



Publicly Accessible Penn Dissertations


Spring 5-16-2011

Unveiling the Molecular Mechanisms Regulating the Activation of the ErbB Family Receptors at Atomic Resolution through Molecular Modeling and Simulations

Andrew Shih

University of Pennsylvania, shihaj@seas.upenn.edu

Follow this and additional works at: <http://repository.upenn.edu/edissertations>

 Part of the [Biophysics Commons](#), [Other Biomedical Engineering and Bioengineering Commons](#), and the [Structural Biology Commons](#)

Recommended Citation

Shih, Andrew, "Unveiling the Molecular Mechanisms Regulating the Activation of the ErbB Family Receptors at Atomic Resolution through Molecular Modeling and Simulations" (2011). *Publicly Accessible Penn Dissertations*. 302.
<http://repository.upenn.edu/edissertations/302>

This paper is posted at ScholarlyCommons. <http://repository.upenn.edu/edissertations/302>
For more information, please contact libraryrepository@pobox.upenn.edu.

Unveiling the Molecular Mechanisms Regulating the Activation of the ErbB Family Receptors at Atomic Resolution through Molecular Modeling and Simulations

Abstract

The EGFR/ErbB/HER family of kinases contains four homologous receptor tyrosine kinases that are important regulatory elements in key signaling pathways. To elucidate the atomistic mechanisms of dimerization-dependent activation in the ErbB family, we have performed molecular dynamics simulations of the intracellular kinase domains of the four members of the ErbB family (those with known kinase activity), namely EGFR, ErbB2 (HER2) and ErbB4 (HER4) as well as ErbB3 (HER3), an assumed pseudokinase, in different molecular contexts: monomer vs. dimer, wildtype vs. mutant. Using bioinformatics and fluctuation analyses of the molecular dynamics trajectories, we relate sequence similarities to correspondence of specific bond-interaction networks and collective dynamical modes. We find that in the active conformation of the ErbB kinases (except ErbB3), key subdomain motions are coordinated through conserved hydrophilic interactions: activating bond-networks consisting of hydrogen bonds and salt bridges. The inactive conformations also demonstrate conserved bonding patterns (albeit less extensive) that sequester key residues and disrupt the activating bond network. Both conformational states have distinct hydrophobic advantages through context-specific hydrophobic interactions. The inactive ErbB3 kinase domain also shows coordinated motions similar to the active conformations, in line with recent evidence that ErbB3 is a weakly active kinase, though the coordination seems to arise from hydrophobic interactions rather than hydrophilic ones. We show that the functional (activating) asymmetric kinase dimer interface forces a corresponding change in the hydrophobic and hydrophilic interactions that characterize the inactivating interaction network, resulting in motion of the α C-helix through allostery. Several of the clinically identified activating kinase mutations of EGFR act in a similar fashion to disrupt the inactivating interaction network. Our molecular dynamics study reveals the asymmetric dimer interface helps progress the ErbB family through the activation pathway using both hydrophilic and hydrophobic interaction. There is a fundamental difference in the sequence of events in EGFR activation compared with that described for the Src kinase Hck.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Bioengineering

First Advisor

Ravi Radhakrishnan

Keywords

ErbB family, kinase activation, molecular dynamics, hydrophilic and hydrophobic interactions

Subject Categories

Biophysics | Other Biomedical Engineering and Bioengineering | Structural Biology

UNVEILING THE MOLECULAR MECHANISMS REGULATING THE ACTIVATION OF
THE ERBB FAMILY RECEPTORS AT ATOMIC RESOLUTION THROUGH
MOLECULAR MODELING AND SIMULATIONS

Andrew J. Shih

A DISSERTATION

in Bioengineering

Presented to the Faculties of the University of Pennsylvania

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2011

Supervisor of Dissertation

Dr. Ravi Radhakrishnan, PhD, Associate Professor of Bioengineering

Graduate Group Chair

Dr. Christopher S. Chen, MD, PhD, Professor of Bioengineering

Dissertation Committee

Dr. Casim A. Sarkar, PhD, Assistant Professor of Bioengineering

Dr. Mark A. Lemmon, PhD, Professor of Biochemistry and Biophysics

Dr. Jeffery G. Saven, PhD, Associate Professor of Chemistry

Acknowledgements

First and foremost my sincerest gratitude to my advisor, Professor Ravi Radhakrishnan, for guiding me and staying with me through the many twists and turns (and sometimes dead ends) in exploring the EGFR kinase that eventually expanded into the other ErbB members. His passion and creativity made it a joy delving deeper into the activation mechanisms of kinases.

A special thanks to Professor Casim Sarkar, Professor Mark Lemmon and Professor Jeff Saven, as my thesis committee members and advisors: for being understanding during my first forays into molecular biology as well as for advice inside and outside my field of expertise, giving me the breadth and scope of knowledge for proper perspective.

I wish to thank the Sung Hee Choi and Fumin Shi of the Lemmon Lab for all the insights into the ErbB family kinases. Also, I greatly appreciate all the lab members of the Radhakrishnan lab for an enjoyable research environment and knowledgeable discussions. In particular, I would like to thank Shannon Telesco and Yingting Liu for help with new programs and processing developments as we intensified our studies of EGFR and other members of the ErbB family.

Beyond the lab, this work could not have possible without Alice Chau and Yuuai, for unquestioned support, keeping me grounded and serving as sounding board to my (sometimes crazy) ideas.

Finally, I want to thank my family for everything they have done to make me who I am.

Abstract

Unveiling the Molecular Mechanisms Regulating the Activation of the ErbB Family Receptors at Atomic Resolution through Molecular Modeling and Simulations

Andrew Shih

Advisor: Ravi Radhakrishnan

The EGFR/ErbB/HER family of kinases contains four homologous receptor tyrosine kinases that are important regulatory elements in key signaling pathways. To elucidate the atomistic mechanisms of dimerization-dependent activation in the ErbB family, we have performed molecular dynamics simulations of the intracellular kinase domains of the four members of the ErbB family (those with known kinase activity), namely EGFR, ErbB2 (HER2) and ErbB4 (HER4) as well as ErbB3 (HER3), an assumed pseudokinase, in different molecular contexts: monomer vs. dimer, wildtype vs. mutant. Using bioinformatics and fluctuation analyses of the molecular dynamics trajectories, we relate sequence similarities to correspondence of specific bond-interaction networks and collective dynamical modes. We find that in the active conformation of the ErbB kinases (except ErbB3), key subdomain motions are coordinated through conserved hydrophilic interactions: activating bond-networks consisting of hydrogen bonds and salt bridges. The inactive conformations also demonstrate conserved bonding patterns (albeit less extensive) that sequester key residues and disrupt the activating bond network. Both conformational states have distinct hydrophobic advantages through context-specific hydrophobic interactions. The inactive ErbB3 kinase domain also shows coordinated

motions similar to the active conformations, in line with recent evidence that ErbB3 is a weakly active kinase, though the coordination seems to arise from hydrophobic interactions rather than hydrophilic ones. We show that the functional (activating) asymmetric kinase dimer interface forces a corresponding change in the hydrophobic and hydrophilic interactions that characterize the inactivating interaction network, resulting in motion of the α C-helix through allostery. Several of the clinically identified activating kinase mutations of EGFR act in a similar fashion to disrupt the inactivating interaction network. Our molecular dynamics study reveals the asymmetric dimer interface helps progress the ErbB family through the activation pathway using both hydrophilic and hydrophobic interaction. There is a fundamental difference in the sequence of events in EGFR activation compared with that described for the Src kinase Hck.

Table of Contents

List of Tables	viii
List of Figures.....	ix
Chapter 1.) Regulation of ErbB Receptor Tyrosine Kinases.....	1
1.1) RTK Structure and Signaling	1
1.2) Regulation of RTK signaling duration at the cell surface	3
1.3) Regulation, Structure and Auto-inhibition of RTKs	5
1.4) Kinase Domain Mutations in Cancer and their Therapeutic Importance	8
Chapter 2.) Computational Methods Related to Reaction Chemistry.....	15
2.1) Introduction	15
2.2) Computational Methods	16
2.2.1) Molecular Dynamics.....	16
2.2.2) Homology Modeling.....	21
2.2.3) Free Energy.....	26
2.2.4) Electronic Structure Methods	29
2.2.5) Methods for Determining Reaction Paths	33
2.2.6) Effect of Force on Biomolecules	38
2.2.7) Limitations and Caveats	43
2.3) Future Directions	45
Chapter 3.) Molecular Systems Biology of ErbB1 Signaling: Bridging the Gap through Multiscale Modeling and High-Performance Computing.....	47
3.1) Introduction	47
3.2) Models and Methods	48
3.3) Rationalizing and Predicting the Effects of Molecular Perturbations on Receptor Kinase Activation.....	50
Chapter 4.) Molecular Dynamics Analysis of Conserved Hydrophobic and Hydrophilic Bond Interaction Networks in ErbB Family Kinases.	57
4.1) Introduction	57
4.2) Methods	62
4.3) Results	65
4.3.1) PCA reveals a tightly coordinated motion in all active ErbB members.....	66
4.3.2) Activation in the asymmetric dimer occurs through disruption of the inactivating bond-network	70
4.3.3) EGFR mutations activate the kinase by disrupting the inactivating bond network	73
4.3.4) Hydrophobic interactions help differentiate active and inactive conformations	74

4.4) Discussion.....	79
Chapter 4A.) Appendix for 4: Molecular Dynamics Analysis of Conserved Hydrophobic and Hydrophilic Bond Interaction Networks in ErbB Family Kinases	86
4A.1) A-loop phosphorylation in monomers promotes coordination of the catalytic sub-domains	90
4A.2) Phosphorylation of the A-loop does not perturb the dimer-mediated allosteric activation mechanism.....	94
4A.3) Hydrophobic interactions are not significantly affected by A-loop phosphorylation.....	95
4A.4) Targeted molecular dynamics studies reveal specific interactions associated with the transition pathway	98
Chapter 5.) A multiscale modeling approach to investigate molecular mechanisms of pseudokinase activation and drug resistance in the HER3/ErbB3 receptor tyrosine kinase signaling network	101
5.1) Introduction	101
5.2) Results	105
5.2.1) PCA reveals that the sub-domain motions of ErbB3 diverge from those of an inactive ErbB kinase	105
5.2.2) Hydrogen bonding analysis indicates a lack of conserved hydrophilic interactions in the ErbB3 catalytic site	109
5.2.3) Hydrophobic analysis of the ErbB kinases identifies a perturbation-sensitive region in ErbB3.....	112
5.3) Discussion.....	116
5.4) Methods	118
Chapter 6.) Future Work	121
6.1) The Asymmetric Dimer Interface achieves Activation through Allosteric Means, Hydrophilic Interactions and Hydrophobic Shielding	124
6.2) Preliminary Results.....	128
6.2.1) RMSD Analysis Reveals the Activation Pathway of the Dimer is Distinct from the Monomer	128
6.2.2) Interkinase Salt Bridges Alter the Conformation of the α C-helix and Promote Partial Extension.....	129
6.2.3) The Dimer Interface Shields the “Receiver” Kinase from Hydrophobic Effects during the Transition to Active.....	134
6.3) Discussion.....	137
6.4) Methods	141
Chapter 7.) Perspectives	142
7.1) Homology modeling of the ErbB3 kinase domain	143

7.2) Clinical Implications of Oncogenic EGFR mutations from a Multiscale Model of ErbB Receptor Signaling	150
7.3) Systems model of ErbB signaling defines a mechanism for ErbB3-mediated TKI resistance.....	154
7.4) Conclusions	158
Appendix A: Hydrophobic interactions help identify perturbation sensitive regions for the ErbB kinases	160
Appendix B: Exhaustive Persistent Bond Tables for EGFR TMD simulations	165
References	167

List of Tables

Table 1.1a: Currently known clinically identified activating cytoplasmic domain mutations in RTKs classified by kinase sub-domains in each of most common tumor types.	10
Table 1.1b: Currently known clinically identified activating cytoplasmic domain mutations in RTKs classified by kinase sub-domains in each of most common tumor types.	11
Table 1.2: FDA approved RTK inhibitors currently in use.	13
Table 4.1: Persistent H-bonds and salt bridges in three homologous ErbB kinase monomer systems (EGFR, ErbB2, ErbB4).	68
Table 4A.1: Exhaustive list of the persistent H-bonds and salt bridges in three homologous ErbB kinase monomer systems (EGFR, ErbB2, ErbB4).	86
Table 4A.2A: Persistent H-bonds and salt bridges in five ErbB kinase dimer systems. ..	87
Table 4A.2B: Persistent H-bonds and salt bridges in five ErbB kinase dimer systems. ..	88
Table 4A.3: Persistent H-bonds and salt bridges in the phosphorylated ErbB kinase monomer systems.	92
Table 5.1a: Hydrogen bonding analysis of the ErbB3 MD trajectory and comparison to the active and inactive conformations of EGFR, ErbB2 and ErbB4.	110
Table 5.1b: Hydrogen bonding analysis of the ErbB3 MD trajectory and comparison to the active and inactive conformations of EGFR, ErbB2 and ErbB4.	111
Table 6.1a: Persistent H-bonds and salt bridges in the EGFR kinase dimer system as it progresses from inactive to active.	131
Table 6.1b: Persistent H-bonds and salt bridges in the EGFR kinase monomer system as it progresses from inactive to active.	132
Table 6.2: Persistent H-bonds and salt bridges between kinase domains in the EGFR kinase dimer systems as it progresses from inactive to active.	133
Table 7.1: Time scales and length scales of common phenomena associated with protein dynamics.	152
Table 7.2: Signaling and inhibition characteristics predicted by our network simulations for wildtype and mutant systems under different conditions.	152
Table B1: Exhaustive list of persistent H-bonds and salt bridges in the EGFR kinase in an asymmetric dimer formation as it progresses from inactive to active.	165
Table B2: Exhaustive list of persistent H-bonds and salt bridges in the EGFR monomer as it progresses from inactive to active.	166

List of Figures

Figure 1.1: Activation scheme for the ErbB family and pertinent subdomains.....	6
Figure 1.2: Non-unique mutations cataloged in cancer samples for the ErbB family in catalytically important kinase sub-domains.....	12
Figure 2.1: Convergence analysis in TPS simulations.....	35
Figure 2.2: Schematic of the effect of the force on the free energy landscape.....	40
Figure 3.1: Hierarchical multiscale modeling scheme for ErbB signaling.....	49
Figure 3.2: Stabilizing network residues in ErbB1 kinase in its active and inactive conformations and the effects of mutations upon them.....	51
Figure 4.1: Alignment and visualization of the four catalytic subdomains comprising the active site in the ErbB kinases.....	58
Figure 4.2: Visualization of the first principal component for the four key subdomains in the inactive and active conformation of the ErbB kinases.....	66
Figure 4.3: Analysis of the mutant L834R and del 729-729 ins S EGFR systems.....	72
Figure 4.4: Normalized water density fluctuations and the Mean surface accessible (exposed) surface area (SASA) values for the C-spine and R-spine.....	74
Figure 4.5: Normalized water density fluctuations and the Mean surface accessible (exposed) surface area (SASA) values for the hydrophobic core, α C- β 4 region and dimer interface.....	76
Figure 4A.1: The RMSD plots for the monomeric systems showing RMSD from the active and inactive conformations.....	89
Figure 4A.2: The RMSD plots for the EGFR symmetric dimer systems showing RMSD from the active and inactive conformations.....	89
Figure 4A.3: Visualization of the first principal component for the four key sub-domains in the inactive and active phosphorylated ErbB systems.....	90
Figure 4A.4: Comparison of the normalized water density fluctuations with the mean solvent accessible surface area for two functionally important sub-regions for the ErbB family in the A-loop phosphorylated systems for the dimer interface and the α C- β 4 region.....	96
Figure 4A.5: Comparison of the normalized water density fluctuations with the solvent accessible surface area for three functionally important sub-regions for the ErbB family in the A-loop phosphorylated systems for the C-spine, the R-spine and the hydrophobic core.....	97
Figure 4A.6: Analysis of the EGFR TMD monomer and the EGFR TMD dimer simulations.....	99
Figure 5.1. Molecular dynamics time course plots of the RMSD for the ErbB3 kinase.	106

Figure 5.2: Motion along the first principal component of the MD trajectory is illustrated for the complete ErbB3 kinase and compared to the active and inactive conformations of EGFR, ErbB2 and ErbB4.....	107
Figure 5.3. Motion along the first principal component of the MD trajectory is illustrated for ErbB3 and compared to the active and inactive conformations of EGFR, ErbB2 and ErbB4 focused on the active site.....	108
Figure 5.4: Correlation between the solvent accessible surface area (SASA) and water density fluctuations in the key hydrophobic regions of the ErbB kinases.....	114
Figure 6.1: RMSD plots tracking the progression of the targeted molecular dynamics of the EGFR dimer and the monomer simulation down the activation pathway.	128
Figure 6.2: Order of events as they occur for the EGFR dimer simulations with enhanced views of the active site.....	134
Figure 6.3: Hydrophobic plots of the solvent accessible surface area (SASA) versus normalized water density fluctuations in the key hydrophobic regions of the ErbB kinases.....	136
Figure 7.1: Multiple sequence alignment of the kinase domains of the ErbB RTK family members.....	145
Figure 7.2. The top 10 refined A-loops for the best ErbB3 models constructed from each ErbB template.	146
Figure 7.3. Enhanced view of unique subdomains in the homology modeled ErbB3 in comparison to EGFR, ErbB4 and multiple templates.....	148
Figure 7.4: Schematic diagram of the EGFR signaling model used as well as the ErbB3 signaling model used.....	151
Figure 7.5 Parameter sensitivity analysis of the ErbB3 signaling model.	156
Figure A1: Hydrophobic plots of the solvent accessible surface area (SASA) versus normalized water density fluctuations in the key hydrophobic regions of the ErbB kinases.....	160

Chapter 1.) Regulation of ErbB Receptor Tyrosine Kinases

Lung Cancer is the number one cause of death (157,000 in 2010) in America and accounts for almost as many deaths as the next three cancers (Breast/Prostate (40,000/32,000 in 2010), Colon (51,000 in 2010) and Pancreas (36,000 in 2010)) in both genders combined (American Cancer Society). About 85-90% of Lung Cancer cases are Non-Small Cell Lung Cancers (NSCLCs), and present a poor prognosis in comparison to breast/prostate and colon cancers. Recent advances in Cancer Therapeutics targeting the EGFR/ErbB family of Receptor Tyrosine Kinases (RTKs) have shown promise in providing an alternate therapy for NSCLC patients with efficacy at least equivalent to chemotherapy in patients harboring mutant forms of EGFR. Here in this chapter we shall focus our attention upon the ErbB family, and in particular the intracellular kinase domain; however we will review overall RTK function and regulation as well.

1.1) RTK Structure and Signaling

The ErbB family kinases are a set of four homologous RTKs: EGFR/ErbB1/HER1, ErbB2/HER2, ErbB3/HER3, ErbB4/HER4. RTKs are single pass transmembrane protein important in intercellular communication, by translating extracellular signals (ligands) into activation of specific cell signaling cascades (reviewed in [1, 2]). In humans, there are currently 58 known RTKs divided into 20 families. RTKs are composed of an extracellular ligand binding domain, a single transmembrane helix, a juxtamembrane

domain, a cytoplasmic kinase domain and a C-terminal tail containing multiple phosphorylatable tyrosines. Activation of the extracellular domain by ligands triggers receptor dimerization and *trans*-phosphorylation of both the cytoplasmic kinase domains as well as the C-terminal tail. Dimerization and phosphorylation of the kinase domains activates the kinase domain which then phosphorylates the tyrosines in the C-terminal tail. The phosphorylated tyrosines serve as a docking site for downstream signaling molecules containing Src homology domain 2 and/or phosphotyrosine-binding domains and in turn relay the signal into the cell. The cell signaling pathways RTKs modulate involve crucial cellular processes such as cell proliferation, differentiation, metabolism, migration, and apoptosis.

The ErbB signaling network presents a “bow-tie” architecture, where multiple inputs and outputs are linked through a set of core processes [3]. The ErbB kinases are capable of binding a variety of ligands resulting in homo and heterodimerization, improving the flexibility and robustness in the ErbB signaling network by increasing response to extracellular signaling as well as allowing for cross-talk between ErbB kinase dimers, compensating for any reduced signaling of an individual ErbB member in a given cell type. For example, ErbB2 does not currently have a known ligand and ErbB3 is missing key residues in the active site, greatly reducing its kinase efficacy: a pseudo kinase [4]. However, ErbB2/ErbB3 heterodimers are extremely relevant in the signaling picture as overexpression is correlated with oncogenic transformation in breast cancers [5, 6]. Extracellular signaling is reduced to the four ErbB family kinases, which is then fed

through the core process: a conserved, relatively small collection of biochemical interactions. The ErbB network then expands again through transcription factors and positive as well as negative feedback mechanisms, eventually leading to the important cell signaling pathways of proliferation, differentiation, etc.

1.2) Regulation of RTK signaling duration at the cell surface

Following the phosphorylation of the RTK kinase domains, there are pathways to modulate the length of time the kinase is active on the cell surface, mainly receptor-mediated endocytosis and phosphatases. Upon ligand induced activation, RTKs are internalized which removes the active RTK as well as the ligand from the cell surface (reviewed in [7-9]). The predominant pathway for endocytosis for RTKs is clathrin-mediated endocytosis, where the RTKs are rapidly endocytosed via clathrin-coated pits. There are also clathrin independent endocytosis pathways for RTK internalization, but they are not as well understood physiologically. One of the members of the ErbB family, ErbB4, has an alternate method of internalization by means of proteolytic cleavage [10], which constitutes a biochemical switch and is involved in proper cardiac and neural development [11, 12].

Protein Tyrosine Phosphatases (PTPs) are the opposite half of the RTKs, removing the phosphate group on phosphotyrosines. The balance of the interplay between the RTKs and PTPs act as a major switch controlling the full activation of RTKs and thereby the

cell decision fates [13]. Before RTK activation, PTPs are in constant activity to reduce any residual phosphorylation from cross-talk and accidental activation. Given significant ligand, the RTKs inhibit local PTP activity and have enhanced signal propagation [14]. Paradoxically, ligand binding also causes recruitment of PTPs that bind to target RTKs, dephosphorylates them, stabilizes the inactive form at the cell surface and inhibits further signaling [15]. The bivalent reaction of PTPs to RTK ligand binding may be from general non-specific PTP function versus RTK specific PTP function, however this is beyond the scope of this thesis.

Adding to activation and regulation of RTKs is the role of recently discovered cytoplasmic proteins cytohesins in EGFR [16] (and the proteins Dok7 in MuSK [17]). The proteins modulate the activity of the proteins in both a positive and negative fashion dependent on concentration. Increased amounts of the proteins activate the RTKs without any ligand binding events, while lacking the proteins prevents the activation of the RTKs even with a ligand binding event. It is clear that cytohesins are important in the scheme of ErbB in dimerization and activation though their specific role in the cell is unclear: as an extra layer of control, where a level of local concentration is necessary for activation or as an activator from the cytoplasmic side, where the cell can trigger the RTK cascade without a ligand binding domain or as a signal propagator, where activation of a single EGFR molecule causes recruitment of cytohesins to further increase the activity or if any combination of the three are true.

1.3) Regulation, Structure and Auto-inhibition of RTKs

At the protein level, the extracellular domains are locked into an auto-inhibitory state preventing dimerization and are released with a ligand binding event. The specifics of how ligand binding facilitates dimerization for each RTK falls in the spectrum of “ligand mediated” dimerization, where the ligands bridge the two receptors without the receptors making direct contact, and “receptor mediated” dimerization, where the ligands make no direct contribution to the dimer interface (all mechanisms are reviewed in [18]). The ErbB family represents the extreme of “receptor mediated” dimerization [19, 20]. The ErbB extracellular domain consists of four domains, with auto-inhibitory interaction between domains II and IV in a tethered conformation (Figure 1.1A) [21-24]. Ligands for the ErbB family are bivalent and bind to Domains I and III which cause a conformational change breaking the tethered conformation and exposing a dimerization arm in domain II allowing the dimerization arm to contact another ErbB RTK molecule.

Each RTK kinase domain is *cis*-autoinhibited in a characteristic fashion (reviewed in [18]) with activation mechanisms being unveiled as each RTK is being studied more in-depth. In RTK signaling, the intracellular kinase domain catalyzes transfer of the γ -phosphate of ATP to tyrosines on both the RTK itself and in other target substrates (reviewed in [1]). Regulation of the RTK kinase domain is thought to involve contributions from several conserved subregions: the catalytic loop (C-loop), the activation loop (A-loop), the glycine-rich nucleotide binding loop (P-loop), and the α C-

helix, which together define the active site in the cleft between the β strand-rich N-lobe and the helical C-lobe. The catalytic loop residues directly participate in phosphoryl transfer. The A-loop and the α C-helix (Figure 1.1B) modulate the activity of the kinase domain by regulating accessibility of the active site to binding and coordinating both ATP and the substrate tyrosine. The A-loop and the α C-helix in ErbB kinases contains one phosphorylatable tyrosine (Y845 in EGFR, Y877 in ErbB2, Y850 in ErbB4, note: there are two numbering schemes for the ErbB family where Y845 is equivalent to Y869). The α C-helix and P-loop must be positioned correctly to coordinate the ATP and the substrate tyrosine for effective phosphoryl transfer.

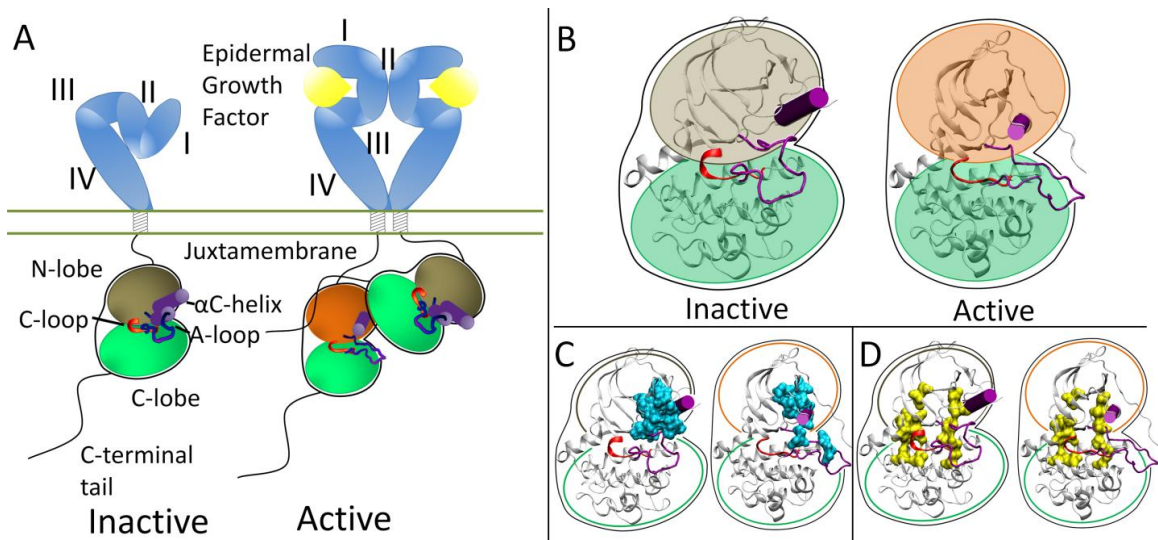


Figure 1.1: (A) Activation scheme for the ErbB family. The inactive kinase (brown N-lobe) is auto-inhibited through the A-loop and α C-helix (Purple). Introduction of the asymmetric dimer interface rotates the α C-helix to the active state (orange N-lobe). (B) Enhanced view of the inactive and active kinase domains (C) Hydrophobic core (cyan) in the inactive and active conformations (D) C-spine (left yellow spine) and R-spine (right yellow spine) in the inactive and active conformations.

In the inactive state there is a small hydrophobic ‘core’ formed between the α C-helix and the A-loop, that helps maintain the kinase in the inactive conformation (Figure 1.1C). Disruption of this hydrophobic core by single point mutations has been shown to activate EGFR [25-29]. Recent structural studies have revealed highly conserved hydrophobic “spines” within active kinases that are considered important in catalysis [30, 31], shown in Figure 1.1D. The regulatory spine (R-spine) consists of four hydrophobic side chains (M742, L753, H811, F832 in EGFR) anchored by an aspartic acid in the α F-helix (D872 in EGFR) and coordinates the motion of the N- and C-lobes of the kinase [30]. The catalytic spine (C-spine) involves eight hydrophobic side-chains (V702, A719, L774, V819, L820, V821, T879, L883 in EGFR) that help support and coordinate the adenine ring of ATP in the active state during phosphotransfer [31].

Many of the kinase domains are inhibited through steric hindrances of a protein segment blocking off the active site and greatly reducing the efficacy of the kinase. Dimerization puts two kinase domains in close proximity to each other and although the kinase efficiency is greatly reduced, it is theorized that each kinase still has enough activity to phosphorylate its dimer partner. Phosphorylation of the protein segment prevents the protein segment from binding into the active site and allows the kinase to fully function. For the Ins [32] and FGFR [33] family of kinases, the A-loop serves as the inhibitory segment, while the juxtamembrane domain serves the same autoinhibitory role as the A-loop in MuSK [34], Flt3 [35], KIT [36] and the Eph [37] family and in the Tie2 [38] kinase, a segment of the C-terminal tail acts as the auto-inhibitor.

The ErbB family has a different method of inactivation as phosphorylation of the A-loop or any other protein segment does activate the kinase; rather than a steric hindrance of the active site, there are collective auto-inhibitory interactions preventing the proper coordination between key loops in the kinase [39, 40]. Activation of the kinase domain is also achieved through dimerization, though in this case, the dimer interface itself serves as the activating mechanism. In the ErbB family, the dimer interface is asymmetric similar to that of the cyclin dependent kinases and cyclin; the C-lobe one kinase, the “activator”, contacts the N-lobe of the other kinase, the “receiver,” with the asymmetric dimer contacts causing a conformational change towards the active state through allosteric methods (Figure 1.1A) [41, 42]. The α C-helix in the inactive ErbB kinase is rotated out preventing key bonds from forming. Introduction of the activating asymmetric dimer interface forces the α C-helix to sample a different conformational space biasing towards the active state. Furthermore, the juxtamembrane domain in EGFR serves as latch to facilitate the asymmetric dimer interface between kinase domains [43-45].

1.4) Kinase Domain Mutations in Cancer and their Therapeutic Importance

Deregulation and mutation of RTKs have been correlated with cancer almost immediately after their discovery and purification in the early 1980s. The v-erbB oncogene in the avian erythroblastosis virus, that was capable of inducing acute leukemia

encoded, was found to encode a constitutively active form of the homologous ErbB kinase protein [46]. With the increased study upon RTKs, the correlation between deregulation of RTKs and a variety of ailments and particularly in cancer has only grown stronger. Deregulation of RTKs in cancers can occur at several points: (1) increased ligand production through enhanced local autocrine activation (2) specific gene translocations to produce kinase fusions with altered signaling profiles (3) RTK overexpression at the cell surface (4) mutation of the RTK protein to modulate activity (5) dysregulation of phosphatase and endocytosis mechanisms to increase RTK signal propagation.

Further exploration of pathway (4) above and specifically clinically identified activating RTK mutations in the intracellular kinase domain have been discovered throughout many cancers (Table 1.1 a&b). The results from Table 1.1 a&b are curated from Catalog of Somatic Mutations In Cancer (COSMIC [47]), which has a much more thorough listing for all the mutations and all cancers. The oncogenic mutations cluster near the characteristic aspects of kinase activation (Table 1.1 a&b). In Kit the predominant clinically identified activating mutations are focused on the juxtamembrane domain and the A-loop, both of which alter how the juxtamembrane domain serves as a steric hindrance to the active site. In the FGFR family, the kinase mutations are around the A-loop which serves as its inactivating segment. EGFR is *cis*-inhibited through autoinhibitory bonds centered around the rotation of the α C-helix and is released by the asymmetric dimer interface. The activating mutations observed in cancers for

	Lung	Colon	Skin	Breast	Prostate	Leukemia
EGFR	PL:[27, 28] α C: [27-29, 48] HC:[27-29, 48] AD1: [48]		α C β 4:[49] AD1:[49, 50]	α C:[51]	AL:[52] α C:[53] AD1:[53]	
ErbB2	α C:[54] α C β 4:[55-57]	α C β 4:[54]	----	α C β 4:[54] α C:[58]	----	----
ErbB4	----	AL: [59]	AL:[60]	AL: [59]	----	----
PDGFR α	JM:[61] CT:[62]	----	----	----	----	AL:[63]
CSF1R/Fms	----	----	----	----	----	CT:[64, 65]
Kit/SCFR	JM:[66]	----	JM:[67-70] α C:[68, 70] AL:[68-70]	----	----	JM:[71, 72] AL:[71-74]
Flt3/Flk2	----	----	----	----	----	JM:[75-77] AL:[78, 79]
VEGFR2/KDR	CT:[61]		----	TM:[80] AL: [80]	----	----
FGFR1	AL:[62]	----	----	----	----	----
FGFR2	----	----	JM:[81] α C:[81] AL:[81]	----	----	----
FGFR3	----	----	AL:[82-84]	----	----	AL:[85-87]
FGFR4	AL:[61]	----	----	----	----	----
Met	JM:[61, 88-90]	α C:[91]	----	----	----	----
EphA1-8,10	AL:[61, 62] CT:[61]	----	----	----	----	----
LTK	AL:[61] CT:[61]	----	----	----	----	----

Table 1.1a: Currently known clinically identified activating cytoplasmic domain mutations in RTKs classified by kinase sub-domains in each of most common tumor types. JM: juxtamembrane domain, PL:P-loop, α C: α C-helix, α C- β 4: α C- β 4 loop, HC: Hydrophobic Core, AL: A-loop, AD1: Asymmetric Dimer Interface in ErbB family, AD2: Asymmetric Dimer Interface in Ret, CT: C-terminal Tail; See also Table 1.1b

EGFR are dominated by two mutations accounting for ~4500 of the 5000 or so total mutations (Figure 1.2): a point mutation (L834R) within the hydrophobic core as well as a small in frame deletion at least involving residues 747 to 751, at the tip of the α C-helix.

	Ovary	Kidney	Thyroid	Gastro Intestinal	Neuro Blastoma
EGFR		α C:[92, 93]	α C:[94] AL:[94] AD1:[95]	----	----
ErbB2	α C β 4:[96-98]	----	----	----	----
ErbB4	----	----	----	α C:[59]	----
PDGFR α	----	----	----	AL:[99-101] JM:[99-101]	----
Kit/SCFR	AL:[102, 103]	----	----	JM:[104-107]	----
FGFR2	AL:[108]	----	----	----	----
Met		AL:[109] α C:[110]	JM [111, 112]	----	----
Ret	----	----	AD2:[113-116] JM:[116, 117] AL:[118]	----	----
ALK	----	----	----	----	AL:[119-122] α C: [120-122] JM:[119, 120, 122]

Table 1.1b: Currently known clinically identified activating cytoplasmic domain mutations in RTKs classified by kinase sub-domains in each of most common tumor types. JM: juxtamembrane domain, PL:P-loop, α C: α C-helix, α C- β 4: α C- β 4 loop, HC: Hydrophobic Core, AL: A-loop, AD1: Asymmetric Dimer Interface in ErbB family, AD2: Asymmetric Dimer Interface in Ret, CT: C-terminal Tail

Both of those mutations alter the motion/conformation of the α C-helix thereby causing activation. The ErbB2 RTK is prevented from forming heterodimers through association with Hsp90 through the uniquely hydrophobic α C- β 4 region [123, 124], which is where the majority of the activating mutations present (Figure 1.2). ErbB4 is not as well studied as EGFR and ErbB2; however it has recently come under scrutiny as potential therapeutic target. However there is debate is to whether constituent activation of ErbB4 functions as an oncogenic promoter [60] or as protection from oncogenic transformation [125].

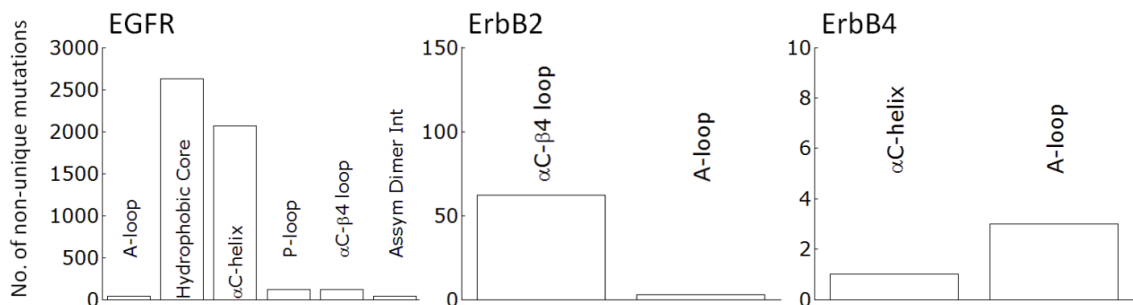


Figure 1.2: Non-unique mutations cataloged in cancer samples for the ErbB family in catalytically important sub-domains, curated from the COSMIC database.

The increased kinase activity increases the dependency of tumor upon the RTK, which becomes “oncogenically addicted [126]” and inhibition of the RTK is a viable route for cancer therapeutics. EGFR and one of its small molecule tyrosine kinase inhibitors (TKIs), Gefitinib, is a canonical example of RTKs, cancer and targeted therapeutics. The initial discovery of Gefitinib in 1994 was met with much excitement as a potential cancer therapeutic since it would be a low-dose targeted oral cancer therapeutic. In two phase II clinical trials of Gefitinib for advanced NSCLC patients after progression of the cancer with chemotherapy, patients overall showed symptom improvement rates around 40% and a 1-year survival rates of 25-35% [127, 128]. The favorable results from the phase II trials gained Gefitinib FDA approval in 2003 prior to phase III clinical trials. However, the phase III clinical trials of Gefitinib versus placebo as a second-line therapy did not show any statistical significance in survival in the overall population, but there was a therapeutic benefit to the sub-group of Asian non-smokers [129]. Examination of the tumors revealed sets of mutations in the EGFR tyrosine kinase domain [27-29]. The subset of the tumors harboring these EGFR mutations are exceptionally sensitive to

inhibition through Gefitinib, so much so that Gefitinib has equal to or greater efficacy than standard chemotherapy treatments in EGFR mutation positive patients [130, 131]. There are several other RTK inhibitors already approved by the FDA and in use in the clinical settings (Table 1.2).

Name	Target	Company	Class
Bevacizumab (Avastin)	VEGF	Genentech	Monoclonal antibody
Cetuximab (Erbixub)	EGFR	Imclone/Bristol-Meyers Squib	Monoclonal antibody
Panitumumab (Vectibix)	EGFR	Amgen	Monoclonal antibody
Ranibizumab (Lucentis)	VEGF	Genentech	Monoclonal antibody
Trastuzumab (Herceptin)	Erb2	Genentech	Monoclonal antibody
Pegaptanib (Macugen)	VEGF	OSI/Pfizer	RNA Aptamer
Dasatinib (Sprycel)	Src/Bcr-Abl	Bristol-Meyers Squib	TKI
Erlotinib (Tarceva)	EGFR	Genentech/OSI	TKI
Gefitinib (Iressa)	EGFR	AstraZeneca	TKI
Imatinib (Gleevec)	Bcr-Abl	Novartis	TKI
Lapatinib (Tykerb)	EGFR/Erb2	GSK	TKI
Nilotinib (Tasigna)	Bcr-Abl	Novartis	TKI
Pazopanib (Votrient)	VEGFR1/2/3 PDGFR/c-kit	GlaxoSmithKline	TKI
Sorafenib (Nexavar)	RAF/VEGFR2/PDGFRB	Onyx/Bayer	TKI
Sunitinib (Sutent)	VEGFR2/PDGFRB c-kit/FLT3	Pfizer	TKI

Table 1.2: FDA approved RTK inhibitors currently in use.

Given the importance of the ErbB family in cancers, it is important to understand their activation mechanisms at the molecular level to help design higher specificity therapeutics. This is especially important in recent light as after sustained use of TKIs the cancers tend to adapt through resistance mutants, in EGFR the main mutation seen after extended treatment with Gefitinib is the T766M mutation [132]. Computational methodologies offer a powerful, quantitative, and complimentary alternative for the study

of intracellular kinase domains which if utilized correctly can predict resistance mutations [133, 134]. The scope of this thesis is to examine the molecular mechanisms of activation for the ErbB family of receptors using computational modeling techniques, particularly molecular dynamics simulation techniques. Chapter 2 reviews common computational modeling techniques utilized for exploring atomic protein properties. Chapter 3 examines the hydrophilic interaction networks in EGFR kinase domain contrasting the inactive and active conformations as well as how the asymmetric dimer interface and mutations affect these networks. Chapter 4 extends the interaction network analysis to the other canonical ErbB kinases, ErbB2 and ErbB4 as well as adding in analysis of hydrophobic interactions. In chapter 5, we examine the interaction networks in the pseudokinase ErbB3. Chapter 6 examines future work especially in regards to umbrella sampling techniques and present preliminary results from targeted molecular dynamics simulations. Finally, chapter 7 places the atomic level analysis in a multi-scale model and links them together with cellular signaling network effects.

Chapter 2.) Computational Methods Related to Reaction Chemistry

2.1) Introduction

This chapter is altered from work presented in a book chapter on simulation techniques for biomaterials [135]. The acceptance of multiscale simulation techniques has helped bridge the gap between theory and experiment [136]. Electronic structure (quantum level or *ab initio*) simulations can reveal how specific molecules assume stable geometrical configurations and charge distributions when subject to specific chemical environment. By examining the charge distributions and structure it is possible to quantify and predict structural properties as well as chemical reactivity pertaining to the molecule, which are particularly pertinent when investigating novel materials. Although the quantum simulations provide a wealth of information regarding structure and reactivity, it is currently not possible to model much more than a few hundred atoms at most. Molecular dynamics (MD) simulations based on classical (empirical) force-fields can model hundred thousands of atoms for nanoseconds and for some systems, up to a microsecond in time. Since MD simulations can be set up at atomic resolution, they are uniquely suited to examine thermodynamic and statistical properties of biomolecules and biomaterials: such properties include (but not limited to) Young's modulus, surface hydration energies, and protein adsorption to different surfaces [137]. Coarse-grained or mesoscale simulations are used to bridge the gap between the atomistic scale of MD

simulations and continuum approaches such as elasticity theory or hydrodynamics at the macroscale (i.e. milliseconds, millimeters and beyond) [136].

2.2) Computational Methods

2.2.1) Molecular Dynamics

Molecular dynamics (MD) simulation techniques are one of the most commonly used to model systems of biomolecules and biomaterials because they can track individual atoms and therefore answer questions pertaining to specific material properties [138, 139]. To perform MD simulations, the starting point is defining the initial coordinates and initial velocities of the atoms characterizing the model system, for example, the desired biomolecule plus the biologically relevant environment; i.e. water molecules or other solvent and/or membranes. The coordinates of the desired biomolecule can usually be found as structural data (X-ray or NMR) deposited into the protein data bank (PDB) [140] (www.pdb.org); otherwise it is possible to derive initial geometry and coordinate data from model building techniques, including homology methods (see section 2.2.2). This step also typically includes the placement and positioning of the environment of the molecules (solvation, ionic strength etc.). The initial velocities are typically derived from the Maxwell-Boltzmann distributions at the desired temperature of the simulation. The potential of interactions of each of the atoms are calculated using a force field, which parameterizes the non-bonded and bonded interaction terms of each atom depending on

its constituent atom connectivity: bond terms, angle terms, dihedral terms, improper dihedral terms, non-bonded Lennard-Jones terms, and electrostatic terms. The potential interactions are summed across all the atoms contained in the system, to compute an overall potential energy function for the system [141-145]:

$$\begin{aligned}
 U(\vec{R}) = & \sum_{bonds} K_b(b - b_0)^2 \\
 & + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi(1 + \cos(\eta\chi - \delta)) \\
 & + \sum_{impropers} K_\phi(\phi - \phi_0)^2 \\
 & + \sum_{nonbonded} \left(\epsilon_{ij} \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{min_{ij}}}{r_{ij}} \right)^6 \right] \right) + \frac{q_i q_j}{\epsilon r_{ij}}
 \end{aligned}$$

Taking the derivative of the potential energy function yields the force, and from Newton's second law, this is equal to mass times acceleration. Although, the process seems simple, the derivative function results in a set of 3N-coupled 2nd order ordinary differential equations that must be solved numerically. The solution consists of a numerical recipe to advance the positions and the velocities by one time step. This process is repeated over and over again to generate MD trajectories of constant energy. Constant temperature dynamics are derived by coupling the system to a thermostat using well established formulations such as the Langevin dynamics or the Nose-Hoover methodologies [146]. Application of MD simulations to biomolecules is facilitated by

several popular choices of force fields such as CHARMM27 [147] (www.charmm.org), AMBER [148] (www.ambermd.org), and GROMOS [149] (www.gromacs.org), as well as dynamic simulations packages and visualization/analysis tools such as NAMD [150] (www.ks.uiuc.edu/Research/namd/) and VMD [151] (www.ks.uiuc.edu/Research/vmd/).

