	2.15	2.29	1.65	2.57
CSV 775	0.71	0	0.10	0.22	0.85	0.36	0	0.80
TTT 790	0.76	0.09	0.06	0.87	0.90	0	0.77	0.75
QQQ 791	1.37	0.27	0.99	1.05	1.13	0.62	0.98	0.91
TTT 854	0.89	0	0	0.03	0.76	0.54	0.11	0.97
sum (exclude N*)	6.03	1.73	2.55	3.93	5.79	3.81	3.51	6.00
	ΔG_b calcd (kcal/mol)[d]							
3WAT protocol	-68.80	-67.61	-67.15	-66.60	-67.82	-65.25	-67.19	-69.21
1WAT protocol	-65.84	-63.56	-63.28	-62.57	-64.53	-62.10	-63.35	-66.18

[a]Wildtype values averaged from 6 MD runs (120,000 frames). [b]Mutation values averaged from 1 MD run (20,000 frames). [c]Averaged H-bond counts employ the 3 waters closest to N* (defined in Table 2.1) with: N*, the entire ligand, residues 775, 790, 791, or 854. [d]Binding energies computed using 3WAT (three closest waters to N*) or 1WAT protocols (one water if within 3 Å of N*). Simulations based on template 1XKK (see Methods).

2.4.7.1 ErbB4 V775C

The first mutation studied, ErbB4 V775C (Figure 2.10h, Table 2.6h), is a hypothetical mutation which was primarily performed to re-affirm the conclusions reached above that the observed differences in experimental activity for the EGFR > HER2 > ErbB4 series is primarily driven by the character of the sidechain at position 775. It was hypothesized that mutating the native nonpolar valine back to a polar cysteine (as in EGFR) would restore the water network and affinity of lapatinib for ErbB4 would improve. As expected, the simple swap at position 775 indeed restores the bridging water at S1 as illustrated visually by a comparison of the before and after hydration patterns (Figure 2.10 vs Table 2.6h). Further, the ErbB4 V775C patterns (Figure 2.10h) are remarkably similar to wildtype EGFR (Figure 2.10a) or wildtype HER2 (Figure 2.10e). Quantitatively (Table 2.6), water counts for the ErbB4 V775C mutant localized to site 775 shows an increase ($0 \rightarrow 0.80$) which yields remarkable accord with values observed in wildtype EGFR (0.71) or HER2 (0.85) simulations. Good agreement is also seen for the increase occurring at the 854 site: ErbB4 V775C ($0.11 \rightarrow 0.97$) vs EGFR (0.89) and HER2 (0.76). Ligand-water values also increase ($1.65 \rightarrow 2.57$) to values which are similar to that observed in the EGFR (2.30) and HER2 (2.15) wildtype systems. Only minor changes are observed at sites 790 and 791, which make physical sense, given that occupancies at the S2 site are not expected to be as drastically altered by this mutation. Energetically, it is gratifying that the accompanying ΔG_b calcd values for the ErbB4 V775C mutant also show favorable increases relative to wildtype ErbB4 (Table 2.6h vs g) as originally hypothesized, using either the 3WAT or 1WAT solvated protocols. And, values for the ErbB4 V775C mutant (3WAT = -69.21 kcal/mol, 1WAT = -66.18 kcal/mol) are close to that of wildtype EGFR for which the V $\rightarrow$ C mutation was designed to mimic (3WAT =

-68.80 kcal/mol, 1WAT = -65.84 kcal/mol). Overall, the simulation results strongly suggest that a valine at position 775 is primary factor that contributes to the selectivity of lapatinib for EGFR and HER2 relative to ErbB4.

2.4.7.2 EGFR C775F

Avizienyte et al[127] proposed that a direct steric clash with the lapatinib aniline group was a likely cause of reduced activity with the EGFR C775F variant identified using in vitro screening. However, based on simulation results of this mutant, positional sampling of lapatinib in the binding site appears similar to that in wildtype EGFR which suggests unfavorable intermolecular interactions are not introduced as a result of the a.a. swap. In fact, a comparison of protein-ligand van der Waals interaction energies, localized to position 775, actually show enhanced interactions for the C775F mutant relative to wildtype (EGFR C775F = -3.24 kcal/mol vs wildtype = -1.56 kcal/mol) which indicates a clash at this position is unlikely. This is not unexpected as the related C to V swap is tolerated at this position in ErbB4.

What does dramatically change, as a result of the increased hydrophobicity and bulk of phenylalanine relative to cystein, is water occupancy at both sites S1 and S2 (Figure 2.10b, Table 2.6b). Relative to wildtype, the EGFR C775F mutant (Table 2.6a vs b) leads to total or nearly total losses in H-bonding with F775 (0.71 to 0), T790 (0.76 to 0.09), T854 (0.89 to 0), or the ligand N* (0.88 to 0). Counts at Q791 (1.37 to 0.27), or with the total ligand (2.30 to 1.37) are also significantly reduced. Losses in overall computed binding energies of between 1.09 and 2.28 kcal/mol (3WAT or 1WAT protocol respectively), again suggest a resistance mechanism involving network waters.

Interestingly, the EGFR C775F mutation was only identified in lapatinib in vitro resistance screenings but not erlotinib.[127]   Although differences in bound inactive vs active EGFR could be a factor, an examination of EGFR crystal structures with lapatinib (1XKK) and erlotinib (1M17) did not reveal any obvious reason why, in principle, C775F could not negatively impact erlotinib. Structurally, an EGFR C775F mutant should negatively affect S1 water occupancy in both active and inactive forms. It should be noted that EGFR T790M and EGFR T854A do negatively affect both lapatinib and erlotinib.[127]   Although not yet reported, the related EGFR C775V mutation would also be expected to be detrimental to both classes of inhibitors.

The analogous mutation for HER2 is S775P. Trowe et al[128] suggests this mutation may act by direct steric interference, by analogy to the spatially homologous mutation V299L change associated with resistance to imatinib in BCR-ABL. However, V299L in this system involves a nonpolar-to-nonpolar change which directly would not affect changes in H-bonding with water in contrast to the polar-to-nonpolar change S775P would introduce into in HER2. Thus, while a steric mechanism might be appropriate to describe the effects of V299L in BCR-ABL, for lapatinib with HER2 S775P, a water-based mechanism is more likely. Additional simulations would be needed to more fully explore this issue.

2.4.7.3 EGFR T790M

The gatekeeper mutation T790M, associated with clinical resistance to gefitnib and erlotinib with EGFR[139,184] has also been identified as a mutation affecting lapatinib.[127]   Although it has been suggested[138] that T790M sterically hinders lapatinib binding with EGFR, again our results indicate changes involving water are much more likely. Results here, from simulations of

EGFR T790M reveal that H-bond counts localized to M790 (0.06), the ligand N* (0), or T854 (0) are essentially zero relative to wildtype (Table 2.6c vs a). The increased bulk in going from T to M, and the fact that sulfur is a weaker H-bond acceptor than oxygen, are the most likely causes.[123] Visually, the accompanying T790M graphic (Figure 2.10c) show S1 waters are no longer present which is well reflected in the total H-bond counts with ligand which goes from 2.30 to 1.40.

Using analysis similar to that reported here, Balius et al[123] showed that mutations at the T790 site (and T854 as described below) would disrupt water-mediated binding. The prior study not only resolved the ambiguity surrounding which residue (T790 or T854) was primarily involved in water-mediated interactions (both are), but provided convincing quantitative evidence why a T790M steric clash mechanism is unlikely. Given the positional similarity for high occupancy water sites in both active and inactive kinase forms, and based on the present results which show disruption of the water-network, lapatinib resistance due to EGFR T790M in the inactive form is also not likely driven by a steric clash. The favorable increase in van der Waals energy observed at the site of the mutation (EGFR T790M -5.85 kcal/mol vs wildtype -3.56 kcal/mol) reaffirms this conclusion. The fact that the mutant also yields less favorable computed binding energies (ΔG_b calcd) relative to wildtype (Table 2.6c vs a), with losses of between 1.65 and 2.56 kcal/mol, when solvent is included (3WAT and 1WAT protocols respectively), provides additional evidence for a water-based mechanism.

2.4.7.4 HER2 T790I

The analogous gate keeper mutation in HER2 is T790I, which was the most frequent mutation identified from in vitro screening by Trowe et al,[128] and is correlated with a high level

of lapatinib resistance. Although direct steric interference was again suggested,[128] based on the present computational results, a water-mediated mechanism is more likely. In agreement with results from the EGFR T790M simulations, occupancy at S1 in the HER2 T790I mutant becomes reduced, and the increase in "favorable" van der Waals interactions occurring at residue 790 (mutant = -3.72 kcal/mol vs wildtype = -3.67 kcal/mol) indicate no clash. Quantitatively, counts for S775, T790, and T854 become zero. As expected, the reduced S1 occupancy also affects S2 waters which are reflected in reduced counts at Q791. Interestingly, counts with the total ligand slightly increase which is somewhat counterintuitive but could be a function of slight differences occurring at, for example, the solvent-exposed methyl sulfone tail. In any event, the reduction at the key ligand N* atom (from 0.75 to 0.54) indicates a lack of S1 water which is consistent with a water-based resistance mechanism.

2.4.7.5 EGFR T854A

The final mutation studied, EGFR T854A, can confer resistance both to lapatinib and erlotinib,[127,140] which, for the latter case was identified from a patient with lung adenocarcinoma.[140] Possible resistance mechanisms proposed by Bean et al[140] for erlotinib include a loss of contact with the inhibitor, altered specificity for ATP, or conformational changes in the protein. Balius et al[123] hypothesized mutations at this site could disrupt erlotinib binding through alteration of the water-mediated network. For lapatinib, the results here suggest both a contact and a water-based mechanism of T854A resistance in the inactive kinase system. Per-residue decomposition shows that van der Waals interactions, localized to position 854, become less favorable in simulations of the EGFR T854A mutant (-2.61 kcal/mol) relative to wildtype (-4.29 kcal/mol). Water H-bond counts with nearby residues also become reduced

(Table 2.6d). For example, counts with N* (0.88 → 0.22), the ligand (2.30 → 1.76), residue 775 (0.71 → 0.22), residue 791 (1.37 → 1.05), and residue 854 (0.89 → 0.03) are all lower relative to wildtype. Interestingly, a slight increase is observed at residue 790 (0.76 → 0.87) which is counterintuitive. A possible explanation involves the fact that, compared to the other mutations studied, T854A involves a swap to a smaller amino acid. As a result, although waters may be less favorably accommodated, they are not sterically blocked from the S1 site (see Figure 2.10d). This could allow water to more freely occupy S2 thereby increasing the possibility of interacting favorably with T790 thus explaining the observed increase. In any event, from a quantitative standpoint (Table 2.6), summed H-bond counts are significantly reduced in simulations of T854A (3.93) vs wildtype (6.03).

2.5 Future Directions

The present results suggest additional studies (both computational and experimental) which could be used to further characterize molecular recognition in these systems. For example, the prediction made here that ErbB4 V775C would improve lapatinib binding was verified computationally, however, the mutation should be tested experimentally. The related mutation ErbB4 V775S, as a surrogate for HER2, should also be examined. Other avenues include use of more quantitative computational methods (free energy perturbation, thermodynamic integration) or alternative force fields (polarizable, quantum-based) to probe the energetic effects of including bound waters at specific locations in the binding site or investigate resistance mutations in finer detail. Additional mutations to study, which are induced by lapatinib, include L747S, R776P, L777Q, L788I/V, K860T, G863S, and R889S.[127]

A growing body of evidence also suggests that use of kinase domain inhibitors leads to upregulation of a fourth ErbB member, HER3, which in turn causes drug resistance through alternative signaling pathways.[185-187] Thus, HER3 could be considered as a potential therapeutic target. Interestingly, HER3 has historically been classified as an inactive kinase although recent studies[188] indicate there is weak catalytic activity which could be clinically relevant. Based on the current simulation results, and comparison with available inactive-form structures of HER3;[189] we hypothesize that lapatinib would bind HER3 only weakly. Similar to ErbB4, the HER3 catalytic domain contains a valine at position 775 and thus should have less water-mediated H-bonding capability compared to EGFR or HER2. Weaker binding and weaker catalytic activity could both contribute to HER3 upregulation. Additional simulations to assess the binding of lapatinib with HER3 are planned for the future.

Finally, from the development point of view, strategies to achieve improved ErbB inhibitors include modifications which yield tighter binding to wildtype enzymes, restore binding lost as a result of resistance mutations, or both. The present results suggest that exploiting differences in the number of binding site waters could be important. As a first step, we performed simulations involving a straightforward modification of lapatinib at the N* position from N* $\rightarrow$ C* (Figure 2.11a). We hypothesized the swap would yield lower affinity for wildtype EGFR as a result of water occupancy becoming reduced at S1 caused by a lack of H-bonding capability at C*. Indeed, a simulation of the analog with wildtype EGFR revealed lower water occupancy at S1 and a less favorable ΔG_b calcd value of -63.59 kcal/mol (analog) versus -65.84 kcal/mol (lapatinib) computed via the 1WAT protocol (Figure 2.11a).
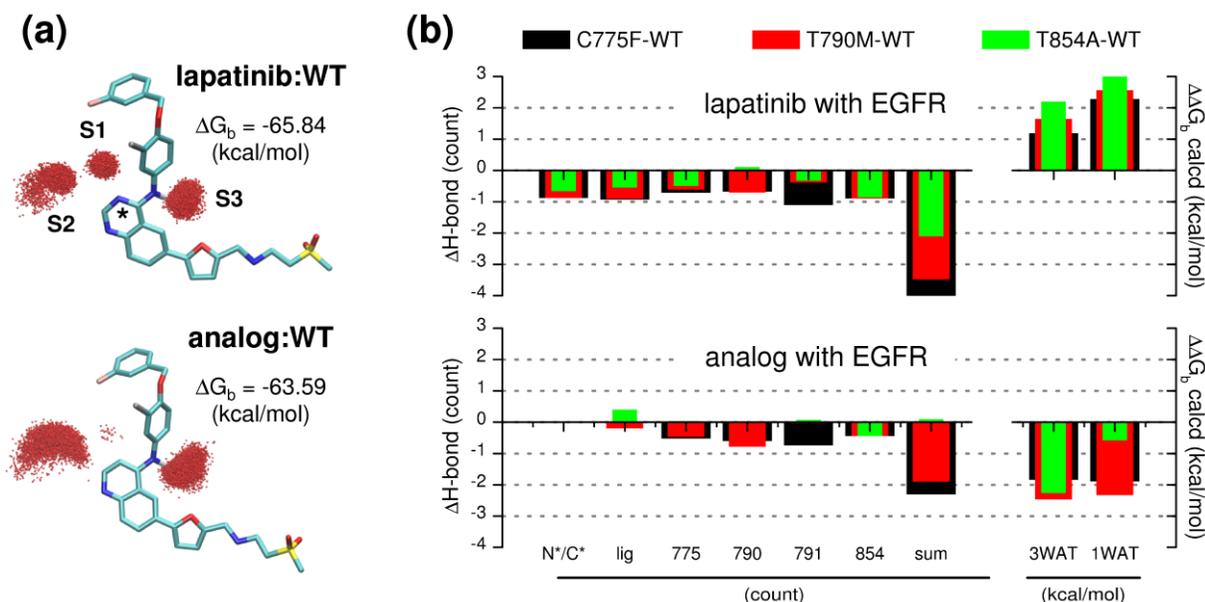
Figure 2.11. (a) Hydration patterns (sites S1-S3) and absolute binding energies (ΔG_b calcd) from wildtype EGFR simulations of lapatinib (top) and a N* → C* analog (bottom). Images show overlaid waters (O atoms) within 3 Å of ligand atom N*/C* (S1), kinase backbone oxygen at residue 791 (S2), and ligand aniline hydrogen (S3) from 10,000 evenly spaced MD snapshots (20 ns trajectories). (b) ΔH-bond and $\Delta\Delta G_b$ calcd values (mutant - wildtype) from simulations of wildtype EGFR, C775F, T790M, and T854A mutants with lapatinb (top) and a N* → C* analog (bottom). Relative H-bonds (count) computed between the 3 waters closest to N*/C* with: N*/C*, the entire ligand, residues 775, 790, 791, 854, and the overall sum excluding N*/C*. Relative binding free energies (kcal/mol) computed using 3WAT (three closest waters to N* / C*) or 1WAT protocols (one water if within 3 Å of N* / C*).

On the other hand, we reasoned the analog would not be as detrimentally affected as lapatinib by mutations that alter S1 occupancy. As illustrated in Figure 2.11b, which shows fold resistance profiles for both ligands, additional simulations of C775F, T854A, and T790M mutants with the analog confirmed the hypothesis. In all three cases (C775F-WT, T790M-WT, T854A-WT), and across both calculation protocols (3WAT, 1WAT), simulations with lapatinib yielded unfavorable fold resistance values (mutant minus wildtype ΔGb calcd) in contrast to the analog which interestingly showed enhanced binding (Figure 2.11b bottom vs top graphs). The increase in binding affinity computed for the analog with the mutants appears to involve overall

more favorable hydrophobic (ΔEvdw + ΔGnonpolar) interactions for T790M and T854A (Figure 2.11b red and green bars) and a combination of more favorable hydrophobic and hydrophilic (ΔEcoul + ΔGpolar) interactions for C775F (Figure 2.11b black bar). As expected, the simulations of lapatinib showed larger losses than for the analog for H-bonding of key waters in the pocket (3 waters closest to N*/C*) with the ligand at N*/C*, the entire ligand, residues 775, 790, 791, TTT 854, and the overall sum (Figure 2.11b top vs bottom). Provided sufficient affinity to wildtype domains could be achieved, analogs that rely less on water at S1 could have utility in combination therapies given the potential for orthogonal resistance profiles. This strategy might also benefit the development of HER3 inhibitors given the expected lower water occupancy at S1 due to the valine at position 775. The design of analogs capable of displacing water at S1, while mimicking the network seen between lapatinib and EGFR or HER2, should also be pursued.

2.6 Conclusions

In this report, all-atom molecular dynamics, free energies of binding, and energy decomposition analyses were performed for the kinase inhibitor lapatinib in complex with wildtype EGFR, HER2 and ErbB4, and relevant mutants including C775F, T854A, T790M (EGFR), T790I (HER2), and V775C (ErbB4). The primary goal was to develop robust computational models consistent with experimental activity data (Table 2.1) and determine how variation in receptor sequence (Figure 2.1) and structure contributes to binding specificity and drug resistance. The simulations employed both crystallographic and crystallographic-derived homology models. Convergence and stability were carefully evaluated using multiple MD runs (N = 6) for each wildtype system for which root-mean-square deviations (Figure 2.5),

64

autocorrelation functions (Figure 2.5), and block-averaged standard errors of the mean (Figure 2.4) demonstrate good system behavior. The remarkable numerical agreement obtained using homology models, derived from one of two different crystallographic templates, is particularly notable (Figure 2.7, Table 2.4).

Analysis of MD trajectories with explicit solvent revealed three high occupancy sites for water termed S1, S2, S3 (see Figure 2.7) that mediate ligand binding in EGFR and HER2. In contrast, ErbB4 showed low occupancy at the S1 site. The results suggested hydration differences could play a role in determining receptor specificity and subsequent free energy calculations in which explicit waters were included confirmed the hypothesis. Relative free energies of binding ($\Delta\Delta G_b$ calcd), using template-averaged data, in which three key waters (3WAT protocol) or one key water (1WAT protocol) are included in the calculations, correctly predict the experimental trend EGFR > HER2 > ErbB4 (Table 2.4). Calculations in which water was not included (0WAT protocol) incorrectly predict the order. Quantitatively, the 1WAT protocol leads to excellent numerical agreement for $\Delta\Delta G_b$: (i) EGFR−HER2 = -0.76 calcd vs -0.87 exptl, (ii) HER2−ErbB4 = -1.37 calcd vs -1.94 exptl, and (iii) EGFR−ErbB4 = -2.13 calcd vs -2.81 exptl which underscores the importance of hydration and in particular the S1 site.

Examination of the underlying ΔG_b calcd energy components with (3WAT, 1WAT) or without (0WAT) explicit waters reveals that electrostatic contributions (ΔE_{coul} and ΔG_{polar}) change much more dramatically than the accompanying steric packing terms (ΔE_{vdw} and $\Delta G_{nonpolar}$), especially for EGFR and HER2 (Table 2.5). Smaller changes for ErbB4 are a function of occupancy at the S1 site and the primary reason why solvated protocols yield the correct experimental ordering. Occupancies at S1 appear to be a function of sequence changes at position 775 (Figure 2.9a) involving a swap from C or S (EGFR, HER2) to V (ErbB4) for which

65

the side chain is bulkier, hydrophobic, and has no H-bonding capability. Otherwise, no specific

residues were identified which might explain specificity through direct modulation of protein-

ligand interactions (Figure 2.8). The fact that ΔG_{polar}, and not ΔE_{coul} or ΔE_{vdw} terms, most closely

tracks with experiment provides additional support that selectivity is a function of differential

hydration in the bound state (Table 2.5).

At position 775, an important quadrifurcated H-bonding network was identified which

involves the ligand, S1 and S2 waters, as well as residues CSV 775, TTT 854, TTT 790, and

QQQ 791 (Figure 2.9a). A similar network was observed in an earlier study of active-form

EGFR.[123] Quantitatively, water H-bond counts and Coulombic interactions with species in the

network are reduced by ca. 40% for ErbB4 relative to EGFR or HER2 (Figure 2.9b). Water H-

bond counts with the ligand N* atom or the total ligand compellingly track with the experimental

ordering EGFR > HER2 > ErbB4 (Figure 2.9b). And, the nearly identical counts across all three

systems, for waters interacting with TTT 854 vs the ligand N* atom, confirm a highly coupled

S1 site (Figure 2.10b). Other groups[139,143,146,148,151] have also discussed the importance of

binding site waters in these and related systems, although to our knowledge, differential

hydration as the primary mechanism for ErbB family specificity has not yet been proposed.

Additional studies to more precisely quantify the energetic impact of including water molecules

in these sites would be worthwhile.

To determine if drug resistance could also be influenced by changes in water-mediated

binding a series of additional simulations were performed (Figure 2.10, Table 2.6) for lapatinib

with mutant kinases identified by in vitro screening. In good qualitative agreement with

experiment, for four deleterious mutations studied (EGFR C775F, EGFR T854A, EGFR T790M,

HER2 T790I), computed free energies of binding using solvated ΔG_b calcd protocols become

less favorable relative to wildtype (Table 2.6). In all cases, S1 occupancies are significantly reduced which suggest specificity as well as resistance share a common mechanism (Table 2.6). In support of this argument, results from a hypothetical mutant (ErbB4 V775C), designed to restore S1 hydration, show a remarkable increase in S1 occupancy and a favorable increase in computed binding free energy (Table 2.6). The computational results also provide a testable prediction for future experimental work. Importantly, in contrast to previously proposed steric clash mechanisms for EGFR C775F[127], EGFR T790M[138], or HER2 T790I[128], the present results indicate a water-based mechanism of resistance. This conclusion is based on several observations: (i) visualization of MD trajectories and ligand RMSDs reveals no identifiable changes in sampling suggestive of a steric clash, (ii) localized per-residue interactions show favorable increases in favorable van der Waals energy at the sites of mutation, (iii) water occupancies at site S1 become significantly reduced, (iv) H-bond counts are lower and Coulombic interactions are less favorable for binding site waters with nearby species, and (v) solvated ΔG_b calcd protocols yield less favorable energies. Somewhat differently, for EGFR T854A, the results are consistent with both a water-based[123] and a previously proposed loss-of-contact[140] mechanism.

A greater understanding of molecular-factors that drives ligand binding for important anti-cancer targets such as EGFR and HER2 will ultimately enable the design of improved drugs with greater clinical utility. The present work has contributed to this goal through development of robust simulation models for lapatinib with inactive-form kinases and providing a physically reasonable water-based mechanism for understanding what drives ErbB family specificity and resistance. The simulations demonstrate that in these systems water is equally as important as any specific residue interaction for binding. The fact that lapatinib affinity for ErbB4 is

disfavored because of a reduction in binding site water is, perhaps fortuitously, likely to be of beneficial outcome.[135,136] Future work could try to exploit this mechanism to further enhance selectivity as a result of hydration differences at S1. However, analogs which further rely on the water network would also be vulnerable to mutations which alter the network. Alternatively, as suggested by the results from simulations of a lapatinib analog (Figure 2.11), inhibitors which do not rely as strongly on water may be less vulnerable. The observation that the network can be altered in different ways (i.e. steric blockage of water, reduced H-bonding capability with water, or both) suggest that orthogonal binding motifs will likely be required to effectively combat drug resistance in the long term.

Chapter 3. Energetic Analysis of a Lapatinib Conformational Change in Complex with EGFR, HER2, and ErbB4

3.1 Introduction

As discussed previously, tyrosine kinases are important anti-cancer drug targets due to their key roles in cellular signaling and growth. In Chapter 2, all-atom MD simulations were performed of the inhibitor lapatinib in complex with three ErbB family members EGFR, HER2, and ErbB4 to probe origins of selectivity. Importantly, the simulations identified a physically unique water-mediated H-bond network which compellingly explains why lapatinib has high affinity for EGFR and HER2 but not ErbB4. In the simulations, all complexes were modeled with a specific lapatinib conformation as seen in the crystal structure 1XKK containing EGFR. Although differential hydration appears to be the driving force for selectivity, an examination of multiple ErbB crystal structures suggests lapatinib could adopt one of two conformations thus alternative ligand conformations might also play a role. For example, although the 1XKK[130] structure shows the ligand fluorophenyl ring pointing towards a CSV motif in the receptor, an alternative structure 3BBT[137] containing lapatinib with ErbB4, show the ring shows flipped by ca 180º. Interestingly, a third structure 2R4B,[159] containing a lapatinib-like inhibitor covalently bound to ErbB4, is consistent with the 1XKK structure. Figure 3.1 show the 3D coordinates of the ligand from these X-ray structures to emphasize the two ring conformations termed here conf1 (pointing towards the receptor CSV motif) and conf2 (flipped 180º). Here, although 1XKK contains a complete set of ligand coordinates, 3BBT is missing the methyl sulfone tail,

and 2R4B contains a covalently-bound lapatinib-like inhibitor. Work in this work chapter employs MD simulations and free energy calculations to determine which of the two crystallographically observed conformations (conf1 or conf2) has the most favorable energy in each of the three receptors.
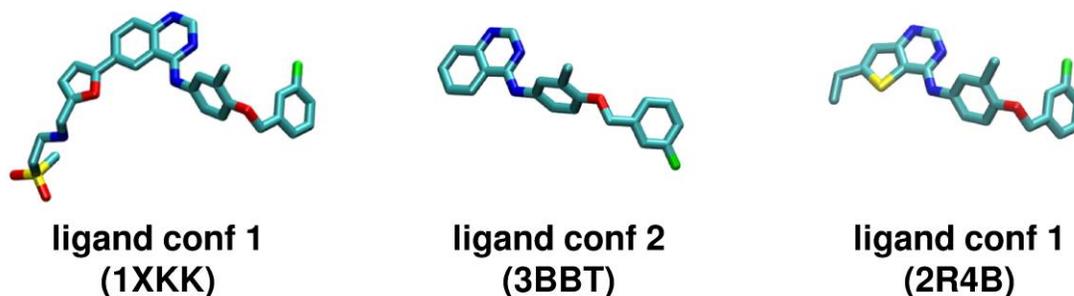
Figure 3.1. Lapatinib conformation 1 and 2 (conf1 and conf2). Left is lapatinib in EGFR complex (pdb 1XKK). Middle is incomplete lapatinib in ErbB4 complex (pdb 3BBT). Right is lapatinib-like ligand in ErbB4 complex (pdb 2R4B).

3.2 Theory (Umbrella Sampling)

To estimate the conformational preferences two general simulations methods were employed: (1) umbrella sampling to derive potentials of mean force (PMFs), and (2) all atom molecular dynamics simulations followed by MM-GBSA free energy analysis. The theory and protocols for MM-GBSA energy analysis were described previously in Chapter 1 so we focus here on umbrella sampling. Due to the large energy barriers that may exist between different conformational states of a system, standard Monte Carlo or molecular dynamics simulation methods may not be able to allow for sufficient and accurate sampling within a reasonable time. This difficulty can prevent accurate calculation of the potential of mean force (PMF) which is the free energy profile (W) determined from the population (p) along a chosen coordinates (ζ), for example the conformational ring flip described above via eq (3.1). The umbrella sampling[190-192] method is a way to increase sampling through a modified potential function which allows the

70

energy barriers to be crossed. The key idea is adding to the Hamiltonian an artificial biasing potential (e.g., a harmonic potential) which flattens the energy landscape and a series of biased window simulations of the chosen structure coordinates (in this case a lapatinib dihedral angle) are performed. For accuracy, there should be significant overlap between the adjacent windows. The weighted histogram analysis method (WHAM)[190,191] then provides a scheme for recombining umbrella sampling results obtained in the presence of the biasing potentials to obtain a single unbiased property distribution.

$$W(\zeta) = -k_B T \ln \langle p(\zeta) \rangle \tag{3.1}$$

3.3 Computational Details

3.3.1 Model Construction

All-atom models were constructed for lapatinib in the conf1 and conf2 states bound to the inactive kinase domains of EGFR, HER2 and ErbB4. While crystal structures 1XKK and 3BBT were available with lapatinib bound to EGFR[130] and ErbB4[137] no crystal structure of lapatinib bound to inactive HER2 was available. However, an active HER2 complex was recently released.[160] Ultimately 3BBT was not used due to 14 residues missing in the activation loop and 1XKK was used for EGFR and to construct homology models for HER2 and ErbB4 using Modeller9v6.[79] Due to high sequence homology (77% ~ 81%) using ClustalW,[161] homology models for the receptors can be constructed with high degree of confidence. The details of receptor construction and evaluation were described in the earlier study.[193] To explore the binding energy and population of lapatinib conf1 and conf2, two ligand conformations were modeled, conf1 was obtained from lapatinib in 1XKK and conf2 was from conf1 with fluorine manually flipped 180º.

71

3.3.2 Potential Mean Force (PMF)

To further explore the energy and the probability of the two different conformations of lapatinib, umbrella sampling[194] was performed using the AMBER8[164] program sander. The reaction coordinate ζ specifies a dihedral angle rotating from -180 (conf1) through 0 (conf2) to +180° with a restrain constant of 100 kcal/(mol*rad^2). In addition, heavy atom positional restrains for lapatinib was enforced with the exception of the dihedral angle about which the fluorophenyl ring was sampled. For comparison, two different methods, employing Berendsen[195] and Langevin[170] thermostats were performed. For the Berendsen method, the initial model was a snapshot from the well equilibrated NPT ensemble obtained from a 20 ns MD explicit solvent simulation of the complex. All 37 reference structures, with specific dihedral angles (-180 to +180) driven every 10° from this initial structure, were simulated for 1 ps equilibration followed by 500 ps of production MD. With regard to the Langevin method, the initial models were generated manually using the software MOE followed by 500 ps equilibration and 1000 ps production run using Langevin NPT dynamics. For both methods, data from the production runs were observed to converge in the PMFs with differences of less than 1.0 kcal/mol. All the results were analyzed using WHAM program[190,191,194] with periodicity across all 37 simulations. WHAM calculations were performed using the convergence tolerance of 0.0001 at a temperature of 298.15K. Ninety bins were used for the histogram to construct the histogram from the number of counts in each bin versus the dihedral angle. Significant overlap between adjacent windows was observed in the histograms which is important for accurate PMF construction.

3.3.3 MM-GBSA

Based on the previous results in Chapter 2, binding energy calculations using the MM-GBSA method without any waters (0WAT protocol) did not achieve satisfactory agreement with experimental activities. In contrast, calculations that included key water molecule in the binding site (1WAT or 3WAT protocol) lead to excellent numerical agreement. Therefore, water molecules were also included in the MM-GBSA calculations in this chapter comparing binding energy of lapatinib in conf1 and conf2. As before, a four-term equation [$\Delta G_b(calcd) = \Delta E_{vdw} + \Delta E_{coul} + \Delta\Delta G_{polar} + \Delta\Delta G_{nonpolar}$] was used consisting of intermolecular van der Waals energy (ΔE_{vdw}), intermolecular Coulombic energy(ΔE_{coul}), changes in polar hydration energy ($\Delta\Delta G_{polar}$), and changes in nonpolar hydration energy ($\Delta\Delta G_{nonplar}$). Changes in solute vibrational, rotational, and translational entropy were omitted. The GB model described by Onufriev et al[113] (igb =5) was used for polar hydration energies using standard constants, mbondi2 radii, and interior and exterior dielectric constants of 1 and 78.5, respectively. Solvent accessible surface areas (SASA) are used to estimate nonpolar hydration energies through $G_{nonplar} = \gamma * SASA + \beta$ ($\gamma = 0.00542$ kcalmol^{-1} Å^{-2} and $\beta = 0.92$ kcal/mol).

3.4 Results and Discussion

3.4.1 Potentials of Mean Force (PMFs) for Driving Conf1 to Conf2

As our first experiment to explore the two conformations, we performed umbrella sampling followed by WHAM analysis to derive PMFs as shown in Figure 3.2. Importantly, both methods reveal that conf1 is more favorable by about 8 kcal/mol with EGFR, by about 5-8 kcal/mol with HER2, and by about 2-3 kcal/mol with ErbB4. Both PMFs show conf1-conf2 ΔG differences for ErbB4 about 2-3 kcal/mol which is about one to two orders of magnitude

difference (1.4-2.8 kcal/mol). In contrast, the conf1-conf2 differences for EGFR or HER2 are much greater (5-8 kcal/mol) suggesting low conf2 population. Interestingly, both thermostats showed that the width of the energy barrier itself for the conf1 to conf2 transition at ca 80 degrees is somewhat narrower and higher for for ErbB4 (Figure 3.2 blue line) versus EGFR or HER2. Notably, despite the fact that two different thermostats were used, it is remarkable that both methods, initiated using different starting structures, and using different protocols, yield very similar shape and trend across all three species.
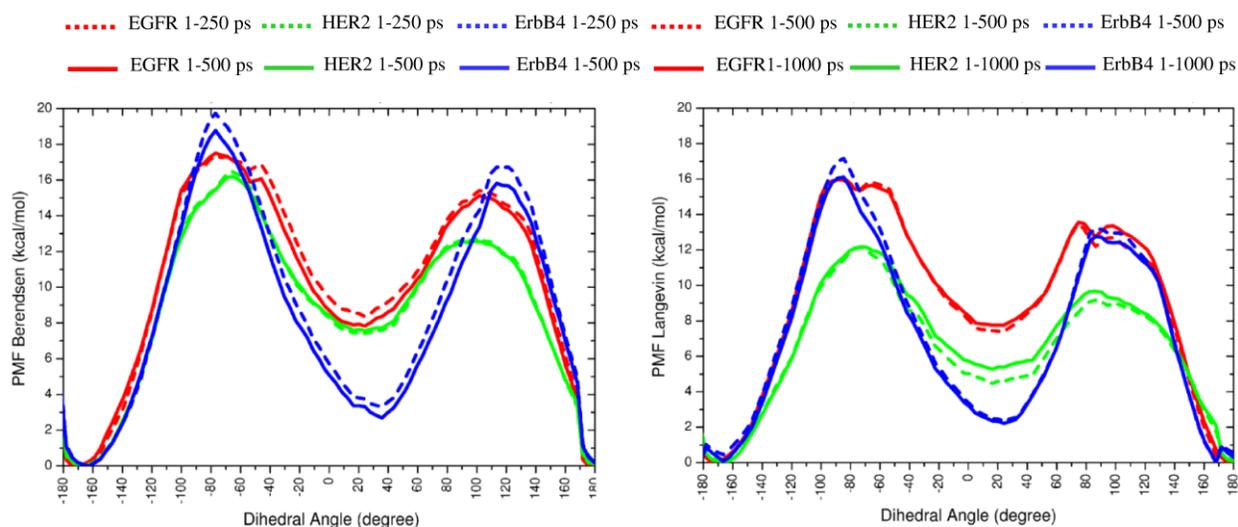
Figure 3.2. WHAM-derived PMFs for the rotation of the fluorophenyl ring of lapatinib while in complex with EGFR (red), HER2 (green), or ErbB4 (blue) using umbrella sampling in 10° increments. The ligand dihedral angles of ca -180 and 20° represents conformation 1 and conformation 2 respectively.

3.4.2 Conf2 Stabilization in ErbB4

In order to probe why conf2 could be stabilized in ErbB4 versus EGFR or HER2, footprint analysis was plotted for per-residue VDW, Coulombic, and H-contributions as shown in Figure 3.4. The results here were derived from the snapshots used in the companion MM-GBSA studies, described in the next section, consisting of six simulations initated in each

conformation (12 independent simulations). Footprints were averaged over 6 simulations and showed on average less than 0.4 kcal/mol standard errors of the mean indicating good convergence. The footprints indicate that only the Columbic interactions show any appreciable changes across the three proteins; all VDW plots are well-overlapped. At position 800 ErbB4 gains about 4 kcal/mol more favorable Columbic energy relative to EGFR and HER2 which is likely due to the sequence variance at this position where ErbB4 has more bulky charged residue glutamic acid E instead of aspartic acid (D) as EGFR and HER2. In addition, visualization of the trajectories shows the lapatinib tail in a somewhat more stable conformation while bound to ErbB4 (Figure 3.5) and in an extended conformation pointing towards residue 800 instead of the internal quinazoline ring. The result is a single ligand tail population versus two populations seen in the other system. The coexisting of both conf1 and conf2 may also contribute to the weaker binding observed experimentally for lapatinib with ErbB4 relative to the other receptors.
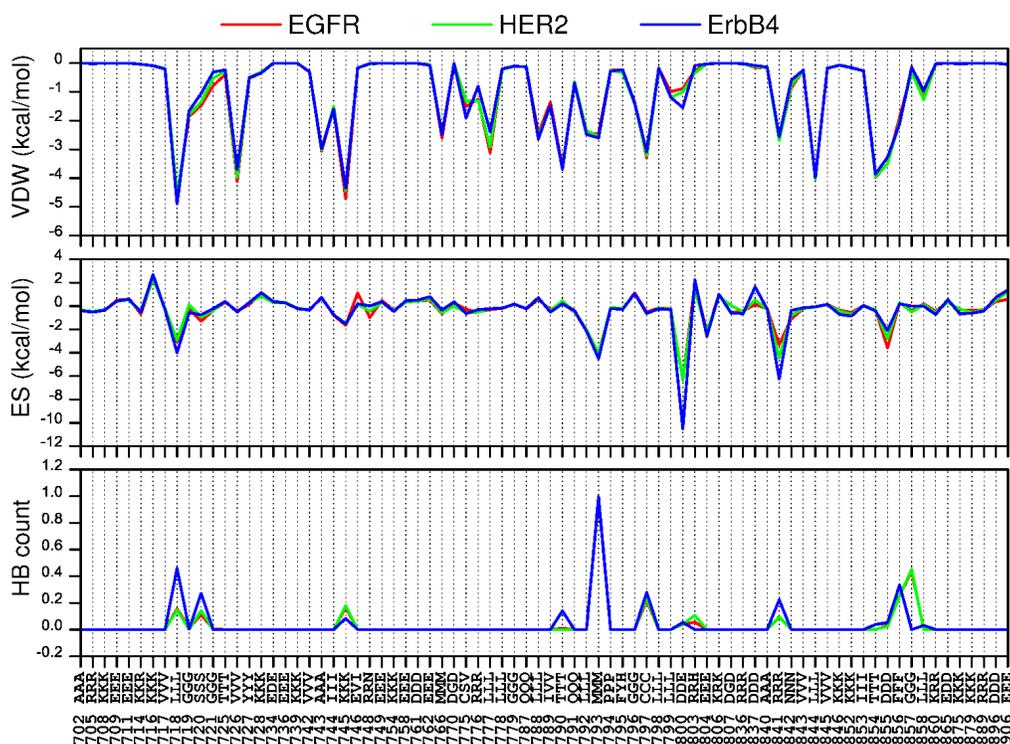
Figure 3.3. Per-residue footprints for lapatinib with EGFR (conf2), HER2 (conf2), and ErbB4 (conf2).
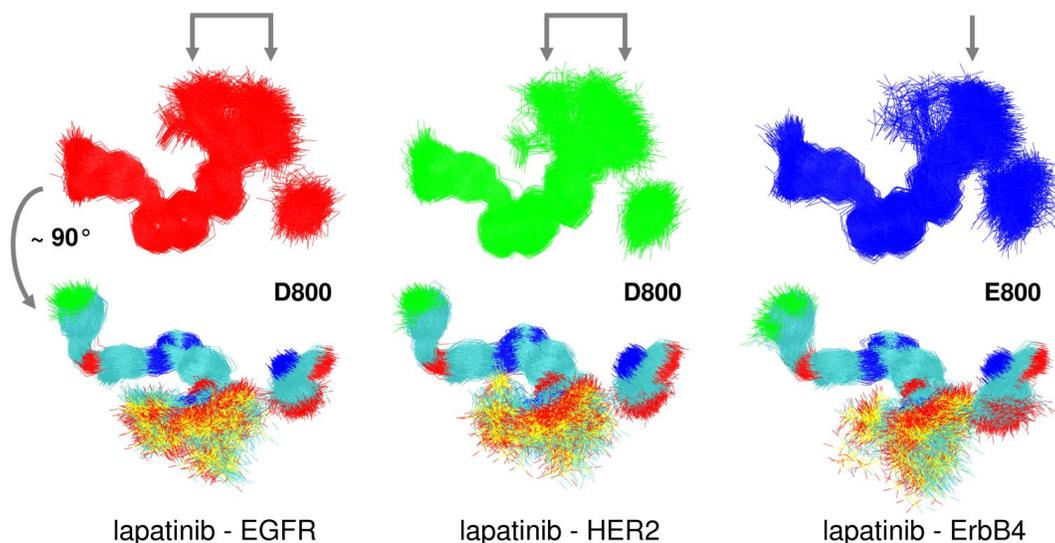
Figure 3.4. Comparison of lapatinib sampling (conformation 2) in complex with EGFR (red), HER2 (green), or ErbB4 (blue). For each system, 1200 evenly spaced MD snapshots are shown which were derived from 6 combined 20 ns simulations. Arrows highlighted the movement of the tail of lapatinib. The fluctuations of lapatinib were shown in two perspectives (current and ~90º).