With analysis of molecular dynamics trajectories, it is possible to calculate statistical properties under a variety of initial and other external conditions [136] such as hydrogen bond analyses for hydrophilic interactions and solvent accessible surface area for hydrophobic interactions. For example, by analyzing the relative positions of the hydrogen bond donors to the hydrogen bond acceptors with a preset cutoff angle and a bond length, the hydrogen bonds present in the majority of a given trajectory can be identified to record permanent stabilizing interactions and differentiate them from transient interactions [152]. Similarly, using the solvent accessible surface area analysis, i.e. by mapping a surface area created using a probe sphere of 1.4 Å and accumulating the statistical data, provides a quantitative metric of hydrophobic stabilization effects.

A popular statistical approach to analyze biomolecular dynamics is principal component analysis (PCA) [153, 154], which provides us with a framework to project out independent motions in an MD trajectory and sort them in the order of their dominance (the strongest motions first). This is achieved by diagonalizing the variance-covariance matrix of atomic fluctuations along the trajectory. PCA solves the eigenvalue equation: $[\sigma - \lambda I]\xi = \mathbf{0}$ to project out principal components (PC) or independent modes of atomic

motion, captured in an MD trajectory and sort them by their variance (in decreasing order). Here σ is a two dimensional variance-covariance matrix of atomic fluctuations about the trajectory average, with elements $\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$ ($i, j = 1, \dots, 3N$, N being the total number of atoms with position given by Cartesian coordinates x); $\xi = (\xi_1, \xi_2, \dots, \xi_{3N})$ are the $3N$ independent (uncorrelated) eigenvectors (PC) with eigenvalues $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{3N})$ sorted in descending order i.e. $\lambda_1 > \lambda_2 > \dots > \lambda_{3N-7} > \lambda_{3N-6}$. All global translations/ rotations about the center of mass are removed prior to evaluating σ and the six eigenvalues corresponding to these degrees of freedom are close to zero. The resulting eigenvectors represent the uncoupled principal components (PCs), (modes orthogonal to each other) and the eigenvalues reflect their magnitude (strength) in the trajectory. Generally, the top 10 principal components contain most of the atomic fluctuations in the MD trajectory (>40-90%); moreover, pair wise correlations between motions of atoms in an extended region of interest such as the active site of an enzyme can provide valuable information in relating structure to function through dynamics (or fluctuations) captured in the MD trajectory [153].

The utility of MD simulations generally depend on the accuracy of the underlying force-field. Since force-fields are created using empirical energy functions, they are parameterized and tuned to specific class of molecules; this introduces a constraint on their transferability to model non-native systems or environments and results must often be compared to experimental results, not only to verify their accuracy, but also to identify

where methodological improvements can be made. Thus, there is continued development of the basic force-field as well as the simulation methodology.

Another important consideration is the ability to perform sufficient sampling of the combinatorially large number of conformations available to even the simplest of biomolecules [155, 156]. In this respect, a potential disadvantage of molecular dynamics calculations is that there is an inherent limitation upon the maximum time step used for the simulation (≤ 2 fs). Solvated systems of protein monomers typically consist of 40000 atoms, and those of higher order complexes such as dimers or membrane-bound proteins can be as large as 200,000 to 500,000 atoms. For such system sizes, with current hardware and software, simulation times extending into the microsecond regime and beyond is an exceedingly difficult and labor intensive endeavor which requires a combination of algorithmic enhancements as well as the utilization of high-performance computing hardware infrastructure. For example, cutoff distances reduce the number of interactions to be computed without loss of accuracy for short-range interactions but not for long-range (electrostatic) interactions; to help maintain accuracy, long-range corrections such as the particle mesh Ewald algorithm [157] along with periodic boundary conditions are typically implemented. Parallelization techniques enable the execution of the simulations on supercomputing resources such as 4096 processors of a networked Linux cluster. Although a cluster of this size is a big investment, its accessibility is feasible through the US National Science Foundations' TeraGrid Initiative (founded in 2001) for academic researchers. TeraGrid resources (www.teragrid.org)

currently include more than a petaflop of computing capability and more than 30 petabytes of online and archival data storage, with rapid access and retrieval over high-performance networks. Capitalizing on advances in hardware architecture, another approach is the creation of custom hardware for MD simulations, and offers one-two orders of magnitude enhancement in performance; examples include MDGRAPE-3 [158, 159] and ANTON [160, 161]. Recently, graphical processing unit (GPU) accelerated computation has come into the forefront to enable massive speed enhancements for easily parallelizable tasks with early data indicating that GPU accelerated computing may allow for the power of a supercomputing cluster in a desktop, see examples [162, 163].

2.2.2) Homology Modeling

Protein structure prediction is a fast-growing research field with applications to biotechnology [164-166]. Current experimental techniques for resolving protein structure through X-ray crystallography or NMR spectroscopy are laborious and thus can solve only a small fraction of proteins sequenced by large-scale genome sequencing endeavors. At present, at least 6,800,000 protein sequences have been deposited in the non-redundant protein sequence database (NR; accessible through the National Center for Biotechnology Information: <ftp://ftp.ncbi.nlm.nih.gov/blast/db/>), yet the Protein Data Bank (PDB; <http://www.rcsb.org/pdb/>) contains fewer than 50,000 protein structures [167]. Given this discrepancy, computational modeling of protein structure has proven to be an invaluable tool for bridging the gap between protein sequence and structure. In

particular, homology modeling, or prediction of an unknown structure by using a related protein with a known structure as a template, has been one of the most successful computational techniques for protein structure prediction [168-171]; see also several computer programs and web servers: Swiss-Model server (<http://www.expasy.ch/swissmod/>); CPH models (<http://www.cbs.dtu.dk/services/CPHmodels/>); MODELLER [172] (<http://salilab.org/modeller/>). Homology modeling typically consists of the following steps: search for homologous protein structures, selection of an appropriate template, target-template alignment, model construction, and model quality assessment.

The search for homologous or highly related protein structures to be used as the template for model-building typically involves querying the Protein Data Bank (PDB) with the target sequence: the target sequence is compared with the sequence of every structure in the database, and potential templates are identified. In order to select the best template for comparative modeling, several factors must be considered. First, the higher the degree of sequence identity between the target and template, the better the quality of the template. Secondly, the similarity between the environment (i.e., the type of solvent, pH, presence of ligands) of the template and the environment of the target should be considered [172]. In addition, the quality of the template structure, which has been experimentally derived, must be taken into account. For example, resolution of a crystallographic structure is an indicator of the structure quality. Once a template structure has been selected, a target-template alignment must be performed, which can be

done by using standard sequence alignment methods [173-175]. However, if the degree of sequence identity between the target and template is below 40%, user intervention is required to correct any gaps or misaligned residues generated in the alignment. More accurate alignments can be generated by incorporating structural information from the template, and some modeling programs, including MODELLER, utilize a combination of sequence and structure information in the alignment algorithm.

Once a target-template alignment is created, several algorithms may be used to build a 3D model of the target protein [164, 166]. One commonly employed method is to use distance geometry to satisfy spatial restraints determined from the target-template alignment [176-178]. MODELLER, for instance, imposes spatial restraints that are derived from two sources: homology-derived restraints on the bond distances and angles in the target structure that are based on its alignment with the template structure, and stereochemical restraints on bond distance and dihedral angle preferences that are obtained from a representative set of all known protein structures. The model is then constructed using molecular dynamics methods to minimize violations of the spatial restraints.

A reliable homology modeling program should allow for modeling of insertions (i.e., loops) during model building. The *ab initio* loop modeling method involves exploring multiple conformations for the specified loop region, and each are then scored by an energy function to identify the most likely loop conformation [179]. Alternatively, a

database approach may be employed, which involves identification of a main chain segment that fits the two stem regions of a loop by searching a database of many known protein structures. A limitation of the database approach is the availability of only a small number of known protein structures [180], whereas the *ab initio* approach is more widely applicable to loop regions bound to ligands or other molecules. Currently, loop regions of up to 12 residues can be modeled accurately using these techniques, provided that the loop environment is well-defined [172].

The final step in homology modeling is model quality assessment. Over the past couple of decades, several techniques have been developed to assess the quality and correctness of protein structural models. These methods analyze the stereochemical quality of the model, including bonds, bond angles, dihedral angles, and non-bonded atom-atom distances. Several programs, including PROCHECK (www.ebi.ac.uk/thornton-srv/software/PROCHECK [181]) and (WhatCheck swift.cmbi.ru.nl/gv/whatcheck [182]), perform this type of analysis. When there is less than approximately 30% sequence identity between the target and template, external assessment methods must be applied to determine whether a correct template was used [183]. Thus, several different alignments for the same template may be tested, in addition to alternative templates. The model can be further analyzed by computing a residue-by-residue energy profile, where peaks in the profile represent model errors. However, a potential pitfall of this method is that a segment of residues may appear to be erroneous, when in fact it is only interacting with

an erroneously modeled region. Therefore, the use of energy profiles should not be the only means of model assessment.

Despite the predictive power and utility of homology modeling, several challenges persist. First, the level of target-template structural conservation and the accuracy of the alignment are key determinants of the quality of the resulting model. If the target-template sequence identity is less than 20%, approximately half of the residues in the model may be misaligned [184]. This issue can be addressed by modeling based on more than one template, to integrate the most relevant features from each template. Another challenge is that homology modeling programs must evolve new techniques to keep pace with the increasing number of known protein structures. Pre-computing structural relationships within the PDB may be helpful in addressing this issue, as many irrelevant structures may be excluded without the need to align them explicitly [184]. An additional challenge in comparative modeling is determination of side-chain conformation. Many side-chain programs are based on rotamer libraries [185], which contain values for side-chain torsional angles for preferred conformations of specific side chains. However, as the number of rotamers increases, the issue of sampling all potential conformations becomes a combinatorial problem. Xiang et al. [186] have recently shown that the use of a rotamer library based on Cartesian coordinates of known structures, rather than optimal bond lengths and angles, can successfully predict side-chain conformation. Thus, local minima for side-chain prediction may be nearly as reliable as the global minimum, and renders the combinatorial problem less of an issue. Despite

these present challenges, homology modeling remains a powerful tool for reliable prediction of protein structure. As the accuracy of the technique increases over the next several years, comparative modeling will further close the gap between the number of known sequences and the number of available structures, as well as deepen our understanding of the relationship between protein structure and function.

2.2.3) Free Energy

The first law of thermodynamics states that natural systems seek a state of minimum free energy at equilibrium. Thus computation of free energy of a system is important in comparing the results of simulation and experiment. Several different methods have been implemented for calculation of the free energy of various biological systems, and here we will discuss two of the more commonly employed techniques, namely the free energy perturbation (FEP) method [187] and umbrella sampling [188].

Free Energy Perturbation (FEP): In molecular systems, the free energy problem is typically presented in terms of computing a free energy difference, ΔF , between two defined thermodynamic states, for example, a ligand-bound versus unbound molecule. The free energy difference between the two states is expressed as [189]:

$$\Delta F = -\frac{1}{\beta} \ln \langle \exp [-\beta \Delta v(x)] \rangle_0; \beta = \frac{1}{k_B T},$$

where the subscript zero indicates configurational averaging over the ensemble of configurations representative of the initial state of the system, k_B is the Boltzmann constant, T is the temperature, and $v(x)$ is the potential energy function that depends on the Cartesian coordinates of the system, $[x]$. ΔF can also be computed by the reverse integration:

$$\Delta F = -\frac{1}{\beta} \ln \langle \exp [-\beta \Delta v(x)] \rangle_1,$$

where the subscript one indicates averaging over the ensemble of configurations representative of the final state of the system. The most straightforward implementation of the FEP method involves defining the potential energy function for each state and performing a molecular dynamics simulation for the initial state of the system, hence calculating the ensemble average. Both forward and backward integrations may be run to obtain an estimate of the statistical uncertainty in ΔF . This uncertainty arises from configurations sampled in the ensemble representative of the initial state but not the final state and vice versa, and is typically small when the initial and final states of the system are very similar (i.e., the free energy difference between the initial and final states is on the order of $2k_B T$, or 1.5 kcal/mol) [190]. However, for systems in which the free energy difference is significantly larger, a series of intermediate states must be defined and must differ by no more than $2k_B T$. The total ΔF can then be computed by summing the ΔF^s between the intermediate states:

$$\Delta F = -\frac{1}{\beta} \sum_{i=1}^{M+1} \ln \langle \exp [-\beta [v(x; \lambda_{i+1}) - v(x; \lambda_i)]] \rangle_{\lambda_i},$$

where M indicates the number of intermediate states and λ is the coupling parameter, a continuous parameter that marks the extent of the transition from the initial to the final state. As λ is varied from 0 (initial state) to 1 (final state), the potential energy function $v(x; \lambda)$ passes from v_0 to v_1 .

A limitation of this method is that the end points of the transformation, which correspond to the creation or elimination of a group of atoms, are subject to VDW clashes that result in end-point catastrophes [191, 192]. To obtain an accurate estimate of the free energy at the diverging end points, the number of windows at the beginning and end of the FEP simulations can be increased to collect data at several points with λ values close to zero or one. Another challenge is particularly relevant to biological systems, as there are often multiple ways that a ligand can bind to a receptor [189]. Indeed, some of these possibilities will result in comparable free energy estimates. However, the current FEP methodology must improve upon ways to distinguish among the alternative possibilities and thoroughly sample the possible conformations at a binding site. Despite these limitations, the FEP method is capable of providing an accurate theoretical estimate of the free energy of numerous biomolecular systems [193-196].

Umbrella Sampling: This procedure enables the calculation of the potential of mean force (free energy density) along an *a priori* chosen set of reaction coordinates or order

parameters, from which free energy changes can be calculated by numerical integration (see for example, [197]). For the free energy calculation, the probability distribution $P(\chi_i)$ is calculated by dividing the range of order parameter χ_i into several windows. The histograms for each window are collected by harvesting and binning trajectories in that window, from which the potential of mean force $\Lambda(\chi_i)$ is calculated; the potential of mean force $\Lambda_i(\chi_i)$ is given by [198, 199],

$$\Lambda_i(\chi_i) = -k_B T \ln(P(\chi_i)) + \text{Constant}; \text{ Then, } \exp(-\beta \Delta F) = \int \exp(-\beta \Lambda_i(\chi_i)) d\chi_i$$

The functions $\Lambda_i(\chi_i)$ in different windows are pieced together by matching the constants such that the Λ_i function is continuous at the boundaries of the windows. Thus, the arbitrary constant associated with each window is adjusted to make the Λ function continuous. The standard deviation in each window of the potential of mean force calculations is estimated by dividing the set of trajectories in two blocks and collecting separate histograms. The calculation of the multi-dimensional potential of mean force (multiple reaction coordinates) using the weighted histogram analysis method (WHAM) reviewed by Roux [200], which enables an easy and accurate recipe for unbiasing and combining the results of umbrella sampling calculations, which simplifies considerably, the task of recombining the various windows of sampling in complex systems and computing ΔF , see examples [199-203].

2.2.4) Electronic Structure Methods

Molecular simulations using the empirical force-field approach are rooted in the validity of classical mechanics. Electronic structure methods [204-206] that relieve this fundamental assumption are often necessary in force-field development as well as in understanding chemical reaction pathways and catalytic mechanisms. While electronic structure methods are computationally rather demanding, recent advances in mixed quantum mechanics molecular mechanics (QM/MM) methods enable us to combine high level electronic structure methods with molecular mechanics [207-218]. This multiscale description of force-fields provides a route to extend the electronic structure methods to the nanometer scale to enable the study of biomolecular systems. Compared to their classical counter-parts [142, 219-243], the quantum approaches (such as the empirical valence bond [207, 208]) and mixed QM/MM approaches are more challenging to apply in the biochemical context, though several successful demonstrations are available [193, 215, 244-247]. Here, we describe the utility of electronic structure methods in the study of catalytic reaction mechanisms.

Quantum Mechanics Molecular Mechanics (QM/MM) Simulations: In the QM/MM simulations, the system is sub-divided into two sub-regions, the quantum mechanical sub-region (QM region) where the reactive events take place, and the molecular mechanical sub-region (which provides the complete environment around the reactive chemistry) [208, 210]. Since electronic structure methods are limited by the number of atoms they can handle (typically 50-500), the QM sub-region is restricted to a small number of atoms

of the total system. For example, in an enzymatic system, the quantum region can consist of Mg^{2+} ions, water molecules within 3 Å of the Mg^{2+} ions, parts of the substrate molecules and the catalytic amino acid residues (such as aspartic acids). The remaining protein and solvent molecules are treated classically using the regular classical force-field (such as CHARMM27).

In QM/MM simulations, wave function optimizations are typically performed in the quantum (or QM) sub-region of the system using an electronic structure method such as density functional theory (DFT) [205]. In this step, the electrostatic coupling between the QM and the MM sub-regions is accounted for: i.e., the charges in the MM sub-region are allowed to polarize the electronic wave functions in the QM sub-region. The forces in the quantum sub-region are calculated using DFT on-the-fly assuming that the system moves on the Born-Oppenheimer surface [210, 248]. That is, we assume a clear timescale of separation between the electronic and nuclear degrees of freedom and the electronic degrees of freedom are in their ground state around the instantaneous configurations of the nuclei. The forces on the classical region are calculated using a classical force-field. In addition, a mixed Hamiltonian (energy function) accounts for the interaction of the classical and the quantum sub-regions. For example, since the QM/MM boundary often cuts across covalent bonds one can use a link atom procedure [213] to satisfy the valences of broken bonds in the QM sub-region. Also, bonded terms and electrostatic terms between the atoms of the QM region and those of the classical region are typically included [211].

From a practitioner's stand-point, QM/MM methods are implemented based on existing interfaces between the electronic structure and the molecular dynamics programs, one implementation is between GAMESS-UK (www.cfs.dl.ac.uk [249]) (an *ab-initio* electronic structure prediction package) and CHARMM [147]. The model system can then be subjected to the usual energy minimization and constant temperature equilibration runs at the desired temperature using the regular integration procedures in operation for pure MM systems; it is customary to carry out QM/MM dynamics runs (typically limited to 10-100 ps because of the computationally intensive electronic structure calculations) using a standard 1 fs time step of integration. The main advantage of the QM/MM simulations is that one can follow reactive events and dissect reaction mechanisms in the active site, while considering the explicit coupling to the extended region. In practice, sufficient experience and care is needed in the choices of the QM sub-region and the many alternative choices of system sizes, as well as the link-atom schemes need to be compared to ensure convergence and accuracy of results [211]. The shorter length of the dynamics runs in the QM/MM simulations (ps) relative to the MM MD simulations (ns) implies that sufficiently high resolution structures are usually necessary for setting up such runs as the simulations only explore a limited conformational space available to the system. Another challenge is an accurate and reliable representation of the mixed QM/MM interaction terms [214]. These challenges are currently being overcome by the suitable design of next generation methods for electronic structure and molecular mechanics simulations [250, 251].

2.2.5) Methods for Determining Reaction Paths

Brute-force molecular dynamics simulations of realistic models are restricted to the nano-microsecond regimes. The development of long-time dynamics and sampling algorithms is a well appreciated central objective of single molecule biophysics as the prediction of biologically relevant properties, which occur in the timescale upwards of a microsecond, has remained a problem. Sampling the complex configurational space of biomolecules is a challenge, but can be partially overcome via smart sampling techniques [252-255], umbrella sampling [194, 200], steered and targeted dynamics [229, 256], meta-dynamics [257, 258], Tsallis statistics [259, 260], adaptive sampling [261], and density of states Monte Carlo [262, 263]. Path-based methodologies seek to describe transition pathways connecting two well defined states [224, 264, 265]; practical applications of this ideology are available through methods such as stochastic path approach [266], nudged elastic band [267-269], finite temperature string [261], and transition path sampling [270-272], which each exploit the separation in timescales in activated processes, namely, the existence of a shorter time scale of relaxation at the kinetic bottle neck or the transition state (τ_{relax}), in comparison to a much longer timescale of activation at the transition state itself (τ_{TS}). Below, we review the path-based method of transition path sampling and the related method of Bolas sampling.

Transition path sampling (TPS) [270, 271] aims to capture rare events (excursions or jumps between metastable basins in the free energy landscape) in molecular processes by essentially performing Monte Carlo sampling of dynamics trajectories; the acceptance or rejection criteria are determined by selected statistical objectives that characterize the ensemble of trajectories. In transition path sampling, time reversible MD trajectories in each transition state region are harvested using the shooting algorithm [272] to connect two metastable states via a Monte Carlo protocol in trajectory space. Essentially, for a given dynamics trajectory, the state of the system (i.e., basin A or B) is characterized by defining a set of order parameters $\chi=[\chi_1,\chi_2,\dots]$. Each trajectory is expressed as a time series of length τ . To formally identify a basin, the population operator $h_A=1$ if and only if a particular molecular configuration associated with a time t of a trajectory belongs to basin A; otherwise $h_A=0$. The trajectory operator $H_B=1$ if and only if the trajectory visits basin B in duration τ , i.e., there is at least one time-slice for which $h_B=1$; otherwise $H_B=0$. The idea in TPS is to generate many trajectories that connect A to B from one such existing pathway. This is accomplished by a Metropolis algorithm that generates an ensemble of trajectories $[\chi^\tau]$ according to a path action $S[\chi^\tau]$ given by: $S[\chi^\tau]=\rho(0)h_A(\chi_0)H_B[\chi^\tau]$, where $\rho(0)$ is the probability of observing the configuration at $t=0$ ($\rho(0)\propto\exp(-E(0)/k_B T)$, in the canonical ensemble). Trajectories are harvested using the shooting algorithm [272]: a new trajectory $\chi^{*\tau}$ is generated from an existing one χ^τ by perturbing the momenta of atoms at a randomly chosen time t in a symmetric manner [272], i.e., by conserving detailed balance. The perturbation scheme is symmetric, i.e., the probability of generating a new set of momenta from the old set is the same as the

reverse probability. Moreover, the scheme conserves the equilibrium distribution of momenta and the total linear momentum (and, if desired, the total angular momentum). The acceptance probability implied by the above procedure is given by $P_{\text{acc}} = \min(1, S[\chi^{*\tau}]/S[\chi^\tau])$. With sufficient sampling in trajectory space, the protocol converges to yield physically-meaningful trajectories passing through the true transition state (saddle) region.

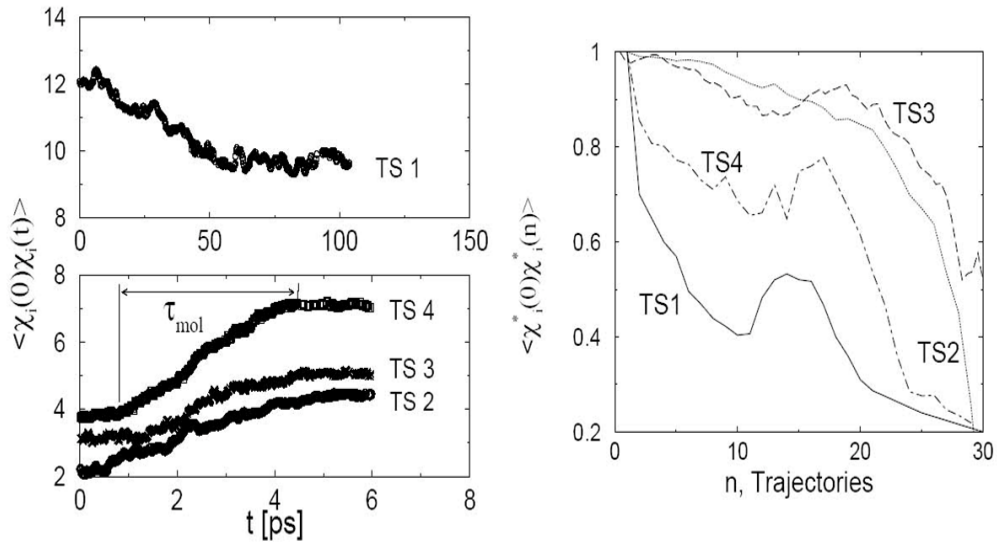


Figure 2.1: Convergence analysis in TPS simulations. Left: order parameter correlation functions transitioning between $\sim\chi_A^2$ and $\sim\chi_A\chi_B$ in a timescale of τ_{mol} . Right: decorrelation of transition paths

The convergence of each sampling run is monitored by calculating the autocorrelation function of the order parameter. The characteristic relaxation time τ_{mol} associated with the crossing of each transition state region is given by the time taken for the gradual transition of the autocorrelation function $\langle \chi_i(0)\chi_i(t) \rangle$, where $\langle \cdot \rangle$ denotes the average over the ensemble of generated trajectories. Order parameter autocorrelation functions

transition from an initial value $\langle \chi_i(0)\chi_i(t) \rangle \approx \chi_A^2$ to a final value $\langle \chi_i(0)\chi_i(\tau) \rangle \approx \langle \chi_A \rangle^* \langle \chi_B \rangle$, to indicate crossing of the barrier region between two metastable regions A and B over time τ . The timescale of barrier relaxation τ_{mol} is inferred from these correlation functions graphically. Shown in Figure 2.1 are these functions for four different sampling runs. The top left panel captures a large subdomain motion while the bottom left panel captures three residue flips during the closing conformational change [273]. The gradual change in the order parameters indicates the decorrelation in each TPS run. In addition to the autocorrelation functions associated with order parameters, an assessment of the quality of our sampling by checking for the decorrelation of order parameters in path (Monte Carlo) space is necessary. This is achieved by calculating the function $\langle \chi_i^*(0)\chi_i^*(n) \rangle$, where n represents the harvested trajectory number, and χ_i^* is the value of the order parameter evaluated at a particular time-slice at the bottleneck of the transition.

In calculating this correlation function, no shifting with respect to the first trajectory is done. This removes the trivial decorrelation because of the shifting moves. Figure 2.1 shows such correlation functions; it is evident from Figure 2.1 that on an average, every 10th to 20th trajectory is statistically decorrelated; therefore the 200 to 300 trajectories that are generated for each TS ensure sufficiently good sampling, see examples [273, 274].

Bolas Sampling for Calculating Free Energies: BOLAS is motivated by the method of transition path sampling. BOLAS generates an ensemble of molecular dynamics

trajectories using a Monte Carlo protocol with an appropriate action S based on shooting perturbations. Below, we define the BOLAS action S and show that using BOLAS, the free energy as a function of a reaction coordinate or order parameter chosen a priori can be computed. The BOLAS path action (different from the TPS path action) is: $S[\chi^\tau]=\rho(0)$. The need to use a modified path action for BOLAS stems from our requirement to compute the unbiased probability distribution of a given order parameter at equilibrium. In principle, configurations contained within the trajectories harvested by TPS are also obtained from the shooting algorithm. However, the bias imposed at the boundaries due to the $h_A(\chi_0)$ and $H_B[\chi^\tau]$ in the TPS action prevents the correct estimation of the equilibrium probability distribution $P(\chi)$. This is because the contribution to $P(\chi)$ comes from six classes of trajectories: trajectories that start in A and visit B in time interval τ ; trajectories that start in B and visit A; trajectories that neither originate in A nor B, but visit both the states in the time interval τ ; trajectories that visit A and not B; trajectories that visit B and not A; and trajectories that neither visit A nor B. The TPS action includes only the first class of trajectories; an action defined by $S[\chi^\tau]=\rho(0)H_A[\chi^\tau]H_B[\chi^\tau]$ includes the first three classes of trajectories; the BOLAS action includes all six classes of trajectories.

Since detailed balance is preserved for the momentum perturbation move of the shooting algorithm, and the individual molecular dynamics trajectories conserve a stationary (equilibrium) distribution ρ , the configurations contained within the ensemble of the

generated trajectories are also distributed according to the equilibrium distribution ρ , (see derivation in [201]). Thus, from the ensemble of trajectories generated using the BOLAS action, the equilibrium probability distribution of the order parameter can be calculated by binning the data from accepted trajectories into histograms of the order parameters. In our implementation, the desired range of χ is divided up in terms of smaller windows and the BOLAS protocol is used to independently sample the configurations in each of these windows. This is equivalent to performing an umbrella sampling. The functions in different windows are then pieced together by using the WHAM algorithm [200]. The validity and application of BOLAS have been illustrated in several applications of protein nucleic acid interactions [201, 273, 275].

2.2.6) Effect of Force on Biomolecules

In the single molecule experiments, a force applied can linearly couple to a reaction coordinate and alter the free energy landscape. If A and B denote the ground and transition states for a given transition (associated with catalysis or a ligand-binding event), then according to transition state theory the equilibrium constant for the system to switch from state A to B in the absence of any external force is: $K_{eq}(0) \propto \exp(-\Delta G/k_B T)$, where ΔG is the free energy difference between A and B [276]. Within the linear response limit, the applied force will shift the ground state equilibrium position $X_A(\mathbf{F})$ and the position of the transition state $X_B(\mathbf{F})$ [276]. Thus, the change in the total (free) energy cost to transition from state A to state B, i.e. $\Delta W=W(0)-$

$W(\mathbf{F})$, under an applied force \mathbf{F} acting along X is $\Delta W(\mathbf{F})=W(0)-0.5k_x[X_B(\mathbf{F})- X_A(\mathbf{F})]^2$, which will alter the equilibrium constant, see Figure 2.2 [276]:

$$K_{eq}(\mathbf{F}) \propto \exp(-[\Delta G - \Delta W(\mathbf{F})]/k_B T).$$

Such a linear response, which assumes a perfect alignment of the applied force and the reaction coordinate, is assumed in the Bell model for receptor-ligand interactions, as well as in models of two-state transitions used to interpret single-molecule experiments [276, 277]. However, the force will change X only if a coupling exists between X and the applied force. We can explore this coupling by carrying out principal component analysis, PCA, see section 2.2.1 [154]. The coupling between the applied force and a coordinate X occurs through the alignment of the force and the PCs and how the PCs impact the coordinate X . While the former can be quantified by projecting the force component along the eigenvectors, the latter can be quantified linearly combining the PCs to describe the motion along X , as we illustrate below [278].

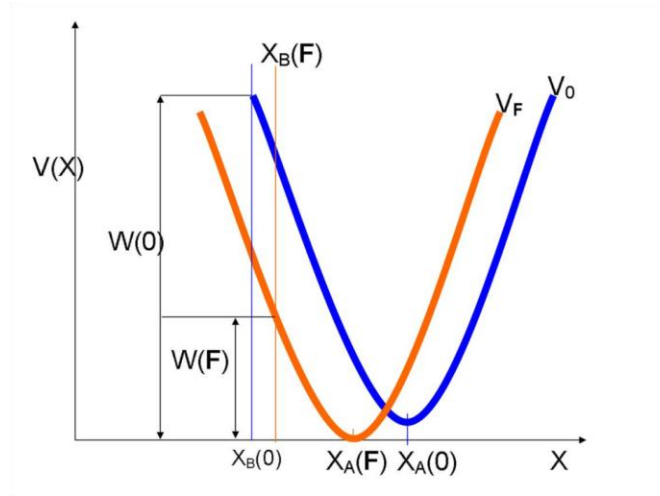


Figure 2.2: Schematic of the effect of the force on the free energy landscape.

An external force applied to a molecule will displace atoms in a given active site. Let $\Delta \mathbf{x} = (\Delta x_1, \Delta y_1, \Delta z_1, \Delta x_2, \Delta y_2, \Delta z_2, \dots, \Delta x_{3N}, \Delta y_{3N}, \Delta z_{3N})$ be the $3N$ dimensional displacement vector which represents the displacement of the N atoms in the active site due to the applied force \mathbf{F} . We can express this displacement vector in terms of the $3N$ normalized PC modes ξ_m which form a complete basis as: $\Delta \mathbf{x} = \sum_m a_m \xi_m$, with expansion coefficients a_m . Under the quasiharmonic approximation and linear response, the Hamiltonian (energy function) for the system is given by:

$$H = \frac{1}{2} \sum_{m=1}^{3N} k_m (a_m \xi_m)^2 - \sum_{m=1}^{3N} \mathbf{F} \cdot (a_m \xi_m).$$

That is, in the quasiharmonic approximation, for each eigenvector ξ_m the spring constant corresponds to $k_m = k_B T / \lambda_m$ [154]. At equilibrium we have $\partial H / \partial a_m = 0$ for each a_m which gives:

$$a_m = \frac{\mathbf{F} \cdot \xi_m}{k_m} = \frac{|\mathbf{F}| \cos \theta_m}{k_m}.$$

Here \mathbf{F} is a 3N dimensional vector representing the force on the active site fragment and θ_m is a generalized angle between \mathbf{F} and ξ_m (PCs are normalized):

$$\cos \theta_m = \frac{\mathbf{F} \cdot \xi_m}{|\mathbf{F}|} = \frac{\sum_{\substack{i=1..N \\ j=x,y,z}} F_i^j \xi_{mi}^j}{\sqrt{\sum_{\substack{i=1..N \\ j=x,y,z}} (F_i^j)^2}}.$$

Where F_i^j denotes the component of applied force acting on the i^{th} atom of the active site fragment in the direction j . In the experiments the force is applied by tethering polymer chains (e.g. DNA) to the ends of the molecule, based on which we can define the applied force to be:

$$F_i^j = F_0(i) n_j \quad \forall i \in [x_T]$$

$$F_i^j = 0 \quad \forall i \notin [x_T]$$

Here $F_0(i)$ is the magnitude of force acting on the i^{th} atom ($|\mathbf{F}| = \sqrt{\sum_{i=1}^{N_T} (F_0(i))^2}$) and $[x_T]$ is a subset of N_T atoms subject to the force. The components n_j ($j=x,y,z$) belong to a unit vector along the applied force direction. We generally assume that $[x_T]$ includes only the heavy atoms and that the region subject to the force is small enough that the same force $F_0(i) = |\mathbf{F}| / \sqrt{N_T}$ acts on all the atoms. Calculations are performed by varying $|\mathbf{F}|$ to get the resultant change in the active site geometry due to the applied force \mathbf{F} : $\mathbf{R}(\mathbf{F}) - \mathbf{R}(0) = \sum_m a_m \xi_m$, where, $\mathbf{R} = (x_1, x_2, \dots, x_{3N})$ is the vector representing the geometry of the active site. Here, $\mathbf{R}(0)$ is the ground state active site geometry at zero force, for which $X = X_A(0)$, and $\mathbf{R}(\mathbf{F})$ is the new active state geometry due to \mathbf{F} for which $X = X_A(\mathbf{F})$. The force along X which causes a displacement, $\Delta X_A(\mathbf{F}) = X_A(\mathbf{F}) - X_A(0)$ is: $F_x(\mathbf{F}) = -k_x \Delta X_A(\mathbf{F})$. The spring constants k_x can simply be obtained from the distribution (histogram) of X values P_x in MD trajectories by fitting a harmonic function to the energy of the distribution $E(X) = (1/2)k_x X^2 = -k_B T \ln(P_x)$. Since the free energy surface projected along the reaction coordinate has a maximum at the transition state, the negative curvature is approximated as $k_{ts} = -\mu(\omega_{ts})^2$, where μ is the reduced mass of the coordinate at the transition state and ω_{ts} is the passage time in transition state theory, $\omega_{ts} = -k_B T / \hbar$. Assuming that the same force acts on X throughout the system's passage from ground to transition state (the dynamic coupling is unaltered), the displacement of the transition state value of X , $\Delta X_B(\mathbf{F}) = X_B(\mathbf{F}) - X_B(0)$, is given as:

$$\Delta X_B(\mathbf{F}) = \frac{F_x(\mathbf{F})}{k_{ts}} = -\frac{k_x}{k_{ts}} \Delta X_A(\mathbf{F}).$$

Note that the displacements $\Delta X_A(\mathbf{F})$ and $\Delta X_B(\mathbf{F})$ have opposite signs due to positive/negative curvatures of the free energy for ground/transition states. We employ the full set of 3N-6 PC modes to calculate the displacements $\Delta X_A(\mathbf{F})$ and $\Delta X_B(\mathbf{F})$ and obtain the ratio $K_{eq}(\mathbf{F})/K_{eq}(0) = \exp(\Delta W(\mathbf{F})/k_B T)$ [278].

2.2.7) Limitations and Caveats

From a structural standpoint, one possible limitation in modeling stems from unresolved regions in the high-resolution crystal structures. This problem can somewhat be addressed through homology modeling. Molecular dynamics simulations suffer from inherent modeling limitations (force-field uncertainties, solvent approximations, limited sampling, finite size effects, etc.). Some of these issues (e.g., finite size effects) are difficult to overcome because of constraints posed by computational resources, while others can be addressed to some degree. It is prudent to test the effect of force-fields on active site geometry by comparing, for example, different force-fields such as CHARMM and AMBER. While solvent effects are taken into account by explicitly treating water molecules, there is in general the problem of determining the protonation states. This can be dealt with at the mean field level by using Debye-Huckel calculations for titratable

side chain residues and by Poisson-Boltzmann (MMPBSA) [279] evaluation of the relative free energies for residues participating in the catalytic reaction. QM/MM calculations can be used to further assess the relative stabilities of the different protonation states. The QM/MM methodologies make some approximations as well. Perhaps the most significant is the choice of the QM/MM boundary. Since the boundary between the MM and QM regions cuts through covalent bonds, single link atom procedure satisfies valences of broken bonds. In particular, the electrostatic terms involving the MM host atoms that connect to the QM region need to be excluded from the Hamiltonian. The test cases [211] have shown that the double link atom method yields better numerical accuracy to the other popular approaches using local self-consistent field (LCSF) formalism and single link atom approach [213]. The study also notes that single link with partially visible MM atoms yields comparable results with the double-link procedure. Other benchmark recipes for link atoms such as the pseudobond method [214] are also available. In order to handle complex chemistry, there is generally a need employ a high-level electronic structure method such as density functional theory, with a reliable energy functional and a high-level set of basis functions often allowing for polarization.

The issue of limited time scales explored in molecular dynamics simulations can be overcome to an extent by smart-sampling algorithms as discussed in section 2.2.5. Of course, these long-time algorithms have shortcomings of their own: these include issues of multiple pathways and choice of order parameters. In general, there is a need to devise

several assessment tools to monitor the quality of sampling and to control the statistical error. The limitations of PCA in extracting dominant modes of protein dynamics arise due to the finite simulation time of the molecular dynamics trajectory [280, 281]. While principal component analysis is not reliable for describing the slow modes of the system beyond what is captured in the dynamics trajectory it is based on, it does provide an approximate description of the slow modes faster than the time scale of the trajectory (t). The error in the eigenvalue propagates as $(tv_i)^{-1/2}$, where v_i is the frequency of the i^{th} mode. Despite the methodological approximations and limitations, we vouch for the notion that the predictions from such simulations can be made to have sufficient accuracy to make meaningful contact with experimental literature.

2.3) Future Directions

As computational modeling techniques are improving, their applications to increasingly complex biological systems are being explored. Therefore, a renewed focus on direct connection with experiments to enable proper validation is critical. One strategic approach is to compare the results of the models and experiments as well as require consistency at multiple length and time scales, which can be realized by comparing to a hierarchy of experiments. For example, structural features can be investigated by means of a combination of experimental approaches such as FTIR, NMR, EPR spectroscopy, AFM, SEM, TEM microscopy, X-ray diffraction etc. One can simultaneously employ thermal (DSC and TG), mechanical (hardness and elastic modulus), and biological

properties for validation and require the modeling to predict a range of properties. In the future, the utility of well validated models and simulation is in directing the rational design and modifications of novel materials. This includes creation of well controlled surfaces that mimic multiple recognition sites for compatibility with the extracellular matrix, quantifying the effect of chemical heterogeneity (OH^- and PO_4^{3-} groups distributed randomly as dangling bonds or the presence of divalent metal ions) in inducing conformational shifts in biomolecules (such as fibronectin), and the effect of applied force mediated through protein-ligand interactions in orchestrating and triggering signaling events in the constituent cells.

Chapter 3.) Molecular Systems Biology of ErbB1 Signaling: Bridging the Gap through Multiscale Modeling and High-Performance Computing.

3.1) Introduction

This chapter has been altered from work published in *Molecular Biosystems* [282]. ErbB family receptors (named because of their homology to the erythroblastoma viral gene product, v-erbB and consisting of the epidermal growth factor receptor or EGFR/ErbB1/HER1, ErbB2/HER2, ErbB3, and ErbB4) signal by activating crucial pathways [283] in response to activation by ligands such as the epidermal growth factor (EGF) and other related peptide growth factors. Through ligand-stimulated formation of various homodimeric and heterodimeric complexes the ErbB receptors are activated leading to the phosphorylation of multiple tyrosine residues on the C-terminal tail of the receptors as well as on other substrate proteins. Through specific interactions of the phospho-tyrosine sites to binding domains, the receptors bind to cytosolic partners that are responsible for the recruitment and activation of multiple down-stream cascades [3, 284-288]. The activation through the mitogen-activated protein kinase (MAPK) cascades of the extracellular signal-regulated kinases (ERKs) is functionally linked to proliferation. The phosphoinositide 3-kinase (PI3K) pathway leads to the activation of the serine/threonine protein kinase Akt (cellular homologue of the viral oncogene v-Akt) which is linked to survival. Other significant pathways mediated by ErbB signaling include activation and nuclear translocation of signal transducers and activators of

transcription proteins (STATs) [289] and clathrin mediated endocytosis [290]. Yet, the molecular context in which ErbB receptors activate and regulate signaling has not been fully recognized. More specifically, it is of great interest to investigate the molecular mechanisms that lead to the dis-regulation of ErbB signaling in several pathologies such as cancer, psoriasis, atherosclerosis, impaired cardiac development, and schizophrenia [291, 292]. With rapid progress in high-performance computing methods and infrastructure¹ we advocate that multiscale modeling offers a powerful, quantitative, and complimentary alternative for the study of functional intracellular modules. We describe the application of a hierarchical multiscale modeling procedure (summarized in section 3.2: Models and Methods) to signaling in the ErbB family receptors to describe how point-mutations in the ErbB receptors can profoundly alter signaling characteristics leading to the onset of oncogenic transformations. Here in this chapter, we only discuss the atomic level fluctuations governing activation of the EGFR kinase domain, and how molecular perturbations (mutations and dimerization) affect them. For how this molecular model fits into a cellular model see Figure 3.1b&c and section 7.2.

3.2) Models and Methods

¹ Our applications multiscale algorithms to ErbB signaling are enabled by high-performance computing infrastructure. As we transition from tera-flops (flops: floating point operations per second) computing to peta (10^{15}) flops computing in the near future, we expect that extension of our molecular and systems modeling to complete cellular pathways will become tractable.