3.4.3 MM-GBSA Result

As an alternative approach, MM-GBSA analysis was also employed to estimate the two conformational preferences. As shown in Table 3.1, based on results using the 3WAT protocol and 1WAT protocols (described in Chapter 2), lapatinib in conf2 with EGFR, HER2, and ErbB4 always yields less favorable binding energies (ca. 2-4 kcal/mol) than conf1. However, in sharp contrast to the results obtained using PMFs, conf2 is shown here to be less energetically favorable when bound to ErbB4. While the inconsistency between the PMF (smaller conf1-conf2 difference for ErbB4), MM-GBSA results (larger conf1-conf2 difference for ErbB4) is perplexing, possible reason for the discrepancy may involve the fact that the umbrella sampling enforces constraints on the both dihedral angle and the general position of lapatinib. Other reasons may be differences associated with changes in entropy. In any event, the MM-GBSA

results consistently support the fact that conf1 is more favorable than conf2 across all three proteins.

Table 3.1. Calculated vs experimental free energies of binding (kcal/mol).

Absolute	3WAT ΔG_b calcd	1WAT ΔG_b calcd
EGFR conf1	-68.80	-65.84
EGFR conf2	-66.90	-63.61
ΔEGFR conf1 – conf2	-1.95	-2.23
HER2 conf1	-67.82	-64.53
HER2 conf2	-65.20	-62.04
ΔHER2 conf1 – conf2	-2.59	-2.49
ErbB4 conf1	-67.19	-62.93
ErbB4 conf2	-63.04	-59.62
ΔErbB4 conf1 – conf2	-4.15	-3.73

Energies in kcal/mol.

In terms of conformational flips, it should be noted that for all MM-GBSA simulations involving EGFR and HER2 all 6 simulations initiated in conf1 stayed in conf1 and all 6 simulations initiated in conf2 stayed in conf2. However, for ErbB4, although the 6 simulations initiated in conf1 stay in conf1, one of the simulations initiated in conf2 flips to conf1. While this could indicate that the alternative lapatinib conformation (conf2) seen in crystal structure in 3BBT is able to sample both conformations, additional computational studies would be required to derive more significant statistics. Interestingly, examination of B factors in the original crystal structures shows much higher average values for the ligand in 3BBT (100.2 Å^2) versus 1XKK (41.8 Å^2) or 2R4B (57.99 Å^2) which suggests high mobility (low stability) for the 3BBT structure in conf2.

3.4.4 Population Analysis

Finally, to explore the relative populations in greater detail, 3D mesh diagrams were plotted using a key lapatinib dihedral angle (C21, C20, C19, and O4) and internal distance (F to Cl) as shown in Figure 3.5. The energy wells represented here were estimated from the free energy relationship $\Delta G = -RT \ln(P)$ from populations derived during the MM-GBSA simulations. As shown in Figure 3.5, in addition to the global free energy minimum local minimums coexist which indicate dynamic behavior and conformational diversity. As expected, lapatinib energy patterns are similar for EGFR and HER2 but are different for ErbB4. Based on these mesh plots, EGFR and HER2 share four similar energy wells, two of which represent conf1 and two of which represent conf2. In contrast, ErbB4 has one well with another very small well for conf1 and three for conf2.
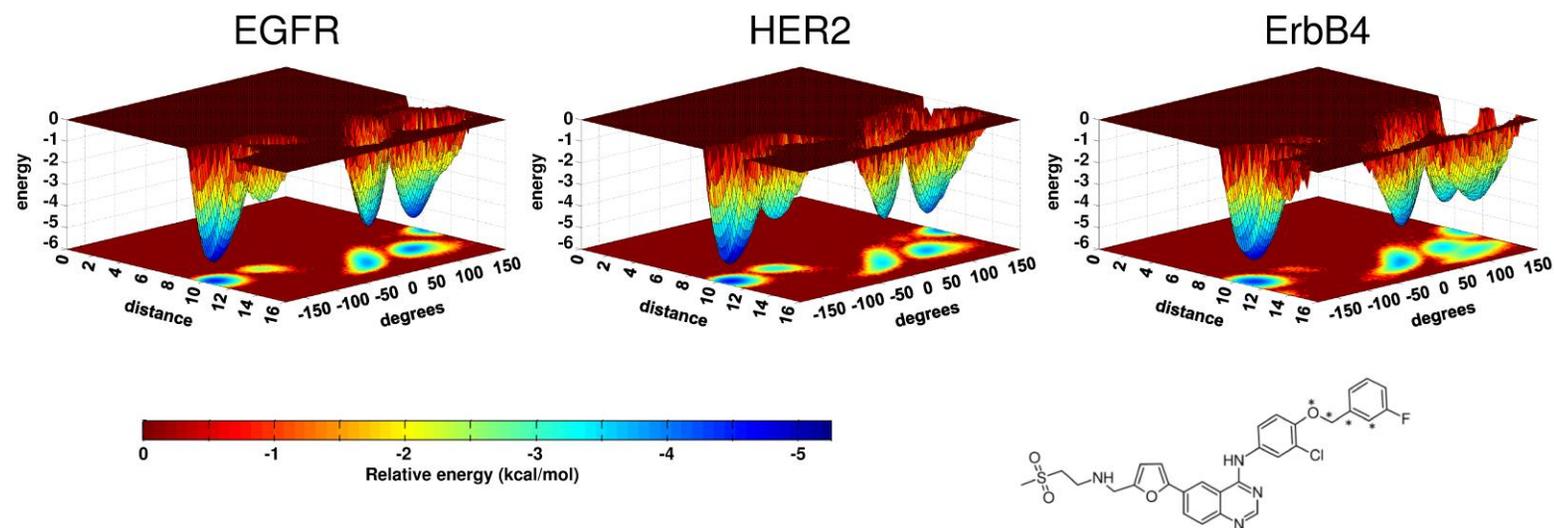
Figure 3.5. A 3-D free energy diagram showing lapatinib conformation transformation pathways. Dihedral angle is plotted between C21, C20, C19, and O4 which are indicated by *. Distance is plotted between F and Cl of lapatinib. -180º and 0º indicate conf1 and conf2, respectively. The distance between F and Cl larger than 15 Å indicating conf2, smaller than 12 Å indicating conf1. The data was obtained from 24 20 ns simulations based on template 1XKK and 2R4B. $\Delta\Delta G$ = -RTln (pi/p).

In agreement with the PMF results, the deeper energy wells for conf1 compared to conf2 correspond to the global energy minimum for these two reaction coordinates in all three proteins. Focusing on conf1, both EGFR and HER2 show two well-formed deep deeper wells while ErbB4 has one deep well but one shallow well (Figure 3.5). With regard to conf2, ErbB4 has one additional well compared to EGFR and HER2 with what appears to be lower energy barriers in this region which suggests the conformational transition from conf2 to conf1 for ErbB4 should be easier than other systems and this could promote ring flips. As potential evidence, among all six simulations starting from conf2, only simulations in ErbB4 ended up showing a flip to conf1, however, no reverse flips were seen.

3.5 Conclusions

In summary, we have performed umbrella sampling and PMF calculations (Figure 3.2) to evaluate the energetic effects of switching between two conformations of lapatinib while bound to three different receptors (EGFR, HER2, and ErbB4). In order to understand why conf2 stabilizes ErbB4 more than EGFR and HER2, footprint analysis was performed. Footprint results (Figure 3.3) indicate that enhanced columbic interactions between the tail of lapatinib with residue E800 appear to be the main reason for conf2 stabilization with ErbB4. Ultimately, however, the more favorable VDW interactions between lapatinib with the glycine-rich loop and residue 855 in the DFG motif likely lead to a more favorable free energy which drives the population towards conf1. In addition, free energy analysis (MM-GBSA) was performed (Table 3.1) using the same molecular dynamic results used in the footprint analysis which provides further evidence that conf1 is in the more favorable conformation validating use of that geometry for the studies in Chapter 2. And, population analysis was performed to explore the relative

populations and energies. The results indicate that conf2 always yields a less favorable energy than conf1 with EGFR, HER2 and ErbB4. The larger conformational diversity of the energy wells observed around conf2 with ErbB4 may help to promote the ring flips observed in the MM-GBSA simmulations.

Chapter 4. Calculation of Relative Binding Free Energies for a Series of Related Inhibitors which Target IGF-IR

4.1 Introduction

As discussed in the introduction and Chapter 2, receptor tyrosine kinases (RTKs) serve as key regulators for a variety of signaling pathways involving differentiation, proliferation, and apoptosis and are valid cancer targets.[196,197] A variety of small drugs have been developed to inhibit one or more RTKs, including imatinib (Gleevec), ertinib (Tarceva), gefitinib (Iressa), lapatinib (Tykerb) and sofafenib (Nexavar).[198,199] The RTK called IGF-IR is also associated with increased the risk of colon, breast, protostate and lung tumors and has become a key oncology target.[200,201]   IGF-IR is a tetrametic transmembrane tyrosine kinase consisting two $\alpha\beta$ heterodimers which are disulphide-bonded.[202] Binding of IGF-1 or IGF-II ligands to the IGF-IR extracellular α subunit triggers a conformational change that leads to phosphorylation of specific tyrosine residues in the cytoplasmic domains and activation. IGF-IR activation leads to autophosphorylation of activation-loop tyrosine residues (Tyr 1135, Tyr 1131 and Tyr 1136)[203] which can enhance the activation of downstream intracellular substrates including IRS-1 and Shc.[28,204,205] This will finally lead to activation of a variety of signaling pathways inducing Ras/Raf/MEK/ERK and PI3K/Akt/mTOR pathways.[28,204,205]

A promising kinase inhibitor in early stage clinical testing is called OSI-906[206] and can inhibit proliferation in a variety of tumor cell lines by inhibiting IGF-IR which blocks ligand-induced activation of a variety of downstream pathways. Importantly, there is a wealth of structure activity relationship information for OSI-906 analogs, containing a common imidazopyrazine scaffold, and there is an available crystallographic structure of an early lead

compound called PQIP. The goal of the work in this chapter is to computationally characterize structure activity relationships for OSI-906 and related ligands using the method of thermodynamic integration (described in the Introduction). The intermediate form of IGF-IR bound with PQIP (pdb code 3D94)[23] was used as the master protein-ligand complex for constructing 12 related complexes containing each of the analogs shown in Figure 4.1.

Specific objectives of the work in this chapter are: (1) develop a general computational protocol for computation of *relative* binding free energy using the thermodynamic integration (TI) method, (2) computationally quantify the structure activity relationship (SAR) for a series compounds of IGF-IR inhibitor with an imidazopyrazine scaffold, and (3) determine how changes in the functional groups lead to variations in binding specificity. Use of well-validated models and TI calculation setups will ultimately enable more accurate prediction of molecular recognition at the atomic level for IGF-IR and other clinically related targets.

			scaffold 1				scaffold 2	Compound PQIP
No.	SBU code	X	R1	R2	R3	R4	Cell IC50(μM)	ΔG_{bind}(kcal/mol)
0	IntA	NH2	N	H	H	H	N/A	N/A
1	Q01	NH2	C	H	H	H	>10.0	>-6.82
2	Q02	OH	C	H	H	Ph	>10.0	>-6.82
3	Q03	NH2	C	H	H	Ph	0.086	-9.63
4	Q04	NH2	N	H	H	Ph	0.335	-8.83
5	Q05	NH2	C	OH	H	Ph	0.089	-9.61
5a	Q05a	NH2	N	OH	H	Ph	N/A	N/A
6	Q06	NH2	C	OH	Methyl	Ph	0.021	-10.47
7	Q07	NH2	N	OH	Methyl	Ph	0.140	-9.34
8	Q08	NH2	C	OH	Ethyl	Ph	0.028	-10.30
9	Q09	NH2	N	OH	Ethyl	Ph	0.279	-8.94
10	Q10	NH2	C	OH	Ph	Ph	0.225	-9.06
11	Q11	NH2	C	OH	CH2-Ph	Ph	>0.500	>-8.59
12	Q12	NH2	C	N/A scaffold 2		Ph	1.04	-8.16

Figure 4.1. A series of compounds used in relative free energy calculations. Ligands from No. 0-11 are build using scaffold 1 and No. 12 uses scaffold 2. IntA is an intermediate structure between Q01 and Q04. Q06 is OSI-906. PQIP 2-D structure is shown on the right panel. Cellular IC50 values for Q01, Q02 and Q12 are obtained from Mulvihill, M.J. et al. *Bioorg. Med. Chem.* 16 (2008) 1359-1375 and others are from Volk, B. et al., 40th Middle Atlantic Regional Meeting (MARM 2008) of the American Chemical Society (poster #313). Experimental binding energy was estimated using ΔGb exptl ≈ RT ln (IC50) at 298.15K.

4.2 Computational Details

4.2.1 Model Construction and Parameter Files

The starting structure for the protein-ligand complex of IGF-IR with PQIP was taken from the 2.3 Å X-ray crystal structure (pdb code 3D94).[23] The missing regions (residues 1066-

1256) in 3D94 was constructed from crystal structure 1JQH[207] using Modeller.[81] The model with the lowest discrete optimized protein energy (DOPE)[87] was selected and subsequently evaluated using PROCHECK[88] which showed ~90% resides were in the most favored regions. PQIP was used as the parent compound for constructing a series of ligands. The scaffold of PQIP (quinolone and imidazopyrazine) was maintained in the same pose in the binding pocket and the remaining part of the ligands were built using MOE.[163] Functional group rotamers, were adjusted to avoid clashes with the protein environment subject to visual inspection. The same forcefield parameters used as in Chapter 2 were also used here: FF99SB[165] for the protein, GAFF[166] for the ligand, and TIP3P[167] for water. Ligand partial charges were computed at the HF/6-31G* level of the theory using Gassian98.[169] All the complexes/ligands were assembled using AMBER and solvated so that each complex/ligand had the exact same periodic box. For all ligand pairs for a given TI setup, the atom names for the same atoms for identical atoms in both states were kept the same and all common atoms were manually adjusted to appear in the same order and the same place in the parameter files. All unique atoms (i.e. perturbed atoms) were placed at the end of each prmtop file.

4.2.2 TI Simulation Setups and MD Details

The AMBER 11 suite[208] was used for molecular dynamics and TI simulations. A nine step of equilibration procedure was used to relax the protein and solvent with gradually decreasing restraints. As shown in Table 4.1, the equilibration consists of 4 steps of minimization (1000 steps of steepest decent) and 5 steps of short molecular dynamic simulations (50 ps). For the bound equilibration, the restraint was enforced on all the atoms except H atoms from 1st to 7th steps while for the last 8th and 9th steps the restraints were only enforced on

backbone atoms. The restraints constants for minimization were gradually reduced from 5.0, 2.0, 0.1 to 0.05 (kcal/mol/$\AA^2$) and for short MD runs were from 5.0, 1.0, 0.5, 0.25 to 0.1 (kcal/mol/$\AA^2$). In the final 8th and 9th step, restraints were also added on ligand atoms C1 and C2 to minimize problems with the inhibitor drifting out of the binding pocket near the endpoints (λ = 0.05 and 0.95) simulations. For production runs, no restraints were added for the bound state runs and for the unbound state runs, a weak restraint of 0.1 kcal/mol/$\AA^2$ was required for the system to avoid the ligand wandering out of the solvent box.

Table 4.1. Restraint used for TI calculation.

Process	Restraint (Bound)	Restraint (Unbound)	Restraint Weight
01 min	everything except H	everything except H	5.0
02 md	everything except H	everything except H	5.0
03 min	everything except H	everything except H	2.0
04 min	everything except H	everything except H	0.1
05 min	everything except H	everything except H	0.05
06 md	everything except H	everything except H	1.0
07 md	everything except H	everything except H	0.5
08 md	protein backbone	atom C1,C2	0.25
09 md	protein backbone	atom C1,C2	0.10
10 md	no restraint	atom C1,C2	0.10

The MD simulations were performed in the NPT ensembles maintaining a constant pressure of 1 bar. Simultaneously, softcore potentials for both the VDW and electrostatic terms were employed which simplified the calculations, and 19 windows were used with λ consecutively ranging from 0.05 to 0.95 at the interval of 0.05. The system temperature was set at a constant temperature of 298.15 K using a Langevin thermostat[170] with the collision frequency of 1 ps^{-1}. Particle mesh Ewald[172] method was used for long-range electrostatic calculation (8 $\AA$ for cutoff) with periodic boundary conditions. Both MD equilibration and

production runs employed a 1 fs time step. For each ligand pair, 2 ns MD production was performed for both the bound state and unbound state.

4.2.3 Relative Binding Free Energy and Error Estimation

Numerical integration for thermodynamic integration can be approximated by using trapezoidal rule shown as in Figure 4.2. The $dV/d\lambda$ (kcal/mol) was calculated at each λ (from 0.05 to 0.95 with interval 0.05) for both the complex and the ligand. It is important to note that the result at endpoints $\lambda = 0$ or 1.0 cannot be obtained directly using AMBER11 (under development for AMBER12) and to estimate the missing free energy, a linear interpolation using the previous 2 data points is used. For example, the $dV/d\lambda$ at $\lambda = 0$ can be derived from the curve for 0.05 and 0.10 (Figure 4.2 red) Relative energy change for the bound state or unbound states can be obtained by summing up all the individual trapezoid areas. The relative binding free energy between any two ligands is obtained by subtracting the unbound from the bound results ($\Delta\Delta G_{bind} = \Delta\Delta G_{bound} - \Delta\Delta G_{unbound}$). For each individual transformation and λ value, the standard deviation of the mean of $dV/d\lambda$ was calculated. Then the standard deviation of bound and unbound states ($error_{com}$ and $error_{lig}$) was obtained by summing the variances and using the trapezoid formula. The error for the complex was calculated using the simple equation

$$error_{bind} = \sqrt{error_{com}^2 + error_{lig}^2} / 2 \,.$$
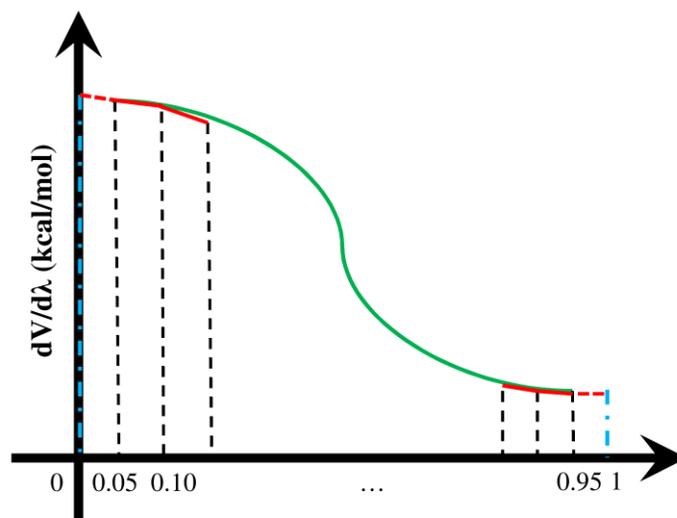
Figure 4.2. Trapezoidal rule for calculating the area under the curve.

4.3 Results and Discussion

4.3.1 Null Transformation

To evaluate our model construction and simulation robustness, null transformations (mutating a ligand to itself which should result in a zero change in free energy) were performed. Several simulations using the same disappearing and appearing functional groups (C3, Methyl, NH2, OH, and Ph) were employed for Q06->Q06 null transformation calculations. Since null transformation should yield no change of relative free energy of binding, any large deviation from 0 in the calculation can be attributed from insufficient sampling during the MD simulation or protocol errors. Encouragingly, the relative binding free energies for these five null transformation in Table 4.1 are close to zero (results ranging from -0.21 to 0.05 kcal/mol). In addition, low errors and uncertainties (≤ 0.13 kcal/mol for the unbound state and ≤ 0.10 kcal/mol for the bound state) are obtained. Interestingly, null transformation for the polar groups NH$_2$ (-0.21) and OH (0.05) yields the larger energy difference and error than other transformations.

88

Table 4.2 Null transformation results (energies in kcal/mol). Disappearing and appearing groups are unique to the initial and final states of the transformation.

Transformation	Disappearig functional group	Appearing functional group	Exptl $\Delta\Delta G_{bind}$	Calcd $\Delta\Delta G_{bind}$	Calcd Transformation Bound	Calcd Transformation Unbound
Q06 -> Q06	C3	C3	0.00	−0.01±0.02	0.00±0.01	0.01±0.02
Q06 -> Q06	Methyl	Methyl	0.00	0.02±0.08	0.01±0.07	−0.00±0.06
Q06 -> Q06	NH$_2$	NH$_2$	0.00	−0.21±0.12	−0.20±0.08	0.01±0.13
Q06 -> Q06	OH	OH	0.00	0.05±0.09	−0.17±0.07	−0.22±0.07
Q06 -> Q06	Ph	Ph	0.00	0.01±0.11	0.01±0.10	0.01±0.08

4.3.2 Ligand Pair Result

For comparison with available experimental activity, Table 4.3 shows the TI results for transformations between 13 ligand pairs. Figure 4.3 plots the results graphically. Among these examination of the results show pairs involving small aliphatic changes are in close agreement with the experimental results. For example, simulations for the pairs Q05 -> Q06 (0.07), Q06->Q05 (0.31), Q06 -> Q08 (-0.44), Q07 -> Q09 (-0.04) are very close to the experimental results with deviation less than ±0.5 kcal/mol from the experimental data. On the other hand, all five polar transformations yield significant errors.

Table 4.3. Relative binding free energy results (kcal/mol) for 13 ligand pairs using TI calculation.

Transformation	Disappearig functional group	Appearing functional group	Experimental $\Delta\Delta G_{exptl}$	Computational $\Delta\Delta G_{calcd}$	Computational Transformation For complex	Computational Transformation For ligand
Q03 -> Q04	C	N	0.80	7.82±0.20	-102.62±0.18	-110.44±0.21
Q03 -> Q02	NH$_2$	OH	>2.81	-6.57±0.10	-25.67±0.13	-19.10±0.15
Q03 -> Q12	H	N/A	1.47	1.89±0.05	-4.31±0.08	-6.20±0.05
Q05 -> Q03	OH	H	-0.02	3.54±0.03	8.15±0.05	4.61±0.04
Q05 -> Q06	H	Methyl	-0.86	-0.93±0.08	48.70±0.11	49.63±0.11
Q06 -> Q05	Methyl	H	0.86	0.55±0.03	-41.30±0.04	-41.85±0.04
Q06 -> Q08	Ethyl	Methyl	0.17	0.61±0.07	-63.00±0.12	-63.61±0.08
Q06 -> Q07	C	N	1.13	6.82±0.01	-107.90±0.20	-114.72±0.20
Q07 -> Q09	Methyl	Ethyl	0.40	0.44±0.08	-53.52±0.12	-53.96±0.12
Q08 -> Q10	Ethyl	Ph	1.24	4.99±0.11	19.44±0.14	24.43±0.16
Q08 -> Q11	Ethyl	CH$_2$-Ph	>1.71	2.05±0.11	115.02±0.16	112.97±0.16
Q09 -> Q08	N	C	-1.36	1.62±0.06	102.71±0.07	101.09±0.09
Q11 -> Q06	CH$_2$-Ph	Methyl	<-1.88	0.12±0.10	-37.66±0.14	-37.78±0.13

Outliers involving polar groups N/C, C/N, NH_2/OH and, OH/H are shown in red and green in Figure 4.3. Among these transformations, although the binding energy is significantly overestimated, it is encouraging that the connect sign (less favorable $\Delta\Delta G$) is obtained in 3 out of 5 cases. Nevertheless, the significant scatters suggest a systematic problem for polar transformations which needs further study.
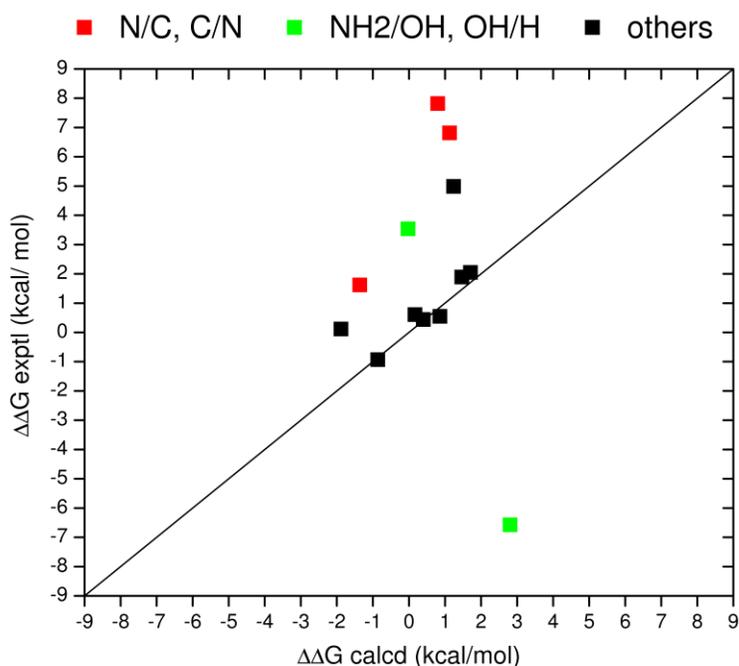
Figure 4.3. The correlation plot for TI calculation in comparison with experimental activity data for 13 ligand pairs. Transformations involving only aliphatic groups shown in black, polar groups shown in red and green.

In addition to null transformation, both forward and backward transformation were performed for the same ligand pair (Q05->Q06 and Q06->Q05) to evaluate the accuracy of the simulation results. This forward and backward transformation yielded similar relative binding free energy results (-0.93 and +0.55 kcal/mol) which are consistent with the experimental results

(-0.86 kcal/mol and +0.86 kcal/mol). Although additional studies are wanted, the transformation direction appears to have little influence on the results.

4.4 Challenges for TI

4.4.1 Initial Setups

Although TI is an elegant theory that in principal should allow for $\Delta\Delta G_{bind}$ to be readily computed, in practice there are a number of setup and protocol challenges. For example, the AMBER package requires that all the atom names be identical across all ligand pairs which is not easily obtained since many software packages may change atom names when building a series of ligands. Additionally, the common atoms between ligand pairs had to be in the same order starting from the same position which, in our present case, must be adjusted manually. In addition, the setup of the water box must be identical for every complex and ligand so that the solvation effect is the same for all ligands. Finally, when the ligand conformation is not known, it can be difficult to determine the robustness for functional groups.

4.4.2 Restraint Mask Selection

Restraint mask selection is a difficult task for thermodynamic integration in that special considerations must be made to prevent the ligand from moving unnaturally during the simulation. For the extreme case when all atom ligands are in the mask (i.e. all atoms are transformed), especially required at the end points of the simulations to prevent the bound state ligands leaving the binding pocket, all atoms except two common atoms were set as the restraint mask so that movement were coupled for the given ligand pair. And, to avoid unbound state ligands leaving the center of the water box, weak restraints were enforced on the scaffold.

However, preliminary studies showed the larger masks can lead to convergence problems for null transformations (>1 kcal/mol) thus a second protocol using much smaller mask (only several atoms are included) were employed. Smaller masks led to low errors and relative binding energy close to 0 kcal/mol for the null transformations however the above noted atom naming and atom ordering challenges were manifested.

4.4.3 Endpoint Singularity

Finally, the AMBER 11 package does not specifically include the endpoint calculations ($\lambda = 0$ and $\lambda = 1$) for thermodynamic integration. The way to overcome this endpoint singularity is to extrapolate the two points closest to each endpoint. Although this will introduce some error in the calculation, multiple windows that are sufficiently close to each end point will help to minimize the error.

4.5 Conclusions

We have calculated the relative binding free energy for 13 ligand pairs with imidazopyrazine scaffold which target the intermediate form of IGF-IR using thermodynamic integration. With 2 ns simulations, we can obtain 0~0.2 kcal/mol relative binding free energy with 0~ 0.13 kcal/mol standard error of the mean for null transformations, suggesting the model construction and simulation protocols are robust. With this standard TI protocol, we can estimate the difference in binding free energy of 6 pairs within 0.5 kcal/mol deviations from the experimental data however 7 pairs yield large errors. Consistent with the null transformation calculation, transformations involving a polar group (N, OH, etc) yield, in general, large deviations from experiment. Although considerable progress was made in establishing a lab

protocol for performing TI, additional studies to identify the source of large errors for polar

transformation should be performed, ideally, using several different SAR data sets.

Chapter 5. Virtual Screening to Identify Small Molecule Drug Leads as "Dual inhibitors" for Combinations of EGFR/HER2, HER2/IGF-IR and EGFR/IGF-IR.

5.1 Introduction

ErbB family members and IGF-IR are important kinase targets for breast cancer, and the cross-talk between these proteins have been implicated in resistance to several inhibitors including trastuzumab (Herceptin, targeting HER2),[24,26-28] gefitinib (Iressa, targeting EGFR),[25,34] and BMS-536924 (targeting IGF-IR) among others. Emerging clinical studies strongly support that co-targeting IGF-IR and ErbB members could contribute to therapies with enhanced resistance profiles and ultimately better patient outcomes.[27,35-37,209-211]   Focusing on small molecules, several series of multi-targeted kinase inhibitors for IGF-IR and ErbB family members are actually being developed.[212] For example, Hubbard et al[212] have reported that inhibitors with a novel pyrazolopyrimidine scaffold can multitarget IGF-IR, EGFR and HER2 with balanced enzyme and cellular inhibitory profiles. Fidanze et al[213] have reported that inhibitors based on an imidazothiazole scaffold has triple inhibitory activity against IGF-IR, EGFR and HER2. Co-targeting HER2 and IGF-IR may provide benefit for patients with trastuzumab-resistant HER2-postive breast cancer and co-targeting EGFR and IGF-IR may provide benefit for patients exhibiting anti-EGFR inhibitor resistance.

Docking-based virtual screening (VS), a widely used computational technique, in modern drug discovery, has been successful in searching for inhibitors for kinases, including BCR-ABL,[214] Chk1,[215] VGFR, [216] FGFR,[217] and CDK[218] among others. Remarkable efforts have been made toward identifying ATP-competitive inhibitors of the ErbB family and FDA-approved drugs include erlotinib (Tarceva) and gefitinib (Iressa) which target an active form of EGFR, and the dual

inihibitor lapatinib (Tykerb) which targets the inactive form of EGFR and HER2. Although no FDA-approved drug for IGF-IR have been reported, a series of compounds are under development (see Chapter 4). Importantly, IGF-IR has three standard forms: termed active, semi-active, and inactive and each of these forms lead to two different conformations of the activation loop. Available crystallographic structures include: 2ZM3,[219] 3F5P[220], and 3QQU[221] (active forms); 3D94,[222] and 3LVP[221] (intermediate form); and 3NW5, 3NW6, 3NW7,[223] 2OJ9,[224] and 3O23[223] (inactive forms). Inhibitors with benzimidazoles[224] (BMS), isoquinolinedions[219] (Wyeth), 3-cyanoquinolines[220] (Wyeth), pyrazolopyrimidines[225] (Abott) and bis-azaindole[221] (Sanofi-aventis) have also been reported. Studies in this chapter target IGF-IR in the intermediate form. The goal is to employ virtual screening to identify for subsequent compatibility analysis with EGFR and HER2 to identify leads with favorable binding profiles to various combinations of kinase receptors.

5.2 Theoretical Methods

5.2.1 Molecular Docking

The program DOCK employed in this chapter for virtual screening is designed to identify molecules based on the concept of shape matching complementary with a drug target (typically a protein). The process consists of 4 steps: (1) receptor and ligand structure preparation; (2) sphere generation and selection; (3) grid generation; (4) docking. DOCK identifies binding sites by employing a collection of overlapping spheres which are determined using the surface of the receptor protein. Spheres (shown in Figure 5.1) are used to guide initial ligand placement in the binding site by matching a subset of ligand atoms (termed anchors) to the sphere centers. An ensemble of anchor orientations can be generated by using different sets of atoms and sphere

centers for matching. Docking evaluates the poses based on force field energy scores which consist of intermolecular VDW and Coulombic energies as described in eq (1.4). In most instances, grids are used to speed up calculations.[102] For every grid point, the VDW and columbic potentials between the receptor and a probe are precalculated within a user-defined cutoff from the center of the binding site. Grid energy can then be approximated using a tri-linear interpolation (TLI) of nearest grid points for the interaction with nearby ligand atoms. Figure 5.1 graphically depicts a docking setup for a representative kinase system.
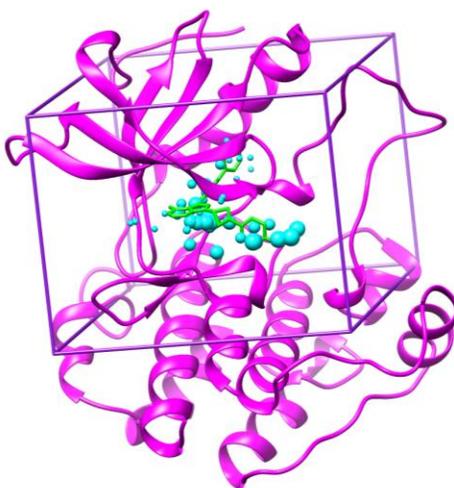
Figure 5.1. Representation of docking setups using crystal structure 3D94 (IGF-IR with PQIP bound). The protein is shown in magenta, the ligand in green, energy grids as purple box, docking spheres as cyan balls.

5.2.2 Footprint Similarity Score (FPS)

Work in this chapter also employs a footprint similarity (FPS) scoring function recently developed in our lab which employs residue-based decompositions of van der Waals energies (VDW), electrostatic energies (ES), and hydrogen bonds (HB) to identify ligands with similar 2-D energy pharmacophores (termed a footprint) as a known reference compound. The known reference compound can be a FDA-approved drug, a native cofactor or substrate, or a known

inhibitor. A vector $\vec{x} = [x_1, x_2, \ldots x_N]$ is used to define a footprint where x_i represents the interaction energy between the ith residue of the receptor and the ligand and all N residues of the receptor are included in the calculation. Footprint similarity is used to quantify the likeness between two footprints vectors $\vec{x}$ and $\vec{y}$ using metrics such as Euclidean distance (d), described by $d = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2}$. Work in this thesis employed normalized Euclidean distance (d_{norm}) which is simply the normalized footprint vectors. For comparison, similarity measures based on the standard Pearson correlation coefficient $r = \mathrm{cov}(\vec{x}, \vec{y}) / \sqrt{\mathrm{var}(\vec{x})\,\mathrm{var}(\vec{y})}$ are also shown.

5.2.3 Virtual Screening.

Virtual screening is an effective paradigm for the prediction of binding geometries (termed poses) and provides a means to rank order compounds in a database through estimation of the interaction energy. As a tool for virtual screening, docking programs are evaluated primarily based on the accuracy of reproducing the crystal pose of known ligands and also the ability to enrich known active ligands from a large dataset of decoy compounds.[98-100] Databases, including ZINC,[46] Available Chemical Directory (ACD),[226] and PubChem[227] contain thousands of purchasable small molecules and are an important source of compounds for virtual screening. Typically, millions of compounds from databases are flexibly docked to a grid, then the single pose for each compound with the most favorable energy is minimized in the context of the receptor. The energy score can be the standard interaction scores (VDW+ES) and/or incorporate the new footprint similarity score (FPS). The docked ligands are clustered and reranked based on several scoring criteria with a variety of descriptors as shown in Figure 5.2.

97
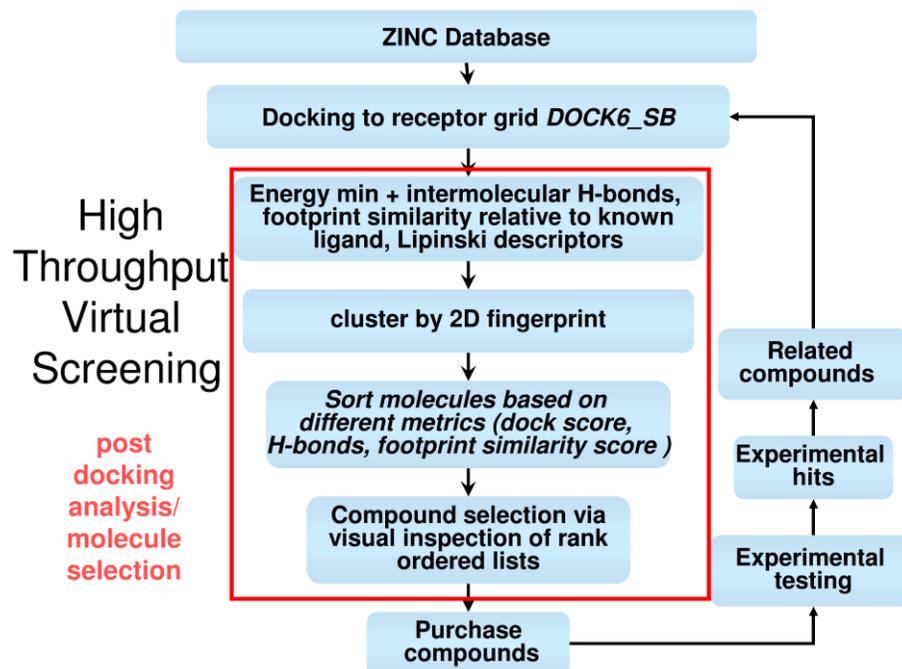
Figure 5.2. The schematic representation of the virtual screening process.

A general requirement is that the receptor structures should be of high-resolution and have low temperature factors. In addition, side-chain protonation states and the different forms of the receptor structures should be taken into account. For example, as noted above, kinases in the active, semi-active, or inactive forms may preferably favor distinct chemical structures and binding poses.[23,130,132] After visual inspection, a subset of top molecules based on each category of scoring function are chosen for further experimental evaluation (typically 100-200 total). In the evaluation process, Lipinksi's rule of five,[228] Oprea's druglikeness,[229] pharmacokinetic and pharmadynamic properties[114] are also typically considered.

5.3 Computational Details

5.3.1 Protein and Ligand Reference Preparation

Protein-ligand complexes 3D94[23] and 2ZM3[230] were obtained from the PDB database, separated into individual receptor and ligand files and saved in MOL2 format using the program MOE.[163] Ligand connectivity, bond orders, atom types, and possible hydrogen bond interactions were assessed using MOE. AM1-BCC[231] charges were added using the AMBER8[103,164] suite of programs. For the receptor, hydrogens, disulfide linkages and force field parameters were assigned using the AMBER8 tleap program. All crystallographic water molecules were removed. A short energy minimization (1000 steps) was performed for the prepared receptor using a stiff 100 kcal/mol/Å^2 restraint on all heavy atoms to allow only the added hydrogen atoms to adjust.

5.3.2 Sphere and Grid Generation

A molecular surface of the receptor was generated using the program DMS.[232] A set of spheres were then generated within 8 Å of ligand heavy atoms in the crystallographic pose by using program SPHGEN.[233] Energy grids were generated using using the accessory program GRID[234] with a 8 Å margin size, a 0.3 Å grid spacing, a $4r$ distance-dependent dielectric constant, and 6-9 van der Waals exponents. A bump filter of 0.75 was used for pruning orientation clashes.

5.3.3 Flexible Docking

Flexible docking (FLX) was performed on the local supercomputer called New York Blue using 512 blocks. Formal charges were restricted within ±2 e. An internal energy function

which only contains a repulsive VDW termed with an exponent of 12 was used for in calculations. For each ligand, 1000 orientations were generated. During simplex minimization, 500 iterations were used for anchor orienting and ligand segment growth. During anchor and grow, an energy cutoff of 100 kcal/mol with 100 conformers were employed. The best scored pose was then minimized off the grid with the restraint constant of 10 kcal/mol/Å^2 and retained for subsequent footprint similarity score calculation (FPS). FPS was calculated using the crystal pose of the ligand in 3D94 or 2ZM3 as the reference. Normalized Euclidean distance was employed in the FPS calculations.

5.3.4 Virtual Screening Setups

A comprehensive virtual screen of 500,000 compounds from the Chemdiv vendor in the ZINC database was performed using DOCK 6.6. The database was downloaded from the ZINC website and used as is (protonation states and partial atomic charges. The database was split into 21 chunks based on the number of rotatable bonds. The IGF-IR intermediate form (pdb 3D94) and active form (pdb 2ZM3) were used for virtual screening. Each compound was flexibly docked into 3D94 and 2ZM3 grids using the DOCK FLX protocol described above and a single lowest-energy pose for each compound was retained. Any duplicates in the database were removed. The top 100,000 scored molecules were then clustered based on MACCS fingerprints[235] using the program MOE[163] with a Tanimoto coefficient of 0.75. The top 250 cluster heads and all families members were selected based on the four different scoring methods: (1) standard DOCK score (DCE_{vdw+es}), (2) van der Waals footprint similarity score (FPS_{vdw}), (3) electrostatic footprint similarity score (FPS_{ES}), and (4) the combined footprint sum (FPS_{vdw+es}). Priority was given if a ligand showed favorable scores in more than one criterion.