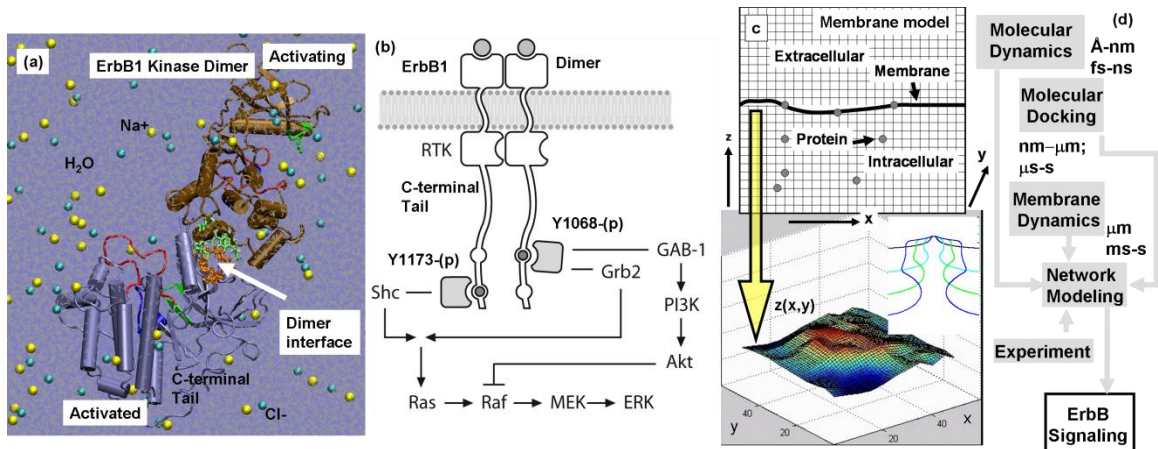


Figure 3.1: Hierarchical multiscale modeling scheme for ErbB signaling. The dimer-mediated receptor activation characteristics of ErbB1 receptor tyrosine kinase is studied using molecular dynamics simulations. The interactions of substrate tyrosine containing peptides derived from the C-terminal tail with the ErbB1 kinase are studied using molecular docking simulations. We have employed a deterministic network-based kinetic modeling scheme to study ErbB1-mediated signaling, and a hybrid discrete/continuum stochastic dynamics protocol to study the initiation of ErbB1 receptor internalization. **(a)** Atomistic model for ErbB1 dimer employed in the molecular dynamics and molecular docking calculations. **(b)** Branched network model for ErbB1-mediated signaling in which phosphorylation of the ErbB1 dimer occurs at either tyrosine Y1068, which can bind GAB-1 or Grb2, or at tyrosine Y1173, which binds Shc. Phosphorylation of the factors Akt and ERK were used as indicators of downstream activation. **(c)** Hybrid stochastic model for ErbB1 internalization. (top) Grids in finite difference scheme for membrane dynamics and the lattice in the kinetic Monte Carlo scheme for protein diffusion; (bottom) snapshot of vesicle bud on the membrane in response to a specific spatial ordering of the curvature-inducing protein, epsin, on the membrane; inset depicts a stabilized vesicle neck. **(d)** Flow of information between different simulation methods.

We provide a brief summary of the hierarchical multiscale modeling scheme we have employed for describing ErbB signaling (see Figure 3.1: a-d). We model the dimer-mediated receptor activation characteristics of the ErbB1 receptor tyrosine kinase using molecular dynamics simulations [293, 294]. Figure 3.1(a) depicts the atomistic model of the explicitly solvated ErbB1 kinase dimer employed in the molecular dynamics: 10-20

ns trajectories of fully atomistic, explicitly solvated systems of wildtype and mutant ErbB1 kinase monomers and dimers are obtained and analyzed for specific stabilizing interactions such as hydrogen bonds and salt-bridges, see also Figure 3.2 [293, 294]. For how the cellular model interfaces with the cellular model see section 7.2.

3.3) Rationalizing and Predicting the Effects of Molecular Perturbations on Receptor Kinase Activation

Small molecule tyrosine kinase inhibitors for ErbB1/2 tyrosine kinase such as gefitinib, erlotinib, and lapatinib, which are ATP analogues, are of significant interest as cancer drugs. While the receptor tyrosine kinase inhibition approach has shown promise in some clinical trials, success has been mixed. In particular, the occurrence of somatic mutations in the ErbB1 kinase domain (L834R: where the leucine residue in position 834 is replaced by an arginine, L837Q: leucine at 837 replaced by a glutamine, G685S, del L723-P729 ins S: deletion of residues 723-729 and an insertion of a serine) as seen in patients of non-small cell lung cancer² [295] renders the cell lines harboring such mutations more sensitive to treatment [295, 296].

² There are 2 main types of lung cancer and they are treated differently: small cell lung cancer and non-small cell lung cancer. About 85% to 90% of all lung cancers are of the non-small cell type with three sub-types. The cells in these sub-types differ in size, shape, and chemical make-up: (1) squamous cell carcinoma linked to smoking, (2) adenocarcinoma usually found in the outer part of the lung, and (3) large-cell (undifferentiated) carcinoma exhibiting rapid growth and spreading.

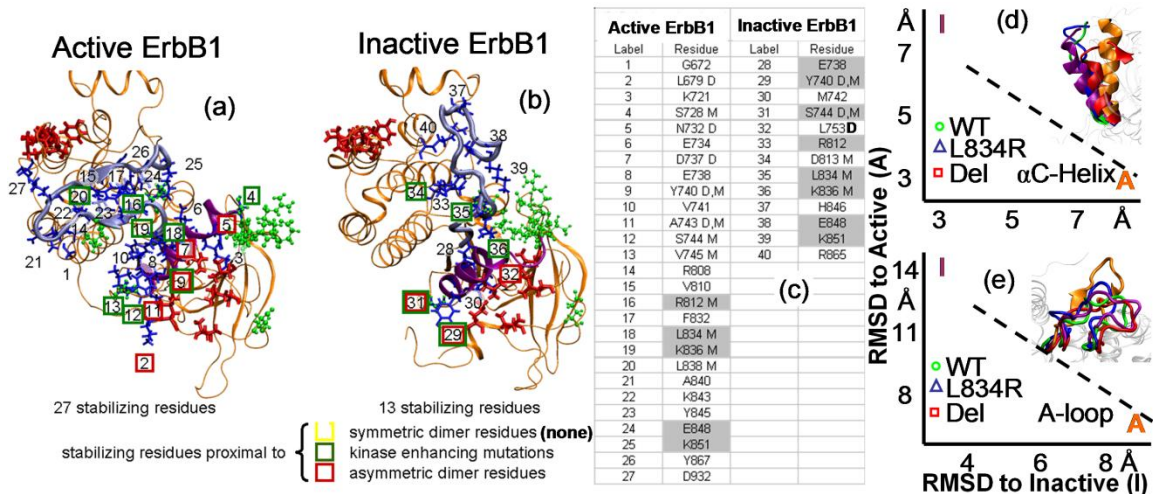


Figure 3.2: (a, b) Stabilizing network residues in ErbB1 kinase in its active and inactive conformations. The numbers in (a, b) correspond to the entries in (c). Compared to inactive, the active state presents a significantly increased number of stabilizing residues, with approximately the same number of residues affected by the asymmetric dimerization (marked D and boxed red) and mutation (marked M and boxed green), so the inactive system is much more susceptible to a conformational shift triggered by the asymmetric dimer or mutation. We identify the residues participating in stabilizing bonds that are also proximal (≤ 3 Å between heavy atoms) to the dimer interface. Cross-referencing this list with the stabilizing residues provides a list of potential residues possibly affected by dimerization: entries marked D in (c) and boxed red in (a,b). (d): Symbols represent trace during dimer molecular dynamics simulations of wildtype and the two mutants L834R and Del L723-P729 ins S or del: (top) Root-mean-squared deviation or RMSD of α C-helix, and (bottom) RMSD of the activation loop (or A-loop). Insets depict the relative positions of the α C-helix and the A-loop at the end of the dynamics runs.

In order to determine how such molecular perturbations can shape cellular fates, we sought to determine how the mutations affect the regulatory mechanisms operational within the kinase domain of ErbB1, ErBb2, and ErbB4. A recent structural and biochemical study involving the ErbB1 by Zhang et al. [40] proposed a new dimer-mediated allosteric activation mechanism of kinase activation according to which ErbB1

receptor tyrosine kinase dimerizes in an asymmetric head-to-tail configuration. While the monomer ErbB1 kinase is stable in an inactive conformation which interferes with ATP binding, in the asymmetric dimer configuration, one ErbB1 kinase domain serves as an activating protein and activates the other ErbB1 kinase in the dimer, through allosteric contacts. The kinase-kinase contact at the asymmetric dimer interface allosterically stabilizes the active conformation. Recently, we hypothesized that an underlying network of stabilizing hydrogen bonds dominates the relative stabilities of the inactive and active conformations and governs the kinase activation. We then performed a hydrogen bond analysis of the molecular dynamics data, focusing on the interactions surrounding the activation-loop and the α C-helix sub-domains in the active and inactive conformations, and identified this network, see Figure 3.2(a,b). The total number of interactions in the active state out-numbers those in the inactive state by 27 to 13, which led us to hypothesize that a particular stimulus such as the asymmetric dimerization preferentially destabilizes the inactive kinase conformation triggering the conformational change to the active state.

In order to consider the effect of ErbB1 kinase dimerization on the network of stabilizing interactions, we identified the protein residues participating in stabilizing bonds that are also proximal (≤ 3 Å between heavy atoms) to the residues involved in the formation of the asymmetric dimer interface, (which are P675, L679, L680, I682, V736, L758, V762 in the kinase undergoing activation). The rationale for the proximity analysis is that the interactions in the stabilizing network could be compromised due to the molecular level

reorganization upon kinase dimerization. The proximity analysis suggests that 3 out of 13 stabilizing residues in the inactive state may be perturbed or lost upon dimerization (these are marked with D in Figure 3.2(c)), which suggests a potential molecular relay mechanism by which the kinase dimerization event activates the kinase domain. Indeed, when we performed 20 ns molecular dynamics simulations of an ErbB1 kinase dimer system (shown in Figure 3.1(a)), we observed a significant rearrangement (change in the root-mean-squared deviation or RMSD of 3 Å) of the α C-helix position: see Figure 3.2(d). The shift in the α C-helix position was accompanied by several changes in the stabilizing network consistent with the predicted bond-patterns in Figures 3.2(a-c). In particular, in the inactive conformation, the bonds between Y740-S744, L834-D813, H846-R865, K851-R812 surrounding the activation loop and the α C-helix were severed. Thus, already in the wildtype dimer, due to the re-configuration of the α C-helix, the bond pattern is found to be shifting significantly toward that observed in the active kinase. The key bonds stabilizing the wildtype inactive ErbB1 kinase are the E738-K836 and E848-R865 salt bridge interactions, as well as the L834-D813 hydrogen bond. In our dimer simulations, the E738-K836 salt bridge has considerably weakened: the fraction of the time this bond was present decreased from >90% in the monomer trajectory to ~60% in the dimer trajectory; moreover this bond has undergone considerable stretching allowing E738 to hydrogen bond to F832, which is one of the bonds seen in the active kinase. The L834-D813 interaction is at the threshold of still being considered a stabilizing hydrogen bond. The residue K851 is hydrogen bonded to E725, moving away from the inactive bond K851-R812 and towards the active salt bridge K851-E734. The E848-R865 salt

bridge is not perturbed significantly due to the formation of the asymmetric dimer interface. Thus, we hypothesize that a few specific bonds act as gatekeepers for each step of the conformational change, namely the E738-K836 salt guards against the α C-helix movement and the E848-R865 salt bridge guards against the activation loop rearrangement. Our dimer simulations re-affirm our notion that the stabilizing network is susceptible to perturbation in the inactive conformation of the kinase, and that formation of the asymmetric dimer will have the effect of directly breaking the network of interactions around the α C-helix, thereby destabilizing the inactive state. The loss of these interactions and the shift of the α C-helix conformation towards the active state will provide the impetus for kinase domain activation. Intriguingly, several of the clinically identified mutations that have been reported to constitutively activate the kinase also directly perturb the stabilizing network by breaking key stabilizing bonds: these are marked by the symbol M in Figure 3.2(c). In addition, the del L723-P729 ins S mutant re-configures the α C-helix in the inactive state to a conformation closer to the active state (see Figure 3.2(d)). Thus, our delineation of the stabilizing hydrogen bond network provides molecular-level insight into the possible mechanisms by which activating mutations of ErbB1 kinase such as L834R and del L723-P729 ins S destabilize the inactive conformation. This preferential destabilization of the inactive conformation renders the receptor kinase constitutively active even as a monomer, producing high basal activation levels of the kinase even in the absence of a growth-factor induced dimerization.

Considering that there is an excellent correlation between the stabilizing network of interactions and the clinically identified activating mutations in ErbB1, our structural studies on kinase activation are well poised to forecast the mutation landscape associated with other ErbB members. We have extended the analysis we have presented for ErbB1 to ErbB2 and ErbB4 kinases in which we have identified similar networks of stabilizing interactions. Based on the similarities between the stabilizing interactions between ErbB1 and ErbB4 kinase domains, we can predict the effect of analogous mutations in ErbB4 on kinase activation: (1) based on the location of the mutations E690G, G700S, the mutants are expected to be activating through directly impacting dimerization (similar to the activating mutants E685G and G695S of ErbB1). (2) Del 728-G733 ins S and S749I are poised to cause a conformation shift of the α C-helix of ErbB4 and hence are predicted to be activating. (3) Mutations F740A, L839R and L842Q in ErbB4 are poised to perturb the bond network of the inactive kinase and hence expected to be activating. We note that in support of our predictions of ErbB4, the F740A and L839R have been tested independently by Qiu et al. [297] and indeed found to be activating. Based on the subtle differences we have noted in the stabilizing bond networks of ErbB1 and ErbB4, we are also able to suggest new activating ErbB4 mutations (that do not have an obvious counter-part in the ErbB1 system). In particular, R841 in ErbB4 is featured prominently in the stabilizing network and mutating R841 to either alanine or aspartic acid is expected to promote activation. Mutation of the two residues blocking the catalytic aspartate, E743 and G838, to residues with smaller side chains, alanine or glycine, is also expected to promote activation.

At the molecular level, considering that there is an excellent correlation between the stabilizing network of interactions and the clinically identified activating mutations in ErbB1, our structural studies on kinase activation are well poised to forecast the mutation landscape associated with other ErbB members. Indeed based on our simulations of ErbB2 and ErbB4, we have identified similar networks of stabilizing residues and are already able to predict activating mutations in these receptors that have not yet been reported clinically [unpublished results] which together with the extensions proposed above can be valuable for evaluating the likely effect mutations on ErbB2 inhibition efficacies in cancer, ErbB4 inhibition in cardiac development and schizophrenia [292]. How the constituent activity of the mutant ErbB kinases affects cell signaling, both as phosphorylation of its C-terminal tail as well as the cell signaling pathways is crucial for understanding how ErbB kinases can alter cell fates; the intersection of the atomic level simulation with cell signaling models is considered in section 7.2

Chapter 4.) Molecular Dynamics Analysis of Conserved Hydrophobic and Hydrophilic Bond Interaction Networks in ErbB Family Kinases.

4.1) Introduction

This chapter is from work that has been submitted and is under review at Biochemical Journal [298]. Receptor tyrosine kinases (RTKs) are transmembrane glycoproteins important in intercellular communication and oncogenesis [299]; they comprise a large ligand-binding extracellular domain, a single transmembrane α -helix, an intracellular tyrosine kinase domain and a C-terminal tail that harbors regulatory tyrosine autophosphorylation sites (reviewed in [1, 2]). The ErbB family consists of four homologous RTKs: the epidermal growth factor receptor (EGFR/ErbB1/HER1), ErbB2 (HER2/Neu), ErbB3 (HER3), and ErbB4 (HER4). Binding of growth factors to the extracellular ligand-binding domains of ErbB receptors promotes their homo and/or heterodimerization, which in turn leads to activation of the cytoplasmic kinase domain. This triggers a multi-layered signaling network of crucial pathways regulating cell proliferation, differentiation, migration, etc [3]. Aberrant signaling by EGFR and ErbB2 is correlated with a variety of diseases, from psoriasis to cancer [300]. In particular, clinically identified mutations in EGFR have been shown to increase the basal activity of the EGFR kinase domain, and non-small-cell lung cancer (NSCLC) patients carrying these mutations respond remarkably to the EGFR RTK inhibitor gefitinib [26, 27, 48]. ErbB2 is the target of the therapeutic Herceptin antibody, and its amplification and

overexpression in breast cancer correlates with a poor prognosis [301, 302]. The loss of ErbB4 signaling in mice has been shown to result in defective heart, nervous system, and mammary gland function [303-305].

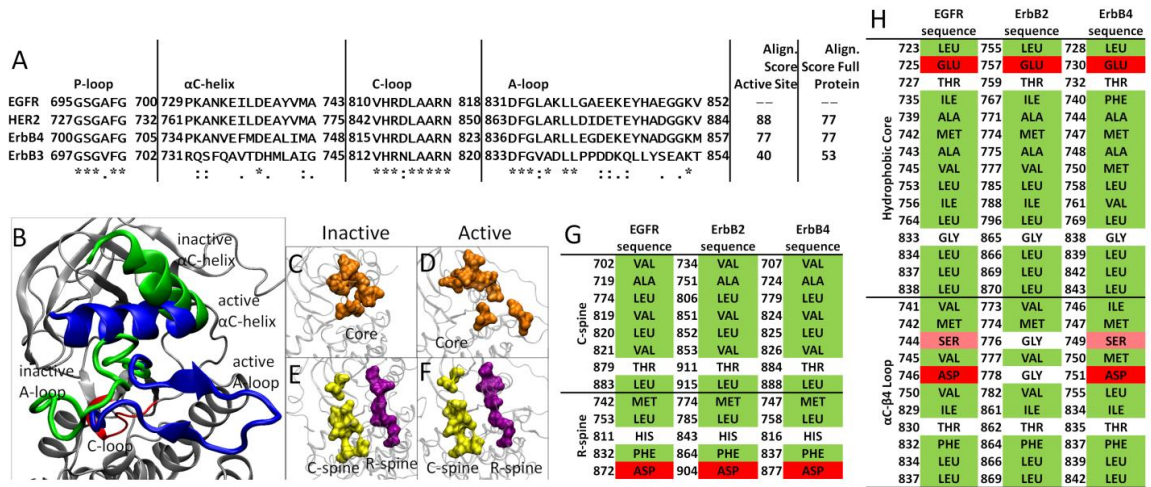


Figure 4.1: (A) Alignment of the four catalytic subdomains comprising the active site in the ErbB kinases, with accompanying alignment scores for the subdomains alone and for the entire kinase. (B-F) Comparison of the inactive and active conformations of the ErbB kinases highlighting the conserved features. (B) The C-loop is highlighted in red with the active conformation shown in green. (C,D) The hydrophobic core is shown in orange, (E,F) while the C-spine is shown in yellow and the R-spine is in purple. (G) The residues comprising the C-spine and R-spine for EGFR, ErbB2 and ErbB4 kinase. The residues highlighted in green are hydrophobic, the residues in red are hydrophilic and those in white are neutral. (H) The residues comprising the Hydrophobic core and the α C- β 4 region for EGFR, ErbB2 and ErbB4 kinase. Serine is shown in pink as it is slightly hydrophilic, though it is still considered neutral.

In RTK signaling, the intracellular kinase domain catalyzes transfer of the γ -phosphate of ATP to tyrosines on both the RTK itself and in other target substrates (reviewed in [1]). Regulation of the RTK kinase domain is thought to involve contributions from several

conserved subregions: the catalytic loop (C-loop), the activation loop (A-loop), the glycine-rich nucleotide binding loop (P-loop), and the α C-helix, which together define the active site in the cleft between the β strand-rich N-lobe and the helical C-lobe. The catalytic loop residues directly participate in phosphoryl transfer. The A-loop, P-loop and α C-helix (marked in Figure 4.1A) modulate the activity of the kinase domain by regulating accessibility of the active site to binding and coordinating both ATP and the substrate tyrosine. The ~20 amino acid A-loop in ErbB kinases contains one phosphorylatable tyrosine (Y845 in EGFR, Y877 in ErbB2, Y850 in ErbB4). In many kinases, e.g. the insulin receptor kinase (IRK), phosphorylation of this A-loop tyrosine triggers conformational changes that allow substrate access to the active site. Thus, A-loop conformational changes constitute a key event in RTK regulation (Figure 4.1B). The α C-helix and P-loop must also be re-positioned to coordinate the ATP and the substrate tyrosine for effective phosphoryl transfer. The four regulatory subdomains are highly conserved across the ErbB family, with 88% sequence identity between EGFR and ErbB2 and 77% sequence identity between EGFR and ErbB4 (Figure 4.1A). ErbB3 is unique in the ErbB family as it lacks specific conserved residues considered to be required for full kinase activity.

Recent structural studies have revealed highly conserved hydrophobic “spines” within kinases that are considered important for defining their catalytic state [30, 31], shown in Figure 4.1E,F. The regulatory spine (R-spine) consists of four hydrophobic side chains (M742, L753, H811, F832 in EGFR) anchored by an aspartic acid in the α F-helix (D872

in EGFR). The R-spine spans several key regulatory subdomains, and coordinates the motion of the N- and C-lobes of the kinase [30]. The catalytic spine (C-spine) involves eight hydrophobic side-chains (V702, A719, L774, V819, L820, V821, T879, L883 in EGFR) that help support and coordinate the adenine ring of ATP in the active state [31]. Similarly, in the inactive state there is a small hydrophobic ‘core’ formed between the α C-helix and the A-loop, which maintains the kinase in the inactive conformation (Figure 4.1C-F). Disruption of this hydrophobic core by single point mutations has been shown to activate EGFR [25-29].

EGFR stands out among RTKs in appearing not to require A-loop phosphorylation for its activity [306]. Mutation of the EGFR A-loop tyrosine Y845 to phenylalanine does not appear to change the level of receptor auto-phosphorylation in cells or *in vitro* [40], in contrast to other kinases such as IRK (although Y845 is phosphorylated by Src in EGFR signaling [307]). Crystal structures have confirmed that the EGFR and ErbB4 kinase domains can adopt active-like conformations even without Y845 (Y850 in ErbB4) phosphorylation [25, 42], and have revealed an allosteric mechanism for kinase domain activation [40]. Activation of the EGFR TKD involves the formation of an asymmetric ‘head-to-tail’ dimer in which one kinase domain (the ‘receiver’) becomes activated through allosteric changes arising from contacts between its N-lobe and the C-lobe of its neighbor (the ‘activator’). The C-lobe of the activator kinase appears to play a cyclin-like function in activating its dimerization partner (the receiver). The importance of the asymmetric dimer interface was confirmed by mutational studies in EGFR and ErbB4

[40, 42]. More recent studies have shown that the intracellular juxtamembrane region of the receptor also contributes to formation of the asymmetric dimer interface, in a manner that is necessary for maximal activation [43-45].

Considering the high degree of sequence similarity and structural homology across the ErbB family members (Figure 4.1A,G,H), we sought to identify the degree to which molecular mechanisms of activation are conserved across the ErbB family, and to identify differences in overall function that arise from variability in primary structure. Recently, we and others have hypothesized the existence of distinct networks of intramolecular non-covalent bonds that characterize the active and inactive conformations of kinases (for Lyn [308, 309], Abl [310], EGFR [152, 310, 311] and ErbB2 [312]), with transitions between the states necessitating a shift in these bond networks. Here, we present bioinformatics and fluctuation analyses of molecular dynamics trajectories of ErbB kinase domains and relate sequence similarities to correspondence of specific bond-interaction networks and resemblances in collective dynamical modes. We investigate how the various stimuli/perturbations such as dimerization, phosphorylation of the A-loop tyrosine, and mutations seen in cancer patients impact both the active and inactive conformations of the ErbB family kinase domains. The solvated systems of the truncated ErbB family kinases we present even have a physiological relevance to cell studies. The protein tyrosine kinases, Src and Abl, have a highly similar active structure to those in receptor tyrosine kinases [1, 313].

Furthermore, ErbB4 is cleaved from the membrane into the s80 protein, a fully active soluble form of the ErbB4 kinase domain [305].

4.2) Methods

Molecular Dynamics (MD) Simulations: Models for ErbB1 (EGFR) kinase were derived from the 1M14 (active) and 2GS7 (inactive) structures [25, 40]. Models for ErbB4 were derived from the structures of Qiu et al., PDB ID: 3BCE and 3BBW [42]. Structures for ErbB2 were constructed using homology modeling following the procedure described in [312]. Models for kinase dimers were constructed based on the asymmetric dimer interface described in [40]. Each system was simulated as a fully atomistic, explicitly solvated-system in NAMD [150], using the CHARMM 27 forcefield [141]. The missing hydrogens in the protein were added using the hbuild plugin in the VMD algorithm [151]. To simulate a physiological pH of 7.0 the histidines were constructed with +1 protonation state on the δ -nitrogen. The entire system consisted of the protein, water molecules (TIP3P model [314]) and ions at 75 mM ionic strength to simulate physiological conditions; the water molecules and ions were placed at locations of electrostatic extrema (maxima for negative Cl^- ions and minima for positive Na^+ ions) as determined by a Debye-Huckel potential using Solvate 1.0 [315]. The placement of the counter-ions was also restricted to 8 Å away from any protein residue. The Rattle algorithm [316] was employed to constrain the hydrogens to allow for a 2 fs timestep of integration. Periodic boundaries were implemented in all three dimensions and long-ranged electrostatics

interactions were accounted for using the Particle Mesh Ewald Algorithm [157]. A conjugate gradient algorithm was employed for all energy minimizations. The solvated system was energy minimized and the volume of each system subsequently equilibrated with constant pressure and temperature (CPT) simulations. Both the temperature and pressure were constrained using a Langevin algorithm [317]. The pressure Langevin piston was set with a reference pressure of 1 atm, a mass of 2000 amu and a collision frequency of 5 (1/ps), while the temperature Langevin piston was set with a reference temperature of 300 K and a mass of 10000 kcal·ps². Following volume equilibration, the energy of the system was equilibrated using constant volume and temperature (NVT) simulations. Each system was simulated for at least 10 ns. For the dimer systems restraints of 1 kcal/Å² were placed on each kinase to their initial state in the inactive asymmetric dimers to maintain the equilibrated structures while removing steric clashes. The kinase systems were equilibrated and then the restraints were reduced to 0.5 kcal/Å² and 0.1 kcal/Å². Following this, the production simulations were performed.

Analyses of MD Simulations: Root-mean-squared deviation (RMSD) calculations were performed using the RMSD tool plugin in VMD by first removing global translation and rotation, and then computing the RMSD of the selected sub-regions (A-loop, C-loop, P-loop and α C-helix) relative to a reference structure (the respective active or inactive crystal structure). Principal component analysis (PCA) was performed using the software Carma [318], by constructing the covariance matrix of atomic fluctuations $\underline{\sigma}$ in Cartesian space and then diagonalizing $\underline{\sigma}$ to obtain the principal components (eigenvalues and

eigenvectors). An analysis of hydrogen bond (H-bond) patterns in the production simulations was performed using CHARMM in conjunction with VMD. The trajectories were first analyzed in CHARMM with a hydrogen bond cutoff of 3.4 Å and a cutoff angle of 150 degrees. CHARMM was used to generate a list of hydrogen bonds that were present in at least 60% of the trajectory; we note that the threshold of 60% was varied from 50-80% in our sensitivity analysis and the results were not altered significantly. These hydrogen bonds were then visualized in VMD and any bonds not providing a consistent, sustained bond were removed to reveal the persistent H-bonds. During the H-bond analysis, acidic and basic residues formed strong hydrogen bonds. Such hydrogen bonds were considered salt bridges if they were between the side chain of an acidic and basic residue, with a bond length less than 1.6 Å and present in the majority (>60% of the trajectory) of the simulation. Statistical hydrogen bonding criteria was chosen to highlight the contributions of specific residues within the protein regardless of its surrounding energetic environment [319]. Solvent accessible surface area (SASA) values were calculated in VMD using the measure SASA module using a probe radius of 1.4 Å larger than the van der Waals radius. The SASA was calculated for each step in the trajectory from which the mean and standard deviation were computed. As an alternative measure of hydrophobicity in heterogeneous environments, following Garde et al. [320, 321], normalized water density fluctuations were computed by recording the ratio of $Vol_N * (\langle N^2 \rangle - \langle N \rangle^2) / \langle N \rangle^2$, where Vol_N is the volume of interest, $\langle N^2 \rangle - \langle N \rangle^2$ is the variance in the fluctuation of water number in Vol_N , and $\langle N \rangle$ is mean associated with the number of water molecules. We choose the region of interest to be within 5 Å of a specified

hydrophobic sub-region in the EGFR monomer kinase. The results are then divided by the water density fluctuations of bulk water for normalization. Although results are presented for a cutoff of 5 Å, other cutoffs ranging from 3 Å-15 Å were investigated and similar trends were recorded.

4.3) Results

Following molecular dynamics simulation of each active or inactive monomeric kinase system for at least 10 ns, the time evolution of the RMSD was used to monitor equilibration and to track any reorganization of the A-loop and α C-helix conformations; no conformational switching towards active or inactive states was observed (Figure 4A.1, chapter 4A). Whereas the majority of the protein backbone, including the C-loop and the P-loop, aligns closely between the inactive and active states, the A-loop and α C-helix conformations differ considerably. In transitioning from the inactive to the active conformation, the α C-helix rotates toward the C-lobe, with the rotating end shifting by ~ 9 Å toward the base of the cleft between the N- and C-lobes (Figure 4.1B). This helix is also extended by 2 turns (involving residues 728-732 in EGFR) in the active conformation compared with the inactive conformation. In the inactive kinase, the A-loop maintains a 'closed' conformation (mainly through inter-region hydrogen bonds) and partially blocks the catalytic site. By contrast, the A-loop appears 'unfurled' in the active kinase, and lies against the C-lobe (blue in Figure 4.1B).

4.3.1) PCA reveals a tightly coordinated motion in all active ErbB members

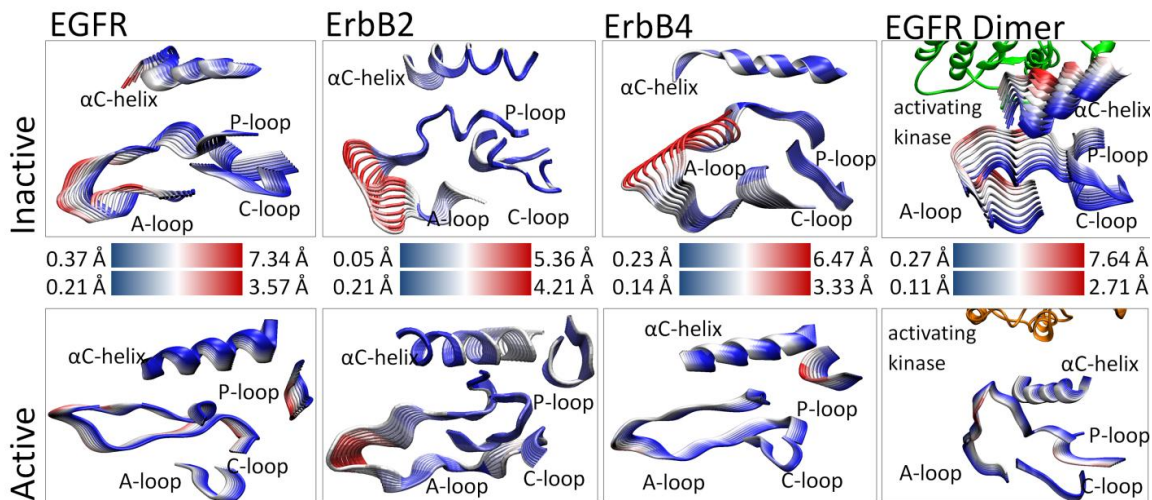


Figure 4.2: Visualization of the first principal component for the four key subdomains in the inactive and active conformation of the ErbB kinases. The motions are overlaid sequentially where the large-amplitude motion in each frame is highlighted in red and the low-amplitude motion is highlighted in blue. For the two dimer systems, the activating dimer is shown in orange and green. The inactive conformations exhibit large, localized motion while the active conformations demonstrate smaller, coordinated motions.

Principal component analysis (PCA) revealed that motions occurring within the active sites of the inactive and active EGFR monomers differ significantly, particularly in the A-loop (see Figure 4.2). The inactive EGFR monomer exhibits large-amplitude motion in both the α C-helix and the A-loop (4.95 and 7.34 Å, respectively), with smaller fluctuations in the P-loop and C-loop (4.17 and 3.65 Å, respectively). We attribute the larger amplitudes to a more flexible protein segment; this is consistent with the observation that part of the activation loop is unresolved in almost all the crystal structures of ErbB kinases to date. Our previous work with the loop modeling program MODELLER [172] discusses the sensitivity of the modeled region (which is absent in the

crystallographic structures). In particular, several candidate structures do not show significant differences in dynamics [312]. Moreover, alternate structures of ErbB4 also using MODELLER are remarkably close to initial simulation structures (data not shown). In contrast, the active EGFR monomer demonstrates a uniform level of motion across all four subdomains of the active site with only low-amplitude fluctuations (2-3Å), and shows no significant local deformations. This implies that the motions are more tightly coordinated across the active site in the active conformation than in the inactive conformation. The kinase monomers of the other ErbB family members (modeled ErbB2 and the ErbB4 structure) demonstrate similar motions, as shown in Figure 4.2. The inactive ErbB4 kinase exhibits a dominant motion in the A-loop (6.46 Å), while the P-loop, C-loop, and the α C-helix undergo smaller lateral motions (4.45 Å, 2.64 Å, and 2.30 Å, respectively). In the active conformation, the ErbB4 kinase presents a fluctuation profile similar to EGFR: the subdomains all have similar small-amplitude motions (2-3 Å) with no large local deformations. Thus, for the three homologous members of the ErbB family, which have a high-degree of sequence similarity, not only are the principal motions conserved across the systems, but the characteristic differences between the inactive and active kinase conformations are maintained in character. This finding suggests large (and possibly similar) differences in the internal network of bonds between the two activity states of each kinase. Indeed, the identified bonding network (Table 4.1, Table 4A.1) shows clear conservation across the members of the ErbB family, and also reflects the differences in principal motions between the inactive and active conformations. In general, the inactive conformations have significantly fewer persistent

bonds in this network when compared to the active conformations, consistent with the larger amplitude motion.

EGFR active	HER2 active	ErbB4 active	EGFR inactive	HER2 inactive	ErbB4 inactive	EGFR active	HER2 active	ErbB4 active	EGFR inactive	HER2 inactive	ErbB4 inactive
aC-helix A-loop bonds						A-loop C-loop bonds					
---	---	---	---	---	E739,R841	---	---	---	---	G865,V842	---
E734,K851	E766,K883	E739,K856	---	---	---	L834,R812	L866,R844	L839,R817	---	---	G838,R817
D737,K836	D769,R868	---	---	---	D742,R841	---	---	L834,D813	---	---	---
E738,F832	---	E743,F837	---	---	---	K836,V810	R868,V842	R841,V815	---	---	---
---	---	---	E738,K836	---	E743,R841	E848,R812	---	---	---	---	---
---	---	---	---	---	E743,R817	---	---	K851,R812	---	---	---
aC-helix C-loop bonds						A-loop bonds					
---	---	---	---	---	E743,R817	---	---	D836,K726	---	D863,K753	D836,K726
aC-helix bonds						---	---	D836,T835	---	---	---
---	A763,S760	---	---	---	---	L838,R808	L870,R840	L843,R813	---	---	---
---	E766,R756	---	---	---	---	---	D871,R840	---	---	---	---
E738,K721	E770,K753	E743,K726	---	---	---	A840,G672	---	---	---	---	---
---	---	---	M742,L753	M774,L785	M747,L758	---	---	---	---	D873,R897	---
A743,L679	---	A748,Q684	---	---	---	---	---	K848,T873	---	---	---
---	---	---	---	---	A748,R757	K843,D932	---	K848,D937	---	---	---
C-loop C-loop bonds						---	E876,R898	---	---	---	---
---	H843,D845	---	---	---	---	---	---	E849,K871	---	---	---
---	---	---	R812,D813	R844,D845	---	Y845,Y867	---	Y850,F872	---	---	---
D813,R817	D845,R849	D818,R822	---	---	---	---	---	H846,R865	---	---	---
D813,N818	---	---	---	---	---	---	A852,R870	---	---	---	---
A815,N818	A847,N850	A820,N823	A815,N818	A847,N850	---	---	D880,R897	D853,R870	E848,R865	D880,R897	D853,R870
---	A848,V851	---	---	---	---	---	---	---	---	---	G855,E730
---	---	---	---	---	---	---	---	---	K883,E757	---	---
---	---	---	---	---	---	---	---	---	---	---	K856,E844

Table 4.1: Persistent H-bonds and salt bridges in the three homologous ErbB kinase monomer systems. The salt bridges are in bold and homologous bonds are aligned. The table presents the subdomains with key conserved bonds across the ErbB family, see Table 4A.1 for an exhaustive list.

Activating bond-network: We find several bonds are conserved across all members of the ErbB family in the active state (EGFR numbering is used in this discussion: see Table 4.1): two salt bridges: E734-K851 and E738-K721, three H-bonds: L834-R812, K836-V810, and L838-R808, as well as the bond D813-R817 which is a salt bridge in EGFR and ErbB4, but an H-bond in ErbB2. The E738-K721 salt bridge is highly conserved across all active kinases and helps coordinate the α and β phosphates of ATP bound in the

active site. The E734-K851 salt bridge connects the A-loop and the α C-helix, coordinating the movements of these two sub-domains and dampening larger fluctuations. Similarly, three conserved H-bonds link the A-loop and the C-loop, coupling the motions of these two loops. These can be regarded as “fastening” H-bonds that maintain the N-terminal side of the A-loop open in its active state – the alternative (in the inactive state) being steric hindrance to the binding of ATP and peptide substrates. A fourth H-bond is seen only in ErbB2: E876-R898 fastens the C-terminal side of the A-loop open [312]. Although there is no analogous bond in EGFR and ErbB4, the Y845-Y867 bond in EGFR and the homologous Y850-F872 bond in ErbB4 serve a similar role. The conserved D813-R817 bond positions the sidechain of the catalytic aspartate D813 in the active site and thus likely facilitates the preorganization of the catalytic site in the active kinase system.

Inactivating bond-network: In contrast with the case for the active configurations, few intramolecular bonds in the inactive kinase conformation are conserved across the ErbB family (Table 4.1 and Table 4A.1). In fact, only one such bond is conserved across EGFR, ErbB2 and ErbB4: a hydrogen bond between M742-L753 that pins the C-terminal end of the α C-helix in its location away from the active conformation. The inactive kinases do share an autoinhibitory pattern in which key residues required for kinase activation are sequestered. Similar to the situation described for Lck [308, 309], E738 of EGFR is salt bridged to K836 in a manner that ‘sequesters’ this glutamate and prevents it from forming the highly conserved E738-K721 salt bridge [312] required for full

activation. For ErbB2 and ErbB4, the homologous residue K836 is replaced with an arginine. In ErbB4, this arginine contacts additional residues (as well as the E738 equivalent), apparently dampening the fluctuations of the α C-helix. In ErbB2 the arginine that replaces the K836 equivalent has flipped away from the α C-helix, so cannot sequester the E738 residue equivalent. In ErbB2 and ErbB4, the other half of the highly conserved salt bridge is sequestered; namely, the homologous K721 interacts with the D831 side chain, and this interaction in turn prevents K721 from forming the coordinating salt bridge.

4.3.2) Activation in the asymmetric dimer occurs through disruption of the inactivating bond-network

We also analyzed fluctuations in the EGFR kinase within the context of the asymmetric dimer described by Kuriyan and colleagues [40]. The fluctuations recorded for active EGFR in this context are very similar to those seen in the active EGFR monomer (Figure 4.2), with the conserved bonds described above being mostly preserved (Table 4A.2). Slight variations in bond patterns in the active configuration include a shift from E734-K851 and L838-R808 in the monomer to E734-K836 and A840-R808 in the dimer, but dimerization has little overall effect on the activating bond network. In contrast, in the inactive dimer the first principal component reveals substantial motion of the α C-helix that is much greater than seen in the inactive monomer system (7.49Å shift in dimer compared to 4.14 Å in the monomer, see Figure 4.2). Simulations of the symmetric

dimer interface of the inactive kinase in the unphosphorylated and phosphorylated states does not result in any conformational switching (see Figure 4A.2), each system is stable in the inactive state. Furthermore, in previous simulations of ErbB2 [312] we used loop modeling to reduce the effect of steric clashes with no significant differences in dynamical motions. Therefore we attribute the motion of the α C-helix uniquely to the introduction of the asymmetric dimer interface. Even in the short timescale of 30 ns of the EGFR dimer trajectory, we observe a rearrangement of the α C-helix position towards the active conformation. Consistent with the allosteric activation mechanism proposed by Zhang et al. [40], several interactions in the inactivating bond network surrounding the A-loop and the α C-helix are indeed disrupted in the dimer trajectory, including Y740-S744, L834-D813, H846-R865, and K851-R812 interactions (Table 4A.2a&b). Some bonds (e.g. E738-K836) are still present, although the population statistics indicate that their survival percentage (fraction present in the trajectory) has decreased from >90% in the inactive monomer trajectory to ~60% in the inactive dimer trajectory over 30ns.

When compared with their monomeric counterparts, the ErbB2 and ErbB4 inactive dimers demonstrate a similar loss of bonds surrounding the α C-helix and the A-loop (see Figure 4A.3, Table 4A.2a&b chapter 4A and [312]). For ErbB4, a list of bonds disrupted upon dimerization includes: E739-R841, D742-R841, E743-R817, G838-R817, G855-E730, and K856-E844. Similar to the E738-K836 salt bridge in EGFR, the E743-R841 salt bridge in ErbB4 shows a marked decrease in survival time from >90% in the monomer trajectory to ~70% in the dimer trajectory. Moreover, two of the bonds broken

(E739-R841 and D742-R841) involve the R841 residue, resulting in a significant weakening of the bonds in the inactivating bond-network discussed above that sequester key side-chains in the inactive state.

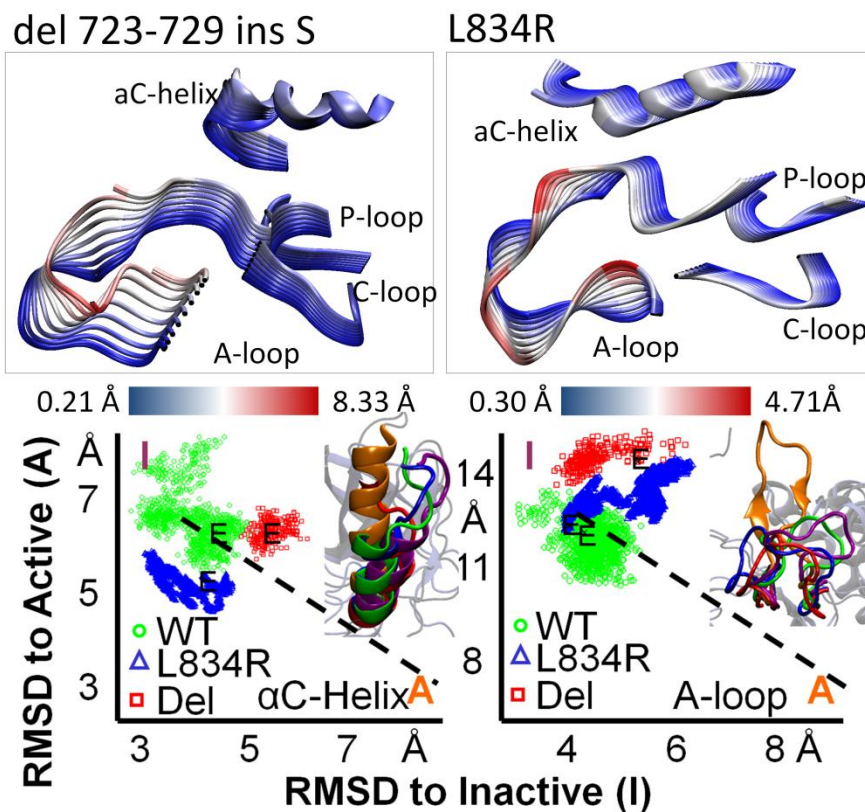


Figure 4.3: (Top) Comparison of the first principal component for the EGFR inactive mutant dimer systems: del 723-729 ins S and L834R. (Bottom) Mapping of the RMSD of the inactive EGFR dimers as a scatter plot with RMSD from inactive on one axis and RMSD from active on the other. The end structures are visualized with an active reference (orange) and an inactive reference (purple). The conformational switching between active and inactive for EGFR of the A-loop and α C-helix are separable.