Selected poses were also minimized in both EGFR and HER2 structure to evaluate whether favorable-scored ligands can be a potential dual inhibitor.

5.4 Preliminary Results and Discussion

Figure 5.3 shows representative virtual screening results based on the intermediate form of IGF-IR. Here, VDW and electrostatic footprints are compared between a reference ligand and three top molecules ZINC06721877 (cyan), ZINC33027966 (blue), and ZINC33027223 (green) obtained using three different metrics. FPS_{VDW+ES} and FPS_{ES} yield the same molecule ZINC33027966. Structurally, the molecules identified with the FPS function ZINC33027966 and ZINC33027223, overlap much better than the molecule ZINC06721877 (obtained based on DOCK score). Molecule ZINC33027223 identified with FPS_{VDW} has a lower Euclidean distance $(0.19 < 0.37)$ and higher correlation coefficient $(0.92 > 0.85)$ than ZINC33027966 for the VDW footprint while ZINC33027966 has a lower Euclidean distance $(0.23 < 1.23)$ and higher correlation coefficient $(0.94 > 0.06)$ than ZINC33027966 in the electrostatic footprint, indicating that the virtual screening results based on FPS scoring is behaving as expected. However, in the VDW footprint plots, ZINC06721877 has a much stronger interaction with LYS48 than either ZINC33027966 or ZINC33027223, which shows that use of different metrics can increase the overall chemical diversity of the compounds. Compound selection is ongoing.
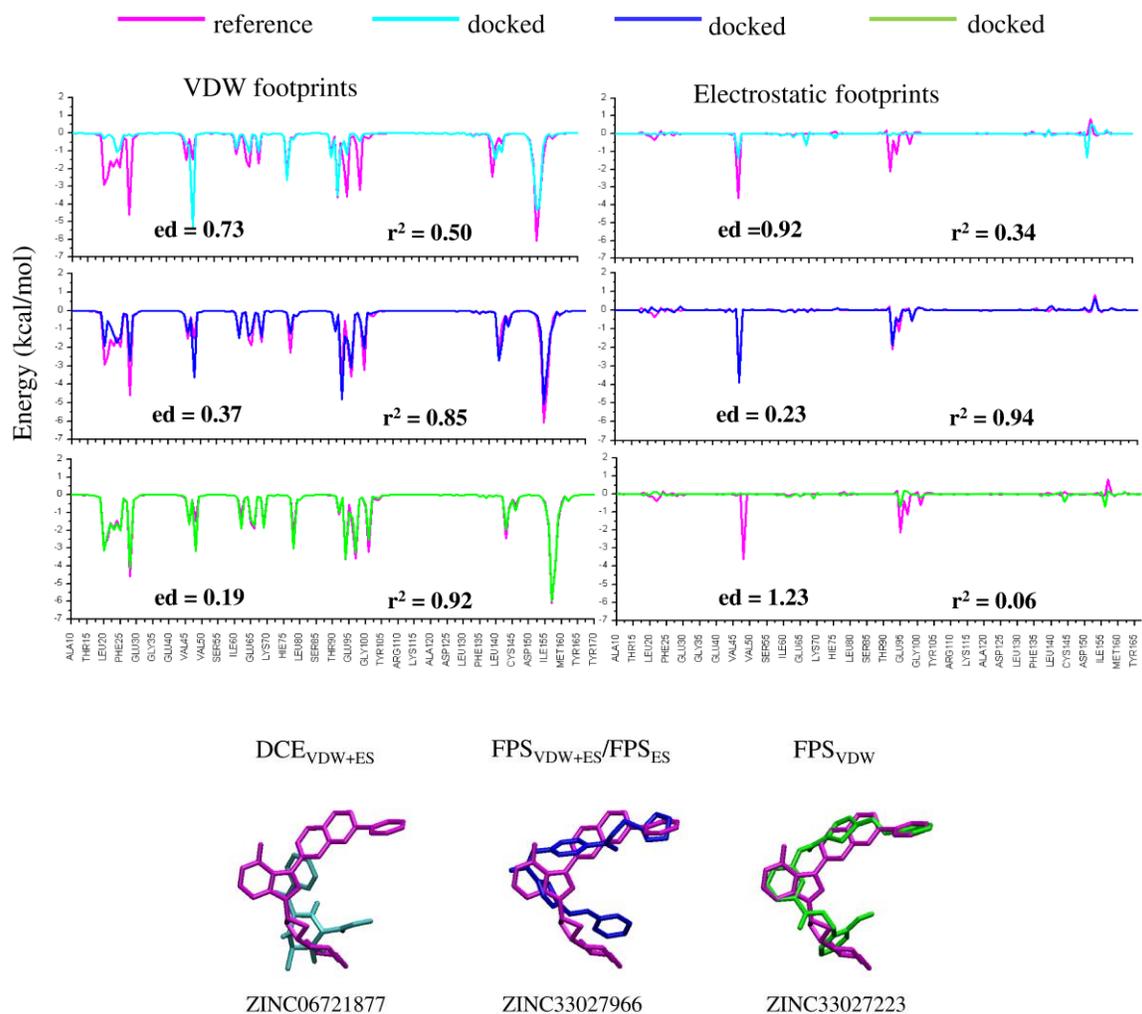
Figure 5.3. Preliminary results of virtual screening from Chemdiv databases using 3D94 receptor. The graph shows 3 molecules ZINC06721877, ZINC33027966, and ZINC33027223 that are obtained based on the best dock energy (DCE_{VDW+ES}), the best footprint similarity score (both FPS_{VDW+ES} and FPS_{ES}), and the best VDW footprint similarity score (FPS_{vdw}), shown in cyan, blue, and green, respectively. The upper panel shows the footprint overlap of the docked molecules and the reference ligand PQIP (in magenta) from residue ALA10 to THR170. The FPS values in units of Euclidean distance (ed) and pearsons correlation (r^2) are also shown. The bottom panel shows the structure overlaps.

In order to characterize properties of the top-scored molecules (using intermediate form of IGF-IR 3D94), population histograms including: (a) DCE_{vdw+es} score, (b) molecular weight, (c) number of rotatable bonds, (d) FPS_{vdw+ES} score, (e) FPS_{vdw} score, and (f) FPS_{ES} score were plotted for the top 250 molecules ranked based on each of the four scoring criteria DCE_{vdw+es} (in black), FPS_{vdw+es} (red), FPS_{vdw} (green), and FPS_{vdw+es} (blue) shown in Figure 5.4. As expected,

102

the top molecules selected based on a particular scoring function have more favorable features (either more favorable energy or smaller euclidean distances) with larger populations than other sets. In Figure 5.4, compounds selected using DCE_{VDW+ES} (in black) have more favorable DCE_{VDW+ES} scores (-90 to -80 kcal/mol) than the other three sets (-80 to -60 kcal/mol). As expected, compounds chosen based on FPS_{VDW+ES} (Figure 5.4 in red), FPS_{VDW} (Figure 5.4 in green) and FPS_{ES} (Figure 5.4 in blue) each have the enhanced populations with smaller euclidean distances relative to other sets in their respective histograms.

Due to the known scoring bias problem in DOCK, molecules chosen based on DCE_{vdw+es} have higher molecular weight and greater number of rotatable bonds. In contrast, FPS criterium leads to compound ensembles with more varied distributions and smaller average molecular weight (FPS_{vdw+ES} score = 452.5 g/mol, FPS_{vdw} = 472.4 g/mol, FPS_{es} = 440.5 g/mol) than those obtained from DCE scores (DCE_{vdw+es} is 499.7 g/mol). In addition, the molecules obtained based on FPS scores have fewer number of rotatable bonds (shown in Figure 5.4 red, green and blue curves versus black one). Thus, molecules obtained from FPS scores may be more attractive for purchase due to their more drug-like properties (lower molecular weight and rotatable bonds).
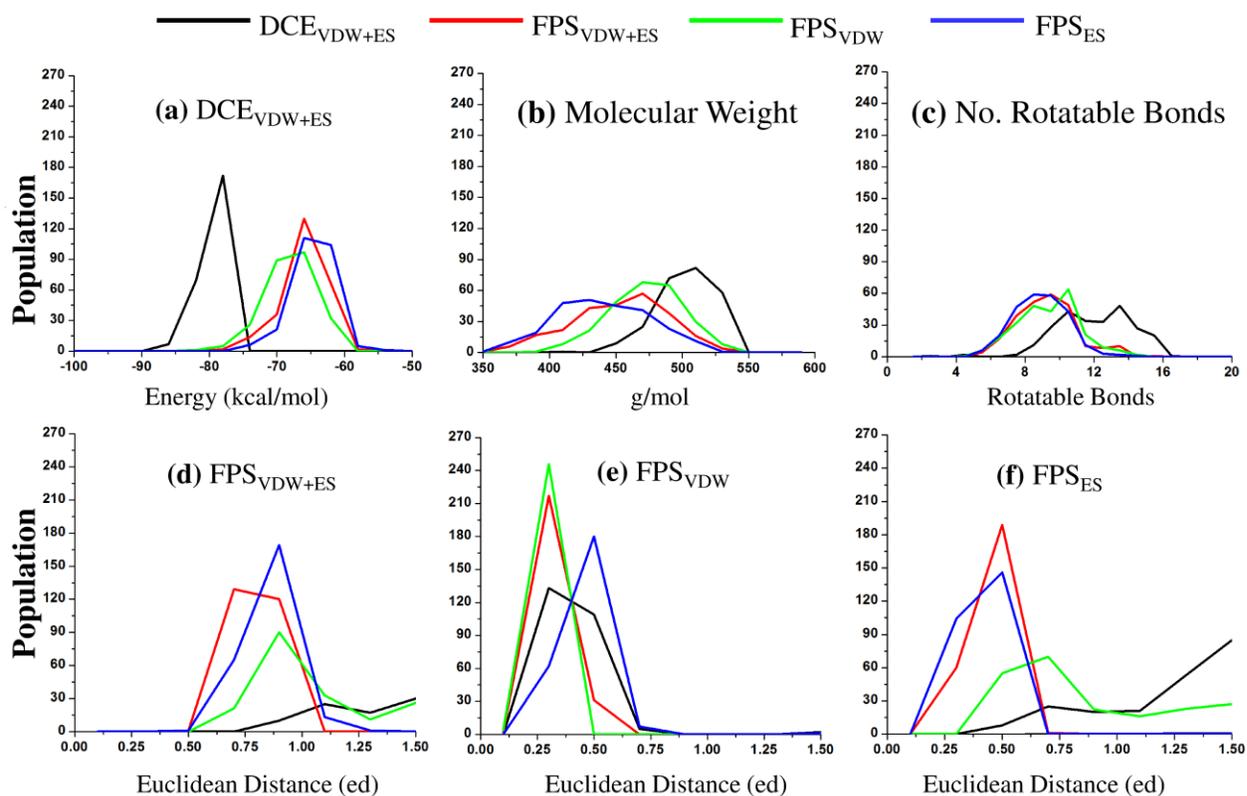
Figure 5.4. The population histograms of the four different ensembles of top 250 molecules obtained from each of the four ranking criteria including DCE_{vdw+es} (in black), FPS_{vdw+es} (in red), FPS_{vdw} (in green), and FPS_{vdw+es} (in blue). Statistics were shown for (a)DCE_{vdw+es} score. (2) molecular weight. (3) number of rotatable bonds. (d) FPS_{vdw+ES} score. (e) FPS_{vdw} score. (f) FPS_{ES} score. The results are based on reference structure 3D94.

In order to visually assess how different ensembles of docked compounds obtained using different criteria compared to the reference, the top 250 molecules selected using DCE_{VDW+ES}, FPS_{VDW+ES}, FPS_{VDW}, and FPS_{ES} scoring were plotted as shown in Figure 5.5. Here, molecules selected based on FPS_{VDW} (Figure 5.5c) more completely occupy the space of the reference ligand PQIP. For the molecules selected based on FPS_{ES} (Figure 5.5d), some molecules (category A, for example ZINC330279566) overlay very well with PQIP but others (category B, for example ZINC12414478) have additional functional groups that the PQIP does not have. Comparing the footprint results of these two categories of molecules, molecules of category B have more favorable Columbic interactions (~ 2 kcal/mol) with residue GLY100 than molecules

of category A and should be explored further. The greater diversity obtained using FPS_{ES} is consistent with the fact that molecular weights shown in Figure 5.4b using FPS_{ES} (350 ~ 550 g/mol) have a greater range than those selected by FPS_{VDW} 400~550 g/mol. In general, these results indicates that a virtual screening using FPS metrics help to maintain important interactions of the reference while allowing for chemical diversity.

(a) DCE_{VDW+ES}　　　(b) FPS_{VDW+ES}　　　(c) FPS_{VDW}　　　(d) FPS_{ES}
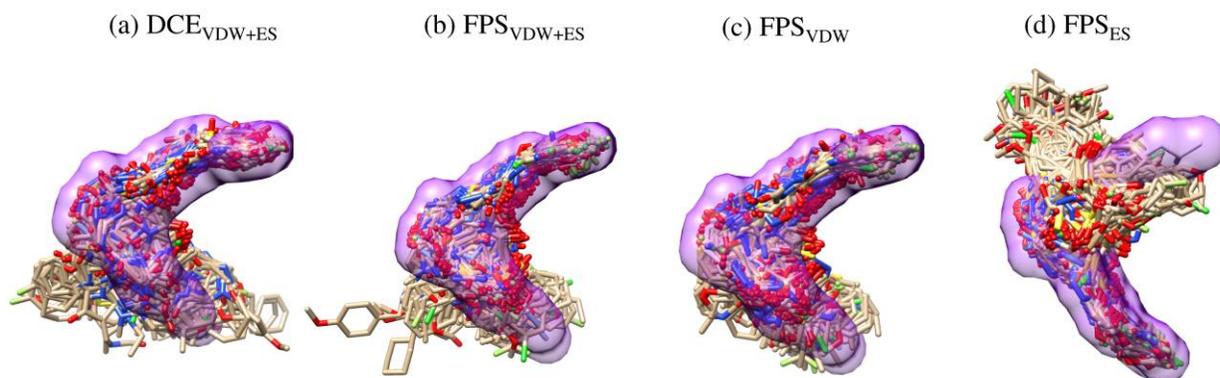
Figure 5.5. Graphical representation of the 250 top ranked poses obtained from docking using (a) DCE_{VDW+ES} (b) FPS_{VDW+ES} (c) FPS_{VDW} and (d) FPS_{VDW+ES} scoring functions. The reference ligand PQIP is shown in purple surface.

We also performed a screen to the active form of IGF-IR. Interestingly, the best pose based on the dock score (DCE) for the active form is also ZINC06721877, however, as shown in Figure 5.6 shows the molecule is in two different poses. Table 5.1 shows the top 5 molecules based on four different metrics using both IGF-IR forms. Favorably-scored compounds (highlighted in red) were identified selected by multiple metrics (DCE, FPS_{vdw}, FPS_{sum}) and should be more closely examined.

Table 5.1. Virtual Screening Results for intermediate form and the active form of IGF-IR.

The intermediate form of IGF-IR			
DCE_{VDW+ES}	FPS_{VDW+ES}	FPS_{ES}	FPS_{VDW}
ZINC06721877	ZINC33027966	ZINC33027966	ZINC33027223
ZINC18120322	ZINC08654669	ZINC33062555	ZINC33027693
ZINC20415006	ZINC33062555	ZINC33062559	ZINC20183341
ZINC11536053	ZINC33010848	ZINC08970236	ZINC08844196
ZINC09747607	ZINC02718465	ZINC01620056	ZINC09406868
The active form of IGF-IR			
ZINC06721877	ZINC08872308	ZINC06473842	ZINC06577743
ZINC11536053	ZINC20286949	ZINC08872308	ZINC20323322
ZINC20136268	ZINC06473842	ZINC20615791	ZINC04374602
ZINC20576342	ZINC18244556	ZINC09202293	ZINC15941250
ZINC20104903	ZINC32983947	ZINC00665425	ZINC09434998

As an initial test, ZINC33027966, identified as the compound with the best FPS_{VDW+ES} and FPS_{ES} score to the IGF-IR receptor, was selected to be minimized in the active structure of EGFR. After minimization, no steric clash was observed, thus this molecule could be a potential starting point to design a dual kinase inhibitor. In order to explore the details of the interaction of ZINC33027966 with both EGFR and IGF-IR, VDW and Coulombic footprints were analyzed as shown in Figure 5.7. Although these two receptors have somewhat different structures, the footprints share strong similarities. Here, ZINC33027966 has favorable VDW and Columbic interactions with both EGFR M793 and IGF-IR M793 which is the residue that forms H-bond in the hinge region of both kinases. And the compound has favorable VDW interactions with the gatekeeper residues EGFR T790 and IGFIR M790 which are critical in modulating kinase selectivity and affinity. As the T790M mutation in EGFR has been shown to negatively affect lapatinib binding and since ZINC33027966 has very favorable interactions with IGF-IR M790, this could be a lead for development of analogs to treat breast cancer patients with the EGFR T790M mutation.
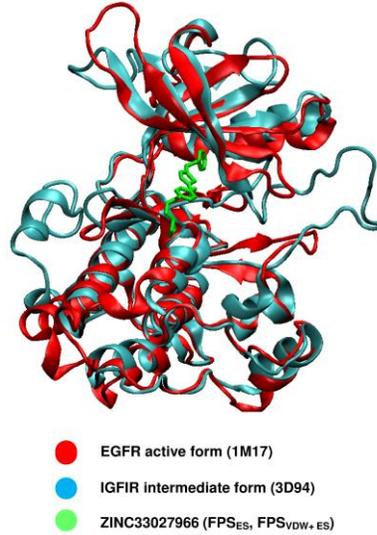
Figure 5.6. Overlap of ZINC33027966 in the intermediate form of IGF-IR and the active form of EGFR.
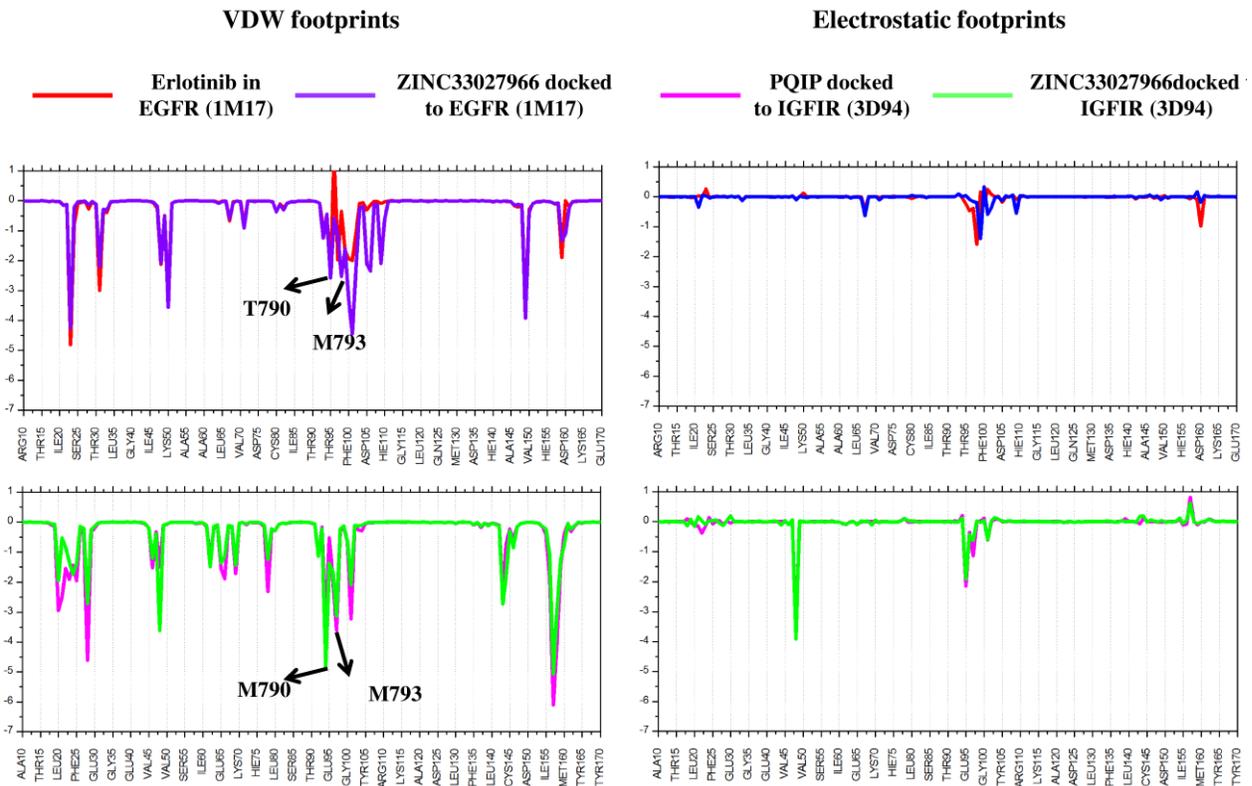
Figure 5.7. VDW and Coulombic footprints of ZINC33027966, erlotinib and PQIP with active EGFR and intermediate form of IGF-IR. The upper panel shows the footprint overlay of erlotinib (in red) and ZINC33027966 (in purple) with EGFR (1M17). The bottom panel shows the footprint overlay of PQIP (in pink) and ZINC33027966 (in green) with IGF-IR (3D94).