4.3.3) EGFR mutations activate the kinase by disrupting the inactivating bond network

In light of the allosteric activation mechanism described in the context of the EGFR dimer system, we examined the effect of EGFR-activating mutations clinically identified in non-small cell lung cancer (NSCLC). The two mutants we examined are those most commonly found in NSCLC: a small in-frame deletion in the α C-helix (del 723-729 ins S) and a point mutation in the A-loop (L834R), see Figure 4.3. The deletion mutant physically shifts the α C-helix, removing residues 723-729 and leading to the shortening of a disordered segment that adjoins the α C-helix. This shifts the α C-helix toward the active state. The shortening of the α C-helix also alters the overall movement of the α C-helix, with a distinctly different motion from all other systems (Figure 4.3). The bond patterns we recorded in the dimer of the deletion mutant are similar to those of the wildtype (WT) inactive EGFR dimer (see Table 4A.2a&b). The majority of the differences between the two systems are seen in the α C-helix: the deletion mutant loses two bonds seen in the WT inactive dimer between the α C-helix and the A-loop: E738-F832 & E738-G833. The L834R mutation alters the conformation of the A-loop slightly; however, our trajectory shows the mutant kinase is still in a distinctly inactive conformation. The L834R mutant dimer system shares a similar bond pattern (Table 4A.2a&b) with the WT inactive EGFR dimer system with some exceptions: namely, G833 has H-bonded with H811, similar to the N-terminal fastening bonds that couple the

A-loop and C-loop in the active state. Moreover, R834 has H-bonded to R865, representing a fastening bond to the C-lobe.

4.3.4) Hydrophobic interactions help differentiate active and inactive conformations

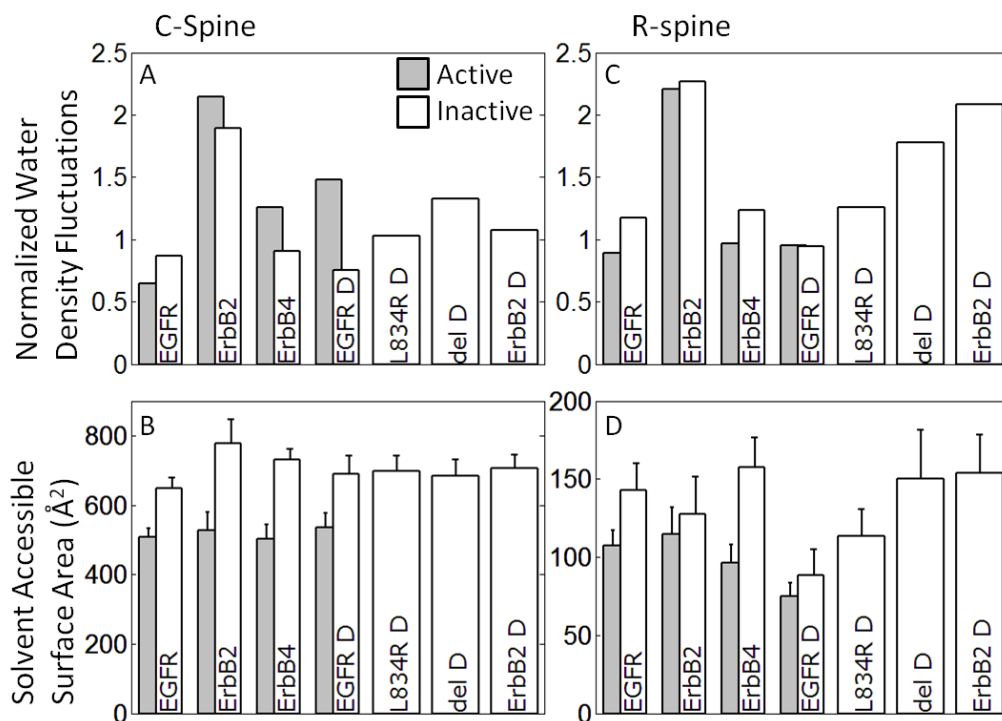


Figure 4.4: Normalized water density fluctuations and the Mean surface accessible (exposed) surface area (SASA) values for functionally important sub-regions, specifically: C-spine (panel A & B), and R-spine. For the normalized water density fluctuations (panels A, C, E) a higher value is correlated with a higher hydrophobicity, which would be expected to bury more SASA. See also Figure 4.5.

To investigate the effect of hydrophobic interactions on the ErbB kinase conformations, we analyzed the hydrophobicity as well as the solvent accessible surface area (SASA) of relevant hydrophobic sub-regions, namely, the C-spine, R-spine, hydrophobic core, and

the α C- β 4 region (Figure 4.1). The four regions have a high percentage of hydrophobic side chains; however, some minor differences between members of the ErbB family exist (Figure 4.1), particularly in the α C- β 4 region. The hydrophobicity of the sub-region is a non-additive quantity, dependent not only upon the primary structure but also on the surrounding environment. Garde et al. [320, 321] have recently proposed an approach for quantifying the hydrophobicity of heterogeneous surfaces using normalized water density fluctuations (see Methods), according to which increased normalized water density fluctuations are used as a signature of a more hydrophobic surface. For an alternate presentation of the hydrophobic interactions, see Appendix A.

The active conformations in the C-spine are noted by the elevated normalized water density fluctuations with the exception of EGFR monomer (where the difference is minor), see Figure 4.4A. The active conformations also consistently expose a mean surface area of around 500 \AA^2 across the members of the ErbB family (represented as a horizontal line), which is consistently less than the exposed surface area in the inactive conformations (mean value of over 700 \AA^2), see Figure 4.4B. Together, these trends reflect a more hydrophobic C-spine region in the active compared to the inactive conformations. The hydrophobicity of the R-spine is not significantly different between the active and inactive conformations (Figure 4.4C). However, the R-spine analysis does reveal that the active conformations have a lower mean SASA value than the inactive conformations (Figure 4.4D); for some systems, the difference is not so clearly delineated in part because it consists of just four residues, so SASA is subject to larger fluctuations.

With similar levels of water density fluctuations between active and inactive conformations, the decreased mean SASA values for the R-spines in the active conformations imply a more stable hydrophobic context, consistent with the finding that the well-formed spines are characteristic of the active kinase conformations. Interestingly, all ErbB2 systems present a higher fluctuation in water density, which is consistent with the notion that hydrophobicity is particularly important in the context of ErbB2 owing to its interaction with Hsp90 known to be mediated by hydrophobic contacts; see also the analysis surrounding the α C- β 4 region, below. We also record a decrease of the R-spine SASA for inactive EGFR in the dimeric form compared to the monomeric form, with similar hydrophobicity (Figure 4.4C,D). The addition of the dimer interface in inactive EGFR reduces the mean SASA for the R-spine from 140\AA^2 to 80\AA^2 , implying that dimerization can provide additional stabilization due to hydrophobic interactions in EGFR.

With respect to the water density fluctuations in the hydrophobic core, the inactive conformation is more hydrophobic than the active conformation (with the exception of ErbB2), see Figure 4.5A. This difference is consistent with the hydrophobic stabilization of the inactive conformation, especially for EGFR and ErbB4 but not for ErbB2. Notably, mutations of hydrophobic residues in the hydrophobic core are reported for EGFR and ErbB4 in clinical studies, (see Discussion, section 4.4), whereas, for ErbB2, such mutations are found surrounding the α C- β 4 region. The analysis in Figure 4.5A also implies that the difference in hydrophobicity between the inactive and active

conformations is reduced in the dimer EGFR compared to monomer. This is consistent with the allosteric activation mechanism, namely that dimerization significantly reduces the hydrophobic advantage and provides a stimulus for activation. As for the SASA results, the hydrophobic core does not show a clear separation between the active and inactive conformations (Figure 4.5B). However, to describe the highly irregular nature of the surface of the hydrophobic core, the water density analysis may be more suitable than the SASA.

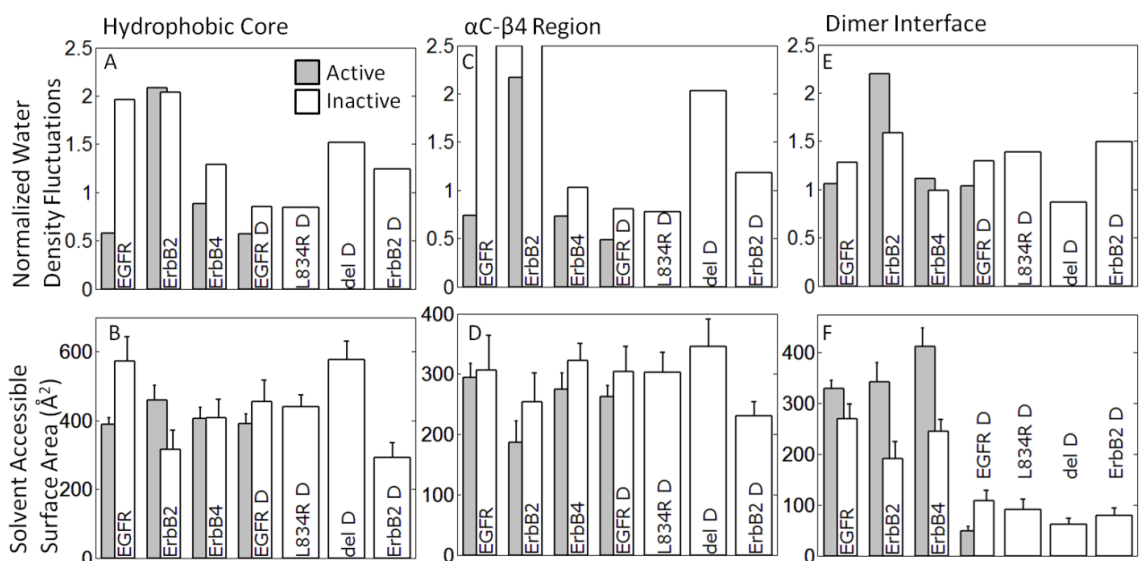


Figure 4.5: Normalized water density fluctuations and the Mean surface accessible (exposed) surface area (SASA) values for functionally important sub-regions, specifically: hydrophobic core (panel A & B), α C- β 4 region (panel C & D) and dimer interface (panel E & F). For the normalized water density fluctuations (panels A, C, E) a higher value is correlated with a higher hydrophobicity, which would be expected to bury more SASA.

The α C- β 4 region is an unstructured span between the α C-helix and the β 4 sheet in RTKs. From a sequence perspective, only in ErbB2 is the α C- β 4 region predominantly hydrophobic (Figure 4.1, see primary sequence; G776 in ErbB2 corresponds to S744 and S749 in EGFR and ErbB4, respectively; G778 in ErbB2 corresponds to D746 and D751

in EGFR and ErbB4, respectively). The water density analysis clearly reflects this trend by singling out the ErbB2 monomer systems are particularly hydrophobic (Figure 4.5C); the mean SASA for the α C- β 4 region in ErbB2 systems is also consistently lower than in other members of the ErbB family (Figure 4.5D). As discussed in [312], this unique feature of ErbB2 is thought to be responsible for its preferential association with the molecular chaperone Hsp90.

The asymmetric dimer interface consists largely of hydrophobic side-chains in the N-lobe of the receiver kinase (L680, I682, L736, L758, and V762) and the C-lobe of the activator kinase (I917, Y920, M921, V924, and M928) [40]. With respect to water density analysis, the dimer interface presents a similar hydrophobicity across all members, with the ErbB2 monomers a little more hydrophobic, particularly in the active conformation (Figure 4.5E). The monomeric systems record a mean SASA value of approximately 350 \AA^2 for the dimer interface residues in the active conformations versus 200 \AA^2 in the inactive conformations (Figure 4.5F). This decrease for the inactive monomers may imply a preference for the inactive state in the monomer context. Notably, the dimeric systems record much lower SASA values of about 75 \AA^2 , implying that hydrophobic stabilization provides a dominant driving force for dimerization; interestingly, for EGFR dimer, the SASA for the inactive state is greater than that for the active case and hence the preference for the inactive conformation is not implied in the context of the dimer. Thus, dimerization also provides a stimulus for activation.

4.4) Discussion

The analysis described here identifies conserved intramolecular non-covalent bond networks across the ErbB family kinases that emerge from sequence homology and lead to conserved dynamic characteristics as well as function. The hydrophilic and hydrophobic networks are cooperative interactions which help differentiate the active and inactive conformations while modulating key loop fluctuations and bonds. Investigation of the networks allows identification of network fragilities which can be exploited through mutation to alter the basal kinase activity.

The bond-networks highlight key conserved residues and bonds that are characteristic to each conformational state. Some of the bonds have been highlighted previously, the highly conserved salt bridge E738-K721 [25] and the fastening bonds L834-R812, K836-V810 and L838-R808 [312]. We add to this another salt bridge coordinating the α C-helix and the A-loop, E734-K851 salt bridge and the H-bond D813-R817 which facilitates the placement of the D813 aspartate side-chain in a catalytically competent orientation. Overall, the six conserved interactions tightly coordinate the N-lobe to the α C-helix, the α C-helix to the A-loop, and the A-loop to the C-loop. The nature of this bond network suggests that interactions among the nearest key subdomains in the ErbB family helps in maintaining the proximity of important catalytic residues, and positioning them so they can contribute directly to kinase activity. The conserved bond network in the inactive conformations of the ErbB family kinases is less extensive (in EGFR, E738-K836 and

D831-K721), but appears to serve a crucial role in sequestering key catalytic residues and thus preventing activity.

Hydrophobic interactions appear to provide context-specific contributions to stability to the active and inactive conformations of ErbB kinases (see Appendix A for an alternate presentation of the hydrophobic interactions). Since a single amino acid change can alter the hydrophobicity of a region, we considered the water fluctuation analysis in conjunction with SASA to present the contributions from hydrophobicity for irregular surfaces characteristic of sub-regions in the kinase. Both SASA and fluctuation analyses of local water densities imply that the inactive monomer conformations of ErbB kinases are preferentially stabilized through hydrophobic interactions associated with the dimer interface. Moreover, consistent across the ErbB family, kinase domain dimerization further reduces the SASA of the dimer interface residues, implying that hydrophobic interactions provide a dominant impetus for dimerization.

The hydrophobic interactions with respect to the C-spine and R-spine regions are found to benefit the active conformations, overall; in particular, the SASA analysis suggests the active conformations benefit from a larger buried surface area). Interestingly, we found ErbB2 to possess a greater hydrophobicity in the spine regions. Thus, the α C- β 4 region as well as the C-spine and R-spine are found to display uniquely hydrophobic character in ErbB2 consistent with its association with Hsp90. Also, in the case of the ErbB2 dimer interface, the active conformation is more hydrophobic than the inactive conformation,

but also exposes more surface area, implying a destabilizing hydrophobic force in the active conformation; considering that ErbB2 does not have a known ligand, this could serve as an auto-inhibitory mechanism against activation. It is also intriguing to note that in the hydrophobic core, there is a large differential between the inactive and active conformations in the water fluctuations, consistent with preferential benefit for the inactive conformation in EGFR and ErbB4, but not for ErbB2. Thus, the analyses for the spine regions and the hydrophobic core collectively lead to the remarkable prediction that while the α C- β 4 region and the R-spine confer hydrophobic benefit to the inactive conformation of ErbB2, the hydrophobic core has the same effect for EGFR and ErbB4. Indeed, this correlates well with clinical studies, where activating point mutations in the hydrophobic core have been found in EGFR and ErbB4 but those in the α C- β 4 region are found in ErbB2, suggesting that the hydrophobic analysis enables the context-specific identification of fragile sub-regions.

The active kinase conformations have a greater number of conserved intramolecular persistent bonds (hydrophilic specific interactions) than the inactive systems, whereas the inactive kinase conformations appear to have a distinct hydrophobic advantage in the dimer interface region. This observation is fully consistent with the view that, for ErbB kinases, the dominant stimuli that activate them operate by modulating the dimer interface (activation mechanism in the wildtype) and the hydrophobic core regions (mode of activation in clinically identified mutation) to destabilize the inactive conformation, which we discuss below.

For the wildtype systems, this finding provides strong support for the allosteric mechanism associated with formation of the asymmetric dimer interface reported by Zhang et al. [40] and Qiu et al. [42], and the recent studies of juxtamembrane domain associations [43-45]. In this respect, the emerging view on allosteric EGFR regulation posits that, rather than forcing the protein to a new conformation, the allosteric interface guides the kinase domain to the active state through pre-existing conformational ensembles [322]. Our results for the EGFR dimer simulations (Figures 4.2, 4.3) are consistent with this view; namely, the allosteric dimer interface causes the α C-helix to alter the conformational space it samples and thus biases it towards a more active conformation. This leads to a metastable intermediate characterized by the α C-helix adopting an active-like conformation, albeit still remaining partially molten. We therefore hypothesize a sequence of events that characterize the EGFR kinase activation pathway (see section 4A.4): (1) Formation of the dimer interface triggers a shift of the α C-helix conformation toward the active state. (2) Formation of the additional helical turns in the α C-helix and the initiation of the A-loop movement, either happen concurrently or sequentially, but the A-loop does not complete its conformational switching until the α C-helix is fully formed. This sequence differs from the mapped pathway of the Src kinase Hck [323-325]; in Hck, following A-loop tyrosine phosphorylation the A-loop can adopt an open (active) conformation while the α C-helix is still in the inactive conformation. In the EGFR the opposite is true, namely, upon

dimerization the α C-helix moves towards the active conformation with the A-loop in the closed (inactive) conformation (see section 4A.4).

Intriguingly, kinase activation in the two clinically identified EGFR activating mutants studied here (L834R and del 723-729 ins S) also appears to be governed by the same principles. The L834R mutation, apart from directly disrupting the hydrophobic core in lieu of the hydrophilic Arg, leads to additional coupling of the A-loop to the C-loop (through the G833-H811 H-bond) and the A-loop to the C-lobe (through the R834-R865 interaction) not seen in other systems. The deletion mutation directly alters the α C-helix to cause a shift toward its active conformation, although it also disrupts the normal dynamics and potential extension of the α C-helix. In the functional studies of Choi et al. [26], the deletion mutant showed an increase in basal phosphorylation, but a decrease in the overall phosphorylation under EGF stimulation. By interfering with the key interactions surrounding the helix, it can potentially stabilize the active conformation of the monomer kinase. However, upon dimerization the active conformation is destabilized because the deformation disrupts both the dimer interface due to its proximity and because the α C-helix cannot fully form. This justifies the dual response of the deletion mutant on kinase phosphorylation activity in the presence/absence of EGF stimulation.

The hydrophilic and hydrophobic bond networks we have identified also enable us to propose a set of mechanisms that increase the basal level of kinase activity for the other EGFR mutations in NSCLC [26, 27, 48]: E685S, G695S, S744I and L837Q, as well as a

set of mutations in ErbB4 found in melanoma [60], namely, E836K, E872K and G936R. Similar to L834R, in the L837Q mutation, replacing the hydrophobic leucine side-chain with the hydrophilic glutamine side-chain is likely to cause it to rotate away from the small hydrophobic core between the α C-helix and the A-loop, thereby disrupting the hydrophobic core and inducing a local steric effect. This reconfiguration may also disrupt the interaction of K836 with E738. The S744I mutation in NSCLC most likely disrupts the M742-L753 bond, allowing transition of the α C-helix into its active position. E685G and G695S are in the N-lobe, close to the asymmetric dimer interface, so mutations here are likely to alter kinase activity either by increasing the dimerization affinity or by reconfiguring the EGFR RTK monomer by partially mimicking the formation of the asymmetric dimer interface. Two of the ErbB4 mutations seen in melanoma, E836K and E872K, are situated around the small hydrophobic core and the change from a negatively charged side chain to a positively charged chain is poised to disrupt the core's stability, thereby providing an activating stimulus. The G936R mutant in ErbB4 is located close to the asymmetric dimer interface in the C-lobe and potentially alters the dimerization affinity. Analogous mutations in ErbB2 have not been reported. Instead, the activating clinical mutations in ErbB2 are in α C- β 4 region [55, 326], shown here to be uniquely hydrophobic in the ErbB family. Disruption of the hydrophobicity of the α C- β 4 region would alter the binding to Hsp90 [124] and increase heterodimerization [123] and thus, activity of ErbB2. Characterization of another set of mutations in ErbB4 found in NSCLC, show two inactivating mutations G802dup and D861Y [125]. D861 is the start of the highly conserved DFG motif in RTKs and mutation of it would reduce activity,

while the G802dup is spatially located near the P-loop and likely affects ATP binding affinity. Similarly in ErbB4, two of the point mutations are located near the active site: E836 is located next to the C-loop, whereas E872 is situated in the A-loop and hence owing to their proximity to the active site, the mutations are also poised to alter substrate binding and phosphorylation; such mechanisms are discussed in a recent computational study in the context of EGFR [327].

As mentioned in the Introduction, A-loop phosphorylation is thought not to be required for ErbB family kinase activation [306]. Consistent with this notion, although we record some context-specific structural changes between systems in which the A-loop is phosphorylated versus unphosphorylated (see section 4A.1-3), our results indicate that this phosphorylation does not provide a dominant activating effect. We reached a similar conclusion in recent studies of ErbB2 using a free energy perturbation approach [312].

In conclusion, our results help to establish the molecular context governing the stability and activation stimuli of ErbB kinases. Given that receptor tyrosine kinases are important signaling elements in cells, and mutations within these kinases cause subtle changes in molecular behavior that result in substantial alterations of downstream signaling [152], this work helps establish the crucial link between molecular mechanisms of kinase activation and the ensuing signaling response.

Chapter 4A.) Appendix for 4: Molecular Dynamics Analysis of Conserved Hydrophobic and Hydrophilic Bond Interaction Networks in ErbB Family Kinases

EGFR active	ErbB2 active	ErbB4 active	EGFR inactive	ErbB2 inactive	ErbB4 inactive
aC-helix aC-helix bonds					
--	P761,K765	--	--	--	--
K730,E734	K762,E766	--	--	K762,E766	--
A731,I735	A763,I767	A736,F740	--	A763,I767	--
N732,S728	N764,S760	N737,G733	--	N764,S760	--
N732,L736	N764,L768	N737,M741	--	--	N737,M741
--	--	--	--	K765,L768	--
K733,D737	K765,D769	V738,D742	K733,D737	--	--
E734,E738	E766,E770	E739,E743	E734,E738	--	--
I735,A739	I767,A771	F740,A744	I735,A739	I767,A771	F740,A744
L736,Y740	L768,Y772	--	L736,Y740	--	--
--	D769,V773	D742,I746	--	--	--
E738,M742	E770,M774	E743,M747	E738,M742	E770,M774	--
--	A771,A775	--	--	--	--
Y740,S744	--	L745,S749	Y740,S744	Y772,G776	L745,S749
V741,V745	V773,V777	I746,M750	--	--	--
C-loop bonds					
H811,D872	--	H816,D877	H811,D872	--	H816,D877
R812,E848	--	--	--	--	--
--	--	--	R812,K851	--	--
R812,D872	--	R817,D877	R812,D872	--	--
R812,S875	--	--	--	--	--
--	--	--	--	A848,L807	--
A816,E882	--	A821,E887	A816,E882	--	A821,E887
N818,T830	N850,T862	N823,T835	N818,T830	--	N823,T835
A-loop A-loop bonds					
--	--	F837,A840	--	--	--
--	--	--	G833,K836	--	--
--	--	--	--	G865,L869	--
--	--	--	--	L866,L870	--
--	--	R841,E847	--	--	--
--	D871,I872	--	--	--	--
--	D871,E874	--	--	--	--
--	--	--	--	--	E844,K856
--	--	--	E841,K843	--	--
--	--	--	E841,E842	--	--
--	--	--	--	--	E847,K848
--	--	--	E844,Y845	--	--
Y845,E848	--	--	--	--	--
--	D880,G881	--	E848,G849	--	--
--	--	--	--	E874,K883	--

Table 4A.1: Persistent H-bonds and salt bridges in the three homologous ErbB kinase monomer systems. The salt bridges are in bold and homologous bonds are aligned.

EGFR WT Active-U Dimer	EGFR L834R Inactive-U Dimer	EGFR deletion Inactive-U Dimer	EGFR WT Inactive-U Dimer	ErbB4 WT Inactive-P Dimer
aC-helix interdimer bond				
--	--	--	--	K689,E912
--	--	--	--	E690,K910
--	--	--	P729,K925	--
--	K730,E939	--	K730,E939	--
--	K733,D918	K733,D918	K733,D918	--
Y740,D918	--	--	--	--
aC-helix A-loop bond				
E734,K836	--	--	--	--
D737,K836	--	--	--	--
--	E738,F832	--	E738,F832	--
--	--	--	E738,G833	--
--	E738,K836	E738,K836	E738,K836	E743, R841
aC-helix bonds				
--	--	K730,F699	--	--
N732,V762	--	--	--	--
E738,K721	--	--	--	--
--	M742,L753	M742,L753	M742,L753	M747,L758
--	A743,R752	--	--	A748,R757
A-loop C-loop bonds				
--	--	D831,N818	D831,N818	--
--	G833,H811	--	--	G838,H816
L834,R812	--	--	--	--
--	--	L834,D813	--	--
K836,V810	--	--	--	--
--	--	--	--	E847,R817
A-loop bonds				
D831,K721	D831,K721	D831,K721	D831,K721	D836,K726
--	L834,R865	--	--	--
A840,R808	--	--	--	--
E841,R808	--	--	--	--
Y845,Y867	--	--	--	--
A847,R865	--	--	--	--
--	E848,R865	E848,R865	E848,R865	--
--	--	--	--	D853,R902
--	--	--	K851,E725	--

Table 4A.2A: Persistent H-bonds and salt bridges in the five ErbB kinase dimer systems. The salt bridges are in bold and homologous bonds are aligned. See also Table 4A.2B

EGFR WT Active-U Dimer	EGFR L834R Inactive-U Dimer	EGFR deletion Inactive-U Dimer	EGFR WT Inactive-U Dimer	ErbB4 WT Inactive-P Dimer
aC-helix aC-helix bond				
N732,S728	--	--	--	--
A731,I735	--	--	--	--
N732,L736	--	--	N732,L736	N737,M741
K733,D737	K733,D737	K733,D737	K733,D737	V738,D742
--	--	E734,E734	E734,E734	--
--	E734,E738	E734,E738	--	E739,E743
I735,A739	I735,A739	I735,A739	I735,A739	F740,A744
L736,Y740	L736,Y740	L736,Y740	--	M741,L745
--	D737,V741	--	--	--
E738,M742	E738,M742	E738,M742	E738,M742	--
A739,A743	A739,A743	--	A739,A743	--
Y740,S744	Y740,S744	Y740,S744	--	L745,S749
V741,V745	V741,V745	--	--	--
A-loop A-loop bonds				
--	D831,K836	--	D831,K836	--
--	F832,K836	--	--	--
--	--	--	G833,K836	--
--	--	--	--	L839,L843
--	--	A835,L837	--	--
--	--	--	G839,Y845	--
--	--	--	--	E844,K848
--	E841,K843	E841,K843	--	--
--	--	--	E848,G849	--
C-loop bonds				
H811,D872	H811,D872	H811,D872	H811,D872	H816,D877
R812,D872	R812,D872	R812,D872	R812,D872	R817,D877
R812,S875	--	--	--	--
A816,E882	A816,E882	A816,E882	A816,E882	A821,E887
N818,T830	N818,T830	N818,T830	--	N823,T835
--	--	--	N818,D831	--
C-loop C-loop bonds				
--	--	--	--	H816,R817
--	R812,D813	--	R812,D813	--
D813,R817	--	--	--	D818,R822
--	A815,N818	A815,N818	A815,N818	A820,N823

Table 4A.2B: Persistent H-bonds and salt bridges in the five ErbB kinase dimer systems. The salt bridges are in bold and homologous bonds are aligned.

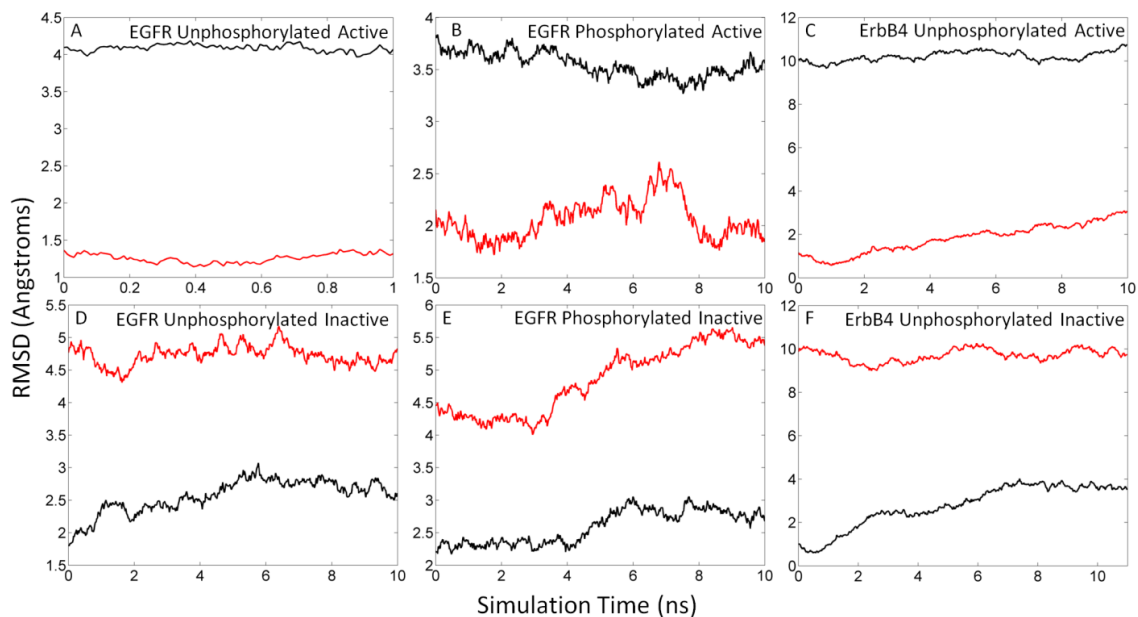


Figure 4A.1: The RMSD plots for the monomeric systems showing RMSD from the active (red) and inactive (black) conformations.

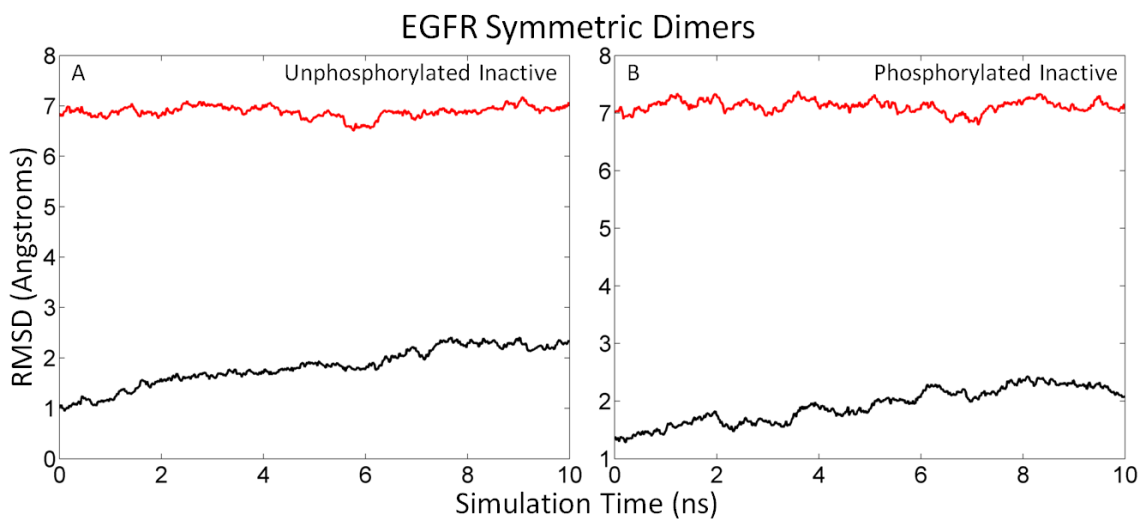


Figure 4A.2: The RMSD plots for the EGFR symmetric dimer systems showing RMSD from the active (red) and inactive (black) conformations.

4A.1) A-loop phosphorylation in monomers promotes coordination of the catalytic sub-domains

To simulate the phosphorylated systems, we used the phosphotyrosine patch from the CHARMM27 forcefield on the A-loop tyrosines [328]. The systems were minimized, equilibrated and simulated as described in the methods. There are two phosphorylated systems created for the EGFR monomers: one in the active conformation and the other in the inactive conformation. Two similar monomer systems for ErbB2 have been previously simulated [312].

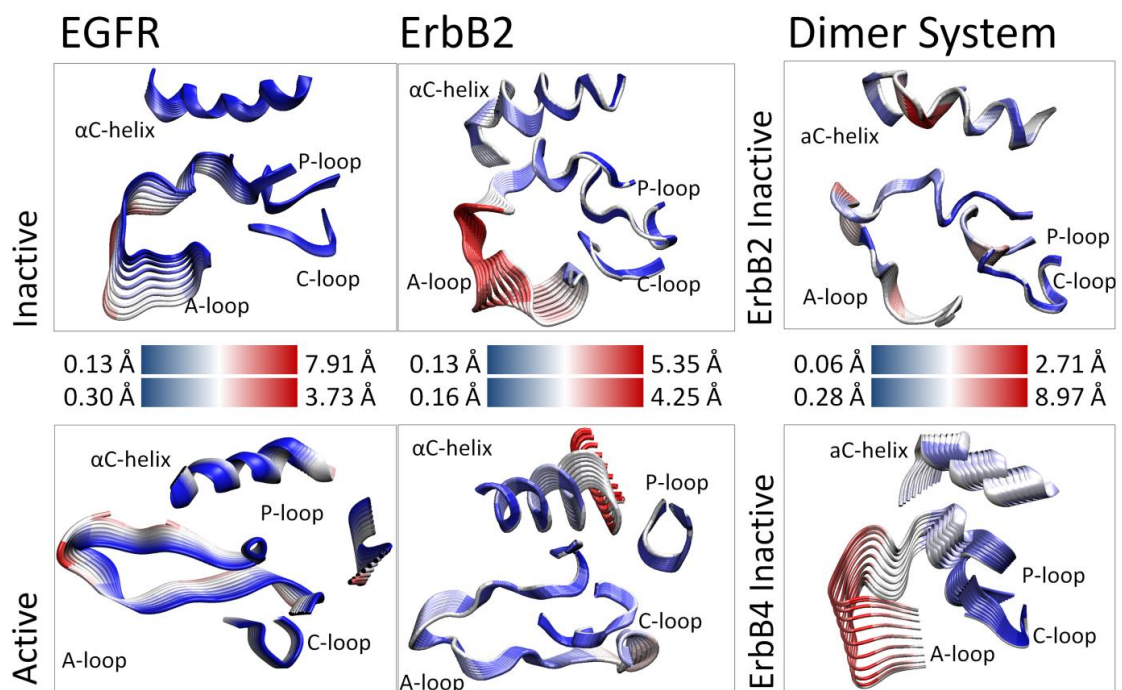


Figure 4A.3: Visualization of the first principal component for the four key sub-domains in the inactive and active phosphorylated ErbB systems. The motions are overlaid sequentially, where the large-amplitude motion in each frame is highlighted in red, and the low-amplitude motion is highlighted in blue.

The added charge from the phosphate tail to the A-loop tyrosine (Y845 in EGFR, Y877 in ErbB2) causes a minor reorganization of the residues in the A-loop (~ 2 Å), but does not greatly alter the rest of the protein. Phosphorylation in the active systems keeps the active site subdomains (A-loop, P-loop, α C-helix and C-loop) coordinated with equivalent levels of fluctuations as the unphosphorylated systems (~ 3 Å). However, A-loop phosphorylation does not have a uniform effect upon the dominant modes of motion in EGFR and ErbB2 (see Figure 4A.3). Compared to the unphosphorylated EGFR system, the inactive has the same motions in the A-loop (~ 7.5 Å), but dampened fluctuations in the other three loops (from ~ 4 Å to ~ 2 Å). In contrast, phosphorylation of the ErbB2 systems does not greatly alter the magnitude of the fluctuations.

Activating Bond Network: The A-loop phosphorylated active systems mostly preserve the conserved bonds seen in the unphosphorylated active systems (E738-K721, E734-K851 salt bridge; L834-R812, K836-V810, L838-R808, D813-R817 H-bonds), see Table 4A.3. The ErbB2-P system loses the E734-K851 salt bridge, because the added charge from phosphorylation causes the K851 residue to bind to the phosphate group (Y877-K883). Both EGFR-P systems gain the D831-K721 bond, as a salt bridge in the active state and as an H-bond in the inactive state. Since K721 is also salt bridged to E738, the D831-K721 bond is a consequence of having all the catalytic residues close together. The active phosphorylated systems also gain several more fastening bonds that maintain the C-terminal side of the A-loop in the open conformation: the ErbB2 systems gain Y877-

EGFR-P active	ErbB2-P active	EGFR-P inactive	ErbB2-P inactive	EGFR-P active	ErbB2-P active	EGFR-P inactive	ErbB2-P inactive
α C-helix bonds				C-loop C-loop bonds			
--	--	N732,A726	--	--	--	--	H843,N850
--	A763,S760	--	--	--	--	R812,D813	R844,D845
N732,V762	--	--	--	D813,R817	D845,R849	--	--
--	E766,R756	--	--	--	D845,N850	--	--
E738,K721	E770,K753	--	--	A815,N818	A847,N850	A815,N818	--
--	M774,L785	--	--	--	A847,V851	--	--
--	--	A743,R752	--	--	--	A816,V819	A848,V851
α C-helix A-loop bonds				C-loop C-loop bonds			
--	--	--	T759,E874	--	--	--	R849,N850
K730,E848	--	--	--	α C-helix α C-helix bonds			
E734,K836	--	--	--	--	P761,K765	--	--
E734,K851	--	--	--	--	K762,E766	K730,E734	--
--	E770,F864	E738,F832	--	A731,I735	A763,I767	--	--
--	--	E738,K836	--	--	N764,S760	--	--
A-loop bonds				N732,L736	N764,L768	--	N764,L768
D831,K721	--	D831,K721	D863,K753	K733,D737	K765,D769	--	K765,D769
--	--	--	G865,H843	E734,E738	E766,E770	--	E766,E770
--	--	--	R868,R840	I735,A739	I767,A771	--	I767,A771
L838,R808	L870,R840	--	--	L736,Y740	L768,Y772	--	--
--	E876,R898	--	--	D737,V741	D769,V773	--	--
--	--	Y845,R865	Y877,R897	E738,M742	E770,M774	E738,M742	--
Y845,Y867	Y877,F899	--	--	A739,A743	A771,A775	A739,A743	A771,A775
A847,R865	A879,R897	--	--	--	--	Y740,S744	Y772,G776
--	--	E848,R865	--	--	V773,V777	V741,V745	--
A-loop C-loop bonds				A-loop A-loop bonds			
--	--	D831,N818	--	--	D863,G865	--	--
L834,R812	L866,R844	--	--	--	--	D831,K836	--
--	--	L834,D813	--	--	--	--	G865,L869
K836,	--	--	--	--	--	G833,K836	--
V810	R868,V842	--	--	--	--	--	L866,L870
Y845,R812	Y877,R844	Y845,R812	Y877,R844	--	R868,Y877	--	--
C-loop bonds				--	D871,E874	--	--
H811,D872	--	H811,D872	--	K843,E844	--	--	--
R812,D872	--	R812,D872	--	--	Y877,K883	--	Y877,K883
A816,E882	--	A816,D882	--	--	D880,G881	--	--
N818,T830	--	N818,T830	--	--	D880,K883	E848,K851	--

Table 4A.3: Persistent H-bonds and salt bridges in the phosphorylated ErbB kinase monomer systems. The salt bridges are in bold and homologous bonds are aligned.

F899, A879-R897 as seen in [312]; the EGFR phosphorylated systems have the homologous fastening bonds in Y845-Y867 as well as A879-R865. All of the phosphorylated systems share a common bond in Y845-R812, which strengthens the coordination between the A-loop and C-loop. The other fastening bonds in the active conformation have an increased survival time and shorter bond length, further increasing the coupling between the A-loop and C-loop.

Inactivating Bond Network: The inactivating bond network for the phosphorylated systems does not present as many conserved bonds. However, similar to the unphosphorylated systems, there are bonds sequestering key catalytic residues. Half of the highly conserved salt bridge E738-K721 [312] is salt bridged to K836 in EGFR while the other half, residue K721, is salt bridged to D831 in ErbB2. The inactive phosphorylated systems have only one common bond not shared by any of the other systems, Y845-R865, which maintains the closed A-loop conformation and thus serves a role opposite of a fastening bond (see Table 4A.3). However, the conserved H-bond in the unphosphorylated inactive systems M745-L753 is not present.

Phosphorylation of the A-loop tyrosine in the ErbB systems provides some alteration to the bond patterns, strengthening the coordination between the loops. However, the shift is not sufficient to cause a conformational change in the short time scales that we employ here. The effects of phosphorylation are more subtle and likely involve more complex protein-protein interactions to promote the full conformational change. To this effect we

have also simulated several phosphorylated systems in the asymmetric dimer configuration to examine whether dimerization and phosphorylation have a synergistic effect as described below.

4A.2) Phosphorylation of the A-loop does not perturb the dimer-mediated allosteric activation mechanism

Two A-loop phosphorylated dimer systems were examined to explore how phosphorylation could affect the dimer-mediated allosteric activation mechanism. An ErbB4 homodimer with the activator kinase in the active conformation and the receiver kinase in the inactive conformation was created and simulated using the same methodology as the monomers. A similar ErbB2 heterodimer has previously been simulated [312].

Addition of the asymmetric dimer interface causes a shift of the α C-helix toward the active conformation, similar to the shift observed in the unphosphorylated systems. The ErbB4 dimer exhibits a comparable motion and directionality in the A-loop and the α C-helix to the EGFR dimer systems. The A-loop-phosphorylated ErbB2 dimer demonstrates slightly more motion than the unphosphorylated dimer, predominantly in the α C-helix (2.56 Å vs. 1.14 Å). Phosphorylation of the A-loop tyrosine binds the A-loop and the α C-helix tightly, increasing the motions between the two subdomains. The increased fluctuations are located near the N-terminal region of the helix (see Figure

4A.3). This region aligns with the molten region of the EGFR and ErbB4 dimer, implying a larger motion allowed if the α C-helix of the ErbB2 dimer were also partially molten.

4A.3) Hydrophobic interactions are not significantly affected by A-loop phosphorylation

Phosphorylation of the A-loop tyrosine in the ErbB family does not significantly alter the hydrophobic interactions of the five hydrophobic sub-regions: the asymmetric dimer interface, the α C- β 4 region, the C-spine, the R-spine and the hydrophobic core, particularly in comparison to the unphosphorylated states. In the dimer interfaces, the water fluctuation analysis shows a similar hydrophobicity in the phosphorylated systems as the unphosphorylated systems (Figure 4A.4A). For the dimer interface residues, the active conformations have the highest mean SASA value of $\sim 360 \text{ \AA}^2$ with the inactive conformations having a significantly lower value of $\sim 220 \text{ \AA}^2$ (Figure 4A.4B). The dimer system is the lowest of all the systems with a mean SASA value of $\sim 75 \text{ \AA}^2$. Considering that the α C- β 4 region is predominantly hydrophobic only in ErbB2, it has the highest overall hydrophobicity of all systems (Figure 4A.4C). Correlated with that, the SASA values for the phosphorylated systems reveals that the ErbB2 systems have a lower mean SASA value of $\sim 220 \text{ \AA}^2$ with the non-ErbB2 systems having a higher mean SASA value of $\sim 315 \text{ \AA}^2$ (Figure 4A.4D).

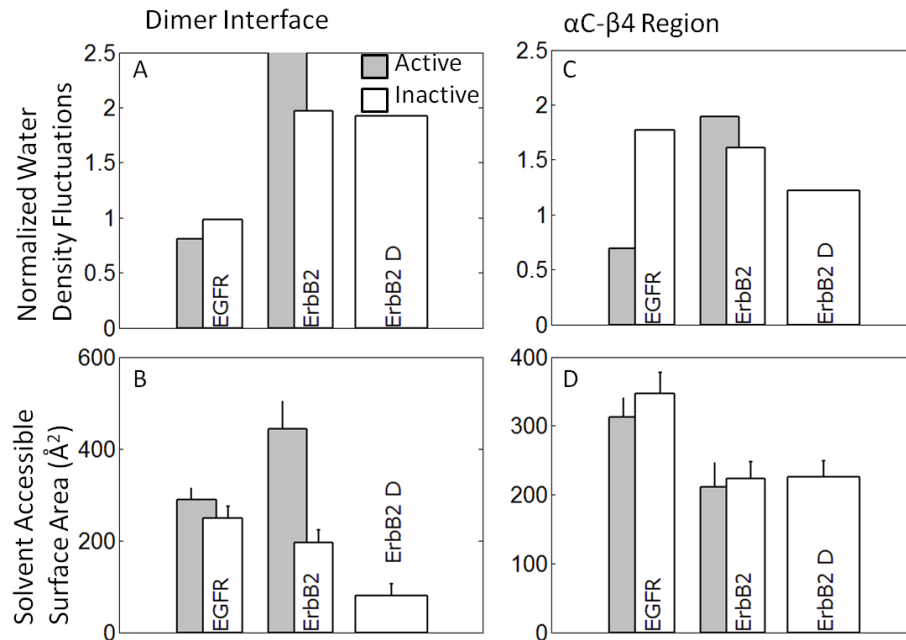


Figure 4A.4: Comparison of the normalized water density fluctuations with the mean solvent accessible surface area for two functionally important subregions for the ErbB family in the A-loop phosphorylated systems: the dimer interface (panels A & B) and the α C- β 4 region (panels C & D). A higher normalized density water fluctuation implies a greater hydrophobicity, which would have less exposed surface area.

The C-spine in the active conformation presents a smaller mean SASA value than the inactive kinase systems. The active conformations have a mean SASA value of 500 \AA^2 with the inactive conformations having greater variability, but a higher mean value of 680 \AA^2 (Figure 4A.5B). In the ErbB2 phosphorylated systems, the hydrophobicity shows a similar trend implying a well formed spine (Figure 4A.5A). However, in the EGFR systems the inactive has a higher exposed surface area with a higher water density fluctuations and therefore higher hydrophobicity, implying a better formed spine in the active conformation. The R-spine has a similar trend of a better formed spine in the active conformation in all systems. The active conformations have a lower SASA value than

the inactive conformations (Figure 4A.5D) with the inactive conformations having an increase water density fluctuation (Figure 4A.5C). For the hydrophobic core (Figure 4A.5F), the active conformations have a higher mean SASA value ($\sim 500 \text{ \AA}^2$) than the inactive conformations ($\sim 350 \text{ \AA}^2$). However, the inactive and active EGFR monomer kinases have approximately equal SASA values, while the ErbB2 active kinase has a higher mean SASA value than the inactive system. While the water density analysis shows the inactive having a higher hydrophobicity than the active conformation in the EGFR monomers, the gap has shrunk in comparison to the unphosphorylated systems implying less of a network fragility in the phosphorylated systems (Figure 4A.5E).

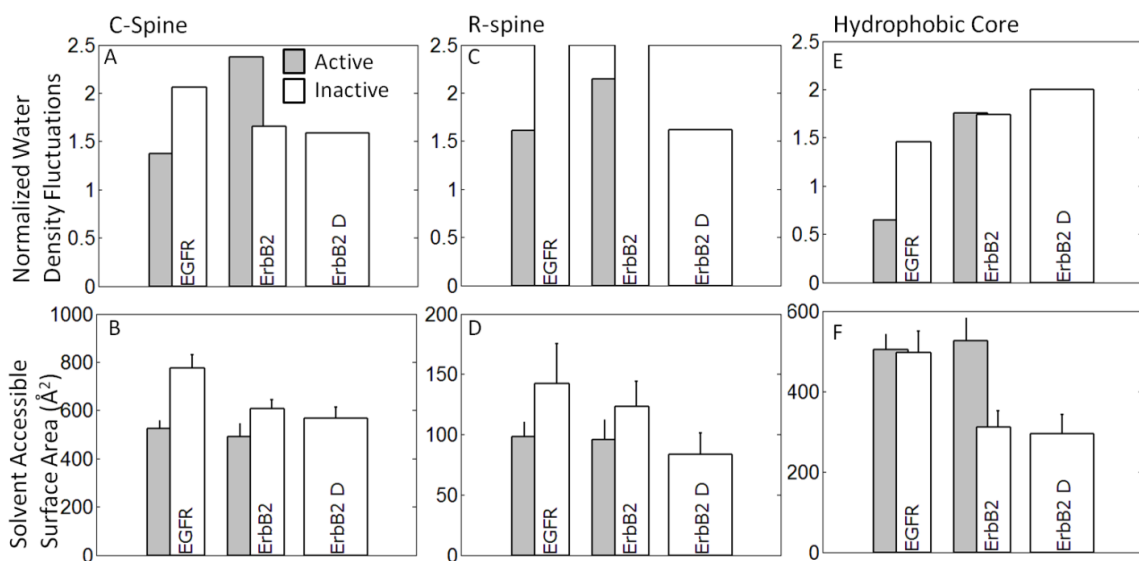


Figure 4A.5: Comparison of the normalized water density fluctuations with the solvent accessible surface area for three functionally important sub-regions for the ErbB family in the A-loop phosphorylated systems: the C-spine (panels A & B), the R-spine (panels C & D) and the hydrophobic core (panels E & F). A higher normalized density water fluctuation implies a greater hydrophobicity, which would have less exposed surface area.

Overall the effect of A-loop phosphorylation in the ErbB family is not clear. Although the phosphorylation of the A-loop tyrosine strengthens specific linkages between the A-loop and C-loop in the active state and promotes a subtle reorganization of the bond network in the inactive state, the stimulus of the added charge is insufficient to cause a conformational shift through hydrophilic bonds. The phosphorylated systems present a similar profile to the unphosphorylated systems with regard to hydrophobic interactions. We have previously performed an analysis on the ErbB2 monomer system with a $\Delta\Delta F$ of -1.1 ± 1.4 kcal/mol difference as a result of phosphorylation of the A-loop tyrosine, and a $\Delta\Delta F$ of 1.2 ± 1.5 kcal/mol difference for dephosphorylation [312].

4A.4) Targeted molecular dynamics studies reveal specific interactions associated with the transition pathway

Targeted molecular dynamics (TMD) simulations were performed in CHARMM [329] using the RMSD of the backbone atoms of the αC -helix and the A-loop as the reaction coordinate (χ_i). The activation pathway is divided into 20 smaller windows and each segment of the window is simulated individually to ensure adequate sampling. To enhance the sampling of low probability events, a harmonic restraint of constant $k=20$ kcal/mol/(\AA^2) is applied along the reaction coordinate, χ_i .

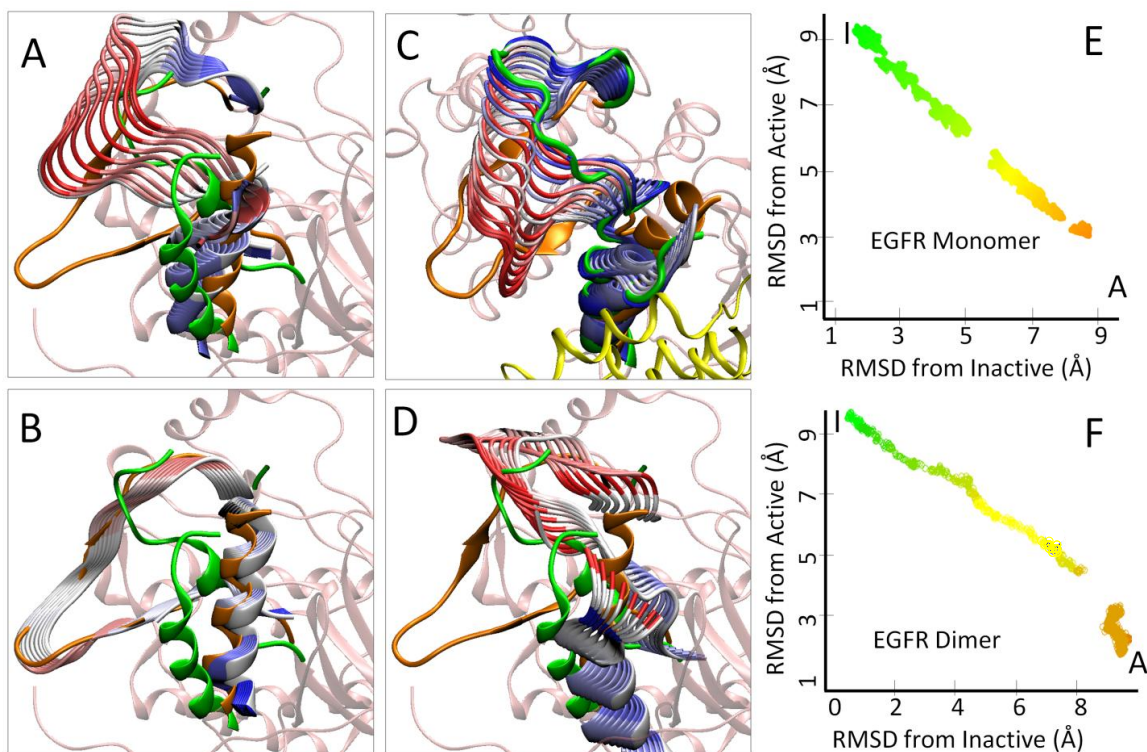


Figure 4A.6: Comparison of the first principal component of the EGFR TMD monomer simulation (panel A) and the EGFR TMD dimer simulation (panel C) vs. the active (Panel B) and inactive EGFR (Panel D) monomer systems, with an active reference (orange) and inactive reference structure (green). The activator kinase is shown in yellow in panel B. The motions are overlaid sequentially, where the large-amplitude motion in each frame is highlighted in red, and the low-amplitude motion is highlighted in blue. TMD pathway for the EGFR monomer (panel E) and EGFR dimer (panel F).

The TMD study of EGFR captured the transition of the kinase from the inactive to the active conformation. The PCA of the TMD trajectory is depicted in Figure 4A.6A. Comparison of the PCA of the active and inactive EGFR unphosphorylated systems to the TMD PCA reveals how each set of fluctuations progresses along the activation pathway. The active unphosphorylated PCA fluctuates closely to the active state without much motion, showing it is stable in the current energy well (see Figure 4A.6B).

However, the inactive state shows some motion toward the active conformation, with both the α C-helix and A-loop moving toward the active state in an indirect fashion (see Figure 4A.6D). The inactive EGFR unphosphorylated system exhibits comparable motion to the TMD EGFR study (7.34 vs. 8.03 Å in the A-loop, 4.95 vs. 5.23 Å in the α C-helix, respectively), though the motions are not sufficiently directed toward the active state. The TMD study did not completely sample the activation pathway; similar to the work of Athanasios et al. [311], we observed several significant molecular events: the breaking of the E738-K836 salt bridge, the formation of the E738-K721 salt bridge and the extension of the α C-helix (Figure 4A.6E). This is similar to what we observed in the EGFR dimer TMD simulations (see Figure 4A.6C). However, the α C-helix extension progresses through a different pathway than that of the monomer (see Figure 4A.6F). The complete pathway has not been mapped, but the data we have collected strongly suggests the α C-helix shifts before the A-loop contrary to the pathway mapped out for the Src kinase Hck [324, 325].

Chapter 5.) A multiscale modeling approach to investigate molecular mechanisms of pseudokinase activation and drug resistance in the HER3/ErbB3 receptor tyrosine kinase signaling network

5.1) Introduction

This chapter is modified from work that has been submitted to Molecular Biosystems and is currently under review [330]. Approximately 10%, or 48 of the 518 protein kinases encoded by the human genome, lack at least one of three catalytic amino acid motifs in the kinase domain, including the Val-Ala-Ile-Lys (VAIK) motif in subdomain II, in which the lysine residue facilitates the positioning of the ATP molecule, the His-Arg-Asp (HRD) motif in subdomain VIb, which contains the catalytic aspartic acid residue responsible for deprotonating the substrate hydroxyl group, and the Asp-Phe-Gly (DFG) motif in the A-loop, which contains the Mg^{2+} -coordinating aspartic acid residue that orients the β and γ phosphates of ATP [4, 331]. These kinases, termed ‘pseudokinases,’ are therefore commonly regarded as catalytically inactive. However, not all kinase domains that lack one or more of the conserved catalytic motifs are inactive. The kinase WNK1 is catalytically active despite lacking the crucial lysine in its VAIK motif, as a lysine residue in subdomain I performs the function of the missing amino acid [332]. In a recent structural and biochemical study, Mukherjee *et al.* [333] resolved the crystal structure of the pseudokinase domain of Ca^{2+} /calmodulin-activated serine threonine kinase (CASK), which lacks both of the canonical Mg^{2+} -coordinating residues, and

determined that CASK employs an unusual phosphorylation mechanism in which the kinase preferentially binds unchelated nucleotides, and is hence capable of catalyzing phosphotransfer despite lacking the conserved DFG motif. It is therefore plausible that other pseudokinases are not truly inactive, but may employ alternative modes of ATP binding and phosphoryl transfer.

In the ErbB family of receptor tyrosine kinases (RTKs), which includes epidermal growth factor receptor (EGFR/ErbB1), ErbB2 (HER2), ErbB3 (HER3) and ErbB4 (HER4), ligand binding followed by receptor homo- or heterodimerization performs a crucial role in regulating critical cellular processes including migration, differentiation, and proliferation [3, 334, 335]. EGFR, ErbB2, and ErbB4 all observe the canonical mechanism of phosphate transfer. ErbB3, however, is the only member of the ErbB family that has long been regarded as an inactive pseudokinase due to amino acid substitutions in the conserved kinase domain. Until recently, ErbB3 activity has been largely undermined, as it contains an asparagine residue (N815) in place of the catalytic aspartic acid residue that is responsible for deprotonating the substrate hydroxyl group [336]. In addition, a crucial glutamate residue in the α C helix, which indirectly participates in ATP-binding, is replaced by a histidine (H740) in ErbB3. Hence it has been widely believed that ErbB3 is dependent on interactions with the other active ErbB receptors for its biological activity.

However, it was recently demonstrated that ErbB3 does, in fact, exhibit robust residual kinase activity that may be crucial for ErbB signaling. Two recent studies by Jura et al. [43] and Shi et al. [337] resolved the crystal structure of the ErbB3 kinase domain, revealing an inactive-like conformation. In addition, Shi et al. [337] demonstrated that, when clustered at a membrane surface, the purified ErbB3 intracellular domain is capable of robust autophosphorylation, albeit at a level 1000-fold weaker than the active EGFR kinase, and that the kinase domain binds ATP with an affinity similar to that of other active kinases. In the same study [337], our quantum mechanics/molecular mechanics (QM/MM) simulations suggested that ErbB3 may catalyze phosphoryl transfer from its inactive-like kinase conformation via an alternative pathway, which explains the 1000-fold slower rate of phosphoryl transfer in ErbB3 compared to EGFR: the alternative pathway involves the migration of the substrate tyrosyl –OH proton directly to the O1 γ oxygen of ATP, hence obviating the requirement for the catalytic aspartate residue. Our study suggests the intriguing possibility that therapeutic targeting of the robust activity of ErbB3, rather than the routinely-targeted EGFR and ErbB2 kinases, may provide a more effective treatment strategy for certain ErbB-driven cancers. However, a major question arising from our work is: what is the physiological relevance, if any, of weak ErbB3 activity to ErbB signaling in the cell?

This question is clinically pertinent as over-expression and activating mutations of the ErbB kinases are implicated in cellular transformation and clinical malignancies including lung and breast cancers [49, 338-342]. Recent studies have demonstrated that

mechanisms of resistance to tyrosine kinase inhibition (TKI) of EGFR and possibly ErbB2 in the treatment of certain human malignancies are mediated by ErbB3 [343-346]. In this scenario, it is hypothesized that incomplete inhibition of ErbB2 catalytic activity restores phosphorylation of ErbB3 in the context of ErbB2/3 heterodimers. As ErbB3 contains six PI3K-binding sites in its C-tail, leaky ErbB3 phosphorylation induces proliferative signaling through the PI3K-AKT pathway [343]. However, the detailed molecular mechanism of this process remains an open question. Indeed, it is possible that ErbB3 catalytic activity is involved in the TKI resistance mechanism.

To explore the implications of ErbB3 activity for ErbB signaling and TKI resistance, a multiscale modeling approach is advantageous. Multiscale computational modeling has been applied to a variety of biological systems [347-351] to help quantify the complexity inherent in intracellular signaling networks. As the biochemical processes within a cell occur on multiple spatial and temporal scales, a multiscale modeling approach is necessary to represent a hierarchy of interactions ranging from the molecular (nm, ns) to cellular signaling (μm , ms) length and time scales. Multiscale modeling provides a powerful and quantitative methodology for studying the effect of molecular perturbations, in our case, ErbB3 catalytic activity, on downstream signaling events, i.e., the ErbB signaling network and development of TKI resistance.

In light of the implication of ErbB3 in TKI resistance mechanisms, the recent interest in catalytic mechanisms of pseudokinases and particularly in the potential for ErbB3

activity demonstrated by Shi et al. [337], we pursue a multiscale modeling study of the ErbB3 kinase system at the molecular and cellular levels. A variety of modeling techniques, ranging from atomic-level molecular dynamics (MD) simulations to systems-level modeling, are applied to investigate the non-canonical catalytic mechanism employed by the ErbB3 kinase and the physiological relevance of this activity to mechanisms of drug resistance in an ErbB-driven tumor cell *in silico*. Here we discuss the results of our MD simulations which suggest that the assembly of the catalytic site in ErbB3 is achieved through a unique network of hydrophobic interactions, in contrast to the extensive set of hydrogen bonds present in the active sites of the other ErbB kinases, and thus the ErbB kinases may have evolved two different modes of achieving the shared goal of prearranging the catalytic machinery for phosphoryl transfer. For how the homology modeling techniques feed into MD simulations as well as the investigation of the molecular level simulation affecting the cell signaling networks, see sections 7.1 and 7.3, respectively.

5.2) Results

5.2.1) PCA reveals that the sub-domain motions of ErbB3 diverge from those of an inactive ErbB kinase

To determine whether the ErbB3 kinase behaves similarly to its ErbB family members at the atomic level, we performed 10 ns molecular dynamics (MD) simulations of the ErbB3

crystal structure (PDB code 3LMG) [337]. The ErbB3 structure was stable for the duration of the simulation, as indicated by the time-course plots of the RMSD of the backbone atoms (Fig. 5.1). Principal component analysis (PCA) was applied to the 10 ns trajectory to characterize the most significant global motions of the ErbB3 kinase. The PCA calculation is based on the diagonalization of the variance-covariance matrix of the atomic fluctuations along the MD trajectory to yield the set of eigenvectors (PCs) and associated eigenvalues. The eigenvectors represent the independent modes of atomic motion, and the eigenvalues reflect the contribution of the corresponding eigenvectors to the global fluctuation of the protein.

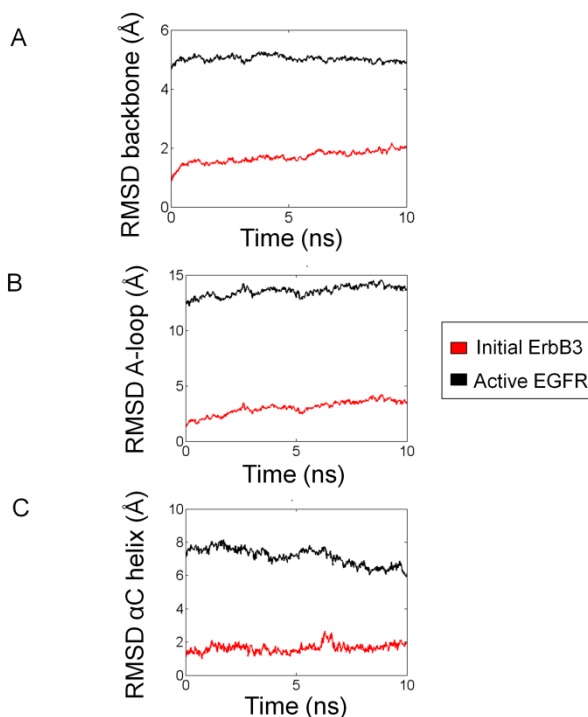


Figure 5.1. Molecular dynamics time course plots of the RMSD for (A) the backbone atoms of the ErbB3 kinase, (B) the backbone atoms of the A-loop and (C) the backbone atoms of the α C helix. The RMSD is plotted in reference to the initial (unsimulated) structure (*red*) as well as the active EGFR structure (*black*), for reference.

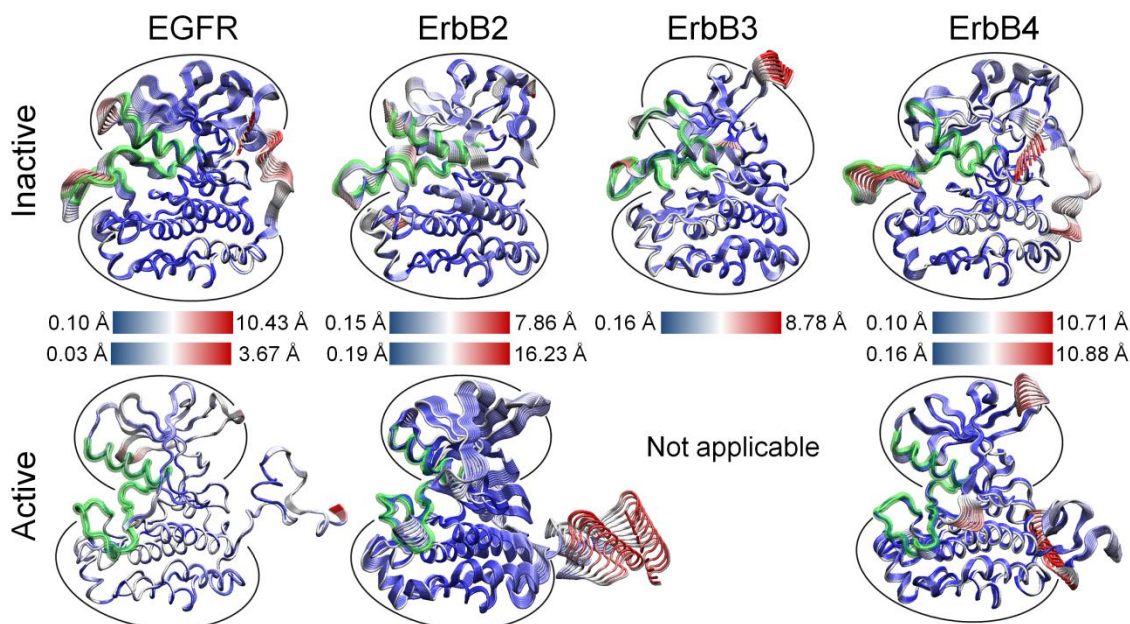


Figure 5.2: Motion along the first principal component of the MD trajectory is illustrated for the complete ErbB3 kinase and compared to the active and inactive conformations of EGFR, ErbB2 and ErbB4. The structures are color-coded according to the RMSD, where red regions indicate large-amplitude fluctuations and blue regions indicate small-amplitude fluctuations. The A-loop and α C helix are highlighted in green for structural reference. Overall, the global motions are conserved across the ErbB family members.

Motion along the first eigenmode for the complete ErbB3 kinase was compared to PCA trajectories for the inactive and active conformations of EGFR, ErbB2 and ErbB4 that had been previously simulated by us [348, 352] (Fig. 5.2); overall the global motions of the kinases were conserved across the ErbB family members. To determine whether the PCA pattern was also conserved in ErbB3's unique active site region, we applied PCA to an active site region which included the A-, C-, and N-loops and the α C helix. Figure 5.3 illustrates that motion along the first eigenmode in the inactive EGFR, ErbB2 and ErbB4 systems is dominated by A-loop movement, with smaller fluctuations in the other

catalytic sub-domains, whereas the active ErbB systems exhibit a uniform level of motion across the sub-domains, with lower-amplitude fluctuations. Despite the crystallization of the ErbB3 kinase in an inactive-like conformation, the PCA pattern in ErbB3 appears to diverge from that of the other inactive ErbB kinases and instead resembles that of the active ErbB kinases, in terms of its concerted and low-amplitude sub-domain fluctuations. We reason that the interactions among the A-, C-, and N-loops and the α C helix in the active ErbB systems (i.e., active EGFR, ErbB2 and ErbB4, and weakly-active ErbB3) are crucial for alignment of the key sub-domains for catalysis. Consistent with this view, our PCA demonstrates that at the atomic level, the sub-domain motions of the ErbB3 system deviate from those of an inactive ErbB kinase.

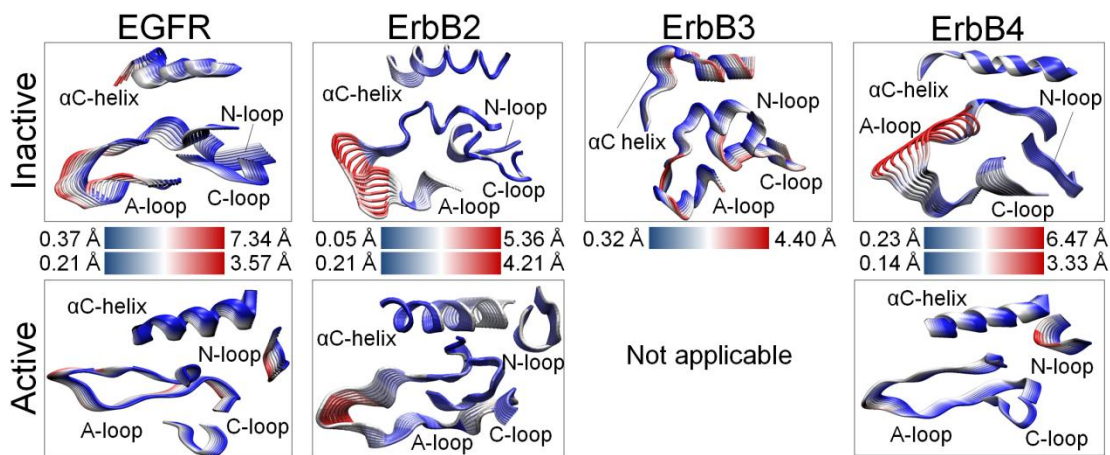


Figure 5.3. Motion along the first principal component of the MD trajectory is illustrated for ErbB3 and compared to the active and inactive conformations of EGFR, ErbB2 and ErbB4 focused on the active site. The structures are color-coded according to the RMSD, where red regions indicate large-amplitude fluctuations and blue regions indicate small-amplitude fluctuations. Despite the crystallization of the ErbB3 kinase in an inactive-like conformation, the ErbB3 PCA pattern resembles that of the active ErbB kinases, in terms of the concerted and low-amplitude fluctuations of its catalytic sub-domains.

5.2.2) Hydrogen bonding analysis indicates a lack of conserved hydrophilic interactions in the ErbB3 catalytic site

To identify specific interactions that could be contributing to the distinct pattern of global motion in the ErbB3 kinase in the inactive-like conformation, individual salt bridges and hydrogen bonds were tabulated for the ErbB3 system through a hydrogen-bonding analysis of the 10 ns MD trajectory (Table 5.1a&b) and compared to the bonds present in the other ErbB systems [348, 352]. The major difference that we observed between the inactive and active ErbB systems is a significantly greater number of bonds that couple the catalytic sub-domains of the kinases in the active systems compared to the inactive systems. Table 5.1a&b highlights several of these bonds (EGFR numbering will be used here): E734-K851 is a key salt bridge which couples the A-loop and α C helix and is conserved across the active conformations of EGFR, ErbB2 and ErbB4, and the L834-R812 and K836-V810 bonds bridge the A-loop and C-loop to maintain the A-loop in its extended, active conformation. We hypothesize that the tight coupling of the sub-domains in the active ErbB systems may help to correctly position the catalytic residues during assembly of the active site for phosphoryl transfer [352].

By contrast, the bonding pattern in ErbB3 is distinct from the active and inactive states of its ErbB family members in its scarcity of hydrogen bonds (Table 5.1a&b). ErbB3 lacks many of the bonds present in the other ErbB kinases due to substitutions in its amino acid

EGFR active	ErbB2 active	ErbB4 active	EGFR inactive	ErbB2 inactive	ErbB3 Inactive	ErbB4 inactive
aC-helix A-loop bonds						
--	--	--	--	--	--	E739,R841
E734,K851	E766,K883	E739,K856	--	--	--	--
D737,K836	D769,R868	--	--	--	--	D742,R841
E738,F832	--	E743,F837	--	--	--	--
--	--	--	E738,K836	--	--	E743,R841
aC-helix C-loop bonds						
--	--	--	--	--	--	E743,R817
A-loop C-loop bonds						
--	--	--	--	G865,V842	--	--
--	--	--	--	--	--	G838,R817
L834,R812	L866,R844	L839,R817	--	--	--	--
--	--	--	L834,D813	--	--	--
K836,V810	R868,V842	R841,V815	--	--	--	--
--	--	--	--	--	D838,R814	--
E848,R812	--	--	--	--	--	--
--	--	--	K851,R812	--	--	--
C-loop C-loop bonds						
--	H843,D845	--	--	--	H813,N815	--
--	--	--	R812,D813	R844,D845	--	--
D813,R817	D845,R849	D818,R822	--	--	--	--
D813,N818	--	--	--	--	N815,N820	--
A815,N818	A847,N850	A820,N823	A815,N818	A847,N850	A817,N820	--
--	A848,V851	--	--	--	--	--

Table 5.1a: Hydrogen bonding analysis of the ErbB3 MD trajectory and comparison to the active and inactive conformations of EGFR, ErbB2 and ErbB4. ErbB3 lacks many of the conserved bonds present in the active EGFR, ErbB2 and ErbB4 structures. The characteristic active bonds are highlighted in green and the sequestering bonds are highlighted in yellow. See also Table 5.1b

EGFR active	ErbB2 active	ErbB4 active	EGFR inactive	ErbB2 inactive	ErbB3 Inactive	ErbB4 inactive
A-loop bonds						
--	--	D836,K726	--	D863,K753	D833,K723	D836,K726
--	--	D836,T835	--	--	--	--
L838,R808	L870,R840	L843,R813	--	--	--	--
--	D871,R840	--	--	--	--	--
A840,G672	--	--	--	--	--	--
--	--	--	--	D873,R897	--	--
--	--	--	--	--	D844,K853	--
--	--	K848,T873	--	--	--	--
K843,D932	--	K848,D937	--	--	--	--
--	E876,R898	--	--	--	--	--
--	--	E849,K871	--	--	--	--
Y845,Y867	--	Y850,F872	--	--	--	--
--	--	--	H846,R865	--	--	--
--	--	--	--	--	K845,E851	--
--	--	A852,R870	--	--	--	--
--	D880,R897	D853,R870	E848,R865	D880,R897	--	D853,R870
--	--	--	--	--	--	G855,E730
--	--	--	--	K883,E757	--	--
--	--	--	--	--	--	K856,E844
aC-helix bonds						
--	A763,S760	--	--	--	--	--
--	E766,R756	--	--	--	--	--
E738,K721	E770,K753	E743,K726	--	--	--	--
--	--	--	M742,L753	M774,L785	--	M747,L758
A743,L679	--	A748,Q684	--	--	--	--
--	--	--	--	--	--	A748,R757

Table 5.1b: Hydrogen bonding analysis of the ErbB3 MD trajectory and comparison to the active and inactive conformations of EGFR, ErbB2 and ErbB4. ErbB3 lacks many of the conserved bonds present in the active EGFR, ErbB2 and ErbB4 structures. The characteristic active bonds are highlighted in green and the sequestering bonds are highlighted in yellow.

sequence: E738 (EGFR numbering), which coordinates the α and β phosphates of ATP by forming a salt bridge with K721 in the α C helix, is mutated to a His in ErbB3. In addition, E734, which couples the A-loop and α C helix by bonding to K851, is an Ala in ErbB3. The few bonds that are present in ErbB3 help to maintain its distinctive active site: the D838-R814 bond, which bridges the 3_{10} helix and the C-loop in ErbB3, does not appear in the other ErbB kinases. ErbB3's unique hydrogen bonding network (or lack thereof) suggests that the molecular mechanism of catalysis in ErbB3 is distinct from that operating in EGFR, ErbB2 and ErbB4.

5.2.3) Hydrophobic analysis of the ErbB kinases identifies a perturbation-sensitive region in ErbB3

In the active conformation, EGFR, ErbB2 and ErbB4 rely on an extensive network of hydrophilic interactions to maintain their active sites in a catalytically competent state, whereas the inactive conformations exhibit characteristic hydrophobic interactions, a feature that helps to differentiate the inactive and active states. In ErbB-mediated signaling, the introduction of the ErbB dimer interface alters the hydrophobicity of the region and, in conjunction with allosteric effects, perturbs the conformational sampling space of the protein to induce a shift toward the active conformation [297, 353, 354].

To quantify these hydrophobic interactions, we computed the solvent accessible surface area (SASA) and water density fluctuations (see Methods, section 5.4) for the ErbB3 MD

trajectory and compared the results to the EGFR, ErbB2 and ErbB4 trajectories previously simulated by us [354]. In particular, we focused on several highly conserved hydrophobic regions which are considered to be important for defining the catalytic state of the ErbB kinases [30, 31]: the catalytic spine (C-spine), regulatory spine (R-spine), hydrophobic core, and β 3- α C loop. Figure 5.4 displays the correlation between the mean SASA and the normalized water density fluctuations for each hydrophobic region in the inactive and active conformations of each ErbB kinase. The SASA analysis is a reliable measure of hydrophobicity for smooth hydrophobic interfaces but does not always correlate perfectly with free energies of solvation of hydrophobic groups near irregular hydrophobic surfaces, which are often found in proteins. The normalized water density fluctuation calculation, which quantifies the hydrophobicity of heterogeneous surfaces, provides an alternative to the SASA analysis and proposes that increased normalized water density fluctuations signify a more hydrophobic region, where a normalized value of 1 indicates a neutral surface [320, 321]. Thus the correlation between the SASA and water density fluctuations produces additional insights into region hydrophobicity that are not revealed by either analysis independently. Specifically, we have designated four ‘quadrants’ in each correlation subplot (Fig. 5.4A): quadrant II represents a hydrophilically favorable region, quadrant IV defines a hydrophobically favorable region, and quadrant I, which is characterized by high SASA (more hydrophilic) and large water density fluctuations, represents a fragile or perturbation-sensitive region. The quadrants facilitate the identification of key patterns that may emerge from a comparison of the hydrophobic interactions in the ErbB systems.

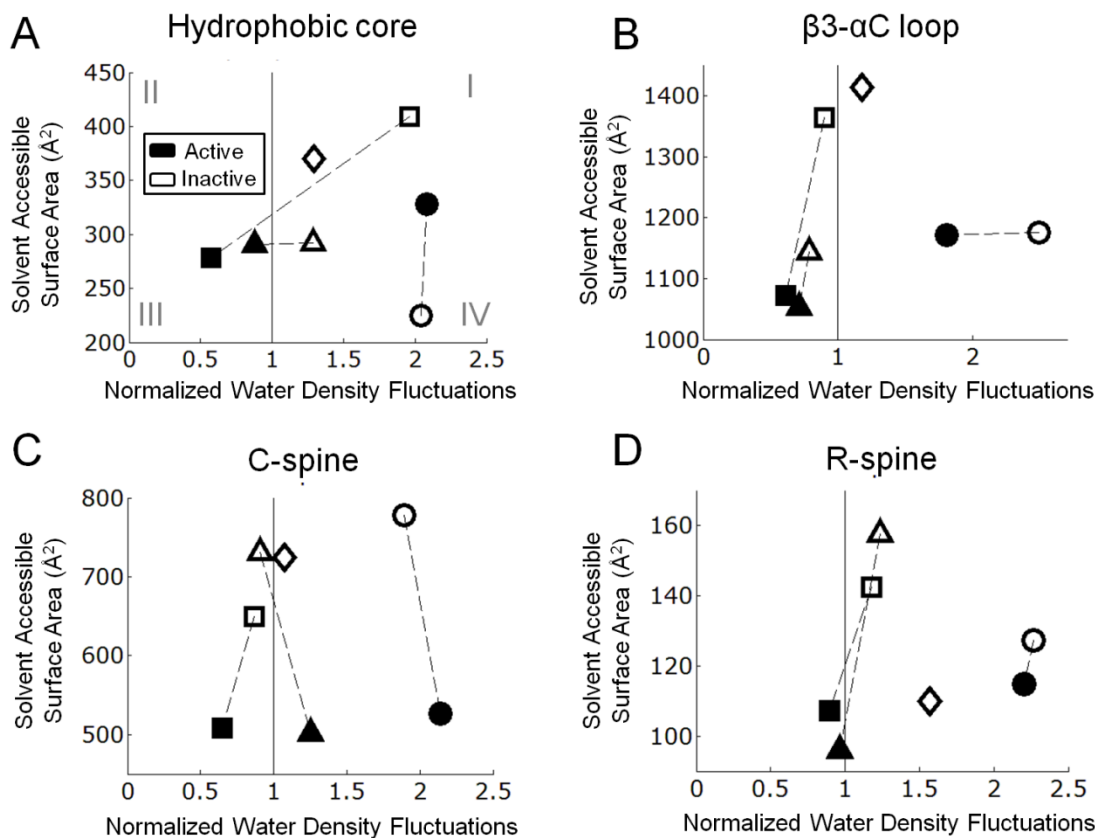


Figure 5.4: Correlation between the solvent accessible surface area (SASA) and water density fluctuations in the key hydrophobic regions of the ErbB kinases: (A) hydrophobic core, (B) $\beta 3$ - αC loop, (C) C-spine and (D) R-spine. Dark and light markers represent the active and inactive ErbB systems, respectively. The ErbB kinase systems are represented as follows: EGFR (*square*), ErbB2 (*circle*), ErbB3 (*diamond*), and ErbB4 (*triangle*). Quadrant I represents a perturbation-sensitive region, quadrant II defines a hydrophilically favorable region, and quadrant IV represents a hydrophobically favorable region.

The ErbB2 system, in general, lies within the hydrophobically favorable region (quadrant IV), which may be rationalized in terms of ErbB2's unique ability to bind the molecular

chaperone Hsp90 through a hydrophobic interface, a point that is elaborated in our previous studies [352, 354]. In the hydrophobic core, which includes segments of the α C helix and the A-loop and helps to maintain the ErbB kinases in the inactive conformation [30, 31], ErbB3 demonstrates characteristic hydrophobic interactions that are similar to the inactive EGFR and ErbB4 systems. Fig. 5.4A illustrates that ErbB3, as well as the inactive EGFR and ErbB4 systems, lies within the ‘perturbation-sensitive’ quadrant, indicating that the hydrophobic core is a potential fragile point of the protein, in which single point mutants would be expected to disrupt the local interactions. Indeed, it has been reported that mutations in the hydrophobic core of EGFR and ErbB4 result in activation of the kinase, due to destabilization of the inactive state [337]. By contrast, an analogous mutation in ErbB3 abolishes ATP-binding and phosphorylation activity [337], indicating that hydrophobic interactions in the core promote ErbB3 activity, rather than maintain an autoinhibited state as they do in EGFR and ErbB4. Furthermore, the β 3- α C loop, which is uniquely extended in ErbB3 and lies proximal to the hydrophobic core, represents a perturbation-sensitive node in ErbB3 (Fig. 5.4B). Thus mutations in this region would also be expected to inactivate the ErbB3 kinase. Our results support the conclusion that ErbB3 critically depends on a coordinated set of hydrophobic interactions for its weak yet robust activity from the inactive-like conformation, whereas EGFR and ErbB4 employ similar interactions to maintain their autoinhibited status from the inactive-like conformation. We propose that these hydrophobic contacts, including the interactions in the β 3- α C loop, account for the tightly coordinated sub-domain motions observed in our PCA analysis of the ErbB3 system, in contrast to the hydrogen bond-

mediated mechanism underlying the coordinated sub-domain motion in the active ErbB systems.

The hydrophobicity of the C-spine and R-spine, whose function is to coordinate the motions of the N- and C-lobes of the EGFR and ErbB4 kinases in the active conformation [30, 31], was also quantified (Fig. 5.4C, D). In the C-spine and R-spine hydrophobicity plots, the inactive and active conformations of EGFR and ErbB4 are clearly delineated with respect to the SASA, with the active systems exhibiting a preferential hydrophobic stabilization (low SASA). The SASA of the ErbB3 C-spine falls within range of the inactive EGFR and ErbB4 systems, reflecting that, despite its weak activity, there is no corresponding ‘fully-active’ state for ErbB3 as for the other ErbB kinases. This inability to ‘fully’ activate can be attributed to the lack of the crucial hydrogen bonding network identified earlier, which is required to stabilize the active-like kinase conformation. The SASA of the ErbB3 R-spine deviates from the values for the inactive EGFR and ErbB4 systems, and instead demonstrates low SASA (high hydrophobicity). This result can be rationalized by the increased hydrophobicity of the R-spine, which includes segments of the truncated α C helix in ErbB3, hence positioning the ErbB3 system in the hydrophobically-favorable quadrant of the R-spine hydrophobicity plot (Fig. 5.4D).

5.3) Discussion

In this work we applied a molecular modeling approach to investigate the non-canonical catalytic mechanism employed by the ErbB kinase ErbB3, a RTK formerly categorized as an inactive pseudokinase, and the physiological relevance of this activity to mechanisms of drug resistance in an ErbB-driven tumor cell *in silico*. At the molecular level, our results suggest that the prearrangement of the catalytic site in ErbB3 is achieved through a unique network of hydrophobic interactions, in contrast to the extensive set of hydrogen bonds present in the active sites of the other ErbB kinases. For the implications of a weakly active ErbB3 kinase at the cellular level, please see section 7.3.

At the atomic level, our molecular dynamics simulations and PCA analysis of the ErbB3 kinase crystal structure revealed that the ErbB3 system exhibits tightly coordinated fluctuations of its catalytic sub-domains, which may facilitate the alignment of key catalytic residues involved in phosphoryl transfer. Despite sharing a pattern of concerted sub-domain motion with its fully-active ErbB family members, the ErbB3 system appeared to achieve this state through a unique mechanism, namely, a tightly coordinated set of hydrophobic contacts. By contrast, an extensive network of hydrophilic interactions is employed by EGFR, ErbB2 and ErbB4. The hydrogen bonding analysis and hydrophobicity calculations demonstrated that the conserved hydrophilic contacts present in the active sites of EGFR, ErbB2 and ErbB4 are replaced in ErbB3 by a unique hydrophobic interface formed by the extended β 3- α C loop, the molten α C helix and the 3_{10} helix in the A-loop [355]. Thus our atomic-level simulations suggest that the ErbB

kinases may have evolved two different modes of achieving the shared goal of prearranging the catalytic machinery for phosphoryl transfer. For the implications of the weakly active ErbB3 kinase domains upon the cell signaling networks, especially the Akt pathway, and its potential role as a TKI resistance mechanism, see section 7.3.

5.4) Methods

Molecular dynamics (MD) simulations: The HER3 kinase crystal structure solved in our previous study (PDB code 3LMG) [356] was prepared for molecular dynamics simulation according to the procedure outlined in [348, 352]. Briefly, the structure was explicitly solvated using the TIP3P model for water [357] and with the buffering distance set to 15 Å. Sodium (Na^+) and chloride (Cl^-) ions were added to achieve net electroneutrality of the system and an ionic strength of 75 mM. The ions were positioned at points of electrostatic extrema using a Debye-Huckel potential calculated within the program Solvate 1.0 [358]. All simulations were performed with NAMD [359] using CHARMM27 force-field parameters. To prepare the system for MD simulation, the solvated structure was energy-minimized using a conjugate gradient algorithm to remove unfavorable contacts. The system was then heated to 300 K using the temperature reassignment method in NAMD. Constant pressure and temperature (NPT) simulations were performed at 300 K and 1 atm to equilibrate the volume of the solvation box. Temperature and pressure were maintained using a Langevin piston coupling algorithm [360]. Following the NPT simulations, constant volume and temperature (NVT)

simulations were performed in NAMD. Finally, a 10 ns production run was completed using the same parameters as in the NVT simulations.

Hydrogen bonding analysis: CHARMM was used to analyze the hydrogen bonds present in the 10 ns trajectory for each system according to the procedure outlined in our previous studies [348, 352]. Hydrogen bonds were defined by a bond length cutoff of 3.4 Å and an angle cutoff of 150°. Bonds that fulfilled these criteria and were present in at least 60% of the trajectory were tabulated in CHARMM. Salt bridges were defined as hydrogen bonds occurring between an acidic and a basic residue and satisfying a bond length cutoff of 1.6 Å. All hydrogen bonds and salt bridges were also visualized in VMD [361] for the duration of the 10 ns simulation.