5.5 Conclusion

In summary, we have performed virtual screening of 100,000 molecules from the ChemDiv database to IGF-IR intermediate and active forms using DOCK 6.6. Preliminary results have shown that our screening protocol is behaving as expected and use of different scoring metrics (DCE_{VDW+ES}, FPS_{VDW+ES}, FPS_{VDW} and FPS_{ES}) can lead to more or less chemical diversity relative to a reference, depending on the function. Molecules showing favorable scores in more than one category (for example ZINC33027966, ZINC33062555) should be examined more closely for purchase. For example, ZINC33027966 has strong FPS_{VDW+ES} and FPS_{ES} scores and favorable interactions in both EGFR (active form) and IGF-IR (intermediate form) and may be a promising dual inhibition lead.

Future work will focus on selection of $150 - 200$ molecules, across the four different metrics used, for experimental testing. Lipinski's rule of five will also be used to prioritize the top-scored molecules from each category. Although some compounds are ranked as the top molecules in the DCE category, they may be ruled out due to larger molecular weights (>500 g/mol). Other metrics, including the number of hydrogen bonds and footprint score between the compounds and important residues in the kinase hinge region will be carefully analyzed. The most promising compounds will be experimentally tested by lab collaborator Dr. Todd Miller for biochemical activity, cellular inhibition of phosphorylation, and longer team for kinase inhibition in mice.

Chapter 6. Dissertation Summary: Scientific Impact, Related Work, and Future Direction.

Computational modeling is becoming an increasingly important component in modern drug discovery. Docking pipelines play a critical role in identifying new lead compounds and free energy calculations enables the leads to be computationally optimized prior to synthesis. This final chapter summarizes the scientific impact of the work presented in this thesis and the future directions.

6.1 Structural and Energetic Analysis of ErbB and IGFIR Simulations

Scientific impact

As described in Chapter 2, through all-atom molecular dynamics simulations, free energies of binding, and energy decomposition analysis, we identified a water-mediated network that is involved in lapatinib specificity and resistance with ErbB family members. The impact of this study is a better understanding of: (1) how receptor sequence and associated structural changes lead to binding energy variations for the highly homologous proteins EGFR, HER2 and ErbB4, (2) how kinase mutations can affect ligand binding by disturbing the water network, and (3) how the network could be useful to design improved ErbB inhibitors that yield tighter binding to wild-type enzymes, restore binding lost due to resistance mutations, or both.

In Chapter 3, umbrella sampling with potential mean force (PMF) calculations were used to quantify the relative population of two lapatinib conformations (conf1 and conf2) completed with three ErbB proteins. The primary impact of this study is: (1) based on PMF results, we demonstrated that conf1 is the global minimum and is the more favorable conformation for

EGFR, HER2 and ErbB4; (2) MM-GBSA results also proved that conf1 always yields more favorable energy than conf2 for all three proteins which is consistent with PMF results; (3) population analysis using the same trajectories for MM-GBSA calculations provided further evidence that conf1 is a more favorable conformation validating the use of conf1 in the studies in Chapter 1. Enhanced sampling methods and energy calculations are able to explore conformations not easily captured by experimental methods, and our analysis provide clues as to why two different conformations are seen in two different crystal structures of ErbB4 but not EGFR on HER2.

In Chapter 4, thermodynamic integration (TI) method was used to examine the structure activity relationship (SAR) for a series of related ligands, based on an imidazopyrazine scaffold, with an intermediate form of IGF-IR. The key finding is that the current protocol is able to achieve null transformation with satisfying accuracy, indicating the robustness of model construction. And with the current protocols, we are able to predict the relative binding free energy of ligand transformations involving small aliphatic functional groups within ±0.5 kcal/mol deviation from experiments. However, transformation involving polar groups are not accurately predicted. Although challenging, this study has provided an example of SAR using TI as an alternative to MM-GBSA calculation. In addition, this attempt has fully tested the current protocol for predicting large transformations (i.e. H to Ph) seldom reported in the literature.

Related work and future direction

Despite the good progress made, challenges still exist in the calculation of accurate binding free energies across multiple kinase systems. For example, binding of a single inhibitor with three different receptors (specificity) cannot reasonably be studied using methods using TI

because the mutation of one receptor to another would be a prohibitively large calculation and more approximate methods such as MM-GBSA are required. An interesting future direction might be to directly compare the performance of MM-GBSA and TI methods for the currently studied IGF-IR systems as well as other forms of IGF-IR. Additional inhibitors with corresponding experimental activities will be helpful to fully test the performance of both methods.

As described in Chapter 2, MM-GBSA calculations in the absence of key waters cannot predict binding that is consistent with the experimental trend, but calculations that include important waters in the binding pocket can achieve the correct prediction. The importance of key waters in other ErbB family members, (i.e IGF-IR), should also be pursued.

In terms of comparing TI and MM-GBSA, the setup time and execution time is also important. MM-GBSA is much simpler to implement than TI in terms of number of simulations, so it is meaningful to account for the time required to set up the systems. However, with MM-GBSA it may be necessary to take additional measures, such as identifying important waters in the binding pocket, which also adds to the time required. The simulation time that leads to convergence multiplied by the number of computer nodes used yields the total CPU time which should be compared for the two methods. In addition, the accuracy of both methods based on correlation coefficient and deviations from the experimental results should be compared. Various categories of the ligand transformations (large structure transformations or large charge transformations) should be included in the test set. In addition, the simulation results starting from random initial conformations using both methods should be pursued. Different charge models, including AM1-BCC, Gasteiger, and RESP charges might also be explored to evaluate convergence.

6.2 Application of Virtual Screening Protocols for Identifying Kinase Inhibitors

Scientific impact

An important component of this thesis is virtual screening to identify new kinase inhibitors using DOCK 6.6. The work described here represents a promising first step in early-stage lead identification of intermediate and active forms of IGF-IR. Favorably-scored compounds with various chemical diversity that derived from the different scoring functions used (standard DOCK score, van der Waals footprint similarity score, electrostatic footprint similarity score, and the combined footprint sum) have been identified for purchase, experimental testing, and subsequent optimization. And, compounds that target both EGFR and IGF-IR have been explored. Importantly, a general protocol for kinase virtual screening has been established.

Related work and future direction

Hydration is considered critical in molecular recognition. As described in Chapter 2, inclusion of important water molecules in binding free energy calculations to assess kinase specificity was needed. Inclusion of water in the footprint similarity score (FPS) might similarly enhance the accuracy of virtual screening for relevant kinase systems. One possible approach would be to include waters in the reference footprint used for screening, thereby facilitating discovery of ligands mimicking water-mediated interactions in the active system.

Receptor flexibility is another important area of research. Multi-grid scoring functions including Multiple Average Receptor (MAR) and Multiple Independent Receptor (MIR) have been recently incorporated into DOCK 6.6 which enable multi-target docking experiments to be

performed. For kinases in particular, targeting multi proteins (eg. EGFR/IGF-IR) or multi conformations (eg. Active/Inactive) simultaneously should be explored.

A preliminary de novo design method has also been implemented into DOCK which would be beneficial for targeting various kinase systems. The advantage of de novo methods is that the scaffolds, linkers, and side-chains used to construct new molecules are all generated from synthesizable molecules that are commercially available and the scoring function could be guided, for example, using solvated footprints to target the kinases studied here.

6.3 Summary

As demonstrated in this dissertation, all-atom computer modeling techniques are a powerful way to approach modern drug discovery. In summary, the most striking results of the current studies are as follows: (1) MD simulations with MM-GBSA and hydration analysis enabled discovery of a novel water-mediated mechanism that explains ErbB family specificity and drug resistance. (2) Umbrella sampling and PMF calculations enabled quantification of free energies and relative populations of different conformations of lapatinib with three ErbB proteins leading to the conclusion that a second conformation seen in the literature is not the global minimum. (3) Thermodynamic integration calculations with the current protocols can achieve satisfactory relative binding free energies for small transformation involving only aliphatic groups but enhanced protocols for polar transformations need to be developed. (4) Virtual screening can be used to identify compounds energetically compatible with both EGFR and IGF-IR. While many challenges remain, especially for TI calculations, it is believed that computational protocols and analysis used in this thesis will aid the community in future structure-based drug design studies to combat breast cancer.

Bibliography

(1)   American Cancer Society: Breast Cancer Facts & Figures 2011-2012. http://www.cancer.org (Accessed Oct 26th 2011).

(2)   Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813-8.

(3)   Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science* **1992**, *257*, 1078-1082.

(4)   NIGMS-Supported Structure-Based Drug Design Saves Lives. NIH/NIGMS Fact Sheet, 2006. http://publications.nigms.nih.gov/factsheets/structure_drugs.html.

(5)   Slamon, D.; Clark, G.; Wong, S.; Levin, W.; Ullrich, A.; McGuire, W. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **1987**, *235*, 177-182.

(6)   Schlotter, C. M.; Vogt, U.; Allgayer, H.; Brandt, B. Molecular targeted therapies for breast cancer treatment. *Breast Cancer Res* **2008**, *10*, 211.

(7)   Harris, A. L.; Nicholson, S.; Sainsbury, J. R. C.; Farndon, J.; Wright, C. Epidermal growth-factor receptors in breast-cancer - association with early relapse and death, poor response to hormones and interactions with neu. *J. Steroid Biochem. Mol. Biol.* **1989**, *34*, 123-131.

(8)   Walker, R. A.; Dearing, S. J. Expression of epidermal growth factor receptor mRNA and protein in primary breast carcinomas. *Breast Canc Res Treat* **1999**, *53*, 167-176.

(9)   Tsuda, H.; Hirohashi, S.; Shimosato, Y.; Hirota, T.; Tsugane, S.; Yamamoto, H.; Miyajima, N.; Toyoshima, K.; Yamamoto, T.; Yokota, J.; Yoshida, T.; Sakamoto, H.; Terada, M.; Sugimura, T. Correlation between long-term survival in breast-cancer patients and amplification of 2 putative oncogene-coamplification units - hst-1/int-2 and c-erbb-2/ear-1. *Cancer Research* **1989**, *49*, 3104-3108.

(10)   Furstenberger, G.; Morant, R.; Senn, H. J. Insulin-Like Growth Factors and Breast Cancer. *Onkologie* **2003**, *26*, 290-294.

(11)   Helle, S. I. The insulin-like growth factor system in advanced breast cancer. *Best Practice & Research Clinical Endocrinology & Metabolism* **2004**, *18*, 67-79.

(12)   Nielsen, T. O.; Andrews, H. N.; Cheang, M.; Kucab, J. E.; Hsu, F. D.; Ragaz, J.; Gilks, C. B.; Makretsov, N.; Bajdik, C. D.; Brookes, C.; Neckers, L. M.; Evdokimova, V.; Huntsman, D. G.; Dunn, S. E. Expression of the Insulin-Like Growth Factor I Receptor and Urokinase Plasminogen Activator in Breast Cancer Is Associated with Poor Survival: Potential for Intervention with 17-Allylamino Geldanamycin. *Cancer Res* **2004**, *64*, 286-291.

(13)   Surmacz, E.; Guvakova, M. A.; Nolan, M. K.; Nicosia, R. F.; Sciacca, L. Type I insulin-like growth factor receptor function in breast cancer. *Breast Canc Res Treat* **1998**, *47*, 255-267.

(14)   Gunter, M. J.; Hoover, D. R.; Yu, H.; Wassertheil-Smoller, S.; Rohan, T. E.; Manson, J. E.; Li, J.; Ho, G. Y. F.; Xue, X.; Anderson, G. L.; Kaplan, R. C.; Harris, T. G.; Howard, B. V.; Wylie-Rosett, J.; Burk, R. D.; Strickler, H. D. Insulin, Insulin-Like Growth Factor-I, and Risk of Breast Cancer in Postmenopausal Women. *J. Natl. Cancer Inst.* **2009**, *101*, 48-60.

(15)   Surmacz, E. Function of the IGF-I Receptor in Breast Cancer. *J Mammary Gland Biol Neoplasia* **2000**, *5*, 95-105.

(16)   http://www.cancer.gov/cancertopics/druginfo/ (Accessed May 14th 2013).

(17)   Yarden, Y.; Sliwkowski, M. X. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol* **2001**, *2*, 127-37.

(18)   Jin, Q.; Esteva, F. J. Cross-talk between the ErbB/HER family and the type I insulin-like growth factor receptor signaling pathway in breast cancer. *J Mammary Gland Biol Neoplasia* **2008**, *13*, 485-98.

(19)   Hynes, N. E.; Lane, H. A. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer* **2005**, *5*, 341-54.

(20)   Moy, B.; Goss, P. E. Lapatinib: current status and future directions in breast cancer. *Oncologist* **2006**, *11*, 1047-57.

(21)   Guix, M.; Granja Nde, M.; Meszoely, I.; Adkins, T. B.; Wieman, B. M.; Frierson, K. E.; Sanchez, V.; Sanders, M. E.; Grau, A. M.; Mayer, I. A.; Pestano, G.; Shyr, Y.; Muthuswamy, S.; Calvo, B.; Krontiras, H.; Krop, I. E.; Kelley, M. C.; Arteaga, C. L. Short preoperative treatment with erlotinib inhibits tumor cell proliferation in hormone receptor-positive breast cancers. *J Clin Oncol* **2008**, *26*, 897-906.

(22)   Ji, Q. S.; Mulvihill, M. J.; Rosenfeld-Franklin, M.; Cooke, A.; Feng, L.; Mak, G.; O'Connor, M.; Yao, Y.; Pirritt, C.; Buck, E.; Eyzaguirre, A.; Arnold, L. D.; Gibson, N. W.; Pachter, J. A. A novel, potent, and selective insulin-like growth factor-I receptor kinase inhibitor blocks insulin-like growth factor-I receptor signaling in vitro and inhibits insulin-like growth factor-I receptor dependent tumor growth in vivo. *Mol Cancer Ther* **2007**, *6*, 2158-67.

(23)   Wu, J.; Li, W.; Craddock, B. P.; Foreman, K. W.; Mulvihill, M. J.; Ji, Q. S.; Miller, W. T.; Hubbard, S. R. Small-molecule inhibition and activation-loop trans-phosphorylation of the IGF1 receptor. *Embo J* **2008**, *27*, 1985-94.

(24)   Ji, Q.-S.; Mulvihill, M. J.; Rosenfeld-Franklin, M.; Buck, E.; Cooke, A.; Eyzaguirre, A.; Mak, G.; O'Connor, M.; Pirritt, C.; Yao, Y.; Gibson, N. W.; A, P. J. Preclinical Characterization of OSI-906: A Novel IGF-1R Inhibitor in Clinical Trials. *AACR-NCI-EORTC International Conference on Molecular Targets and Cancer Therapeutics* **2007**, *Abstract # C192*.

(25)   Riedemann, J.; Sohail, M.; Macaulay, V. M. Dual silencing of the EGF and type 1 IGF receptors suggests dominance of IGF signaling in human breast cancer cells. *Biochem Biophys Res Commun* **2007**, *355*, 700-6.

(26)   Riedemann, J.; Takiguchi, M.; Sohail, M.; Macaulay, V. M. The EGF receptor interacts with the type 1 IGF receptor and regulates its stability. *Biochem Biophys Res Commun* **2007**, *355*, 707-14.

(27)   Knowlden, J. M.; Hutcheson, I. R.; Barrow, D.; Gee, J. M.; Nicholson, R. I. Insulin-like growth factor-I receptor signaling in tamoxifen-resistant breast cancer: a supporting role to the epidermal growth factor receptor. *Endocrinology* **2005**, *146*, 4609-18.

(28)   Adams, T. E.; McKern, N. M.; Ward, C. W. Signalling by the type 1 insulin-like growth factor receptor: interplay with the epidermal growth factor receptor. *Growth Factors* **2004**, *22*, 89-95.

(29)   Jones, H. E.; Goddard, L.; Gee, J. M.; Hiscox, S.; Rubini, M.; Barrow, D.; Knowlden, J. M.; Williams, S.; Wakeling, A. E.; Nicholson, R. I. Insulin-like growth factor-I receptor signalling and acquired resistance to gefitinib (ZD1839; Iressa) in human breast and prostate cancer cells. *Endocr Relat Cancer* **2004**, *11*, 793-814.

(30)   Kauraniemi, P.; Kallioniemi, A. Activation of multiple cancer-associated genes at the ERBB2 amplicon in breast cancer. *Endocr Relat Cancer* **2006**, *13*, 39-49.

(31)   Lu, Y.; Zi, X.; Zhao, Y.; Mascarenhas, D.; Pollak, M. Insulin-like growth factor-I receptor signaling and resistance to trastuzumab (Herceptin). *J Natl Cancer Inst* **2001**, *93*, 1852-7.

(32)   Nahta, R.; Yuan, L. X.; Zhang, B.; Kobayashi, R.; Esteva, F. J. Insulin-like growth factor-I receptor/human epidermal growth factor receptor 2 heterodimerization contributes to trastuzumab resistance of breast cancer cells. *Cancer Res* **2005**, *65*, 11118-28.

(33)   Balana, M. E.; Labriola, L.; Salatino, M.; Movsichoff, F.; Peters, G.; Charreau, E. H.; Elizalde, P. V. Activation of ErbB-2 via a hierarchical interaction between ErbB-2 and type I insulin-like growth factor receptor in mammary tumor cells. *Oncogene* **2001**, *20*, 34-47.

(34)   Chakravarti, A.; Loeffler, J. S.; Dyson, N. J. Insulin-like growth factor receptor I mediates resistance to anti-epidermal growth factor receptor therapy in primary human glioblastoma cells through continued activation of phosphoinositide 3-kinase signaling. *Cancer Res* **2002**, *62*, 200-7.

(35)   Knowlden, J. M.; Jones, H. E.; Barrow, D.; Gee, J. M.; Nicholson, R. I.; Hutcheson, I. R. Insulin receptor substrate-1 involvement in epidermal growth factor receptor and insulin-like growth factor receptor signalling: implication for Gefitinib ('Iressa') response and resistance. *Breast Cancer Res Treat* **2008**, *111*, 79-91.

(36)   Camirand, A.; Zakikhani, M.; Young, F.; Pollak, M. Inhibition of insulin-like growth factor-1 receptor signaling enhances growth-inhibitory and proapoptotic effects of gefitinib (Iressa) in human breast cancer cells. *Breast Cancer Res* **2005**, *7*, R570-9.

(37)   Lu, D.; Zhang, H.; Koo, H.; Tonra, J.; Balderes, P.; Prewett, M.; Corcoran, E.; Mangalampalli, V.; Bassi, R.; Anselma, D.; Patel, D.; Kang, X.; Ludwig, D. L.; Hicklin, D. J.; Bohlen, P.; Witte, L.; Zhu, Z. A fully human recombinant IgG-like bispecific antibody to both the epidermal growth factor receptor and the insulin-like growth factor receptor for enhanced antitumor activity. *J Biol Chem* **2005**, *280*, 19665-72.

(38)   Collins, I.; Workman, P. New approaches to molecular cancer therapeutics. *Nat Chem Biol* **2006**, *2*, 689-700.

(39)   Dickson, M.; Gagnon, J. P. Key factors in the rising cost of new drug discovery and development. *Nat Rev Drug Discov* **2004**, *3*, 417-29.

(40)   Congreve, M.; Murray, C. W.; Blundell, T. L. Structural biology and drug discovery. *Drug Discov Today* **2005**, *10*, 895-907.

(41)   Scapin, G. Structural biology and drug discovery. *Curr Pharm Des* **2006**, *12*, 2087-97.

(42)   Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res* **2000**, *28*, 235-242.

(43)   Celebrating structural biology. *Nat Struct Mol Biol* **2011**, *18*, 1304-1316.

(44)   Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. Calculation of the relative change in binding free energy of a protein-inhibitor complex. *Science* **1987**, *235*, 574-6.

(45)   Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem Rev* **1993**, *93*, 2395-2417.

(46)   Irwin, J. J.; Shoichet, B. K. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* **2005**, *45*, 177-82.

(47)   Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* **2001**, *15*, 411-28.

(48)   Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. Development and validation of a modular, extensible docking program: DOCK 5. *J Comput Aided Mol Des* **2006**, *20*, 601-19.

(49)   Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA* **2009**, *15*, 1219-30.

(50)   Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: Applications of AutoDock. *J Mol Recogn* **1996**, *9*, 1-5.

(51)   Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Protein Struct Funct Genet* **1999**, *37*, 228-241.

(52)   Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004**, *47*, 1739-49.

(53)   Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* **2004**, *47*, 1750-9.

(54)   Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Protein Struct Funct Genet* **2003**, *52*, 609-623.

(55)   Jorgensen, W. L. Free energy calculations: a breakthrough for modeling organic chemistry in solution. *Accounts Chem Res* **1989**, *22*, 184-189.

(56)   Steinbrecher, T.; Case, D. A.; Labahn, A. A multistep approach to structure-based drug design: studying ligand binding at the human neutrophil elastase. *J Med Chem* **2006**, *49*, 1837-44.

(57)   Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts Chem Res* **2000**, *33*, 889-897.

(58)   Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J Phys Chem B* **2000**, *104*, 3712-3720.

(59)   Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., 3rd. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J Comput Chem* **2004**, *25*, 265-84.

(60)   Blundell, T. L.; Sibanda, B. L.; Sternberg, M. J.; Thornton, J. M. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* **1987**, *326*, 347-52.

(61)   Sali, A. Modeling mutations and homologous proteins. *Curr Opin Biotechnol* **1995**, *6*, 437-51.

(62)   Bajorath, J.; Stenkamp, R.; Aruffo, A. Knowledge-based model building of proteins: concepts and examples. *Protein Sci* **1993**, *2*, 1798-810.

(63)   Marti-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* **2000**, *29*, 291-325.

(64)   Sanchez, R.; Sali, A. Advances in comparative protein-structure modelling. *Curr Opin Struct Biol* **1997**, *7*, 206-14.

(65)   http://www.rcsb.org/pdb/statistics/holdings.do (Accessed April 15th 2013).

(66)   Hubbard, T. J. P.; Ailey, B.; Brenner, S. E.; Murzin, A. G.; Chothia, C. SCOP: a Structural Classification of Proteins database. *Nucleic Acids Research* **1999**, *27*, 254-256.

(67)   Holm, L.; Sander, C. Protein folds and families: sequence and structure alignments. *Nucleic Acids Res* **1999**, *27*, 244-7.

(68)   Orengo, C. A.; Pearl, F. M.; Bray, J. E.; Todd, A. E.; Martin, A. C.; Lo Conte, L.; Thornton, J. M. The CATH Database provides insights into protein structure/function relationships. *Nucleic Acids Res* **1999**, *27*, 275-9.

(69)   Pearson, W. R. Empirical statistical estimates for sequence similarity searches. *J Mol Biol* **1998**, *276*, 71-84.

(70)   Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **1990**, *215*, 403-10.

(71)   Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **1997**, *25*, 3389-402.

(72)   Bowie, J. U.; Luthy, R.; Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **1991**, *253*, 164-70.

(73)   Jones, D. T.; Taylor, W. R.; Thornton, J. M. A new approach to protein fold recognition. *Nature* **1992**, *358*, 86-9.

(74)   Saqi, M. A.; Russell, R. B.; Sternberg, M. J. Misleading local sequence alignments: implications for comparative protein modelling. *Protein Eng* **1998**, *11*, 627-30.

(75)   Rost, B. Twilight zone of protein sequence alignments. *Protein Eng* **1999**, *12*, 85-94.

(76)   Thompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F.; Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **1997**, *25*, 4876-82.

(77)   Barton, G. J.; Sternberg, M. J. A strategy for the rapid multiple alignment of protein sequences. Confidence levels from tertiary structure comparisons. *J Mol Biol* **1987**, *198*, 327-37.

(78)   Taylor, W. R.; Flores, T. P.; Orengo, C. A. Multiple protein structure alignment. *Protein Sci* **1994**, *3*, 1858-70.

(79)   Sali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J Mol Biol* **1993**, *234*, 779-815.

(80)   Fiser, A.; Do, R. K.; Sali, A. Modeling of loops in protein structures. *Protein Sci* **2000**, *9*, 1753-73.

(81)   Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A. Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* **2007**, *Chapter 2*, Unit 2 9.

(82)   Sali, A.; Overington, J. P. Derivation of rules for comparative protein modeling from a database of protein structure alignments. *Protein Sci* **1994**, *3*, 1582-96.

(83)   Finkelstein, A. V.; Badretdinov, A.; Gutin, A. M. Why do protein architectures have Boltzmann-like statistics? *Proteins* **1995**, *23*, 142-50.

(84)   Miyazawa, S.; Jernigan, R. L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J Mol Biol* **1996**, *256*, 623-44.

(85)   Sippl, M. J. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J Mol Biol* **1990**, *213*, 859-83.

(86)   MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B* **1998**, *102*, 3586-3616.

(87)   Shen, M. Y.; Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci* **2006**, *15*, 2507-24.

(88)   Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **1993**, *26*, 283-291.

(89)   Morris, A. L.; MacArthur, M. W.; Hutchinson, E. G.; Thornton, J. M. Stereochemical quality of protein structure coordinates. *Proteins* **1992**, *12*, 345-64.

(90)   Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc* **1996**, *118*, 11225-11236.

(91)   Mackerell, A. D., Jr. Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem* **2004**, *25*, 1584-604.

(92)   Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field

for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* **1995**, *117*, 5179-5197.

(93)   Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* **2003**, *32*, 335-73.

(94)   Ewing, T. J.; Kuntz, I. D.; University of California at San Francisco: San Francisco, CA, 1999.

(95)   May, A.; Zacharias, M. Protein−Ligand Docking Accounting for Receptor Side Chain and Global Flexibility in Normal Modes: Evaluation on Kinase Inhibitor Cross Docking. *J Med Chem* **2008**, *51*, 3499-3506.

(96)   Gianti, E.; Zauhar, R. J. Modeling Androgen Receptor Flexibility: A Binding Mode Hypothesis of CYP17 Inhibitors/Antiandrogens for Prostate Cancer Therapy. *J Chem Inform Model* **2012**, *52*, 2670-2683.

(97)   Leach, A. R. Ligand docking to proteins with discrete side-chain flexibility. *J Mol Biol* **1994**, *235*, 345-356.

(98)   Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *Journal of the American Chemical Society* **2005**, *49*, 5912-5931.

(99)   Mukherjee, S.; Balius, T. E.; Rizzo, R. C. Docking validation resources: protein family and ligand flexibility experiments. *J Chem Inf Model* **2010**, *50*, 1986-2000.

(100)   Balius, T. E.; Mukherjee, S.; Rizzo, R. C. Implementation and evaluation of a docking-rescoring method using molecular footprint comparisons. *J Comput Chem* **2011**, *32*, 2273-2289.

(101)   Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. Ligand solvation in molecular docking. *Protein Struct Funct Genet* **1999**, *34*, 4-16.

(102)   Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **1992**, *13*, 505-524.

(103)   Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J Comput Chem* **2005**, *26*, 1668-88.

(104)   Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm - a Program For Macromolecular Energy, Minimization, and Dynamics Calculations. *J Comput Chem* **1983**, *4*, 187-217.

(105)   Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van

Gunsteren, W. F. The GROMOS software for biomolecular simulation: GROMOS05. *J Comput Chem* **2005**, *26*, 1719-51.

(106)   Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J Comput Chem* **2005**, *26*, 1781-1802.

(107)   Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr Opin Struct Biol* **2011**, *21*, 150-60.

(108)   Hong, G.; Cornish, A. J.; Hegg, E. L.; Pachter, R. On understanding proton transfer to the biocatalytic [Fe-Fe](H) sub-cluster in [Fe-Fe]H(2)ases: QM/MM MD simulations. *Biochim Biophys Acta* **2011**, *1807*, 510-7.

(109)   Cieplak, P.; Dupradeau, F. Y.; Duan, Y.; Wang, J. Polarization effects in molecular mechanical force fields. *J Phys Condens Matter* **2009**, *21*, 333102.

(110)   Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J Chem Phys* **2004**, *120*, 11919-29.

(111)   Yang, J.; Wang, Y.; Chen, Y. GPU accelerated molecular dynamics simulation of thermal conductivities. *J. Comput. Phys.* **2007**, *221*, 799-804.

(112)   Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation For Molecular Mechanics and Dynamics. *J Am Chem Soc* **1990**, *112*, 6127-6129.

(113)   Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct. Funct. Bioinformatics* **2004**, *55*, 383-94.

(114)   Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. *J Phys Chem* **1994**, *98*, 1978-1988.

(115)   Baron, R.; Hünenberger, P. H.; McCammon, J. A. Absolute Single-Molecule Entropies from Quasi-Harmonic Analysis of Microsecond Molecular Dynamics: Correction Terms and Convergence Properties. *J Chem Theor Comput* **2009**, *5*, 3150-3160.

(116)   Steinbrecher, T.; Mobley, D. L.; Case, D. A. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J Chem Phys* **2007**, *127*, 214108-13.

(117)   Klijn, J. G. M.; Berns, P. M. J. J.; Schmitz, P. I. M.; Foekens, J. A. The Clinical Significance of Epidermal Growth Factor Receptor (EGF-R) in Human Breast Cancer: A Review on 5232 Patients. *Endocr Rev* **1992**, *13*, 3-17.

(118)   Pegram, M. D.; Pauletti, G.; Slamon, D. J. HER-2/neu as a predictive marker of response to breast cancer therapy. *Breast Cancer Res Treat* **1998**, *52*, 65-77.

(119) Drugs@FDA website. http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm (Accessed Oct 26th 2011).

(120) Konecny, G. E.; Pegram, M. D.; Venkatesan, N.; Finn, R.; Yang, G.; Rahmeh, M.; Untch, M.; Rusnak, D. W.; Spehar, G.; Mullin, R. J.; Keith, B. R.; Gilmer, T. M.; Berger, M.; Podratz, K. C.; Slamon, D. J. Activity of the dual kinase inhibitor lapatinib (GW572016) against HER-2-overexpressing and trastuzumab-treated breast cancer cells. *Cancer Res* **2006**, *66*, 1630-9.

(121) Blackwell, K. L.; Pegram, M. D.; Tan-Chiu, E.; Schwartzberg, L. S.; Arbushites, M. C.; Maltzman, J. D.; Forster, J. K.; Rubin, S. D.; Stein, S. H.; Burstein, H. J. Single-agent lapatinib for HER2-overexpressing advanced or metastatic breast cancer that progressed on first- or second-line trastuzumab-containing regimens. *Annals of Oncology* **2009**, *20*, 1026-1031.

(122) Xia, W.; Bacus, S.; Hegde, P.; Husain, I.; Strum, J.; Liu, L.; Paulazzo, G.; Lyass, L.; Trusk, P.; Hill, J.; Harris, J.; Spector, N. L. A model of acquired autoresistance to a potent ErbB2 tyrosine kinase inhibitor and a therapeutic strategy to prevent its onset in breast cancer. *Proc Natl Acad Sci* **2006**, *103*, 7795-7800.

(123) Balius, T. E.; Rizzo, R. C. Quantitative Prediction of Fold Resistance for Inhibitors of EGFR. *Biochemistry* **2009**, *48*, 8435-8448.

(124) Eichhorn, P. J. A.; Gili, M.; Scaltriti, M.; Serra, V.; Guzman, M.; Nijkamp, W.; Beijersbergen, R. L.; Valero, V.; Seoane, J.; Bernards, R.; Baselga, J. Phosphatidylinositol 3-Kinase Hyperactivation Results in Lapatinib Resistance that Is Reversed by the mTOR/Phosphatidylinositol 3-Kinase Inhibitor NVP-BEZ235. *Cancer Res* **2008**, *68*, 9221-9230.

(125) Huang, C.; Park, C. C.; Hilsenbeck, S. G.; Ward, R.; Rimawi, M. F.; Wang, Y. C.; Shou, J.; Bissell, M. J.; Osborne, C. K.; Schiff, R. beta1 integrin mediates an alternative survival pathway in breast cancer cells resistant to lapatinib. *Breast Cancer Res* **2011**, *13*, R84.

(126) Rexer, B. N.; Ham, A. J. L.; Rinehart, C.; Hill, S.; de Matos Granja-Ingram, N.; Gonzalez-Angulo, A. M.; Mills, G. B.; Dave, B.; Chang, J. C.; Liebler, D. C.; Arteaga, C. L. Phosphoproteomic mass spectrometry profiling links Src family kinases to escape from HER2 tyrosine kinase inhibition. *Oncogene* **2011**, *30*, 4163-4174.

(127) Avizienyte, E.; Ward, R. A.; Garner, A. P. Comparison of the EGFR resistance mutation profiles generated by EGFR-targeted tyrosine kinase inhibitors and the impact of drug combinations. *Biochem J* **2008**, *415*, 197-206.

(128) Trowe, T.; Boukouvala, S.; Calkins, K.; Cutler, R. E.; Fong, R.; Funke, R.; Gendreau, S. B.; Kim, Y. D.; Miller, N.; Woolfrey, J. R.; Vysotskaia, V.; Yang, J. P.; Gerritsen, M. E.; Matthews, D. J.; Lamb, P.; Heuer, T. S. EXEL-7647 Inhibits Mutant Forms of ErbB2 Associated with Lapatinib Resistance and Neoplastic Transformation. *Clin Cancer Res* **2008**, *14*, 2465-2475.

(129) Rusnak, D. W.; Lackey, K.; Affleck, K.; Wood, E. R.; Alligood, K. J.; Rhodes, N.; Keith, B. R.; Murray, D. M.; Knight, W. B.; Mullin, R. J.; Gilmer, T. M. The Effects of the Novel, Reversible Epidermal Growth Factor Receptor/ErbB-2 Tyrosine Kinase Inhibitor, GW2016, on

the Growth of Human Normal and Tumor-derived Cell Lines in Vitro and in Vivo. *Mol Canc Therapeut* **2001**, *1*, 85-94.

(130)   Wood, E. R.; Truesdale, A. T.; McDonald, O. B.; Yuan, D.; Hassell, A.; Dickerson, S. H.; Ellis, B.; Pennisi, C.; Horne, E.; Lackey, K.; Alligood, K. J.; Rusnak, D. W.; Gilmer, T. M.; Shewchuk, L. 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 Res* **2004**, *64*, 6652-9.

(131)   Zhang, X.; Gureasko, J.; Shen, K.; Cole, P. A.; Kuriyan, J. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* **2006**, *125*, 1137-49.

(132)   Stamos, J.; Sliwkowski, M. X.; Eigenbrot, C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J Biol Chem* **2002**, *277*, 46265-72.

(133)   Yun, C. H.; Boggon, T. J.; Li, Y.; Woo, M. S.; Greulich, H.; Meyerson, M.; Eck, M. J. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* **2007**, *11*, 217-27.

(134)   Lackey, K. E. Lessons from the drug discovery of lapatinib, a dual ErbB1/2 tyrosine kinase inhibitor. *Curr Top Med Chem* **2006**, *6*, 435-60.

(135)   Sartor, C. I.; Zhou, H.; Kozlowska, E.; Guttridge, K.; Kawata, E.; Caskey, L.; Harrelson, J.; Hynes, N.; Ethier, S.; Calvo, B.; Earp, H. S., III. HER4 Mediates Ligand-Dependent Antiproliferative and Differentiation Responses in Human Breast Cancer Cells. *Mol. Cell. Biol.* **2001**, *21*, 4265-4275.

(136)   Barnes, N. L.; Khavari, S.; Boland, G. P.; Cramer, A.; Knox, W. F.; Bundred, N. J. Absence of HER4 expression predicts recurrence of ductal carcinoma in situ of the breast. *Clin Cancer Res* **2005**, *11*, 2163-8.

(137)   Qiu, C.; Tarrant, M. K.; Choi, S. H.; Sathyamurthy, A.; Bose, R.; Banjade, S.; Pal, A.; Bornmann, W. G.; Lemmon, M. A.; Cole, P. A.; Leahy, D. J. Mechanism of Activation and Inhibition of the HER4/ErbB4 Kinase. *Structure* **2008**, *16*, 460-467.

(138)   Gilmer, T. M.; Cable, L.; Alligood, K.; Rusnak, D.; Spehar, G.; Gallagher, K. T.; Woldu, E.; Carter, H. L.; Truesdale, A. T.; Shewchuk, L.; Wood, E. R. Impact of Common Epidermal Growth Factor Receptor and HER2 Variants on Receptor Activity and Inhibition by Lapatinib. *Cancer Res* **2008**, *68*, 571-579.

(139)   Kobayashi, S.; Boggon, T. J.; Dayaram, T.; Jänne, P. A.; Kocher, O.; Meyerson, M.; Johnson, B. E.; Eck, M. J.; Tenen, D. G.; Halmos, B. EGFR Mutation and Resistance of Non–Small-Cell Lung Cancer to Gefitinib. *New Engl J Med* **2005**, *352*, 786-792.

(140)   Bean, J.; Riely, G. J.; Balak, M.; Marks, J. L.; Ladanyi, M.; Miller, V. A.; Pao, W. Acquired resistance to epidermal growth factor receptor kinase inhibitors associated with a novel

T854A mutation in a patient with EGFR-mutant lung adenocarcinoma. *Clin Cancer Res* **2008**, *14*, 7519-25.

(141)   Gorre, M. E.; Mohammed, M.; Ellwood, K.; Hsu, N.; Paquette, R.; Rao, P. N.; Sawyers, C. L. Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science* **2001**, *293*, 876-80.

(142)   Yun, C. H.; Mengwasser, K. E.; Toms, A. V.; Woo, M. S.; Greulich, H.; Wong, K. K.; Meyerson, M.; Eck, M. J. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci* **2008**, *105*, 2070-5.

(143)   Wissner, A.; Berger, D. M.; Boschelli, D. H.; Floyd, M. B., Jr.; Greenberger, L. M.; Gruber, B. C.; Johnson, B. D.; Mamuya, N.; Nilakantan, R.; Reich, M. F.; Shen, R.; Tsou, H. R.; Upeslacis, E.; Wang, Y. F.; Wu, B.; Ye, F.; Zhang, N. 4-Anilino-6,7-dialkoxyquinoline-3-carbonitrile inhibitors of epidermal growth factor receptor kinase and their bioisosteric relationship to the 4-anilino-6,7-dialkoxyquinazoline inhibitors. *J Med Chem* **2000**, *43*, 3244-56.

(144)   Kamath, S. A.; Buolamwini, J. K. Asp746 to glycine change may have a greater influence than Cys751 to serine change in accounting for ligand selectivity between EGFR and HER-2 at the ATP site. *J Comput Aided Mol Des* **2005**, *19*, 287-91.

(145)   Scaltriti, M.; Verma, C.; Guzman, M.; Jimenez, J.; Parra, J. L.; Pedersen, K.; Smith, D. J.; Landolfi, S.; Ramon y Cajal, S.; Arribas, J.; Baselga, J. Lapatinib, a HER2 tyrosine kinase inhibitor, induces stabilization and accumulation of HER2 and potentiates trastuzumab-dependent cell cytotoxicity. *Oncogene* **2008**, *28*, 803-814.

(146)   Hou, T.; Zhu, L.; Chen, L.; Xu, X. Mapping the binding site of a large set of quinazoline type EGF-R inhibitors using molecular field analyses and molecular docking studies. *J Chem Inf Comput Sci* **2003**, *43*, 273-87.

(147)   Chen, H. F. Computational study of the binding mode of epidermal growth factor receptor kinase inhibitors. *Chem Biol Drug Des* **2008**, *71*, 434-46.

(148)   Cavasotto, C. N.; Ortiz, M. A.; Abagyan, R. A.; Piedrafita, F. J. In silico identification of novel EGFR inhibitors with antiproliferative activity against cancer cells. *Bioorg Med Chem Lett* **2006**, *16*, 1969-74.

(149)   Liu, B.; Bernard, B.; Wu, J. H. Impact of EGFR point mutations on the sensitivity to gefitinib: insights from comparative structural analyses and molecular dynamics simulations. *Proteins* **2006**, *65*, 331-46.

(150)   Liu, Y.; Purvis, J.; Shih, A.; Weinstein, J.; Agrawal, N.; Radhakrishnan, R. A multiscale computational approach to dissect early events in the Erb family receptor mediated activation, differential signaling, and relevance to oncogenic transformations. *Ann Biomed Eng* **2007**, *35*, 1012-25.

(151)   Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Energetics of Displacing Water Molecules from Protein Binding Sites: Consequences for Ligand Optimization. *J Am Chem Soc* **2009**, *131*, 15403-15411.