Principal component analysis: A principal component analysis (PCA) was applied to the 10 ns trajectory to identify the main eigenvectors ($3N$ directions) along which the majority of the complex motion is defined. The calculation is based on the diagonalization of the variance-covariance matrix of the atomic fluctuations along each MD trajectory to yield the set of eigenvectors (PCs) and associated eigenvalues. The eigenvectors represent the independent modes of atomic motion, and the eigenvalues reflect the contribution of the corresponding eigenvectors to the global fluctuation of the protein. PCA computes the covariance matrix as

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$$

where $(i,j=1,\dots,3N)$, and N is the total number of atoms with positions given by Cartesian coordinates x . The resulting matrix is diagonalized to compute the $3N$ independent (uncorrelated) eigenvectors, $\{\xi_i\}$, and the eigenvalues, $\{\lambda_i\}$, sorted in descending order. The angle brackets denote the time average over the entire trajectory. PCA was performed on the entire protein backbone (*i.e.* all C_α atoms), and also an active site region that comprises all catalytic sub-domains, including the A-, C-, and N-loops and the αC helix. The program CARMA [362] was used to project the atomic fluctuations along the MD trajectory.

Solvent accessible surface area (SASA) and water density fluctuation analysis: SASA values were calculated in VMD [361] using the measure SASA module with a probe radius 1.4 Å larger than the van der Waals radius. The SASA was calculated for each step in the trajectory, from which the mean and standard deviation were computed. As an alternative measure of hydrophobicity in heterogeneous environments, following the procedure established in [320, 321], normalized water density fluctuations were computed by recording the ratio of $\sigma_N/\langle N \rangle$, where σ_N and $\langle N \rangle$ are the standard deviation and mean associated with the number of water molecules within 5 Å of a specified hydrophobic sub-region in the HER3 kinase. Although results are presented for a cutoff of 5 Å, other cutoffs ranging from 3-15 Å were investigated and similar trends in $\sigma_N/\langle N \rangle$ were recorded.

Chapter 6.) Future Work

RTKs are pieces of molecular machinery important in normal cell function and the dysregulation of which leads to disease states, particularly cancer. Understanding the activation mechanisms allows a better regulation of the RTKs and in turn offers better therapeutic options. Computational methods provide a unique perspective in understanding activation at the atomic level by allowing us to construct, visualize and analyze multiple proteins as they move in a physiologically relevant environment. Also by constructing the proteins in different conformations as well as using directed simulation techniques it is possible to observe the proteins as they function, offering a wealth of information into their molecular mechanisms. In this thesis we have examined how hydrophilic and hydrophobic interaction networks help the activation mechanisms of the ErbB family kinases using molecular dynamics simulation techniques. Still, there is much to learn about the ErbB kinases, not to mention the rest of the RTKs. Future work presented here is either wide-spread application of the techniques used here across the entire kinase domain and full-length RTK, improving the techniques used here to allow easier translation of the results to different scales or methodological improvements to capture rare events.

We have presented a method to examine hydrophilic/hydrophobic interaction networks and shown their importance in kinase activation by comparing the results across homologous proteins. To reduce the computational burden and ensure validated results,

we focused on subregions shown previously to be biologically relevant. The active kinase domains are highly conserved structurally, so by expanding the scope of the analysis to the entire intracellular kinase domain, it would be possible to highlight hydrophilically and hydrophobically important residues and regions not explicitly observed through experiments. The hydrophobic interaction analysis can be automated especially with the normalized water density fluctuations in conjunction with the Solvent Accessible Surface Area to provide regions that are hydrophobically sensitive, giving potential mutational hotspots. Similarly, the hydrophilic interaction analysis with a wider focus would provide a larger list of bonds and by comparing the lists across the ErbB family a characteristic network of bonds through the entire kinase can be highlighted.

Refinement of the interaction network analysis gives a better picture of the atomic level characteristics of each protein. The hydrophilic interaction analysis can be refined beyond mean survival time and mean bond length to observe the bond characteristics fitted to a statistical distribution. The hypothesis here is the characteristic bonds defining a conformation would be stronger and therefore have a shifted distribution in comparison to those that are transient. Since we have already identified sets of bonds important in each conformation they would serve as a control for the refinement. Characterizing the hydrophobic stability of a system is difficult and is its own field of research within computational chemistry. The normalized water density fluctuation analysis is one current technique that correlates with hydration free energy. To provide a better analysis using standard molecular dynamics forcefields, collaborations with groups like the Garde

group [320, 321] would be necessary so as to not reinvent the wheel. However, once the method is refined, it would facilitate identifying important subregions significantly easier.

Even though we have employed modern molecular dynamics simulation techniques, there are still time and length scale limitations in the microseconds and micrometers regime respectively. Since most of the biological processes exist on milliseconds to minutes to even longer, simulating for hundreds of years using the same technology would not breach even the millisecond threshold. However the techniques are constantly improving and there are two techniques that would offer enormous improvement into both length and timescales: coarse graining and specialized hardware. Coarse graining groups atoms together into a functional group and simulates them as one single “atom”. By grouping atoms, the length scale is enhanced and the smallest vibrational frequency (hydrogen vibration) is removed so the minimum time step is increased by orders of magnitude. The currently accepted coarse grain force-field is MARTINI [363] offering simulation times into microseconds as the norm, with improvements to come. There are also sets of specialized computer hardware that is designed and tuned to provide a similar speed up for molecular dynamics simulation while maintaining the same accuracy, particularly MDGRAPE-3 [158, 159] and ANTON [160, 161]. Extending the simulations would give access to a wider range of conformational space mapped out and give better ensemble data as well as allowing us to explicitly capture and analyze rarer events.

An alternate technique to sample rare events is biased or directed sampling simulations, which apply a potential to the protein in simulation to transition it to the desired conformation, usually down the activation pathway (see section 2.2.5). Some of the techniques reviewed are more complex to explore pathways if the final structure is not known. However, since there are beginning and ending crystal structures for the EGFR kinase domain, free energy simulations are a better option (see section 2.2.3). Umbrella sampling simulations not only transition the protein but also measure the free energy change necessary for the conformation change. Here we present preliminary results on Targeted Molecular Dynamics (TMD) simulations performed on monomer and dimer trajectories of the EGFR kinase domain mapping out the activation pathway in preparation for umbrella sampling simulations.

6.1) The Asymmetric Dimer Interface achieves Activation through Allosteric Means, Hydrophilic Interactions and Hydrophobic Shielding

The EGFR kinase is a member of the EGFR/ErbB/HER family of kinases which contain four homologous Receptor Tyrosine Kinases (RTKs) that are important regulatory elements in key signaling pathways. RTKs are transmembrane proteins important in translating extracellular signals (ligands) into intracellular nuclear responses leading to cell proliferation, differentiation, migration and the cell cycle (reviewed in [1, 2, 18]). They consist of a large ligand-binding extracellular domain, a single transmembrane α -helix, a juxtamembrane domain, an intracellular tyrosine kinase domain and a C-terminal

tail that contains several phosphorylatable tyrosines [1, 2]. Considering the cell signaling pathways it regulates, understanding how the EGFR kinase domain activates is crucial. Clinically identified activating mutations in the EGFR kinase domain found in Non-Small Cell Lung Cancer (NSCLC) patients have been shown increase the basal activity of the EGFR kinase domain [26, 27, 48] and respond to the EGFR RTK inhibitor gefitinib with efficacy rivaling that of chemotherapy [130, 131].

The intracellular kinase domain of RTKs transfers the γ -phosphate of ATP to tyrosines on both the RTK itself as well as other target substrates [1]. The active conformations of kinases are all remarkably similar, with two separate lobes, the β strand-rich N-lobe plus the α -helix rich C-lobe and the active site situated in the cleft between N-lobe and C-lobe. There are also several conserved subregions in the RTK kinase domain thought to be important in regulation of kinase activity: the catalytic loop (C-loop), the activation loop (A-loop), the glycine-rich nucleotide binding loop (P-loop), and the α C-helix. The catalytic loop contains the conserved HRD motif, with the aspartate directly participating in the phosphoryl transfer. The A-loop defined at the start by the DFG motif and the α C-helix modulate the activity of the kinase domain by regulating accessibility of the active site to binding and coordinating both ATP and the substrate tyrosine. The A-loop consists of approximately 20 amino acids and in the EGFR kinase contains one phosphorylatable tyrosine (Y845 in EGFR, note: there are two numbering schemes for the EGFR where Y845 is equivalent to Y869). The α C-helix and P-loop must be positioned correctly to coordinate the ATP and the substrate tyrosine for effective phosphoryl transfer.

Beyond the structural similarities, recent studies upon the spatial location of non-sequential residues in the kinases have revealed highly conserved hydrophobic “spines” that are thought to help coordinate the motions of key domains during activity [30, 31]. The regulatory spine (R-spine: M742, L753, H811, F832 in EGFR) helps maintain the motion of the N- and C-lobes of the kinase [30]. The catalytic spine (C-spine: V702, A719, L774, V819, L820, V821, T879, L883 in EGFR) helps coordinate the adenine ring of ATP [31]. In the inactive state there is a small hydrophobic ‘core’ formed between the α C-helix and the A-loop, which when disrupted by single point mutations have been shown to activate EGFR [25-29].

There have been activation pathway studies done upon the Abl kinase [161] and Src family kinase Hck [323, 325], but our previous studies on the ErbB family show a fundamentally different pathway of activation [364]. The Abl kinase inactive conformation compares well to the active conformation, with a well formed active site and many of the conserved kinase bonds in place [365]. However the sidechain of the coordinating aspartate (D381, the D in the conserved DFG motif) is in an “out” position pointing away from the active site. In the active conformation D381 has flipped into an “in” position pointing towards the N-lobe, and forms a salt bridge. In the transition between the two states the α C-helix rotates away from the active site, creating a hydrophobic pocket at the base of the N-lobe and F382 slips into this pocket causing the flip of D381. The final step is for the α C-helix to rotate back into its original position.

The Src family kinase Hck is inactivated by autoinhibitory interactions in the unphosphorylated A-loop causing steric hindrances to the active site. In addition, there are SH3 and SH2 domain interactions in the Hck kinase fairly distal to the active site that result in outwards rotation of the α C-helix and disrupt some conserved active kinase bonds (reviewed in [366]). Phosphorylation of these regions release their autoinhibitory interactions allowing full activation of the kinase [233], where the A-loop moves towards the active state before the α C-helix rotates into its active state [323].

In terms of inactive states, the EGFR kinase domain is more similar to Hck kinase or to the intermediate state of the Abl kinase, where the α C-helix is rotated out in comparison to the active state, preventing key bonds from forming. The activation mechanism though, is not from phosphorylation; the ErbB family is not dependent on phosphorylation on the A-loop or any part of the intracellular kinase domain for activation. Instead, an asymmetric dimer interface similar to that seen in cyclin dependent kinases serves as the allosteric activation mechanism [41]. In the ErbB family, the dimer interface is asymmetric; the C-lobe of one kinase, the “activator”, contacts the N-lobe of the other kinase, the “receiver,” with the asymmetric dimer contacts causing a conformational change towards the active state through allosteric methods. Our previous simulations of the asymmetric dimer have shown that introduction of the interface causes a rearrangement of the α C-helix towards the active state before any motion of the A-loop [364]. Here we present an analysis of the hydrophilic and hydrophobic interaction networks of catalytically important subdomains as the EGFR kinase progresses through

the activation pathway in both the monomer and dimer states to elucidate the precise mechanisms of the allosteric, activating asymmetric dimer interface.

6.2) Preliminary Results

6.2.1) RMSD Analysis Reveals the Activation Pathway of the Dimer is Distinct from the Monomer

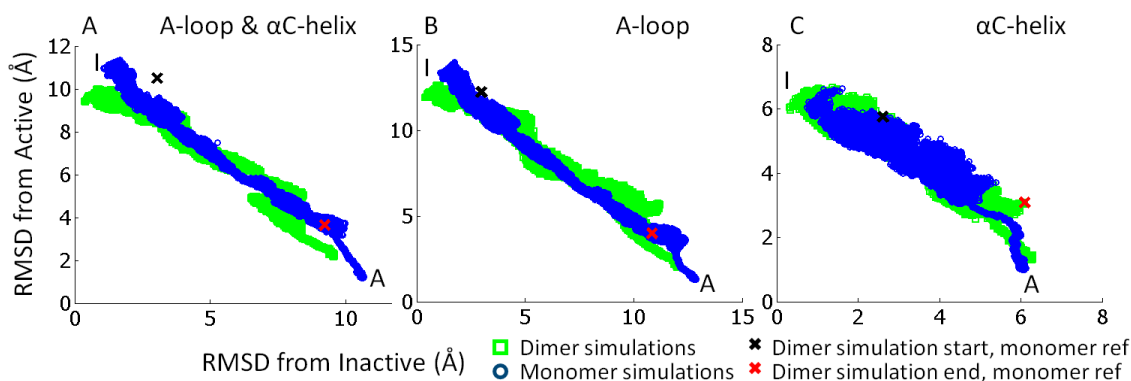


Figure 6.1: RMSD plots tracking the progression of the targeted molecular dynamics of the dimer (green squares) and the monomer (blue circles) simulation down the activation pathway for the (A) combined A-loop and α C-helix as well as the (B) A-loop alone and (C) the α C-helix alone.

The RMSD calculation of the TMD trajectory in both the monomer and dimer shows a complete capture of the pathway between the inactive and active conformations (Figure 6.1). The predominant motions involved in the reorganization of the active site involved the α C-helix and A-loop without much motion in the P-loop of C-loop, so we focused upon the motions of the α C-helix and A-loop as the reaction coordinates. In the

monomer TMD dynamics, the simulations run through a fairly linear pathway from the inactive to the active states in the A-loop and α C-helix, with the A-loop transitioning from a closed conformation over the C-loop to laying open within the N-lobe and C-lobe as well as the α C-helix rotating in and extending the helical conformation. There is a small section at the end of the simulation for the α C-helix (Figure 6.1C), starting at 6Å from the inactive and 2Å from the active, that deviates from this path, which is associated with the extension of the α C-helix. The dimer TMD simulations also follow a fairly linear path mapping from inactive to active and seem to align remarkably well with monomer simulations, but without the small non-linearity at the end of the simulation seen in the monomer for the α C-helix. However the reference structures for the dimer TMD are the equilibrated dimer conformations. Marked on Figure 6.1 are black X's as the starting point of the dimer simulations using the equilibrated monomer structures and red X's as the end point of the dimer simulations using the same structures. Although the pathways in the monomer and dimer are very similar, there is shift in the pathway because of the introduction of the asymmetric dimer interface, particularly in the α C-helix (Figure 6.1C).

6.2.2) Interkinase Salt Bridges Alter the Conformation of the α C-helix and Promote Partial Extension

We have previously identified several bonds that are conserved across the canonical kinase members of the ErbB family in the active state: two salt bridges: E734-K851

(E734-K836 in the dimer) and E738-K721, three H-bonds: L834-R812, K836-V810, and L838-R808 and the bond D813-R817 [298, 364]. The D813-R817 bond helps position the aspartate side chain in proper orientation for phosphotransfer. The E738-K721 salt bridge is highly conserved across all active kinases and helps coordinate the phosphates in ATP. The E734-K851 salt bridge connects the A-loop and the α C-helix, coordinating the movements of these two sub-domains and dampening larger fluctuations. Similarly, the three conserved H-bonds link the A-loop and the C-loop, coupling the motions of these two loops. These can be regarded as “fastening” H-bonds that maintain the N-terminal side of the A-loop open in its active state. The inactive bond network is less extensive and not well conserved, but in EGFR the E738-K836 salt bridge sequesters residues in the key E738-K721 salt bridge thereby preventing activation by keeping the salt bridge from forming.

The transition between the E738-K836 salt bridge and the formation of the E738-K721 salt bridge is a crucial one for activity of the kinase. However previous TMD studies of EGFR have seen the two events happen almost simultaneously [311, 364]. To better resolve the pathway and the interaction networks of the monomer and dimer simulations, we broke up each TMD pathway simulation into smaller windows of 10 ns a piece (5 windows for the monomer and 6 windows for the dimer) and performed the interaction network analysis on each window independently. The hydrophilic bond networks between the monomer and the dimer simulations are fairly similar (Table 6.1a&b and Appendix B).

EGFR Dimer Inactive	Window 1	Window 2	Window 3	Window 4	Window 5	Window 6	EGFR Dimer Active
aC-helix A-loop bond							
--	--	--	--	--	--	E734,K836	E734,K836
--	--	--	--	--	--	--	D737,K836
E738,F832	E738,F832	E738,F832	E738,F832	E738,F832	E738,F832	E738,F832	--
E738,G833	E738,G833	E738,G833	E738,G833	--	E738,G833	--	--
E738,K836	E738,K836	E738,K836	E738,K836	E738,K836	E738,K836	--	--
aC-helix bonds							
--	--	--	--	--	--	--	N732,V762
--	--	--	--	--	--	E738,K721	E738,K721
M742,L753	M742,L753	M742,L753	M742,L753	--	--	--	--
--	A743,R752	A743,R752	--	--	--	--	--
A-loop C-loop bonds							
--	T830,N818	T830,N818	T830,N818	T830,N818	T830,N818	--	T830,N818
D831,N818	D831,N818	D831,N818	D831,N818	D831,N818	D831,N818	--	--
--	--	--	--	--	--	--	L834,R812
--	L834,D813	L834,D813	--	L834,D813	L834,D813	L834,D813	--
--	--	--	--	--	--	K836,V810	K836,V810
--	--	--	--	--	--	--	--
--	--	--	--	--	--	--	--
A-loop bonds							
D831,K721	D831,K721	D831,K721	D831,K721	--	--	--	D831,K721
--	--	--	--	--	--	--	L838,R808

Table 6.1a: Persistent H-bonds and salt bridges in the EGFR kinase dimer system as it progresses from inactive to active with the salt bridges in bold, and bonds consistently present but not considered persistent in grey. Characteristic inactive bonds are boxed in orange while characteristic active bonds are boxed in green. The common bonds between the monomer and dimer break and form in a similar timeframe; see Table 6.1b.

In the monomer and dimer the sequestering salt bridge E738-K836 breaks in a similar position in the pathway, the E738-K836 is weakened severely in window 4 in the monomer and broken between window 4 and window 5, while in the dimer the salt bridge is broken between window 5 and window 6. After the E738-K836 salt bridge is broken we see the beginning of formation of the characteristic active kinase salt bridge E738-

K721 weakly in the penultimate window in both the monomer and the dimer which finalizes once the kinase completes its transition to active. All of the fastening bonds seen in the ErbB active kinases do not appear in the monomer or the dimer simulations until the active conformation is fully achieved, except for K836-V810 that appears in window 6 in the dimer.

EGFR Monomer Inactive	Window 1	Window 2	Window 3	Window 4	Window 5	EGFR Monomer Active
aC-helix A-loop bonds						
--	--	--	--	--	E734,K851	E734,K851
--	--	--	--	--	D737,K836	D737,K836
--	--	--	--	--	E738,F832	E738,F832
--	--	--	--	--	E738,G833	--
E738,K836	E738,K836	E738,K836	E738,K836	E738,K836	--	--
aC-helix bonds						
--	--	--	--	--	--	--
--	--	--	--	--	E738,K721	E738,K721
M742,L753	M742,L753	M742,L753	M742,L753	M742,L753	M742,L753	--
--	A743,R752	A743,R752	A743,R752	A743,R752	--	A743,L679
A-loop C-loop bonds						
T830,N818	T830,N818	T830,N818	T830,N818	T830,N818	--	T830,N818
--	--	--	D831,N818	D831,N818	D831,N818	--
--	--	--	--	--	--	L834,R812
L834,D813	L834,D813	L834,D813	L834,D813	--	--	--
--	--	--	--	--	--	K836,V810
--	--	--	--	--	--	E848,R812
K851,R812	--	--	K851,R812	K851,R812	--	--
A-loop bonds						
--	D831,K721	D831,K721	--	--	--	--
--	--	--	--	--	--	L838,R808

Table 6.1b: Persistent H-bonds and salt bridges in the EGFR kinase monomer system as it progresses from inactive to active with the salt bridges in bold and bonds consistently present but not considered persistent in grey. Characteristic inactive bonds are boxed in blue while characteristic active bonds are boxed in red. The common bonds between the monomer and dimer break and form in a similar timeframe.

EGFR Dimer Inactive	Window 1	Window 2	Window 3	Window 4	Window 5	Window 6	EGFR Dimer Active
A-loop Interdimer bonds							
--	--	--	--	E841,R949	E841,R949	--	--
aC-helix interdimer bond							
P729,K925	P729,K925	--	--	--	--	--	--
K730,E939	K730,E939	K730,E939	K730,E939	K730,E939	--	--	--
K733,D918	K733,D918	--	K733,D918	K733,D918	--	--	--
--	--	--	--	--	K733,E943	--	--
--	--	--	--	--	--	--	Y740,D918

Table 6.2: Persistent H-bonds and salt bridges between kinase domains in the EGFR kinase dimer systems as it progresses from inactive to active. The salt bridges are in bold and bonds consistently present but not considered persistent in grey. Important salt bridges are boxed in purple.

The main differences in the hydrophilic bond networks between the monomer and dimer simulations are those formed between the two kinase molecules which can only be present in the dimer simulations and not in the monomer simulations (Table 6.2). In the inactive dimer, we see the formation of three inter dimer bonds, the P729-K925 H-bond plus the K730-E939 and the K733-D918 salt bridges. All of these residues are at the tip of the α C-helix and help alter the conformation of the helix in the dimer into a partial extension as the simulation progresses (Figure 6.2C). These bonds are all broken by the fifth window and only one H-bond Y740-D918 is present in the active conformation, which is near the end of the α C-helix. Between the inactive and active conformations (window 4 and window 5) we see the formation of two transient bonds which are not present in either the inactive and active conformations: the salt bridge E841-R949 and the H-bond K733-E943. The E841 residue is in the middle of the A-loop and forms a salt bridge E841-K843 which makes a small curl within the A-loop in the inactive kinase simulations (Appendix B) of both the monomer and the dimer systems and the K733

bond is at the tip of the α C-helix. There is overlap in the salt bridges in the α C-helix and the A-loop opening salt bridge in window 4, which may indicate a potential intermediate state.

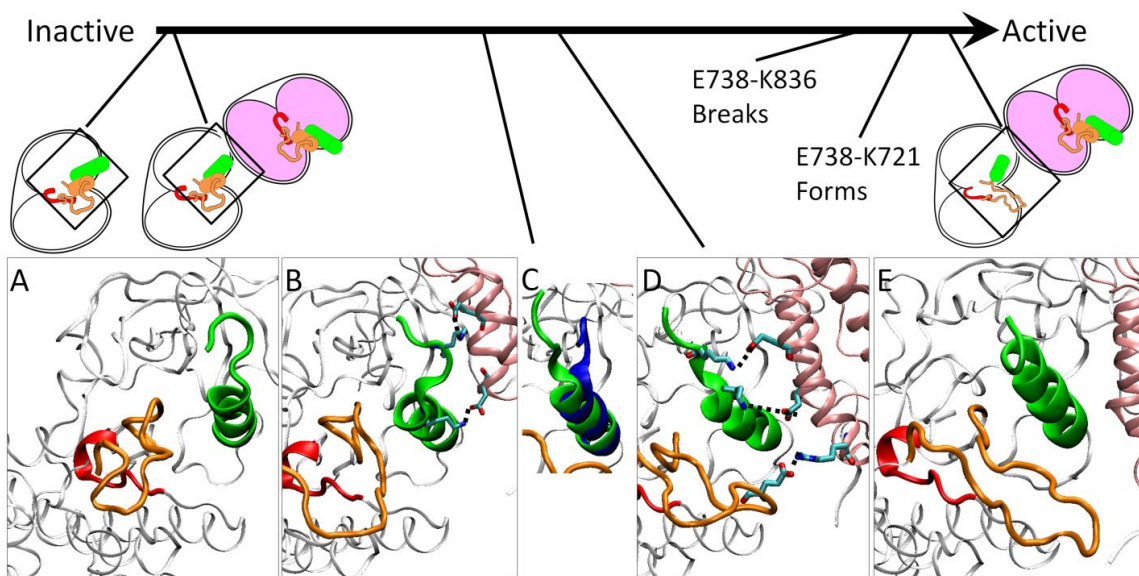


Figure 6.2: Order of events as they occur for the dimer simulations with enhanced views of the active site; the α C-helix is in green, A-loop is in orange, and the C-loop is in red. (A) The inactive monomer alone (B) introduction of the asymmetric dimer and formation of the K730-E939 and K733-E939 salt bridges (C) partial extension of the α C-helix in dimer (green) not present in monomer (blue) (D) formation of the E841-R949 salt bridge keeping the A-loop partially open and presents a potential “active” intermediate state (E) full extension of the α C-helix.

6.2.3) The Dimer Interface Shields the “Receiver” Kinase from Hydrophobic Effects during the Transition to Active

To investigate the effect of hydrophobic interactions as the EGFR kinase transitions between the monomer and dimer, we analyzed the water density fluctuations as well as the solvent accessible surface area (SASA). The hydrophobicity of a sub-region is a non-additive quantity, where the primary structure and the surrounding environment both affect the overall hydrophobicity. Recent studies by Garde et al. [320, 321] have shown quantifying the hydrophobicity of heterogeneous surfaces using water density fluctuations correlates with hydration free energy: increased water density fluctuations are signature of a more hydrophobic surface. We have normalized the water density fluctuations so that a neutral surface is at a value of one. Plotting the normalized water density fluctuations versus SASA allows us to divide the plot into four quadrants labeled on Figure 6.3A. Quadrant I represents a hydrophilically favorable region with low hydrophobicity and high SASA, quadrant IV represents a hydrophobically favorable region with high hydrophobicity and low SASA and quadrant II represents a fragile or perturbation sensitive region with high hydrophobicity but also with a high SASA.

We analyzed the effects of hydrophobic interactions of relevant hydrophobic sub-regions, namely the dimer interface (Figure 6.3A) the hydrophobic core (Figure 6.3B) the R-spine (Figure 6.3C) and the C-spine (Figure 6.3D) for each simulation: dimer (green border) and monomer (black border) progress from the inactive (white filled squares) to active (black filled squares); each window is connected sequentially with dotted lines on Figure 6.3. The asymmetric dimer helps shields the effects of hydrophobicity by minimizing the SASA in comparison to the monomer in particular to the dimer interface (mean 96.4 \AA^2

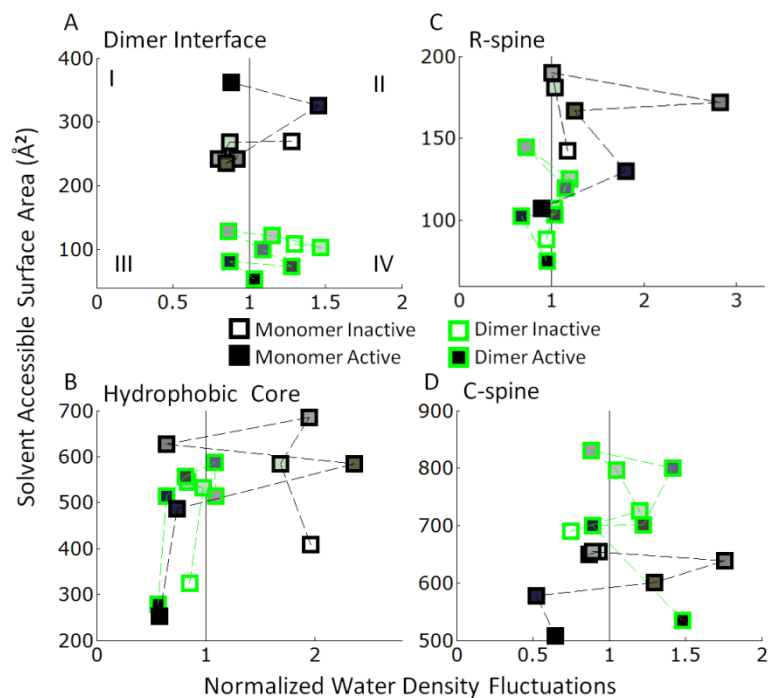


Figure 6.3: Hydrophobic plots of the solvent accessible surface area (SASA) versus normalized water density fluctuations in the key hydrophobic regions of the ErbB kinases: (A) Dimer Interface (B) Hydrophobic Core (C) R-spine and (D) C-spine. The edge colors represent monomer (black) and dimer (green) systems. The shape internal color represents active (black) and inactive (white), with the gradient between the two a progressive step.

in the dimer vs. mean 277.8 \AA^2 in the monomer), but also the hydrophobic core (mean 481.6 \AA^2 in the dimer vs. mean 540.7 \AA^2 in the monomer) and the R-spine (mean 108.3 \AA^2 in the dimer vs. mean 155.7 \AA^2 in the monomer). The C-spine is more exposed (722.4 \AA^2 in the dimer vs. 611.9 \AA^2 in the monomer) in the dimer during the transition, but the C-spine is thought to help coordinate ATP by providing a hydrophobic environment for the adenine ring that is not present in the simulation. The monomer simulations' hydrophobic path progresses through the perturbation sensitive regions whereas the dimer path tends to sit in the hydrophobically stable regions, excepting the C-spine, indicating

the dimer helps shield from any negative hydrophobic interactions, to ease the transition to active. We present the idea of the path length of the normalized water density with SASA as a measure of the rearrangement (protein and environmental) necessary for the change in hydrophobicity, with a lower path length representing an easier transition. The SASA was normalized so that the SASA and normalized water density each had equal contribution to the path lengths. The dimer simulations had a lower mean path length for the hydrophobic core (0.5703 in the dimer vs. 1.2954 in the monomer) dimer interface (0.2773 in the dimer vs. 0.3842 in the monomer) and the R-spine (0.2773 in the dimer vs. 0.3842 in the monomer). The C-spine had roughly equal mean path length in the dimer (0.5051) and the monomer (0.4873). The dimer interface then not only reduces the amount of SASA and time spend in perturbation sensitive regions, but also reduces the amount of reorganization necessary for the change in hydrophobic interactions.

6.3) Discussion

The fastening bonds do not appear until the active conformation is fully formed and well defined. This is indicative of the hypothesis that they help maintain a well formed active site, keeping key loops close together while dampening large fluctuations, and are necessary for activity but they themselves are not an activating stimulus. The hydrophilic bond network also helps keep the kinase inactive by sequestering key residues from forming bonds necessary for kinase activity, the key one being E738-K836 in EGFR.

The allosteric activation mechanism shown by Zhang et al. [41] moves the α C-helix into a more active like conformation, and also causing the α C-helix of the dimer to move down a different pathway in comparison to the monomer in our simulations. The dimer interface helps promote the active conformations not only allosterically, but also through hydrophilic and hydrophobic interaction networks. Introduction of the dimer interface hydrophobically shields the “receiver” kinase in its transition to active while reducing the amount of reorganization during the transition and promotes several bonds not seen in the monomer.

As the EGFR kinase transitions between inactive and active because of the dimer interface, we propose several events occur in a specific order (Figure 6.2): (1) the introduction of the dimer interface shifts the α C-helix and causes the formation of the K730-E939 and the K733-D918 salt bridges that help maintain the α C-helix in a partially “in” conformation (2) the flexible A-loop moves freely until it forms the E841-R949 salt bridge reducing its motions and keeping the A-loop partially open; which we hypothesize is an “active” intermediate (3) the E738-K836 salt bridge is broken thereby freeing E738 from sequestration (4) the formation of the conserved active E738-K721 salt bridge (5) extension of the α C-helix places the α C-helix in the active conformation and allows the fastening bonds to form, making a well formed and coordinated active site.

We hypothesize the combined effect of the allosteric mechanism as well as the added hydrophilic bonds from the dimer interface promotes a partially active state, while the

active kinase needs the formation of the fastening bonds for full activity. These two states in the pathway fits well into the “three phases” of autophosphorylation in kinase proposed by Lemmon and Schlessinger [18]. The partially active conformation of the EGFR dimer represents the “first phase” steps of autophosphorylation in kinase activity enhancement following a ligand-binding event, where it creates the local increase in kinase activity. The “first phase” of autophosphorylation in FGFR correlates with a 10-50 times increase in kinase activity [367]. A study by Shi et al. showed the ErbB3 kinase, which is missing crucial residues, previously thought to be a dead kinase is still competent to bind ATP and phosphorylate tyrosines, though at a greatly reduced kinase activity [337]. Since EGFR is a fully competent kinase, the opening of the active site and partial alignment of α C-helix potentially increases the kinase activity 10-50 times.

The “second phase” is *trans*-phosphorylation of its kinase partner’s tyrosines, in the C-terminal tail as well as in the A-loop to create phosphotyrosine binding sites to recruit downstream signaling molecular containing Src homology-2 SH2 domains and phosphotyrosine binding domain PTB containing proteins. Considering the arrangement of the EGFR asymmetric dimer, the “activator” kinase is in position for phosphorylation of its C-terminal tail. Recent studies have shown that the FGFR kinase domains also arrange into an asymmetric dimer, not as an activating stimulus but to phosphorylate specific tyrosine residues [368]. The “third phase” of autophosphorylation follows the phosphorylation of the C-terminal tail and increases the kinase activity to 500-1000 times the basal activity of the kinase. There is evidence that the C-terminal tail in EGFR forms

autoinhibitory interactions keeping the kinase domain in the inactive conformation [43, 369] with two residues in the start of the C-terminal tail, D960 and E961, forming an auto-inhibitory interaction with the S787 in the kinase domain [370]. Since the D960 and E961 are both negatively charged, the added negative charge from phosphorylation of C-terminal tail tyrosines could break the inhibitory interaction in the “activator” kinase and allowing another increase in kinase activity. The transition into the active conformation of both kinase domains in the dimer pair, including the formation of the well coordinated active site through the set of fastening bonds (L834-R812, K836-V810, and L838-R808) and coordinating bonds (E734-K851/K836, E738-K721 and D813-R817) could represent the “third phase” activation into a fully active kinase.

In conclusion, our results presented here have helped establish the molecular context the asymmetric dimer interface has upon the EGFR kinase. Furthermore, through the timeline of specific events of activation we have evidence of the formation of a partially active intermediate that fits in well with the current thoughts of kinase activation via autophosphorylation. Since ErbB4 and ALK kinase domains have a similar inactive crystal structures to EGFR, expansion of this methodology to the ErbB4 kinase as well as the ALK kinase domain and contrasting the results from the EGFR TMD studies would be interesting to elucidate differences in primary structure and crystal structures affecting the activation mechanism of each kinase.

6.4) Methods

Simulation techniques: Models for EGFR kinase monomer and dimer systems in the inactive conformation were constructed, minimized and equilibrated [364]. Targeted molecular dynamics (TMD) simulations were performed in NAMD using the RMSD of the heavy atoms of the α C-helix and the A-loop as the reaction coordinate (χ_i). The activation pathway is divided into 5 smaller windows for the monomer and 6 windows for the dimer. Each window is simulated individually for 10 ns to ensure adequate sampling. To enhance the sampling of low probability events, a harmonic restraint of constant $k=20$ kcal/mol/(\AA^2) is applied along the reaction coordinate, χ_i .

Analyses of MD Simulations: Root-mean-squared deviation (RMSD) calculations were performed using the RMSD tool plugin in VMD by first removing global translation and rotation, and then computing the RMSD of the selected reaction coordinate (A-loop and α C-helix) relative to a reference structure (the respective active or inactive crystal structure). An analysis of hydrogen bond (H-bond) patterns in the TMD simulations was performed using CHARMM in conjunction with VMD similar to previous studies [152, 298, 312, 330]. Solvent accessible surface area (SASA) values and normalized water density fluctuations were used to measure the hydrophobicity as performed previously [298, 330]

Chapter 7.) Perspectives

The molecular dynamics strategies we have presented in this thesis so far have been exceptionally useful in examining atomic level protein characteristics. While there has been a lot of controversy on the accuracy of molecular dynamics simulations, the phenomena we are examining (side chain motions, hydration dynamics and loop and collective motions) are all within the timescales (nanoseconds) we are simulating (Table 7.1).

Time Scales	Length Scales	Event
short femto to pico		
$10^{-15} \text{ s} - 10^{-12} \text{ s}$	$0.001 \text{ \AA} - 0.1 \text{ \AA}$	<ul style="list-style-type: none">• bond stretching, angle bending• constraint dihedral motion
medium pico to nano		
$10^{-12} \text{ s} - 10^{-9} \text{ s}$	$0.1 \text{ \AA} - 10 \text{ \AA}$	<ul style="list-style-type: none">• surface side chain motion• loop motion, collective motion• hydration dynamics
long nano to micro		
$10^{-9} \text{ s} - 10^{-6} \text{ s}$	$1 \text{ \AA} - 100 \text{ \AA}$	<ul style="list-style-type: none">• folding in small peptides• helix coil transition
very long micro to second		
$10^{-6} \text{ s} - 1 \text{ s}+$	$10 \text{ \AA} - 100 \text{ \AA}$	<ul style="list-style-type: none">• protein folding• cellular processes

Table 7.1: Time scales and length scales of common phenomena associated with protein dynamics.

The results we show have a striking correlation in highlighting perturbation sensitive regions with the regions mutated in clinically identified oncogenic kinase mutations: the hydrophobic core in the EGFR and ErbB4 kinase domains as well as the α C- β 4 region in

ErbB2 kinase domain. We also have preliminary results resolving the sequence of events in the activation of the EGFR kinase domain. However, the start of molecular dynamics simulations is a high resolution crystal structure. When a crystal structure is not available, then homology modeling techniques can create a crystal structure with which to allow simulation as well as providing insights into similarities in primary structure affecting the crystal structure. Furthermore, by linking together the atomic level protein fluctuation data we can see the effects of the kinase activity upon the cell signaling networks. Such a multi-scale model then can not only link kinase domain fluctuations with signaling effects to help contextualize mutations found in oncogenic cell lines, but refine therapeutic strategies to target key perturbation sensitive regions of kinase signaling, at the molecular level as well as the cellular level. Here in this chapter we review methods and results performed by others in the Radhakrishnan lab that helps link the MD simulations with the multi-scale model (for an example see Figure 3.1). The ErbB3 homology (section 7.1) and signaling models (section 7.3) were done by Shannon Telesco which was submitted to *Molecular Biosystems* and is under review [330], while the EGFR signaling models (section 7.2) were performed by Jeremy Purvis and the work is published in *Molecular Biosystems* [282].

7.1) Homology modeling of the ErbB3 kinase domain

Given the recently-resolved crystal structure of the ErbB3 kinase domain [337] (as well as the structure by Jura et al. [43]), we chose to use this structure in our MD simulations.

However, as the ErbB3 kinase domain shares a relatively high level of sequence identity with its family members (54%, 55% and 63% identity with EGFR, ErbB2, and ErbB4, respectively) (Fig. 7.1), we first constructed several homology models of the ErbB3 kinase based on the other ErbB kinase structures and validated them with respect to the ErbB3 crystal structure, as a verification of our homology modeling method and also as a means of deriving insights into a structural comparison of the ErbB kinase active sites. As the crystal structures of the EGFR and ErbB4 kinase domains have been determined [25, 297, 353], we selected EGFR (PDB code 2GS7) [353] and ErbB4 (PDB code 3BBT) [297] as templates for homology modeling of the ErbB3 kinase in the inactive-like conformation. In addition, we applied the multiple templates (MT) algorithm in MODELLER to generate a ErbB3 kinase model based on a combination of the EGFR and ErbB4 crystal structures. To verify that EGFR and ErbB4 were the most suitable templates available for modeling ErbB3, a search was performed in MODELLER's internal database for other proteins that might have high sequence identity with ErbB3. The ErbB receptor kinases were confirmed to be the most homologous structures for modeling of the ErbB3 kinase domain.

The protein sequence selected for alignment of the kinase domains included residues 678-957 (EGFR) and 683-962 (ErbB4); we opted to exclude the flexible C-tail from the alignment, as its sequence is highly variable among the ErbB kinases. A total of 50

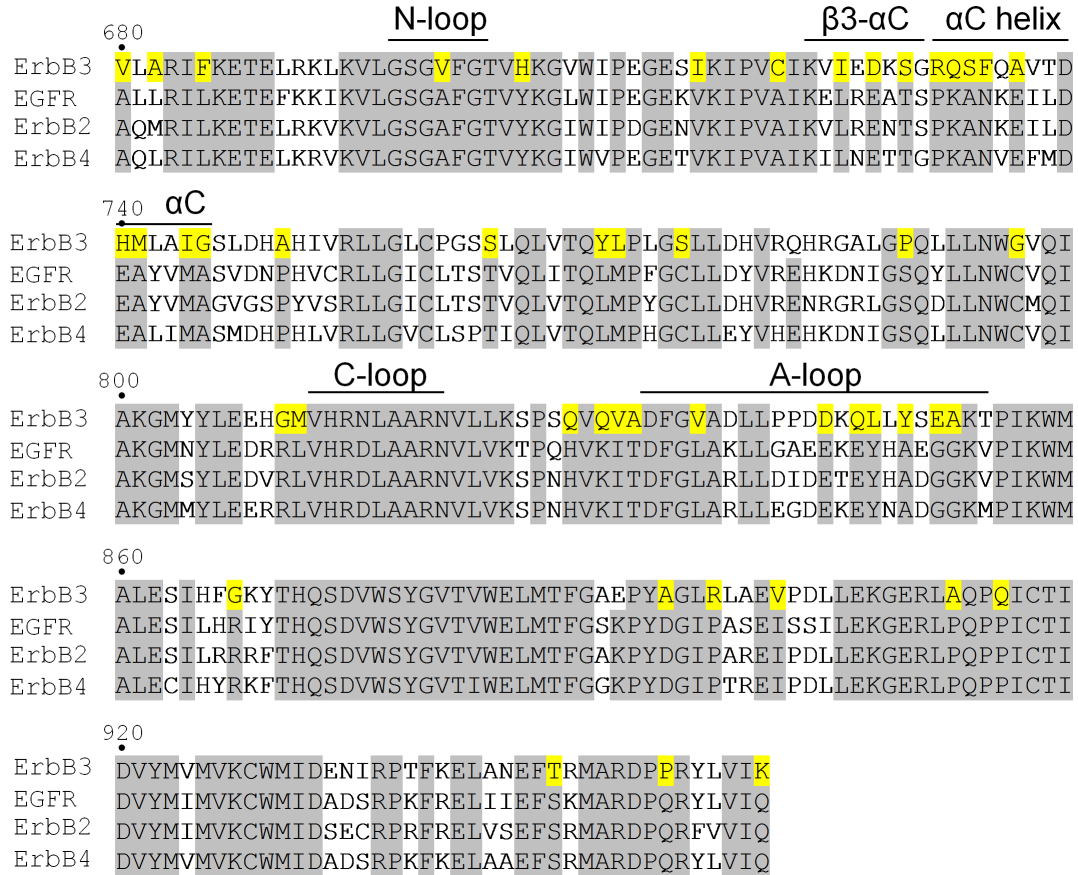


Figure 7.1: Multiple sequence alignment of the kinase domains of the ErbB RTK family members. Residues highlighted in gray are conserved among all four ErbB kinases, whereas residues highlighted in yellow are unique to ErbB3. The ErbB3 kinase shares a relatively high level of sequence identity with its family members (54%, 55% and 63% identity with EGFR, ErbB2, and ErbB4, respectively). The key catalytic sub-domains (A-loop, C-loop, N-loop, αC-helix, and β3-αC loop) are labeled.

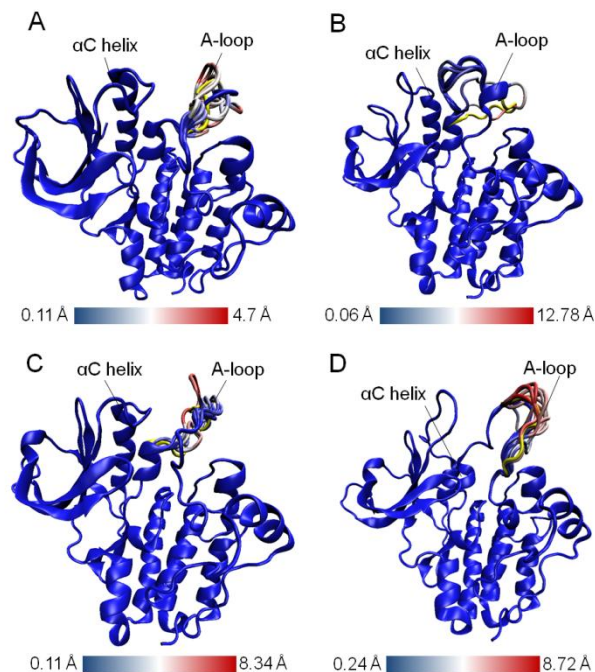


Figure 7.2. The top 10 refined A-loops for the best ErbB3 models constructed from each ErbB template. The top structures are shown for the ErbB3 models based on (A) the EGFR template, (B) the ErbB4 template, (C) Multiple templates and (D) the loop-modeled ErbB3 crystal structure. The structures are colored-coded according to the RMSD, where red regions indicate large RMSD values and blue regions represent small RMSD values. The original unrefined model for each ErbB3 structure is shown in yellow. The top structures form a dominant cluster of conformations in each ErbB3 model, indicating a pronounced energy minimum and a higher level of accuracy in the best structural prediction.

models were generated from each of the templates (EGFR, ErbB4, and multiple templates) by satisfying a set of static and dynamic spatial restraints in MODELLER. The stereochemical quality of each model was evaluated using the Discrete Optimized Protein Energy (DOPE) method [371], which is an atomic distance-dependent statistical potential optimized for model assessment in MODELLER. A combination of stereochemical parameters and DOPE energies were considered in order to determine the most

energetically favorable ErbB3 models derived from each template. Furthermore, the RMSD among the top model A-loops was computed, as minimal variation among the low energy conformations correlates with a more pronounced free energy minimum and a higher level of accuracy in the best structural prediction [179]. The superposition of the top 10 models from each template resulted in a dominant cluster of conformations, increasing our confidence in the reliability of the top structures (Fig. 7.2). The top 10 MT models exhibited the smallest RMSD, or minimal variation.

A comparison of the most energetically favorable models derived from each template revealed several similarities in structure and in DOPE energies: overall, the C-lobes of the kinase models are similar, as are the DOPE profiles, although the DOPE scores in the A-loop region are lowest for the ErbB4- and MT-based models (Fig. 7.3b). However, in comparing the top homology-modeled structures to our ErbB3 crystal structure, we observed several conformational differences, which can be better understood by first describing the major structural features that distinguish the ErbB3 crystal structure from the EGFR and ErbB4 structures. A unique feature of the catalytic site in the ErbB3 crystal structure is the truncated N-terminus of the α C helix (Fig. 7.3a); in the EGFR and ErbB4 kinase crystal structures, the helix is fully formed. The molten terminus of the ErbB3 α C helix forms a loop, referred to as the β 3- α C loop, which interacts with specific sub-regions of the catalytic site, including the short 3_{10} helix in the A-loop, to form a hydrophobic interface that maintains the weakly-active conformation of ErbB3 and is not observed in EGFR or ErbB4 [355]. F734, which is located in the β 3- α C loop and

corresponds to an Asn in EGFR and ErbB4, appears to nucleate the hydrophobic interface (Fig. 7.3a) [355], and V836, L839 and L840, which are located in the 3_{10} helix, form hydrophobic contacts with the truncated α C helix. The extensive set of residues stabilizing the hydrophobic interface and the 3_{10} helix are unique to ErbB3 and function to maintain the distinctive catalytic site conformation resolved in the ErbB3 crystal structure.

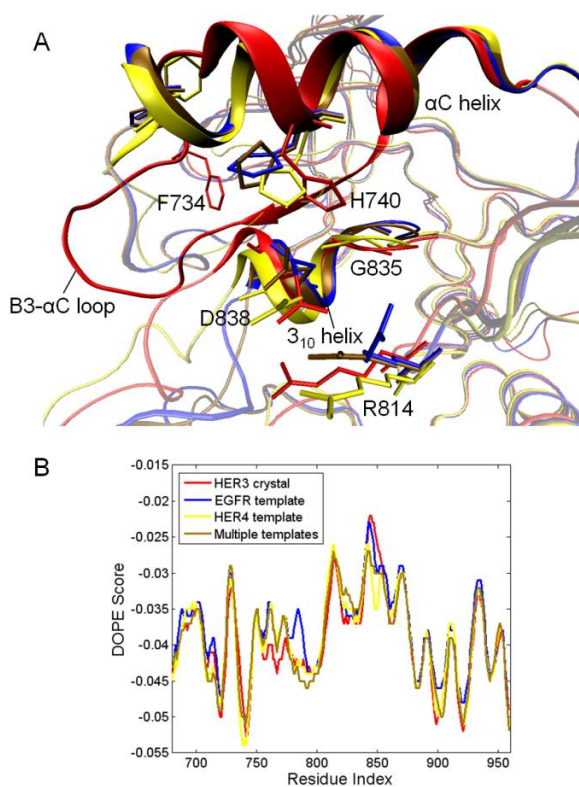


Figure 7.3. (A) Superposition of the top ErbB3 models constructed from the ErbB templates: EGFR (*blue*), ErbB4 (*yellow*) and multiple templates (*brown*) as well as the ErbB3 crystal structure (*red*) (PDB code 3LMG). Key residues contributing to the hydrophobic interface in ErbB3 are highlighted. (B) DOPE energy plots for the top ErbB3 models based on each ErbB template after A-loop refinement. The ErbB4- and MT-based models exhibit lower DOPE energies in certain regions, including the sequence spanning residues 775-800. The DOPE energy plot for the ErbB3 crystal structure (PDB code 3LMG) is illustrated for reference.

In comparing our top homology-modeled structures to the ErbB3 crystal structure, we focused on the unique set of interactions in the β 3- α C loop, α C helix and 3_{10} helix (Fig. 7.3a). The side-chain conformations of the aliphatic residues in the 3_{10} helix, including V836, L839, and L840, were correctly predicted in the models derived from EGFR, ErbB4, and MT. This result is not entirely surprising, as the inactive EGFR and ErbB4 templates also contain the 3_{10} helix in the A-loop. However, F734, which serves to nucleate the hydrophobic interface in the ErbB3 crystal structure, points away from the hydrophobic pocket in the homology models. The reason for the altered F734 conformation is depicted in Fig. 3a, which illustrates that the homology models contain a fully formed α C helix in place of the truncated helix and extended β 3- α C loop (which contains F734) in the ErbB3 crystal structure. In addition, the H740 ring in the α C helix, which stabilizes homology models. The D838-R814 bond, which bridges the 3_{10} helix and the C-loop in the ErbB3 crystal structure, is conserved in the ErbB4- and MT-based models, whereas the residues are unbonded in the EGFR-based model. Furthermore, the ErbB4-based model more closely mimics the extended β 3- α C loop in the ErbB3 crystal structure, as the apex of the loop projects away from the catalytic site.

Although none of the homology models correctly predict the molten α C helix, several features of the ErbB4- and MT-based models, including the D838-R814 bond, approximate the ErbB3 crystal structure more closely than does the model derived from the EGFR template. Additionally, the DOPE plots for the top structures (Fig. 7.3b) reveal

smaller (more favorable) DOPE energies for the ErbB4- and MT-based models, especially for residues 775-800, which exhibit an energy peak in the EGFR-based model, and residues 840-860 in the A-loop. These results may be attributed to the closer evolutionary ties between ErbB3 and ErbB4, as ErbB3 and ErbB4 are thought to have emerged from a gene duplication event separate from the one that led to EGFR and ErbB2 [372]. Indeed, Jura et al. [43] postulate that the N-lobe dimer observed in their resolved ErbB3 and ErbB4 crystal structures but not in their EGFR structure may be a consequence of the evolutionary homology between ErbB3 and ErbB4. Our results emphasize the importance of selecting the best available template for homology modeling of even highly related proteins, and indicate that the application of multiple templates in the sequence alignment may improve the quality of homology models in certain cases [373]. Furthermore, the homology modeling analysis provides a framework for comparing the ErbB kinase active sites and identifying the molecular features that contribute to ErbB3's unique catalytic conformation.

7.2) Clinical Implications of Oncogenic EGFR mutations from a Multiscale Model of ErbB Receptor Signaling

Previously in this thesis, we have extensively explored atomic level characterization of the ErbB family kinase domains. Here in this section as well as section 7.3 we link those molecular results to cell signaling models into a multiscale model. For cellular homeostasis, pro-survival signals are balanced by pro-apoptotic signals, with both being

triggered and balanced by a variety of interacting intracellular pathways. Using a simplified model (Figure 7.4) for the effect of AKT activation on cell response, we showed that preferential AKT activation is conducive for the cell to rely on and be addicted to [126, 374]) for generation of pro-survival signals [282]. Our simplified model illustrates a mechanism by which inhibition of the dominant source of pro-survival signals shifts the cellular state to one devoid of pro-survival signals, representing a perturbation sensitive point in the cellular signaling network which can account for a remarkable inhibitor sensitivity [375].

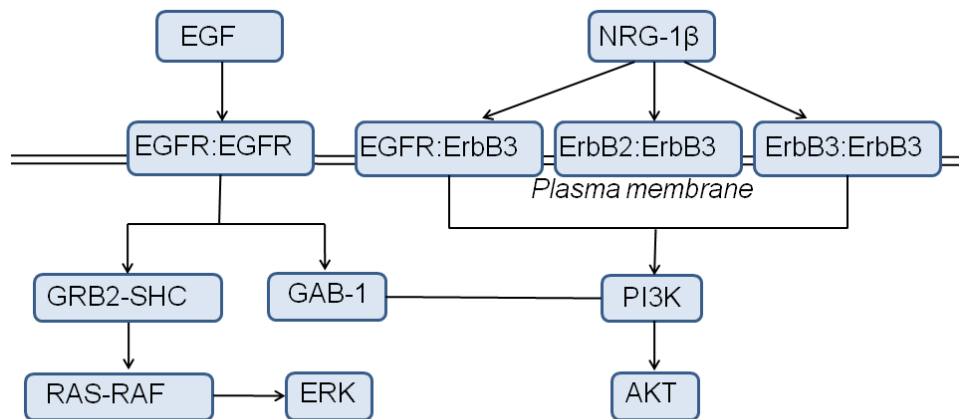


Figure 7.4: Schematic diagram of the EGFR signaling model used as well as the ErbB3 signaling model used. The differential ligands (EGF and NRG-1β) causes differential heterodimerization between members of the ErbB family triggering their respective cell signaling pathways.

We hypothesized the mechanisms that lead to inhibitor hypersensitivity (Gefitinib for EGFR) attack these perturbation sensitive points of network hypersensitivity and fragility. Since preferential AKT activation is a hall mark of the hyper-sensitive mutants

as well as the efficacy of the inhibitors, we determined, through a global sensitivity analysis [375, 376], the combinations of model parameter perturbations that drive enhanced production of pAKT and pERK (Table 7.2B). The top components that produce maximum sensitivity in terms of changes to the pAKT and pERK levels were PI3K, Ras, Gab-1, MEK, Raf which have all been observed in several human cancers [377-381]. Moreover, it has been established in screened breast and colorectal cancer patients that the GAB-1, MEK, and Ras mutations are non-random and likely arise from selective evolutionary pressures that give the cancer cells a survival advantage [381].

	Wildtype	L834R	Del	Rate & Binding Constants
ERK-(p) in nM for normal ErbB1 expression				k_f :Y1068; k_f : Y1173
-EGF/+EGF	0.2/5.0	1.0/1.0	8.0/8.0	Ki: Inhibitor; KM: ATP●RTK
Akt-(p) in nM for normal ErbB1 expression				KM : GAB-1●ErbB1-(p)
-EGF/+EGF	1.0/15.0	13.0/13.0	20.0/20.0	Initial Concentrations
ERK-(p) in nM for ErbB1 over-expression				[Raf●Ras●GTP]
-EGF/+EGF	3.0/3.0	5.0/5.0	1.0/1.0	[Pase3] phosphatase for ERK-(p)
Akt-(p) in nM for ErbB1 over-expression				[Pase4] phosphatase for Akt-(p)
-EGF/+EGF	20.0/20.0	22.0/22.0	19.0/19.0	[MEK-(p)]
Cellular EC50 in nM for inhibition of ERK-(p) and Akt-(p) for normal ErbB1 expression/ErbB1 over-expression				[PI3K inactive]
ERK-(p)	700/4000	100/1000	-/-	[MEK●Raf active]
Akt-(p)	1200/500000	300/50000	-/-	[ErbB1●Shc●Grb2●SOS●RasGTP]
				[inhibitor]

Table 7.2: (A, left) Comparison of signaling and inhibition characteristics predicted by our network simulations for wildtype and mutant systems under different conditions; Del stands for the deletion mutant del L723-P729 ins S. **(B, right)** Parameter variations constituting the top three principal components for network hyper-sensitivity calculated through global sensitivity analysis: k_f , turnover for phosphorylation; (●) denotes bound complex; the square brackets represent concentrations.

With reference to the clinically identified EGFR mutants found in non-small cell lung cancer patients, mainly L834R and del L723-P729 ins S, we found the mutations had altered affinities for phosphorylation of specific tyrosines in the C-terminal tail: Y1068 and Y1173. The preferential binding characteristics of different cytosolic substrates to different phospho-tyrosine locations of the ErbB family kinases has been reported [382]. Thus, differences in the phosphorylation kinetics associated the different tyrosine sites of the cytoplasmic C-terminal tail of the EGFR kinase can induce differential patterns of downstream signaling leading to differences in the activation of cell signaling networks. The effect of altered affinities of the Y1068 and Y1173 sites to the catalytic domain of the EGFR is that the L834R under normal EGFR expression sees a ~5-fold decrease in ERK activation and a smaller ~15% decrease in AKT activation (Table 7.2A). The del L723-P729 ins S mutant however, shows sustained ERK as well as AKT activation relative to wildtype. For EGFR over-expressed cells, both ERK and AKT activation characteristics show relative insensitivity to EGFR as a result of signal saturation. Furthermore, the mutants can continue to signal even in the absence of the growth factor. In addition, the mutant signaling can be different due to changes in the ATP affinity. However, both these factors do not introduce any differential characteristics (in terms of preferring Y1068 to Y1173) and cause a differential in overall activation levels of ERK and AKT.

The perturbation of the phosphotyrosine kinetics of Y1068 and Y1173 through mutations (L834R and del L723-P729 ins S) is directly responsible for the differential signaling

leading to preferential AKT activation [282]. The restoration of signaling has also been reported through a double mutation of L834R/T766M [133, 383]. This double mutant increases receptor phosphorylation (Y1068 and Y1173) kinetics 100-fold [384] while simultaneously decreasing inhibitor affinity [383]. Another drug resistance mechanism related to Y1068 kinetics that circumvents Y1068 has been identified. In the presence of ErbB3, a branch of signaling analogous to that through Y1068 becomes available through ErbB hetero-dimerization, directly resulting in PI3K recruitment on ErbB3 and subsequent AKT activation, which is discussed below.

7.3) Systems model of ErbB signaling defines a mechanism for ErbB3-mediated TKI resistance

Previous experimental studies have demonstrated that ErbB3 is a key mediator of resistance to various tyrosine kinase inhibitors (TKIs) currently in use [343, 344, 385-387]. Some postulated mechanisms include leaky ErbB2-catalyzed phosphorylation of ErbB3, e.g., incomplete inhibition of ErbB2 catalytic activity by the TKI [343, 386, 388]. Indeed, previous experimental studies have demonstrated that leaky ErbB2 phosphorylation of ErbB3 in TKI-bound ErbB2/3 heterodimers is amplified by additional resistance mechanisms, such as inhibition of cellular phosphatases by TKI-mediated production of reactive oxygen species (ROS), and increased expression of ErbB3 at the plasma membrane [343, 345, 389]. However, Shi et al. [337] have recently shown the assumed inactive ErbB3 pseudokinase is, in fact, a weakly active kinase with ~1000 fold

weaker phosphorylation than the canonical kinase members of the ErbB family [337]. Therefore, though the ErbB2 kinase is a viable route of resistance, here we consider ErbB3 catalytic activity in the ErbB signaling network as a potential TKI resistance mechanism [330].

To translate our observations of the weak, yet robust, activity of the ErbB3 kinase into a physiologically relevant context and investigate the implications of ErbB3 activation for ErbB signaling dynamics, we constructed a systems-level model for ErbB3 based on that of Schoeberl et al. [390], with added ErbB3 phosphorylation rate constants derived from experiments [337] (Figure 7.4). We simulated stimulation of the ErbB signaling network through NRG-1 β , which signals through the ErbB2/ErbB3/ErbB4 kinases (though ErbB4 kinase is omitted from the model since ErbB4 signaling is weak or absent in many cancer cell lines).

A parameter sensitivity analysis identified the key proteins that direct signaling in our model of the ErbB network with respect to AKT activation *via* stimulation of NRG-1 β . ErbB3 and NRG-1 β represent the most sensitive species in the signaling network, followed by ErbB2 concentration (Figure 7.5A). EGFR is not a strong determinant of the extent of AKT phosphorylation, as expected from the weak ability of NRG-1 β to elicit EGFR dimers. PTEN (the PIP₃ phosphatase) and the ErbB phosphatase exhibited a negative sensitivity in the analysis, as these phosphatases negatively regulate the signaling network through dephosphorylation of key molecular species.

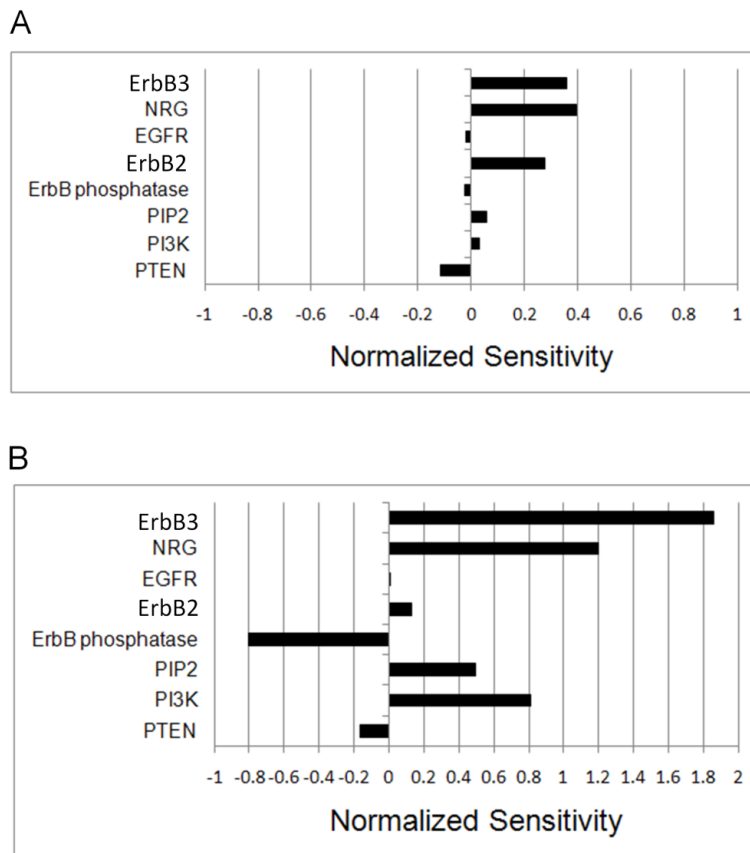


Figure 7.5 Parameter sensitivity analysis of the ErbB3 signaling model. The normalized, time-integrated sensitivity of pAKT to key molecular species was computed in response to NRG1- β =25 nM by making a 0.1% change in each species concentration for (A) the model representing weak ErbB3 activity, and (B) the model representing weak ErbB3 activity in the presence of the EGFR/ErbB2 inhibitor lapatinib. Sensitivity to certain species, including ErbB3, ErbB2 and the ErbB phosphatase, changed upon addition of lapatinib.

Upon incorporation of the TKI lapatinib, which inhibits EGFR and ErbB2 catalytic activity, into our model of ErbB3, the ErbB signaling network relies more heavily upon ErbB3 activity for AKT induction (Figure 7.5B). With completely inhibited EGFR and ErbB2 signaling via introduction of lapatinib into the model, ErbB3 and AKT signals still persist at maximal inhibition by lapatinib. The results of pAKT sensitivity analysis of the

lapatinib-treated model to those of the inhibitor-free model show that sensitivity to ErbB3 and NRG-1 β increases, whereas sensitivity to EGFR and ErbB2 decreases, as lapatinib sequesters EGFR and ErbB2 molecules. The negative normalized sensitivity to the ErbB phosphatase also increases, as the pool of ErbB dimers has diminished due to sequestration of EGFR and ErbB2 by lapatinib. Thus a single alteration to the signaling model (in this case, addition of lapatinib) significantly redefines the most perturbation-sensitive nodes in the network.

Although the pAKT signal induced by ErbB3 phosphorylation in our *in silico* lapatinib-treated cell is relatively weak, in an actual physiological context, a tumor cell may employ several resistance mechanisms at once [343, 386, 391, 392]. One of the sensitive nodes of NRG-1 β -driven ErbB signaling is ErbB phosphatase levels, and simulations showed decreased activity of phosphatases resulted in an amplification of the level of AKT signaling induced by ErbB3. It has been demonstrated that in certain cases of TKI resistance, the tumor cell responds to the reduction in pAKT levels by upregulating vesicular transport of ErbB3 from the cytoplasm to the plasma membrane [343, 389]. Our results show for a 2-fold increase in surface ErbB3 level, AKT signaling is similarly amplified. For 25 nM NRG-1 β , the pAKT signal is restored to nearly 60% of its no-inhibitor control level, and pAKT levels are nearly 100% regained for 100 nM NRG-1 β , effectively recreating drug resistance *in silico*.

Our data parallels the experimental studies performed by Sergina and colleagues [343], which describe ErbB3-mediated resistance and pAKT signaling in various TKI-treated tumor cell lines. Thus our model demonstrates that even a weak level of ErbB3 signaling, as suggested by our previous results [337], are physiologically relevant in the context of an ErbB-driven tumor cell, and illustrates several routes through which ErbB3 signaling may be compounded by other previously postulated resistance mechanisms to generate TKI resistance.

7.4) Conclusions

As multi-scale modeling techniques become more sophisticated, the application of simulations will become more widespread necessitating a stronger link between experiments and simulations. This thesis has aimed to present a thorough exploration of the ErbB kinase activation at the atomic level, highlighting hydrophilic and hydrophobic interaction networks that help differentiate the active and inactive state. We have also highlighted perturbation sensitive regions which correlate well with clinically identified activating mutations in cancer as well as hypothesized the mechanism and the sequence of events that occur from the activating asymmetric dimer interface. Furthermore, by linking the atomic level data with cell signaling networks we can also highlight perturbation sensitive regions at the network level from altered signaling characteristics of the mutants and thereby predict potential therapeutic targets. Therefore, use of a multiscale model can investigate biological systems at multiple timescales and length

scales, providing detailed and predictive information purely *in silico*. Hopefully this work has shown the interesting and complex biological problems that can be addressed through computational methods.

Appendix A: Hydrophobic interactions help identify perturbation sensitive regions for the ErbB kinases

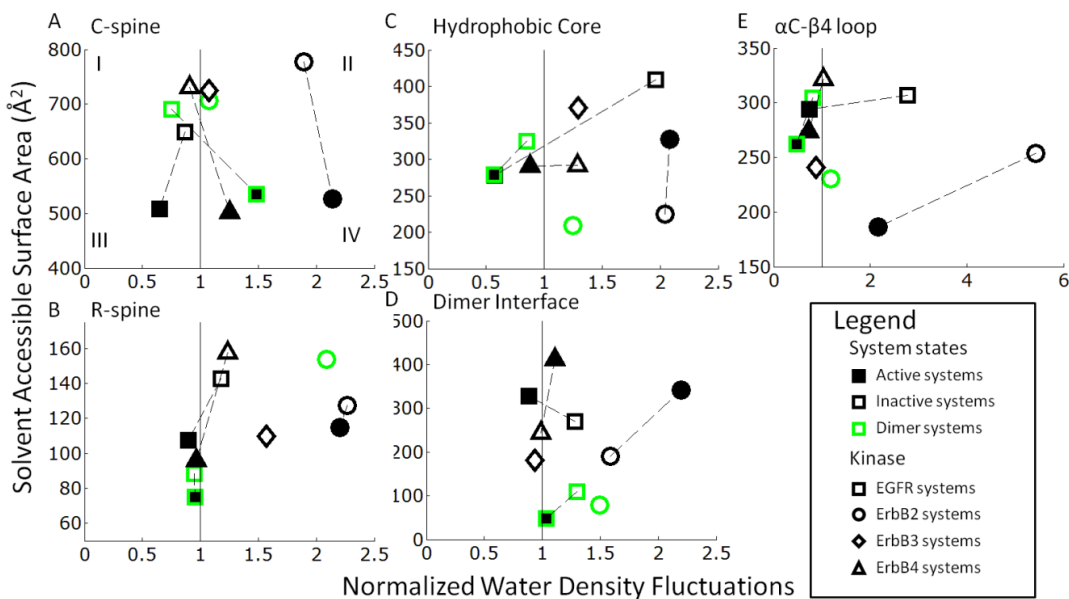


Figure A1: Hydrophobic plots of the solvent accessible surface area (SASA) versus normalized water density fluctuations in the key hydrophobic regions of the ErbB kinases: (A) C-spine (B) R-spine (C) Hydrophobic Core (D) Dimer Interface and (E) αC-β4 loop. The edge colors represent monomer (black) and dimer (green) systems. The shape internal color represents active (black) and inactive (white). The shape itself represents a different member of the ErbB family EGFR (square), ErbB2 (circle) and ErbB4 (triangle). The quadrants represent different hydrophobic interaction regions (I) hydrophilically favorable region, (II) perturbation-sensitive region and (IV) hydrophobically favorable region.

Hydrophobic interactions appear to provide context-specific contributions to stability to the active and inactive conformations of ErbB kinases. To investigate the effect of hydrophobic interactions on the ErbB kinase conformations, we analyzed the hydrophobicity as well as the solvent accessible surface area (SASA) of relevant hydrophobic sub-regions, namely, the C-spine, R-spine, hydrophobic core, and the αC-β4

region. The four regions have a high percentage of hydrophobic side chains; however, some minor differences between members of the ErbB family exist, particularly in the α C- β 4 region. Garde et al. [320, 321] have recently proposed an approach for quantifying the hydrophobicity of heterogeneous surfaces using water density fluctuations, according to which increased normalized water density fluctuations are used as a signature of a more hydrophobic surface. We have normalized the water density fluctuations so that 1 is indicative of a neutral region and plotted it against the SASA. By splitting the graph into four quadrants (Figure A1.A) we can see quadrant I represents a hydrophilically favorable region with low hydrophobicity and high SASA, while quadrant IV represents a hydrophobically favorable region with high hydrophobicity and low SASA. Quadrant II represents a fragile or perturbation sensitive region with high hydrophobicity but also with a high SASA.

The active conformations in the C-spine show a clear delineation between the active and inactive conformations regardless of dimer or mutational state (Figure A1.A). The active conformations minimize the SASA (mean of 500\AA^2) in comparison to the inactive (mean value of 700\AA^2). In addition, in the transition between the inactive to active, the active conformations increase in hydrophobicity and settle into the hydrophobically favorable region, implying the active conformations have a better formed C-spine correlating with the observation that the C-spine helps in coordinating the loops in the active conformation. The SASA of the ErbB3 C-spine falls within range of the inactive EGFR and HER4 systems, reflecting that, despite its weak activity, there is no corresponding

‘fully-active’ state for HER3 as for the other ErbB kinases. This inability to ‘fully’ activate can be attributed to the lack of the crucial hydrogen bonding network identified earlier, which is required to stabilize the active-like kinase conformation. The hydrophobicity plots of the R-spine show a similar difference between the active and inactive conformations but not as drastically as the C-spine (Figure A1.B). For each system the active systems expose less surface area than the inactive systems, but only for each system locally (dotted lines on Figure A1). The SASA of the ErbB3 R-spine deviates from the values for the inactive EGFR and ErbB4 systems, and instead demonstrates low SASA (high hydrophobicity). This result can be rationalized by the increased hydrophobicity of the R-spine, which includes segments of the truncated α C helix in ErbB3.

With respect to the hydrophobicity plots of the hydrophobic core, the monomer EGFR and ErbB4 inactive systems are situated within the perturbation sensitive region with their respective active conformations out of the perturbation sensitive region (Figure A1.C). For ErbB2 the reverse is true which is consistent with sensitivity of the inactive conformation to hydrophobic perturbation, especially for EGFR and ErbB4 but not for ErbB2. Therefore single point mutations would be expected to disrupt local conformations contributing to hydrophobicity. Notably, mutations of hydrophobic residues in the hydrophobic core are reported for EGFR and ErbB4 in clinical studies [27-29], whereas, for ErbB2, such mutations are found surrounding the α C- β 4 region [56, 57]. By contrast, an analogous mutation in ErbB3 abolishes ATP-binding and

phosphorylation activity [337], indicating that hydrophobic interactions in the core promote ErbB3 activity, rather than maintain an autoinhibited state as they do in EGFR and ErbB4.

So far all the hydrophobic plots have shown preferential hydrophobic interactions in the active systems versus the inactive systems. However, examination of the asymmetric dimer interface reveals a clear delineation in hydrophobicity in preference for the inactive systems as well as dimerization reducing this hydrophobic benefit implying an activating stimulus (Figure A1.D). The asymmetric dimer interface consists largely of hydrophobic side-chains in the N-lobe of the receiver kinase (L680, I682, L736, L758, and V762) and the C-lobe of the activator kinase (I917, Y920, M921, V924, and M928) [40]. The active monomer systems trend to the perturbation sensitive quadrant, while the inactive monomer systems show a reduced SASA in comparison to the active removing them from the perturbation sensitive quadrant. This decrease for the inactive monomers may imply a preference for the inactive state in the monomer context. Notably, the dimeric systems record much lower SASA, pushing all the dimer systems into the hydrophobically favorable quadrant, implying that hydrophobic interactions provides a dominant driving force for dimerization; interestingly, for EGFR dimer, the inactive state is very similar to the active and hence the preference for the inactive conformation is not implied in the context of the dimer. Thus, dimerization also provides a stimulus for activation.

Interestingly, all ErbB2 monomer systems present a very high hydrophobicity (black bordered circles on Figure A1) across all of the hydrophobic subregions while the ErbB2 dimer as a reduced hydrophobicity in comparison to the monomers (green circles on Figure A1). This is consistent with the notion that hydrophobicity is particularly important in the context of ErbB2 owing to its interaction with Hsp90 known to be mediated by hydrophobic contacts, particularly the α C- β 4 region which is an unstructured span between the α C-helix and the β 4 sheet in RTKs. From a sequence perspective, only in ErbB2 is the α C- β 4 region predominantly hydrophobic and exceptionally so. Both the inactive and active conformations of the ErbB2 monomer systems reflect the trend by singling out the ErbB2 monomer systems are particularly hydrophobic for the α C- β 4 region (Figure A1.E). The mean SASA for the α C- β 4 region in ErbB2 systems is also consistently lower than in other members of the ErbB family. As discussed in [312], this unique feature of ErbB2 is thought to be responsible for its preferential association with the molecular chaperone Hsp90.

Thus, the analyses for the spine regions and the hydrophobic core collectively lead to the remarkable prediction that while the α C- β 4 region is perturbation sensitive for the inactive conformation of ErbB2, the hydrophobic core has the same effect for EGFR and ErbB4. Indeed, this correlates well with clinical studies, where activating point mutations in the hydrophobic core have been found in EGFR and ErbB4 but those in the α C- β 4 region are found in ErbB2, suggesting that the hydrophobic analysis enables the context-specific identification of fragile sub-regions.

Appendix B: Exhaustive Persistent Bond Tables for EGFR TMD simulations

EGFR Dimer Inactive	Window 1	Window 2	Window 3	Window 4	Window 5	Window 6	EGFR Dimer Active
aC-helix aC-helix bond							
--	--	--	--	--	K730,E734	K730,E734	--
--	--	--	--	--	--	--	A731,I735
--	--	--	--	--	--	--	N732,S728
N732,L736	N732,L736	N732,L736	N732,L736	--	--	N732,L736	N732,L736
K733,D737	K733,D737	K733,D737	K733,D737	K733,D737	K733,D737	K733,D737	K733,D737
--	--	E734,E738	E734,E738	E734,E738	E734,E738	E734,E738	--
I735,A739	I735,A739	I735,A739	I735,A739	I735,A739	I735,A739	I735,A739	I735,A739
--	L736,Y740	L736,Y740	L736,Y740	L736,Y740	L736,Y740	L736,Y740	L736,Y740
E738,M742	E738,M742	E738,M742	E738,M742	E738,M742	E738,M742	E738,M742	E738,M742
A739,A743	A739,A743	A739,A743	A739,A743	A739,A743	A739,A743	A739,A743	A739,A743
--	--	Y740,S744	Y740,S744	Y740,S744	Y740,S744	Y740,S744	Y740,S744
--	--	--	--	V741,V745	V741,V745	V741,S744	V741,V745
A-loop bonds							
--	--	--	--	--	--	--	E841,R808
--	--	--	--	--	--	--	Y845,Y867
--	--	--	--	--	--	--	A847,R865
E848,R865	E848,R865	E848,R865	E848,R865	E848,R865	E848,R865	E848,R865	--
--	--	--	--	E848,Y867	E848,Y867	--	--
K851,E725	--	--	--	--	--	--	--
C-loop bonds							
H811,D872	H811,D872	H811,D872	H811,D872	H811,D872	H811,D872	H811,D872	H811,D872
R812,D872	R812,D872	R812,D872	R812,D872	R812,D872	R812,D872	R812,D872	R812,D872
--	--	--	--	--	--	--	R812,S875
A816,E882	A816,E882	A816,E882	A816,E882	A816,E882	A816,E882	--	A816,E882
C-loop C-loop bonds							
R812,D813	R812,D813	R812,D813	R812,D813	R812,D813	R812,D813	R812,D813	--
--	--	R812,L814	R812,L814	R812,L814	--	R812,L814	--
--	--	--	--	--	--	--	D813,R817
--	--	--	--	--	D813,N818	D813,N818	--
A815,N818	A815,N818	A815,N818	A815,N818	A815,N818	A815,N818	A815,N818	--
--	--	A816,V819	A816,V819	A816,V819	A816,V819	--	--
A-loop A-loop bonds							
D831,K836	D831,K836	D831,K836	D831,K836	D831,K836	D831,K836	--	--
G833,K836	G833,K836	G833,K836	G833,K836	G833,K836	G833,K836	G833,K836	--
G839,Y845	--	--	--	--	--	--	--
--	E841,K843	E841,K843	E841,K843	--	--	--	--
--	--	--	--	K843,E844	--	--	--
--	--	--	--	--	H846,E848	--	--
E848,G849	E848,G849	E848,G849	E848,G849	E848,G849	--	E848,G849	--
--	--	E848,K851	E848,K851	--	--	--	--
--	--	--	--	--	--	G850,V852	--

Table B1: Exhaustive list of persistent H-bonds and salt bridges in EGFR kinases in an asymmetric dimer formation as it progresses from inactive to active. The salt bridges are in bold and identical bonds are aligned. Weaker bonds are in grey, to track the progression of bonds.

EGFR Monomer Inactive	Window 1	Window 2	Window 3	Window 4	Window 5	EGFR Monomer Active
aC-helix aC-helix bonds						
--	--	--	--	--	--	K730,E734
--	--	--	--	--	A731,I735	A731,I735
--	N732,S728	N732,S728	N732,S728	N732,S728	--	N732,S728
--	N732,L736	N732,L736	N732,L736	N732,L736	N732,L736	N732,L736
K733,D737	K733,D737	K733,D737	K733,D737	K733,D737	K733,D737	K733,D737
E734,E738	E734,E738	E734,E738	E734,E738	E734,E738	E734,E738	E734,E738
I735,A739	I735,A739	I735,A739	I735,A739	I735,A739	I735,A739	I735,A739
L736,Y740	L736,Y740	L736,Y740	L736,Y740	L736,Y740	L736,Y740	L736,Y740
--	D737,V741	--	--	D737,V741	D737,V741	--
E738,M742	E738,M742	E738,M742	E738,M742	E738,M742	E738,M742	E738,M742
--	A739,A743	A739,A743	A739,A743	A739,A743	A739,A743	--
Y740,S744	Y740,S744	--	--	Y740,S744	Y740,S744	Y740,S744
--	--	V741,V745	V741,V745	--	--	V741,V745
A-loop bonds						
--	--	--	--	--	--	A840,G672
--	--	--	--	E841,R808	--	--
--	--	--	--	--	--	K843,D932
--	--	--	--	--	Y845,Y867	Y845,Y867
--	--	--	--	--	H846,R865	--
E848,R865	E848,R865	E848,R865	E848,R865	E848,R865	E848,R865	--
--	--	--	--	--	E848,Y867	--
C-loop bonds						
H811,D872	H811,D872	H811,D872	H811,D872	H811,D872	H811,D872	H811,D872
--	--	--	--	--	--	R812,E848
R812,K851	--	--	R812,K851	R812,K851	--	--
R812,D872	R812,D872	R812,D872	R812,D872	R812,D872	R812,D872	R812,D872
--	--	--	--	--	--	R812,S875
--	--	--	--	A816,V819	--	--
--	A816,L775	A816,L775	A816,L775	--	A816,L775	--
A816,E882	--	--	--	A816,E882	A816,E882	A816,E882
--	R817,D776	R817,D776	R817,D776	R817,D776	--	--
C-loop C-loop bonds						
R812,D813	R812,D813	R812,D813	R812,D813	R812,D813	R812,D813	--
--	--	--	--	--	--	D813,R817
--	--	--	--	--	--	D813,N818
A815,N818	A815,N818	A815,N818	A815,N818	A815,N818	A815,N818	A815,N818
A-loop A-loop bonds						
G833,K836	G833,K836	G833,K836	G833,K836	--	--	--
E841,E842	E841,E842	--	--	--	--	--
E841,K843	E841,K843	--	--	--	--	--
--	--	E841,Y845	--	--	--	--
E844,Y845	--	--	--	--	--	--
--	--	--	--	--	--	Y845,E848
E848,G849	E848,G849	--	--	--	E848,G849	--
--	--	--	--	--	E848,G850	--

Table B2: Exhaustive list of persistent H-bonds and salt bridges in the EGFR kinase monomer as it progresses from inactive to active. The salt bridges are in bold and identical bonds are aligned. Weaker bonds are in grey, to track the progression of bonds.

References

- 1 Hubbard, S. R. and Till, J. H. (2000) PROTEIN TYROSINE KINASE STRUCTURE AND FUNCTION. *Annual Review of Biochemistry*. **69**, 373-398
- 2 Schlessinger, J. (2000) Cell Signaling by Receptor Tyrosine Kinases. *Cell*. **103**, 211-225
- 3 Citri, A. and Yarden, Y. (2006) EGF-ERBB signalling: towards the systems level. *Nat Rev Mol Cell Biol*. **7**, 505-516
- 4 Boudeau, J., Miranda-Saavedra, D., Barton, G. J. and Alessi, D. R. (2006) Emerging roles of pseudokinases. *Trends Cell Biol*. **16**, 443-452
- 5 Alimandi, M., Romano, A., Curia, M. C., Muraro, R., Fedi, P., Aaronson, S. A., Di Fiore, P. P. and Kraus, M. H. (1995) Cooperative signaling of ErbB3 and ErbB2 in neoplastic transformation and human mammary carcinomas. *Oncogene*. **10**, 1813-1821
- 6 Holbro, T., Beerli, R. R., Maurer, F., Koziczak, M., Barbas, C. F., 3rd and Hynes, N. E. (2003) The ErbB2/ErbB3 heterodimer functions as an oncogenic unit: ErbB2 requires ErbB3 to drive breast tumor cell proliferation. *Proc Natl Acad Sci U S A*. **100**, 8933-8938
- 7 Sorkin, A. and von Zastrow, M. (2009) Endocytosis and signalling: intertwining molecular networks. *Nat Rev Mol Cell Biol*. **10**, 609-622
- 8 Doherty, G. J. and McMahon, H. T. (2009) Mechanisms of Endocytosis. *Annual Review of Biochemistry*. **78**, 857-902
- 9 Ramanan, V., Agrawal, N. J., Liu, J., Engles, S., Toy, R. and Radhakrishnan, R. Systems Biology and Physical Biology of Clathrin-Mediated Endocytosis: An Integrative Experimental and Theoretical Perspective. Submitted to *Integrative Biology*
- 10 Vecchi, M., Baulida, J. and Carpenter, G. (1996) Selective Cleavage of the Heregulin Receptor ErbB-4 by Protein Kinase C Activation. *J. Biol. Chem*. **271**, 18989-18995
- 11 Gassmann, M. (1995) Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor. *Nature*. **378**, 390-394
- 12 Hahn, C. G., Wang, H. Y., Cho, D. S., Talbot, K., Gur, R. E., Berrettini, W. H., Bakshi, K., Kamins, J., Borgmann-Winter, K. E., Siegel, S. J., Gallop, R. J. and Arnold, S. E. (2006) Altered neuregulin 1-erbB4 signaling contributes to NMDA receptor hypofunction in schizophrenia. *Nat Med*. **12**, 824-828
- 13 Tonks, N. K. (2006) Protein tyrosine phosphatases: from genes, to function, to disease. *Nat Rev Mol Cell Biol*. **7**, 833-846
- 14 Reynolds, A. R., Tischer, C., Verveer, P. J., Rocks, O. and Bastiaens, P. I. H. (2003) EGFR activation coupled to inhibition of tyrosine phosphatases causes lateral signal propagation. *Nat Cell Biol*. **5**, 447-453
- 15 Tarcic, G., Boguslavsky, S. K., Wakim, J., Kiuchi, T., Liu, A., Reinitz, F., Nathanson, D., Takahashi, T., Mischel, P. S., Ng, T. and Yarden, Y. (2009) An Unbiased Screen Identifies DEP-1 Tumor Suppressor as a Phosphatase Controlling EGFR Endocytosis. *Current biology : CB*. **19**, 1788-1798
- 16 Bill, A., Schmitz, A., Albertoni, B., Song, J.-N., Heukamp, L. C., Walrafen, D., Thorwirth, F., Verveer, P. J., Zimmer, S., Meffert, L., Schreiber, A., Chatterjee, S., Thomas, R. K., Ullrich, R. T., Lang, T. and Famulok, M. (2010) Cytohesins Are Cytoplasmic ErbB Receptor Activators. *Cell*. **143**, 201-211
- 17 Inoue, A., Setoguchi, K., Matsubara, Y., Okada, K., Sato, N., Iwakura, Y., Higuchi, O. and Yamanashi, Y. (2009) Dok-7 Activates the Muscle Receptor Kinase MuSK and Shapes Synapse Formation. *Sci. Signal*. **2**, ra7-
- 18 Lemmon, M. A. and Schlessinger, J. (2010) Cell Signaling by Receptor Tyrosine Kinases. *Cell*. **141**, 1117-1134
- 19 Garrett, T. P. J., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Zhu, H.-J., Walker, F., Frenkel, M. J., Hoyne, P. A., Jorissen, R. N., Nice, E. C., Burgess, A. W. and Ward, C. W.