(152)   Daub, H.; Specht, K.; Ullrich, A. Strategies to overcome resistance to targeted protein kinase inhibitors. *Nat Rev Drug Discov* **2004**, *3*, 1001-10.

(153)   Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Free energy of ligand binding to protein: Evaluation of the contribution of water molecules by computational methods. *Curr Med Chem* **2004**, *11*, 3093-3118.

(154)   Lu, Y.; Wang, R.; Yang, C.-Y.; Wang, S. Analysis of Ligand-Bound Water Molecules in High-Resolution Crystal Structures of Protein-Ligand Complexes. *J Chem Inform Model* **2007**, *47*, 668-675.

(155)   Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding. *Proc Natl Acad Sci* **2007**, *104*, 808-813.

(156)   de Beer, S. B. A.; Vermeulen, N. P. E.; Oostenbrink, C. The role of water molecules in computational drug design. *Curr Top Med Chem* **2010**, *10*, 55-66.

(157)   Varghese, J. N.; Epa, V. C.; Colman, P. M. Three-dimensional structure of the complex of 4-guanidino-Neu5Ac2en and influenza virus neuraminidase. *Protein Sci* **1995**, *4*, 1081-7.

(158)   Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.; et al. High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat Struct Biol* **1995**, *2*, 293-302.

(159)   Wood, E. R.; Shewchuk, L. M.; Ellis, B.; Brignola, P.; Brashear, R. L.; Caferro, T. R.; Dickerson, S. H.; Dickson, H. D.; Donaldson, K. H.; Gaul, M.; Griffin, R. J.; Hassell, A. M.; Keith, B.; Mullin, R.; Petrov, K. G.; Reno, M. J.; Rusnak, D. W.; Tadepalli, S. M.; Ulrich, J. C.; Wagner, C. D.; Vanderwall, D. E.; Waterson, A. G.; Williams, J. D.; White, W. L.; Uehling, D. E. 6-Ethynylthieno[3,2-d]- and 6-ethynylthieno[2,3-d]pyrimidin-4-anilines as tunable covalent modifiers of ErbB kinases. *Proc Natl Acad Sci* **2008**, *105*, 2773-2778.

(160)   Aertgeerts, K.; Skene, R.; Yano, J.; Sang, B. C.; Zou, H.; Snell, G.; Jennings, A.; Iwamoto, K.; Habuka, N.; Hirokawa, A.; Ishikawa, T.; Tanaka, T.; Miki, H.; Ohta, Y.; Sogabe, S. Structural analysis of the mechanism of inhibition and allosteric activation of the kinase domain of HER2 protein. *J Biol Chem* **2011**, *286*, 18756-65.

(161)   Thompson, J. D.; Higgins, D. G.; Gibson, T. J. Clustal W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res* **1994**, *22*, 4673-4680.

(162)   Shen, M.-Y.; Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Science* **2006**, *15*, 2507-2524.

(163)   Chemical Computing Group Inc., Montreal, Canada, 2008.

(164)   Case, D. A.; Darden, T. A.; Cheatham, T.E., 3rd; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A.; University of California: San Francisco, CA, 2004.

(165)   Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct. Funct. Bioinformatics* **2006**, *65*, 712-25.

(166)   Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **2004**, *25*, 1157-74.

(167)   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **1983**, *79*, 926-935.

(168)   Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361-373.

(169)   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A.; Gaussian Inc.: Pittsburgh PA, 1998.

(170)   Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys* **1995**, *103*, 4613-4621.

(171)   Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comp Phys* **1977**, *23*, 327-341.

(172)   Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J Chem Phys* **1993**, *98*, 10089-10092.

(173)   Srinivasan, J.; Cheatham, T. E., 3rd; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J Am Chem Soc* **1998**, *120*, 9401-9409.

(174) Strockbine, B.; Rizzo, R. C. Binding of antifusion peptides with HIVgp41 from molecular dynamics simulations: quantitative correlation with experiment. *Proteins: Struct. Funct. Bioinformatics* **2007**, *67*, 630-42.

(175) Chachra, R.; Rizzo, R. C. Origins of Resistance Conferred by the R292K Neuraminidase Mutation via Molecular Dynamics and Free Energy Calculations. *J. Chem. Theory Comput.* **2008**, *4*, 1526-1540.

(176) Carrascal, N.; Rizzo, R. C. Calculation of binding free energies for non-zinc chelating pyrimidine dicarboxamide inhibitors with MMP-13. *Bioorg Med Chem Lett* **2009**, *19*, 47-50.

(177) McGillick, B. E.; Balius, T. E.; Mukherjee, S.; Rizzo, R. C. Origins of Resistance to the HIVgp41 Viral Entry Inhibitor T20. *Biochemistry* **2010**, *49*, 3575-3592.

(178) Grossfield, A.; Zuckerman, D. M. Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu Rep Comput Chem* **2009**, *5*, 23-48.

(179) Hess, B. Determining the shear viscosity of model liquids from molecular dynamics simulations. *J Chem Phys* **2002**, *116*, 209-217.

(180) Spackova, N.; Cheatham, T. E., 3rd; Ryjacek, F.; Lankas, F.; Van Meervelt, L.; Hobza, P.; Sponer, J. Molecular dynamics simulations and thermodynamics analysis of DNA-drug complexes. Minor groove binding between 4',6-diamidino-2-phenylindole and DNA duplexes in solution. *J Am Chem Soc* **2003**, *125*, 1759-69.

(181) Masukawa, K. M.; Kollman, P. A.; Kuntz, I. D. Investigation of Neuraminidase-Substrate Recognition Using Molecular Dynamics and Free Energy Calculations. *J Med Chem* **2003**, *46*, 5628-5637.

(182) Wong, S.; Amaro, R. E.; McCammon, J. A. MM-PBSA Captures Key Role of Intercalating Water Molecules at a Protein-Protein Interface. *J. Chem. Theory Comput.* **2009**, *5*, 422-429.

(183) Liu, Y.; Gray, N. S. Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol* **2006**, *2*, 358-364.

(184) Pao, W.; Miller, V. A.; Politi, K. A.; Riely, G. J.; Somwar, R.; Zakowski, M. F.; Kris, M. G.; Varmus, H. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* **2005**, *2*, e73.

(185) Sergina, N. V.; Rausch, M.; Wang, D.; Blair, J.; Hann, B.; Shokat, K. M.; Moasser, M. M. Escape from HER-family tyrosine kinase inhibitor therapy by the kinase-inactive HER3. *Nature* **2007**, *445*, 437-41.

(186) Hsieh, A. C.; Moasser, M. M. Targeting HER proteins in cancer therapy and the role of the non-target HER3. *Br J Cancer* **2007**, *97*, 453-7.

(187) Garrett, J. T.; Olivares, M. G.; Rinehart, C.; Granja-Ingram, N. D.; Sanchez, V.; Chakrabarty, A.; Dave, B.; Cook, R. S.; Pao, W.; McKinely, E.; Manning, H. C.; Chang, J.; Arteaga, C. L. Transcriptional and posttranslational up-regulation of HER3 (ErbB3) compensates for inhibition of the HER2 tyrosine kinase. *Proc Natl Acad Sci U S A* **2011**, *108*, 5021-6.

(188) Shi, F.; Telesco, S. E.; Liu, Y.; Radhakrishnan, R.; Lemmon, M. A. ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation. *Proc Natl Acad Sci U S A* **2010**, *107*, 7692-7.

(189) Jura, N.; Shan, Y.; Cao, X.; Shaw, D. E.; Kuriyan, J. Structural analysis of the catalytically inactive kinase domain of the human EGF receptor 3. *Proc Natl Acad Sci U S A* **2009**, *106*, 21608-13.

(190) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. Multidimensional free-energy calculations using the weighted histogram analysis method. *J Comput Chem* **1995**, *16*, 1339-1350.

(191) Roux, B. The calculation of the potential of mean force using computer simulations. *Comput Phys Comm* **1995**, *91*, 275-282.

(192) Kottalam, J.; Case, D. A. Dynamics of ligand escape from the heme pocket of myoglobin. *J Am Chem Soc* **1988**, *110*, 7690-7697.

(193) Huang, Y.; Rizzo, R. C. A Water-Based Mechanism of Specificity and Resistance for Lapatinib with ErbB Family Kinases. *Biochemistry* **2012**, *51*, 2390-2406.

(194) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* **1992**, *13*, 1011-1021.

(195) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J Chem Phys* **1984**, *81*, 3684-3690.

(196) Gschwind, A.; Fischer, O. M.; Ullrich, A. The discovery of receptor tyrosine kinases: targets for cancer therapy. *Nat Rev Cancer* **2004**, *4*, 361-370.

(197) Noble, M. E. M.; Endicott, J. A.; Johnson, L. N. Protein Kinase Inhibitors: Insights into Drug Design from Structure. *Science* **2004**, *303*, 1800-1805.

(198) Arora, A.; Scholar, E. M. Role of Tyrosine Kinase Inhibitors in Cancer Therapy. *J Pharmacol Exp Therapeut* **2005**, *315*, 971-979.

(199) Tibes, R.; Trent, J.; Kurzrock, R. Tyrosine kinase inhibitors and the dawn of molecular cancer therapeutics *Annu Rev Pharmacol Toxicol* **2004**, *45*, 357-384.

(200) Hm, K.; McCutcheon, I. E.; McCutcheon Ie; Flyvbjerg, A.; Friend, K. E. - The effects of insulin-like growth factors on tumorigenesis and neoplastic growth. *Endocr Rev* **2000**, *21*, 215-44.

(201) Mn, P.; Schernhammer, E. S.; Es, S.; Hankinson, S. E. - Insulin-like growth factors and neoplasia. *Nat Rev Cancer* **2004**, *4*, 505-18.

(202) Y, O.; Muller, H. L.; Hl, M.; Neely, E. K.; Ek, N.; Lamson, G.; Rosenfeld, R. G. - New concepts in insulin-like growth factor receptor physiology. *Growth Regul* **1993**, *3*, 113-23.

(203) Kato, H.; Faria, T. N.; Stannard, B.; Roberts, C. T., Jr.; LeRoith, D. Essential role of tyrosine residues 1131, 1135, and 1136 of the insulin-like growth factor-I (IGF-I) receptor in IGF-I action. *Mol Endocrinol* **1994**, *8*, 40-50.

(204) Butler; Yakar, S.; I, G.; Karas, M.; Y, O.; LeRoith, D. - Insulin-like growth factor-I receptor signal transduction: at the interface between physiology and cell biology. *Comp Biochem Physiol B* **1998**, *121*, 19-26.

(205) Baserga, R.; Hongo, A.; Rubini, M.; Prisco, M.; Valentinis, B. The IGF-I receptor in cell growth, transformation and apoptosis. *Biochim Biophys Acta* **1997**, *1332*, F105-26.

(206) Mulvihill, M. J.; Cooke, A.; Rosenfeld-Franklin, M.; Buck, E.; Foreman, K.; Landfair, D.; O'Connor, M.; Pirritt, C.; Sun, Y.; Yao, Y.; Arnold, L. D.; Gibson, N. W.; Ji, Q.-S. Discovery of OSI-906: a selective and orally efficacious dual inhibitor of the IGF-1 receptor and insulin receptor. *Future Medicinal Chemistry* **2009**, *1*, 1153-1171.

(207) Pautsch, A.; Zoephel, A.; Ahorn, H.; Spevak, W.; Hauptmann, R.; Nar, H. Crystal Structure of Bisphosphorylated IGF-1 Receptor Kinase: Insight into Domain Movements upon Kinase Activation. *Structure* **2001**, *9*, 955-965.

(208) Case, D. A.; Darden, T. A.; Cheatham, T.E., 3rd; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A.; University of California: San Francisco, CA, 2010.

(209) Nahta, R.; Yuan, L. X. H.; Du, Y.; Esteva, F. J. Lapatinib induces apoptosis in trastuzumab-resistant breast cancer cells: effects on insulin-like growth factor I signaling. *Mol Canc Therapeut* **2007**, *6*, 667-674.

(210) Wilsbacher, J. L.; Zhang, Q.; Tucker, L. A.; Hubbard, R. D.; Sheppard, G. S.; Bamaung, N. Y.; Fidanze, S. D.; Wang, G. T.; Hu, X.; Davidsen, S. K.; Bell, R. L.; Wang, J. Insulin-like Growth Factor-1 Receptor and ErbB Kinase Inhibitor Combinations Block Proliferation and Induce Apoptosis through Cyclin D1 Reduction and Bax Activation. *J Biol Chem* **2008**, *283*, 23721-23730.

(211) Haluska, P.; Carboni, J. M.; TenEyck, C.; Attar, R. M.; Hou, X.; Yu, C.; Sagar, M.; Wong, T. W.; Gottardis, M. M.; Erlichman, C. HER receptor signaling confers resistance to the insulin-like growth factor-I receptor inhibitor, BMS-536924. *Mol Cancer Ther* **2008**, *7*, 2589-98.

(212) Hubbard, R. D.; Bamaung, N. Y.; Fidanze, S. D.; Erickson, S. A.; Palazzo, F.; Wilsbacher, J. L.; Zhang, Q.; Tucker, L. A.; Hu, X.; Kovar, P.; Osterling, D. J.; Johnson, E. F.; Bouska, J.; Wang, J.; Davidsen, S. K.; Bell, R. L.; Sheppard, G. S. Development of multitargeted

inhibitors of both the insulin-like growth factor receptor (IGF-IR) and members of the epidermal growth factor family of receptor tyrosine kinases. *Bioorg Med Chem* **2009**, *19*, 1718-1721.

(213)   Steve D. Fidanze; Scott A. Erickson; Robert D. Hubbard; Gary T. Wang; Robert A. Mantei; Nwe Y. BaMaung; Richard F. Clark; Bryan K. Sorensen; Peter Kovar; William Pappano; Qian Zhang; Julie L. Wilsbacher; Xiaoming Hu; Lora A. Tucker; Min Cheng; Kerren K. Swinger; Kent D. Stewart; Jieyi Wang; Randy L. Bell; Steven K. Davidsen; Sheppard, G. S. Imidazo[2,1-b]thiazole and imidazo[1,2-a]pyridine amides as novel inhibitors of the insulin-like growth factor (IGF1R) and members of the epidermal growth factor family of tyrosine kinases. *Mol Cancer Ther* **2009**, poster #A245.

(214)   Peng, H.; Huang, N.; Qi, J.; Xie, P.; Xu, C.; Wang, J.; Yang, C. Identification of novel inhibitors of BCR-ABL tyrosine kinase via virtual screening. *Bioorg Med Chem* **2003**, *13*, 3693-3699.

(215)   Foloppe, N.; Fisher, L. M.; Howes, R.; Kierstan, P.; Potter, A.; Robertson, A. G. S.; Surgenor, A. E. Structure-Based Design of Novel Chk1 Inhibitors: Insights into Hydrogen Bonding and Protein−Ligand Affinity. *J Med Chem* **2005**, *48*, 4332-4345.

(216)   Ms, E.; El-Araby, M. E.; Serya, R. T.; Abouzid, K. A. M. - Virtual screening and synthesis of new chemical scaffolds as VEGFR-2 kinase inhibitors. *Arzneimittelforschung* **2012**, *62*, 554-60.

(217)   Rabal, O.; Schneider, G.; Borrell, J. I.; Teixidó, J. Structure-based virtual screening of FGFR inhibitors: cross-decoys and induced-fit effect. *BioDrugs* **2007**, *21*, 31-45.

(218)   Vadivelan, S.; Sinha, B. N.; Irudayam, S. J.; Jagarlapudi, S. A. R. P. Virtual Screening Studies to Design Potent CDK2-Cyclin A Inhibitors. *J Chem Inform Model* **2007**, *47*, 1526-1535.

(219)   Mayer, S. C.; Banker, A. L.; Boschelli, F.; Di, L.; Johnson, M.; Kenny, C. H.; Krishnamurthy, G.; Kutterer, K.; Moy, F.; Petusky, S.; Ravi, M.; Tkach, D.; Tsou, H.-R.; Xu, W. Lead identification to generate isoquinolinedione inhibitors of insulin-like growth factor receptor (IGF-1R) for potential use in cancer treatment. *Bioorg Med Chem* **2008**, *18*, 3641-3645.

(220)   Miller, L. M.; Mayer, S. C.; Berger, D. M.; Boschelli, D. H.; Boschelli, F.; Di, L.; Du, X.; Dutia, M.; Floyd, M. B.; Johnson, M.; Kenny, C. H.; Krishnamurthy, G.; Moy, F.; Petusky, S.; Tkach, D.; Torres, N.; Wu, B.; Xu, W. Lead identification to generate 3-cyanoquinoline inhibitors of insulin-like growth factor receptor (IGF-1R) for potential use in cancer treatment. *Bioorg Med Chem* **2009**, *19*, 62-66.

(221)   Nemecek, C.; Metz, W. A.; Wentzler, S.; Ding, F.-X.; Venot, C.; Souaille, C.; Dagallier, A.; Maignan, S.; Guilloteau, J.-P.; Bernard, F.; Henry, A.; Grapinet, S.; Lesuisse, D. Design of Potent IGF1-R Inhibitors Related to Bis-azaindoles. *Chem Biol Drug Des* **2010**, *76*, 100-106.

(222)   Wu, J.; Li, W.; Craddock, B. P.; Foreman, K. W.; Mulvihill, M. J.; Ji, Q.-s.; Miller, W. T.; Hubbard, S. R. Small-molecule inhibition and activation-loop trans-phosphorylation of the IGF1 receptor. *Embo J* **2008**, *27*, 1985-1994.

(223)   Lesuisse, D.; Mauger, J.; Nemecek, C.; Maignan, S.; Boiziau, J.; Harlow, G.; Hittinger, A.; Ruf, S.; Strobel, H.; Nair, A.; Ritter, K.; Malleron, J.-L.; Dagallier, A.; El-Ahmad, Y.; Guilloteau, J.-P.; Guizani, H.; Bouchard, H.; Venot, C. Discovery of the first non-ATP competitive IGF-1R kinase inhibitors: Advantages in comparison with competitive inhibitors. *Bioorg Med Chem* **2011**, *21*, 2224-2228.

(224)   Velaparthi, U.; Wittman, M.; Liu, P.; Stoffan, K.; Zimmermann, K.; Sang, X.; Carboni, J.; Li, A.; Attar, R.; Gottardis, M.; Greer, A.; Chang, C. Y.; Jacobsen, B. L.; Sack, J. S.; Sun, Y.; Langley, D. R.; Balasubramanian, B.; Vyas, D. Discovery and initial SAR of 3-(1H-benzo[d]imidazol-2-yl)pyridin-2(1H)-ones as inhibitors of insulin-like growth factor 1-receptor (IGF-1R). *Bioorg Med Chem* **2007**, *17*, 2317-2321.

(225)   Hubbard, R. D.; Bamaung, N. Y.; Palazzo, F.; Zhang, Q.; Kovar, P.; Osterling, D. J.; Hu, X.; Wilsbacher, J. L.; Johnson, E. F.; Bouska, J.; Wang, J.; Bell, R. L.; Davidsen, S. K.; Sheppard, G. S. Pyrazolo[3,4-d]pyrimidines as potent inhibitors of the insulin-like growth factor receptor (IGF-IR). *Bioorg Med Chem* **2007**, *17*, 5406-5409.

(226)   Available Chemicals Directory. www.mdli.com/products/acd.html (Accessed May 14th 2013).

(227)   Bolton, E.; Wang, Y.; Thiessen, P.; Bryant, S. PubChem: integrated platform of small molecules and biological activities. *Annu Rep Comput Chem* **2008**, *4*, 217 - 241.

(228)   Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **2001**, *46*, 3-26.

(229)   Head, R. D.; Smythe, M. L.; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R. VALIDATE: A New Method for the Receptor-Based Prediction of Binding Affinities of Novel Ligands. *J. Am. Chem. Soc.* **1996**, *118*, 3959-3969.

(230)   Mayer, S. C.; Banker, A. L.; Boschelli, F.; Di, L.; Johnson, M.; Kenny, C. H.; Krishnamurthy, G.; Kutterer, K.; Moy, F.; Petusky, S.; Ravi, M.; Tkach, D.; Tsou, H. R.; Xu, W. Lead identification to generate isoquinolinedione inhibitors of insulin-like growth factor receptor (IGF-1R) for potential use in cancer treatment. *Bioorg Med Chem Lett* **2008**, *18*, 3641-5.

(231)   Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* **2002**, *23*, 1623-41.

(232)   University of California at San Francisco Computer Graphics Laboratory: San Francisco, CA. http://www.cgl.ucsf.edu/Overview/software.html (Accessed May 14th 2013).

(233) DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J Med Chem* **1988**, *31*, 722-9.

(234)   Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit* **1996**, *9*, 1-5.

(235)   Brown, R. D.; Martin, Y. C. Use of structure Activity data to compare structure-based clustering methods and descriptors for use in compound selection. *J Chem Inform Comput Sci* **1996**, *36*, 572-584.