- (2002) Crystal Structure of a Truncated Epidermal Growth Factor Receptor Extracellular Domain Bound to Transforming Growth Factor [alpha]. *Cell*. **110**, 763-773
- 20 Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.-H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M. and Yokoyama, S. (2002) Crystal Structure of the Complex of Human Epidermal Growth Factor and Receptor Extracellular Domains. *Cell*. **110**, 775-787
- 21 Bouyain, S., Longo, P. A., Li, S., Ferguson, K. M. and Leahy, D. J. (2005) The extracellular region of ErbB4 adopts a tethered conformation in the absence of ligand. *Proc Natl Acad Sci U S A*. **102**, 15024-15029
- 22 Burgess, A. W., Cho, H. S., Eigenbrot, C., Ferguson, K. M., Garrett, T. P. J., Leahy, D. J., Lemmon, M. A., Sliwkowski, M. X., Ward, C. W. and Yokoyama, S. (2003) An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Molecular Cell*. **12**, 541-552
- 23 Cho, H. S. and Leahy, D. J. (2002) Structure of the extracellular region of HER3 reveals an interdomain tether. *Science*. **297**, 1330-1333
- 24 Ferguson, K. M., Berger, M. B., Mendrola, J. M., Cho, H.-S., Leahy, D. J. and Lemmon, M. A. (2003) EGF Activates Its Receptor by Removing Interactions that Autoinhibit Ectodomain Dimerization. *Molecular Cell*. **11**, 507-517
- 25 Stamos, J., Sliwkowski, M. X. and Eigenbrot, C. (2002) Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J Biol Chem*. **277**, 46265-46272
- 26 Choi, S. H., Mendrola, J. M. and Lemmon, M. A. (2007) EGF-independent activation of cell-surface EGF receptors harboring mutations found in gefitinib-sensitive lung cancer. *Oncogene*. **26**, 1567-1576
- 27 Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A. and Brannigan, B. W. (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med*. **350**, 2129-2139
- 28 Paez, J. G., Janne, P. A., Lee, J. C., Tracy, S., Greulich, H. and Gabriel, S. (2004) EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*. **304**, 1497-1500
- 29 Pao, W., Miller, V., Zakowski, M., Doherty, J., Politi, K. and Sarkaria, I. (2004) EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl Acad. Sci. USA*. **101**, 13306-13311
- 30 Kornev, A. P., Haste, N. M., Taylor, S. S. and Ten Eyck, L. F. (2006) Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proceedings of the National Academy of Sciences*. **103**, 17783-17788
- 31 Kornev, A. P., Taylor, S. S. and Ten Eyck, L. F. (2008) A helix scaffold for the assembly of active protein kinases. *Proceedings of the National Academy of Sciences*. **105**, 14377-14382
- 32 Hubbard, S. R., Wei, L., Ellis, L. and Hendrickson, W. A. (1994) Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature*. **372**, 746-754
- 33 Mohammadi, M., Schlessinger, J. and Hubbard, S. R. (1996) Structure of the FGF receptor tyrosine kinase domain reveals a novel autoinhibitory mechanism. *Cell*. **86**, 577-587
- 34 Till, J. H., Becerra, M., Watty, A., Lu, Y., Ma, Y., Neubert, T. A., Burden, S. J. and Hubbard, S. R. (2002) Crystal Structure of the MuSK Tyrosine Kinase: Insights into Receptor Autoregulation. *Structure*. **10**, 1187-1196
- 35 Griffith, J., Black, J., Faerman, C., Swenson, L., Wynn, M., Lu, F., Lippke, J. and Saxena, K. (2004) The structural basis for autoinhibition of FLT3 by the juxtamembrane domain. *Molecular Cell*. **13**, 169-178
- 36 Mol, C. D., Dougan, D. R., Schneider, T. R., Skene, R. J., Kraus, M. L., Scheibe, D. N., Snell, G. P., Zou, H., Sang, B. C. and Wilson, K. P. (2004) Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *Journal of Biological Chemistry*. **279**, 31655-31663
- 37 Wybenga-Groot, L. E., Baskin, B., Ong, S. H., Tong, J. F., Pawson, T. and Sicheri, F. (2001) Structural basis for autoinhibition of the EphB2 receptor tyrosine kinase by the unphosphorylated juxtamembrane region. *Cell*. **106**, 745-757

- 38 Shewchuk, L. M., Hassell, A. M., Ellis, B., Holmes, W. D., Davis, R., Horne, E. L., Kadwell, S. H., McKee, D. D. and Moore, J. T. (2000) Structure of the Tie2 RTK Domain: Self-Inhibition by the Nucleotide Binding Loop, Activation Loop, and C-Terminal Tail. *Structure*. **8**, 1105-1113
- 39 Knowles, P. P., Murray-Rust, J., Kjøfod, S., Scott, R. P., Hanrahan, S., Santoro, M., Ibáñez, C. F. and McDonald, N. Q. (2006) Structure and Chemical Inhibition of the RET Tyrosine Kinase Domain. *Journal of Biological Chemistry*. **281**, 33577-33587
- 40 Zhang, X., Gureasko, J., Shen, K., Cole, P. A. and Kuriyan, J. (2006) An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell*. **125**, 1137-1149
- 41 Zhang, X. W., Gureasko, J., Shen, K., Cole, P. A. and Kuriyan, J. (2006) An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell*. **125**, 1137-1149
- 42 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. and Leahy, D. J. (2008) Mechanism of Activation and Inhibition of the HER4/ErbB4 Kinase. *Structure*. **16**, 460-467
- 43 Jura, N., Endres, N. F., Engel, K., Deindl, S., Das, R., Lamers, M. H., Wemmer, D. E., Zhang, X. and Kuriyan, J. (2009) Mechanism for Activation of the EGF Receptor Catalytic Domain by the Juxtamembrane Segment. *Cell*. **137**, 1293-1307
- 44 Red Brewer, M., Choi, S. H., Alvarado, D., Moravcevic, K., Pozzi, A., Lemmon, M. A. and Carpenter, G. (2009) The Juxtamembrane Region of the EGF Receptor Functions as an Activation Domain. *Molecular Cell*. **34**, 641-651
- 45 Thiel, K. W. and Carpenter, G. (2007) Epidermal growth factor receptor juxtamembrane region regulates allosteric tyrosine kinase activation. *Proceedings of the National Academy of Sciences*. **104**, 19238-19243
- 46 Downward, J., Yarden, Y., Mayes, E., Scrace, G., Totty, N., Stockwell, P., Ullrich, A., Schlessinger, J. and Waterfield, M. D. (1984) Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences. *Nature*. **307**, 521-527
- 47 Forbes, S. A., Tang, G., Bindal, N., Bamford, S., Dawson, E., Cole, C., Kok, C. Y., Jia, M., Ewing, R., Menzies, A., Teague, J. W., Stratton, M. R. and Futreal, P. A. (2010) COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Research*. **38**, D652-D657
- 48 Sordella, R., Bell, D. W., Haber, D. A. and Settleman, J. (2004) Gefitinib-Sensitizing EGFR Mutations in Lung Cancer Activate Anti-Apoptotic Pathways. *Science*. **305**, 1163-1167
- 49 Willmore-Payne, C., Holden, J. A. and Layfield, L. J. (2006) Detection of EGFR- and HER2-activating mutations in squamous cell carcinoma involving the head and neck. *Mod Pathol*. **19**, 634-640
- 50 Ridd, K. and Bastian, B. C. (2009) Somatic Mutation of Epidermal Growth Factor Receptor in a Small Subset of Cutaneous Squamous Cell Carcinoma. *J Invest Dermatol*. **130**, 901-903
- 51 Da Silva, L., Simpson, P., Smart, C., Cocciardi, S., Waddell, N., Lane, A., Morrison, B., Vargas, A., Healey, S., Beesley, J., Pakkiri, P., Parry, S., Kurniawan, N., Reid, L., Keith, P., Faria, P., Pereira, E., Skalova, A., Bilous, M., Balleine, R., Do, H., Dobrovic, A., Fox, S., Franco, M., Reynolds, B., Khanna, K., Cummings, M., Chenevix-Trench, G. and Lakhani, S. (2010) HER3 and downstream pathways are involved in colonization of brain metastases from breast cancer. *Breast Cancer Research*. **12**, R46
- 52 de Muga, S., Hernandez, S., Agell, L., Salido, M., Juanpere, N., Lorenzo, M., Lorente, J. A., Serrano, S. and Lloreta, J. (2010) Molecular alterations of EGFR and PTEN in prostate cancer: association with high-grade and advanced-stage carcinomas. *Mod Pathol*. **23**, 703-712
- 53 Cho, K. S., Lee, J. S., Cho, N. H., Park, K., Ham, W. S. and Choi, Y. D. (2008) Gene amplification and mutation analysis of epidermal growth factor receptor in hormone refractory prostate cancer. *The Prostate*. **68**, 803-808
- 54 Lee, J. W., Soung, Y. H., Seo, S. H., Kim, S. Y., Park, C. H., Wang, Y. P., Park, K., Nam, S. W., Park, W. S., Kim, S. H., Lee, J. Y., Yoo, N. J. and Lee, S. H. (2006) Somatic mutations of ERBB2 kinase domain in gastric, colorectal, and breast carcinomas. *Clinical Cancer Research*. **12**, 57-61
- 55 Stephens, P., Hunter, C., Bignell, G., Edkins, S., Davies, H., Teague, J., Stevens, C., O'Meara, S., Smith, R., Parker, A., Barthorpe, A., Blow, M., Brackenbury, L., Butler, A., Clarke, O., Cole, J., Dicks, E., Dike, A., Drozd, A., Edwards, K., Forbes, S., Foster, R., Gray, K., Greenman, C., Halliday, K., Hills, K.,

- Kosmidou, V., Lugg, R., Menzies, A., Perry, J., Petty, R., Raine, K., Ratford, L., Shepherd, R., Small, A., Stephens, Y., Tofts, C., Varian, J., West, S., Widaa, S., Yates, A., Brasseur, F., Cooper, C. S., Flanagan, A. M., Knowles, M., Leung, S. Y., Louis, D. N., Looijenga, L. H., Malkowicz, B., Pierotti, M. A., Teh, B., Chenevix-Trench, G., Weber, B. L., Yuen, S. T., Harris, G., Goldstraw, P., Nicholson, A. G., Futreal, P. A., Wooster, R. and Stratton, M. R. (2004) Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature*. **431**, 525-526
- 56 Shigematsu, H., Takahashi, T., Nomura, M., Majmudar, K., Suzuki, M., Lee, H., Wistuba, I. I., Fong, K. M., Toyooka, S., Shimizu, N., Fujisawa, T., Minna, J. D. and Gazdar, A. F. (2005) Somatic Mutations of the HER2 Kinase Domain in Lung Adenocarcinomas. *Cancer Research*. **65**, 1642-1646
- 57 Buttitta, F., Barassi, F., Fresu, G., Felicioni, L., Chella, A., Paolizzi, D., Lattanzio, G., Salvatore, S., Campese, P. P., Rosini, S., Iarussi, T., Mucilli, F., Sacco, R., Mezzetti, A. and Marchetti, A. (2006) Mutational analysis of the HER2 gene in lung tumors from Caucasian patients: Mutations are mainly present in adenocarcinomas with bronchioloalveolar features. *International Journal of Cancer*. **119**, 2586-2591
- 58 Shah, S. P., Morin, R. D., Khattra, J., Prentice, L., Pugh, T., Burleigh, A., Delaney, A., Gelmon, K., Guliany, R., Senz, J., Steidl, C., Holt, R. A., Jones, S., Sun, M., Leung, G., Moore, R., Severson, T., Taylor, G. A., Teschendorff, A. E., Tse, K., Turashvili, G., Varhol, R., Warren, R. L., Watson, P., Zhao, Y., Caldas, C., Huntsman, D., Hirst, M., Marra, M. A. and Aparicio, S. (2009) Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*. **461**, 809-813
- 59 Soung, Y. H., Lee, J. W., Kim, S. Y., Wang, Y. P., Jo, K. H., Moon, S. W., Park, W. S., Nam, S. W., Lee, J. Y., Yoo, N. J. and Lee, S. H. (2006) Somatic mutations of the ERBB4 kinase domain in human cancers. *International Journal of Cancer*. **118**, 1426-1429
- 60 Prickett, T. D., Agrawal, N. S., Wei, X., Yates, K. E., Lin, J. C., Wunderlich, J. R., Cronin, J. C., Cruz, P., Rosenberg, S. A. and Samuels, Y. (2009) Analysis of the tyrosine kinome in melanoma reveals recurrent mutations in ERBB4. *Nat Genet*. **41**, 1127-1132
- 61 Ding, L., Getz, G., Wheeler, D. A., Mardis, E. R., McLellan, M. D., Cibulskis, K., Sougnez, C., Greulich, H., Muzny, D. M., Morgan, M. B., Fulton, L., Fulton, R. S., Zhang, Q., Wendl, M. C., Lawrence, M. S., Larson, D. E., Chen, K., Dooling, D. J., Sabo, A., Hawes, A. C., Shen, H., Jhangiani, S. N., Lewis, L. R., Hall, O., Zhu, Y., Mathew, T., Ren, Y., Yao, J., Scherer, S. E., Clerc, K., Metcalf, G. A., Ng, B., Milosavljevic, A., Gonzalez-Garay, M. L., Osborne, J. R., Meyer, R., Shi, X., Tang, Y., Koboldt, D. C., Lin, L., Abbott, R., Miner, T. L., Pohl, C., Fewell, G., Haiepek, C., Schmidt, H., Dunford-Shore, B. H., Kraja, A., Crosby, S. D., Sawyer, C. S., Vickery, T., Sander, S., Robinson, J., Winckler, W., Baldwin, J., Chirieac, L. R., Dutt, A., Fennell, T., Hanna, M., Johnson, B. E., Onofrio, R. C., Thomas, R. K., Tonon, G., Weir, B. A., Zhao, X., Ziaugra, L., Zody, M. C., Giordano, T., Orringer, M. B., Roth, J. A., Spitz, M. R., Wistuba, I. I., Ozenberger, B., Good, P. J., Chang, A. C., Beer, D. G., Watson, M. A., Ladanyi, M., Broderick, S., Yoshizawa, A., Travis, W. D., Pao, W., Province, M. A., Weinstock, G. M., Varmus, H. E., Gabriel, S. B., Lander, E. S., Gibbs, R. A., Meyerson, M. and Wilson, R. K. (2008) Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*. **455**, 1069-1075
- 62 Davies, H., Hunter, C., Smith, R., Stephens, P., Greenman, C., Bignell, G., Teague, J., Butler, A., Edkins, S., Stevens, C., Parker, A., O'Meara, S., Avis, T., Barthorpe, S., Brackenbury, L., Buck, G., Clements, J., Cole, J., Dicks, E., Edwards, K., Forbes, S., Gorton, M., Gray, K., Halliday, K., Harrison, R., Hills, K., Hinton, J., Jones, D., Kosmidou, V., Laman, R., Lugg, R., Menzies, A., Perry, J., Petty, R., Raine, K., Shepherd, R., Small, A., Solomon, H., Stephens, Y., Tofts, C., Varian, J., Webb, A., West, S., Widaa, S., Yates, A., Brasseur, F., Cooper, C. S., Flanagan, A. M., Green, A., Knowles, M., Leung, S. Y., Looijenga, L. H. J., Malkowicz, B., Pierotti, M. A., Teh, B. T., Yuen, S. T., Lakhani, S. R., Easton, D. F., Weber, B. L., Goldstraw, P., Nicholson, A. G., Wooster, R., Stratton, M. R. and Futreal, P. A. (2005) Somatic Mutations of the Protein Kinase Gene Family in Human Lung Cancer. *Cancer Research*. **65**, 7591-7595
- 63 Lierman, E., Michaux, L., Beullens, E., Pierre, P., Marynen, P., Cools, J. and Vandenberghe, P. (2009) FIP1L1-PDGFR[alpha] D842V, a novel panresistant mutant, emerging after treatment of FIP1L1-PDGFR[alpha] T674I eosinophilic leukemia with single agent sorafenib. *Leukemia*. **23**, 845-851

- 64 Ridge, S. A., Worwood, M., Oscier, D., Jacobs, A. and Padua, R. A. (1990) FMS mutations in myelodysplastic, leukemic, and normal subjects. *Proceedings of the National Academy of Sciences of the United States of America*. **87**, 1377-1380
- 65 Tobal, K., Pagliuca, A., Bhatt, B., Bailey, N., Layton, D. and Mufti, G. (1990) Mutation of the human FMS gene (M-CSF receptor) in myelodysplastic syndromes and acute myeloid leukemia. *Leukemia*. **4**, 486-489
- 66 Boldrini, L., Ursino, S., Gisfredi, S., Faviana, P., Donati, V., Camacci, T., Lucchi, M., Mussi, A., Basolo, F., Pingitore, R. and Fontanini, G. (2004) Expression and Mutational Status of c-kit in Small-Cell Lung Cancer. *Clinical Cancer Research*. **10**, 4101-4108
- 67 Willmore-Payne, C., Holden, J. A., Tripp, S. and Layfield, L. J. (2005) Human malignant melanoma: detection of BRAF- and c-kit-activating mutations by high-resolution amplicon melting analysis. *Human Pathology*. **36**, 486-493
- 68 Curtin, J. A., Busam, K., Pinkel, D. and Bastian, B. C. (2006) Somatic Activation of KIT in Distinct Subtypes of Melanoma. *Journal of Clinical Oncology*. **24**, 4340-4346
- 69 Beadling, C., Jacobson-Dunlop, E., Hodi, F. S., Le, C., Warrick, A., Patterson, J., Town, A., Harlow, A., Cruz, F., Azar, S., Rubin, B. P., Muller, S., West, R., Heinrich, M. C. and Corless, C. L. (2008) KIT Gene Mutations and Copy Number in Melanoma Subtypes. *Clinical Cancer Research*. **14**, 6821-6828
- 70 Torres-Cabala, C. A., Wang, W.-L., Trent, J., Yang, D., Chen, S., Galbincea, J., Kim, K. B., Woodman, S., Davies, M., Plaza, J. A., Nash, J. W., Prieto, V. G., Lazar, A. J. and Ivan, D. (2009) Correlation between KIT expression and KIT mutation in melanoma: a study of 173 cases with emphasis on the acral-lentiginous/mucosal type. *Mod Pathol*. **22**, 1446-1456
- 71 Furitsu, T., Tsujimura, T., Tono, T., Ikeda, H., Kitayama, H., Koshimizu, U., Sugahara, H., Butterfield, J. H., Ashman, L. K. and Kanayama, Y. (1993) Identification of mutations in the coding sequence of the proto-oncogene c-kit in a human mast cell leukemia cell line causing ligand-independent activation of c-kit product. *The Journal of Clinical Investigation*. **92**, 1736-1744
- 72 Kitayama, H., Kanakura, Y., Furitsu, T., Tsujimura, T., Oritani, K., Ikeda, H., Sugahara, H., Mitsui, H., Kanayama, Y. and Kitamura, Y. (1995) Constitutively activating mutations of c-kit receptor tyrosine kinase confer factor-independent growth and tumorigenicity of factor-dependent hematopoietic cell lines. *Blood*. **85**, 790-798
- 73 Nagata, H., Worobec, A. S., Oh, C. K., Chowdhury, B. A., Tannenbaum, S., Suzuki, Y. and Metcalfe, D. D. (1995) Identification of a point mutation in the catalytic domain of the protooncogene c-kit in peripheral blood mononuclear cells of patients who have mastocytosis with an associated hematologic disorder. *Proceedings of the National Academy of Sciences of the United States of America*. **92**, 10560-10564
- 74 Longley, B. J., Tyrrell, L., Lu, S.-Z., Ma, Y.-S., Langley, K., Ding, T.-g., Duffy, T., Jacobs, P., Tang, L. H. and Modlin, I. (1996) Somatic c-KIT activating mutation in urticaria pigmentosa and aggressive mastocytosis: establishment of clonality in a human mast cell neoplasm. *Nature Genetics*. **12**, 312 - 314
- 75 Nakao, M., Yokota, S., Iwai, T., Kaneko, H., Horiike, S., Kashima, K., Sonoda, Y., Fujimoto, T. and Misawa, S. (1996) Internal tandem duplication of the *flt3* gene found in acute myeloid leukemia. *Leukemia*. **10**, 1911-1918
- 76 Yokota, S., Kiyoi, H., Nakao, M., Iwai, T., Misawa, S., Okuda, T., Sonoda, Y., Abe, T., Kahsima, K., Matsuo, Y. and Naoe, T. (1997) Internal tandem duplication of the *FLT3* gene is preferentially seen in acute myeloid leukemia and myelodysplastic syndrome among various hematological malignancies. A study on a large series of patients and cell lines. *Leukemia*. **11**, 1605-1609
- 77 Kiyoi, H., Naoe, T., Yokota, S., Nakao, M., Minami, S., Kuriyama, K., Takeshita, A., Saito, K., Hasegawa, S., Shimodaira, S., Tamura, J., Shimazaki, C., Matsue, K., Kobayashi, H., Arima, N., Suzuki, R., Morishita, H., Saito, H., Ueda, R. and Ohno, R. (1997) Internal tandem duplication of *FLT3* associated with leukocytosis in acute promyelocytic leukemia. *Leukemia Study Group of the Ministry of Health and Welfare (Kohseisho)*. *Leukemia*. **11**, 1447-1452
- 78 Yamamoto, Y., Kiyoi, H., Nakano, Y., Suzuki, R., Kodaera, Y., Miyawaki, S., Asou, N., Kuriyama, K., Yagasaki, F., Shimazaki, C., Akiyama, H., Saito, K., Nishimura, M., Motoji, T., Shinagawa, K.,

- Takeshita, A., Saito, H., Ueda, R., Ohno, R. and Naoe, T. (2001) Activating mutation of D835 within the activation loop of FLT3 in human hematologic malignancies. *Blood*. **97**, 2434-2439
- 79 Abu-Duhier, F. M., Goodeve, A. C., Wilson, G. A., Care, R. S., Peake, I. R. and Reilly, J. T. (2001) Identification of novel FLT-3 Asp835 mutations in adult acute myeloid leukaemia. *British Journal of Haematology*. **113**, 983-988
- 80 Antonescu, C. R., Yoshida, A., Guo, T., Chang, N.-E., Zhang, L., Agaram, N. P., Qin, L.-X., Brennan, M. F., Singer, S. and Maki, R. G. (2009) KDR Activating Mutations in Human Angiosarcomas Are Sensitive to Specific Kinase Inhibitors. *Cancer Research*. **69**, 7175-7179
- 81 Gartside, M. G., Chen, H., Ibrahim, O. A., Byron, S. A., Curtis, A. V., Wellens, C. L., Bengston, A., Yudit, L. M., Eliseenkova, A. V., Ma, J., Curtin, J. A., Hyder, P., Harper, U. L., Riedesel, E., Mann, G. J., Trent, J. M., Bastian, B. C., Meltzer, P. S., Mohammadi, M. and Pollock, P. M. (2009) Loss-of-Function Fibroblast Growth Factor Receptor-2 Mutations in Melanoma. *Molecular Cancer Research*. **7**, 41-54
- 82 Logie, A., Dunois-Larde, C., Rosty, C., Levrel, O., Blanche, M., Ribeiro, A., Gasc, J.-M., Jorcano, J., Werner, S., Sastre-Garau, X., Thiery, J. P. and Radvanyi, F. (2005) Activating mutations of the tyrosine kinase receptor FGFR3 are associated with benign skin tumors in mice and humans. *Human Molecular Genetics*. **14**, 1153-1160
- 83 Hafner, C., van Oers, J. M. M., Hartmann, A., Landthaler, M., Stoehr, R., Blaszyk, H., Hofstaedter, F., Zwarthoff, E. C. and Vogt, T. (2006) High Frequency of FGFR3 Mutations in Adenoid Seborrhic Keratoses. *J Invest Dermatol*. **126**, 2404-2407
- 84 Hafner, C., Hartmann, A., Real, F. X., Hofstaedter, F., Landthaler, M. and Vogt, T. (2007) Spectrum of FGFR3 Mutations in Multiple Intraindividual Seborrhic Keratoses. *J Invest Dermatol*. **127**, 1883-1885
- 85 Chesi, M., Nardini, E., Brents, L. A., Schröck, E., Ried, T., Kuehl, W. M. and Bergsagel, P. L. (1997) Frequent translocation t(4;14)(p16.3;q32.3) in multiple myeloma is associated with increased expression and activating mutations of fibroblast growth factor receptor 3. *Oncogene*. **16**, 260 - 264
- 86 Intini, D., Baldini, L., Fabris, S., Lombardi, L., Ciceri, G., Maiolo, A. T. and Neri, A. (2001) Analysis of FGFR3 gene mutations in multiple myeloma patients with t(4;14). *British Journal of Haematology*. **114**, 362-364
- 87 Ronchetti, D., Greco, A., Compasso, S., Colombo, G., Dell'Era, P., Otsuki, T., Lombardi, L. and Neri, A. (2001) Deregulated FGFR3 mutants in multiple myeloma cell lines with t(4;14): comparative analysis of Y373C, K650E and the novel G384D mutations. *Oncogene*. **20**, 3553-3562
- 88 Ma, P. C., Kijima, T., Maulik, G., Fox, E. A., Sattler, M., Griffin, J. D., Johnson, B. E. and Salgia, R. (2003) c-MET Mutational Analysis in Small Cell Lung Cancer. *Cancer Research*. **63**, 6272-6281
- 89 Ma, P. C., Jagadeeswaran, R., Jagadeesh, S., Tretiakova, M. S., Nallasura, V., Fox, E. A., Hansen, M., Schaefer, E., Naoki, K., Lader, A., Richards, W., Sugarbaker, D., Husain, A. N., Christensen, J. G. and Salgia, R. (2005) Functional Expression and Mutations of c-Met and Its Therapeutic Inhibition with SU11274 and Small Interfering RNA in Non-Small Cell Lung Cancer. *Cancer Research*. **65**, 1479-1488
- 90 Kong-Beltran, M., Seshagiri, S., Zha, J., Zhu, W., Bhawe, K., Mendoza, N., Holcomb, T., Pujara, K., Stinson, J., Fu, L., Severin, C., Rangell, L., Schwall, R., Amler, L., Wickramasinghe, D. and Yauch, R. (2006) Somatic Mutations Lead to an Oncogenic Deletion of Met in Lung Cancer. *Cancer Research*. **66**, 283-289
- 91 Lorenzato, A., Olivero, M., Patan , S., Rosso, E., Oliaro, A., Comoglio, P. M. and Di Renzo, M. F. (2002) Novel Somatic Mutations of the MET Oncogene in Human Carcinoma Metastases Activating Cell Motility and Invasion. *Cancer Research*. **62**, 7025-7030
- 92 Franco-Hernandez, C., Martinez-Glez, V., Arjona, D., de Campos, J. M., Isla, A., Gutierrez, M., Vaquero, J. and Rey, J. A. (2007) EGFR sequence variations and real-time quantitative polymerase chain reaction analysis of gene dosage in brain metastases of solid tumors. *Cancer Genetics and Cytogenetics*. **173**, 63-67
- 93 Iyevleva, A. G., Novik, A. V., Moiseyenko, V. M. and Imyanitov, E. N. (2009) EGFR mutation in kidney carcinoma confers sensitivity to gefitinib treatment: A case report. *Urologic Oncology: Seminars and Original Investigations*. **27**, 548-550

- 94 Masago, K., Asato, R., Fujita, S., Hirano, S., Tamura, Y., Kanda, T., Mio, T., Katakami, N., Mishima, M. and Ito, J. (2009) Epidermal growth factor receptor gene mutations in papillary thyroid carcinoma. *International Journal of Cancer*. **124**, 2744-2749
- 95 Mitsiades, C. S., Kotoula, V., Poulaki, V., Sozopoulos, E., Negri, J., Charalambous, E., Fanourakis, G., Voutsinas, G., Tseleni-Balafouta, S. and Mitsiades, N. (2006) Epidermal Growth Factor Receptor as a Therapeutic Target in Human Thyroid Carcinoma: Mutational and Functional Analysis. *J Clin Endocrinol Metab*. **91**, 3662-3666
- 96 Anglesio, M. S., Arnold, J. M., George, J., Tinker, A. V., Tothill, R., Waddell, N., Simms, L., Locandro, B., Fereday, S., Traficante, N., Russell, P., Sharma, R., Birrer, M. J., Group, A. S., deFazio, A., Chenevix-Trench, G. and Bowtell, D. D. L. (2008) Mutation of ERBB2 Provides a Novel Alternative Mechanism for the Ubiquitous Activation of RAS-MAPK in Ovarian Serous Low Malignant Potential Tumors. *Molecular Cancer Research*. **6**, 1678-1690
- 97 Lassus, H., Sihto, H., Leminen, A., Joensuu, H., Isola, J., Nupponen, N. and Butzow, R. (2006) Gene amplification, mutation, and protein expression of <i>EGFR</i> and mutations of <i>ERBB2</i> in serous ovarian carcinoma. *Journal of Molecular Medicine*. **84**, 671-681
- 98 Nakayama, K., Nakayama, N., Kurman, R., Cope, L., Pohl, G., Samuels, Y., Velculescu, V., Wang, T. and Shih, I. (2006) Sequence mutations and amplification of PIK3CA and AKT2 genes in purified ovarian serous neoplasms. *Cancer Biol Ther*. **5**, 779-785
- 99 Heinrich, M. C., Corless, C. L., Duensing, A., McGreevey, L., Chen, C.-J., Joseph, N., Singer, S., Griffith, D. J., Haley, A., Town, A., Demetri, G. D., Fletcher, C. D. M. and Fletcher, J. A. (2003) PDGFRA Activating Mutations in Gastrointestinal Stromal Tumors. *Science*. **299**, 708-710
- 100 Heinrich, M. C., Corless, C. L., Demetri, G. D., Blanke, C. D., von Mehren, M., Joensuu, H., McGreevey, L. S., Chen, C.-J., Van den Abbeele, A. D., Druker, B. J., Kiese, B., Eisenberg, B., Roberts, P. J., Singer, S., Fletcher, C. D. M., Silberman, S., Dimitrijevic, S. and Fletcher, J. A. (2003) Kinase Mutations and Imatinib Response in Patients With Metastatic Gastrointestinal Stromal Tumor. *Journal of Clinical Oncology*. **21**, 4342-4349
- 101 Hirota, S., Ohashi, A., Nishida, T., Isozaki, K., Kinoshita, K., Shinomura, Y. and Kitamura, Y. (2003) Gain-of-function mutations of platelet-derived growth factor receptor [alpha] gene in gastrointestinal stromal tumors. *Gastroenterology*. **125**, 660-667
- 102 Hoei-Hansen, C., Kraggerud, S., Abeler, V., Kaern, J., Rajpert-De Meyts, E. and Lothe, R. (2007) Ovarian dysgerminomas are characterised by frequent KIT mutations and abundant expression of pluripotency markers. *Molecular Cancer*. **6**, 12
- 103 Tian, Q., Frierson, H. F., Jr., Krystal, G. W. and Moskaluk, C. A. (1999) Activating c-kit Gene Mutations in Human Germ Cell Tumors. *Am J Pathol*. **154**, 1643-1647
- 104 Hirota, S., Isozaki, K., Moriyama, Y., Hashimoto, K., Nishida, T., Ishiguro, S., Kawano, K., Hanada, M., Kurata, A., Takeda, M., Muhammad Tunio, G., Matsuzawa, Y., Kanakura, Y., Shinomura, Y. and Kitamura, Y. (1998) Gain-of-Function Mutations of c-kit in Human Gastrointestinal Stromal Tumors. *Science*. **279**, 577-580
- 105 Lasota, J., Jasinski, M., Sarlomo-Rikala, M. and Miettinen, M. (1999) Mutations in Exon 11 of c-KIT Occur Preferentially in Malignant versus Benign Gastrointestinal Stromal Tumors and Do Not Occur in Leiomyomas or Leiomyosarcomas. *Am J Pathol*. **154**, 53-60
- 106 Moskaluk, C., Tian, Q., Marshall, C., Rumpel, C., Franquemont, D. and Frierson, H. (1999) Mutations of c-kit JM domain are found in a minority of human gastrointestinal stromal tumors. *Oncogene*. **18**, 1897-1902
- 107 Taniguchi, M., Nishida, T., Hirota, S., Isozaki, K., Ito, T., Nomura, T., Matsuda, H. and Kitamura, Y. (1999) Effect of c-kit Mutation on Prognosis of Gastrointestinal Stromal Tumors. *Cancer Research*. **59**, 4297-4300
- 108 Pollock, P. M., Gartside, M. G., Dejeza, L. C., Powell, M. A., Mallon, M. A., Davies, H., Mohammadi, M., Futreal, P. A., Stratton, M. R., Trent, J. M. and Goodfellow, P. J. (2007) Frequent activating FGFR2 mutations in endometrial carcinomas parallel germline mutations associated with craniosynostosis and skeletal dysplasia syndromes. *Oncogene*. **26**, 7158-7162

- 109 Schmidt, L., Duh, F.-M., Chen, F., Kishida, T., Glenn, G., Choyke, P., Scherer, S. W., Zhuang, Z., Lubensky, I., Dean, M., Allikmets, R., Chidambaram, A., Bergerheim, U. R., Feltis, J. T., Casadevall, C., Zamarron, A., Bernues, M., Richard, S., Lips, C. J. M., Walther, M. M., Tsui, L.-C., Geil, L., Orcutt, M. L., Stackhouse, T., Lipan, J., Slife, L., Brauch, H., Decker, J., Niehans, G., Hughson, M. D., Moch, H., Storkel, S., Lerman, M. I., Linehan, W. M. and Zbar, B. (1997) Germline and somatic mutations in the tyrosine kinase domain of the MET proto-oncogene in papillary renal carcinomas. *Nat Genet.* **16**, 68-73
- 110 Park, W. S., Dong, S. M., Kim, S. Y., Na, E. Y., Shin, M. S., Pi, J. H., Kim, B. J., Bae, J. H., Hong, Y. K., Lee, K. S., Lee, S. H., Yoo, N. J., Jang, J. J., Pack, S., Zhuang, Z., Schmidt, L., Zbar, B. and Lee, J. Y. (1999) Somatic Mutations in the Kinase Domain of the Met/Hepatocyte Growth Factor Receptor Gene in Childhood Hepatocellular Carcinomas. *Cancer Research.* **59**, 307-310
- 111 Wasenius, V.-M., Hemmer, S., Karjalainen-Lindsberg, M.-L., Nupponen, N. N., Franssila, K. and Joensuu, H. (2005) MET Receptor Tyrosine Kinase Sequence Alterations in Differentiated Thyroid Carcinoma. *The American Journal of Surgical Pathology.* **29**, 544-549
- 112 Ricarte-Filho, J. C., Ryder, M., Chitale, D. A., Rivera, M., Heguy, A., Ladanyi, M., Janakiraman, M., Solit, D., Knauf, J. A., Tuttle, R. M., Ghossein, R. A. and Fagin, J. A. (2009) Mutational Profile of Advanced Primary and Metastatic Radioactive Iodine-Refractory Thyroid Cancers Reveals Distinct Pathogenetic Roles for BRAF, PIK3CA, and AKT1. *Cancer Research.* **69**, 4885-4893
- 113 Hofstra, R. M. W., Landsvater, R. M., Ceccherini, I., Stulp, R. P., Stelwagen, T., Luo, Y., Pasini, B., Hoppener, J. W. M., van Amstel, H. K. P., Romeo, G., Lips, C. J. M. and Buys, C. H. C. M. (1994) A mutation in the RET proto-oncogene associated with multiple endocrine neoplasia type 2B and sporadic medullary thyroid carcinoma. *Nature.* **367**, 375-376
- 114 Zedenius, J., Wallin, G., Hamberger, B., Nordenskjöld, M., Weber, G. n. and Larsson, C. (1994) Somatic and MEN 2A de novo mutations identified in the RET proto-oncogene by screening of sporadic MTC: s. *Human Molecular Genetics.* **3**, 1259-1262
- 115 Eng, C., Mulligan, L. M., Smith, D. P., Healey, C. S., Frilling, A., Raue, F., Neumann, H. P. H., Pfragner, R., Behmel, A., Lorenzo, M. J., Stonehouse, T. J., Ponder, M. A. and Ponder, B. A. J. (1995) Mutation of the RET protooncogene in sporadic medullary thyroid carcinoma. *Genes, Chromosomes and Cancer.* **12**, 209-212
- 116 Romei, C., Elisei, R., Pinchera, A., Ceccherini, I., Molinaro, E., Mancusi, F., Martino, E., Romeo, G. and Pacini, F. (1996) Somatic mutations of the ret protooncogene in sporadic medullary thyroid carcinoma are not restricted to exon 16 and are associated with tumor recurrence. *J Clin Endocrinol Metab.* **81**, 1619-1622
- 117 Alemi, M., Lucas, S., Sällström, J., Bergholm, U., Akerström, G. and Wilander, E. (1997) A complex nine base pair deletion in RET exon 11 common in sporadic medullary thyroid carcinoma. *Oncogene.* **14**, 2041-2045
- 118 Uchino, S., Noguchi, S., Yamashita, H., Sato, M., Adachi, M., Yamashita, H., Watanabe, S., Ohshima, A., Mitsuyama, S., Iwashita, T. and Takahashi, M. (1999) Somatic Mutations in RET Exons 12 and 15 in Sporadic Medullary Thyroid Carcinomas: Different Spectrum of Mutations in Sporadic Type from Hereditary Type. *Cancer Science.* **90**, 1231-1237
- 119 Chen, Y., Takita, J., Choi, Y. L., Kato, M., Ohira, M., Sanada, M., Wang, L., Soda, M., Kikuchi, A., Igarashi, T., Nakagawara, A., Hayashi, Y., Mano, H. and Ogawa, S. (2008) Oncogenic mutations of ALK kinase in neuroblastoma. *Nature.* **455**, 971-974
- 120 George, R. E., Sanda, T., Hanna, M., Frohling, S., Ii, W. L., Zhang, J., Ahn, Y., Zhou, W., London, W. B., McGrady, P., Xue, L., Zozulya, S., Gregor, V. E., Webb, T. R., Gray, N. S., Gilliland, D. G., Diller, L., Greulich, H., Morris, S. W., Meyerson, M. and Look, A. T. (2008) Activating mutations in ALK provide a therapeutic target in neuroblastoma. *Nature.* **455**, 975-978
- 121 Janoueix-Lerosey, I., Lequin, D., Brugieres, L., Ribeiro, A., de Pontual, L., Combaret, V., Raynal, V., Puisieux, A., Schleiermacher, G., Pierron, G., Valteau-Couanet, D., Frebourg, T., Michon, J., Lyonnet, S., Amiel, J. and Delattre, O. (2008) Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. *Nature.* **455**, 967-970
- 122 Mosse, Y. P., Laudenslager, M., Longo, L., Cole, K. A., Wood, A., Attiyeh, E. F., Laquaglia, M. J., Sennett, R., Lynch, J. E., Perri, P., Laureys, G., Speleman, F., Kim, C., Hou, C., Hakonarson, H.,

- Torkamani, A., Schork, N. J., Brodeur, G. M., Tonini, G. P., Rappaport, E., Devoto, M. and Maris, J. M. (2008) Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature*. **455**, 930-935
- 123 Citri, A., Gan, J., Mosesson, Y., Vereb, G., Szollosi, J. and Yarden, Y. (2004) Hsp90 restrains ErbB-2/HER2 signalling by limiting heterodimer formation. *EMBO Rep*. **5**, 1165-1170
- 124 Xu, W., Yuan, X., Xiang, Z., Mimnaugh, E., Marcu, M. and Neckers, L. (2005) Surface charge and hydrophobicity determine ErbB2 binding to the Hsp90 chaperone complex. *Nat Struct Mol Biol*. **12**, 120-126
- 125 Tvorogov, D., Sundvall, M., Kurppa, K., Hollmen, M., Repo, S., Johnson, M. S. and Elenius, K. (2009) Somatic Mutations of ErbB4. *Journal of Biological Chemistry*. **284**, 5582-5591
- 126 Weinstein, I. B. (2002) Cancer. Addiction to oncogenes--the Achilles heel of cancer. *Science*. **297**, 63-64
- 127 Kris, M. G., Natale, R. B., Herbst, R. S., Lynch, T. J., Prager, D., Belani, C. P., Schiller, J. H., Kelly, K., Spiridonidis, H., Sandler, A., Albain, K. S., Cella, D., Wolf, M. K., Averbuch, S. D., Ochs, J. J. and Kay, A. C. (2003) Efficacy of Gefitinib, an Inhibitor of the Epidermal Growth Factor Receptor Tyrosine Kinase, in Symptomatic Patients With Non-Small Cell Lung Cancer. *JAMA: The Journal of the American Medical Association*. **290**, 2149-2158
- 128 Fukuoka, M., Yano, S., Giaccone, G., Tamura, T., Nakagawa, K., Douillard, J.-Y., Nishiwaki, Y., Vansteenkiste, J., Kudoh, S., Rischin, D., Eek, R., Horai, T., Noda, K., Takata, I., Smit, E., Averbuch, S., Macleod, A., Feyereislova, A., Dong, R.-P. and Baselga, J. (2003) Multi-Institutional Randomized Phase II Trial of Gefitinib for Previously Treated Patients With Advanced Non-Small-Cell Lung Cancer. *Journal of Clinical Oncology*. **21**, 2237-2246
- 129 Thatcher, N., Chang, A., Parikh, P., Rodrigues Pereira, J., Ciuleanu, T., von Pawel, J., Thongprasert, S., Tan, E. H., Pemberton, K., Archer, V. and Carroll, K. (2005) Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *The Lancet*. **366**, 1527-1537
- 130 Niho, S., Ichinose, Y., Tamura, T., Yamamoto, N., Tsuboi, M., Nakagawa, K., Shinkai, T., Jiang, H., Nishiwaki, Y. and Fukuoka, M. (2007) Results of a randomized phase III study to compare the overall survival of gefitinib (IRESSA) versus docetaxel in Japanese patients with non-small cell lung cancer who failed one or two chemotherapy regimens. *J Clin Oncol (Meeting Abstracts)*. **25**, LBA7509-
- 131 Douillard, J.-Y., Kim, E., Hirsh, V., Mok, T., Socinski, M., Gervais, R., Wu, Y.-L., Li, L., Sellers, M. and Lowe, E. (2007) Gefitinib (IRESSA) versus docetaxel in patients with locally advanced or metastatic non-small-cell lung cancer pre-treated with platinum-based chemotherapy: a randomized, open-label Phase III study (INTEREST): PRS-02. *Journal of Thoracic Oncology*. **2**, S305-S306
310.1097/1001.JTO.0000283087.0000271346.0000283019
- 132 Pao, W., Miller, V. A., Politi, K. A., Riely, G. J., Somwar, R., Zakowski, M. F., Kris, M. G. and Varmus, H. (2005) Acquired Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib Is Associated with a Second Mutation in the EGFR Kinase Domain. *PLoS Med*. **2**, e73
- 133 Blencke, S., Ullrich, A. and Daub, H. (2003) Mutation of Threonine 766 in the Epidermal Growth Factor Receptor Reveals a Hotspot for Resistance Formation against Selective Tyrosine Kinase Inhibitors. *Journal of Biological Chemistry*. **278**, 15435-15440
- 134 Frey, K. M., Georgiev, I., Donald, B. R. and Anderson, A. C. (2010) Predicting resistance mutations using protein design algorithms. *Proceedings of the National Academy of Sciences*. **107**, 13707-13712
- 135 Shih, A. J., Telesco, S. E., Liu, Y., Venkatramani, R. and Radhakrishnan, R. (2010) *Computational methods related to reaction chemistry*. Elsevier London
- 136 Redondo, A. and LeSar, R. (2004) MODELING AND SIMULATION OF BIOMATERIALS. *Annual Review of Materials Research*. **34**, 279-314
- 137 Giuseppina, R. and Fabio, G. (2007) Understanding the Performance of Biomaterials through Molecular Modeling: Crossing the Bridge between their Intrinsic Properties and the Surface Adsorption of Proteins. *Macromolecular Bioscience*. **7**, 552-566

- 138 Karplus, M. and Kuriyan, J. (2005) Molecular dynamics and protein function. *Proc Natl Acad Sci U S A.* **102**, 6679-6685
- 139 Karplus, M., Brunger, A. T., Elber, R. and Kuriyan, J. (1987) Molecular dynamics: applications to proteins. *Cold Spring Harb Symp Quant Biol.* **52**, 381-390
- 140 Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235-242
- 141 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. and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B.* **102**, 3586-3616
- 142 Ferrara, P., Apostolakis, J. and Caflisch, A. (2000) Computer simulations of protein folding by targeted molecular dynamics. *Proteins.* **39**, 252-260
- 143 (2000) All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comp. Chem.* **21**, 105-120
- 144 (1999) A modified version of the Cornell \textit{et al.} force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* **16**, 845-862
- 145 Wang, W., Donini, O., Reyes, C. M. and Kollman, P. A. (2001) Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 211-243
- 146 Glenn, J. M., Douglas, J. T. and Michael, L. K. (1994) Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics.* **101**, 4177-4189
- 147 Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M. (1983) Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry.* **4**, 187-217
- 148 Weiner, P. W. and Kollman, P. A. (1981) AMBER: assisted model building with energy refinement. *J. Comput. Chem.* **2**, 287-303
- 149 Scott, W. R. P. (1999) The GROMOS biomolecular simulation program package. *J. Phys. Chem. A.* **103**, 3596-3607
- 150 Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L. and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry.* **26**, 1781-1802
- 151 Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics.* **14**, 33--38
- 152 Shih, A. J., Purvis, J. and Radhakrishnan, R. (2008) Molecular Systems Biology of ErbB1 Signaling: Bridging the Gap through Multiscale Modeling and High-Performance Computing Molecular Biosystems. **4**, 1151-1159
- 153 Amadei, A., Linssen, A. B. M. and Berendsen, H. J. C. (1993) Essential Dynamics of Proteins. *Proteins-Structure Function and Genetics.* **17**, 412-425
- 154 Amadei, A., Linssen, A. B. M., deGroot, B. L., Aalten, D. M. F. v. and Berendsen, H. J. C. (1996) An Efficient Method for Sampling the Essential Subspace of Proteins. *J. Biomol. Struct. Dynam.* **13**, 615-625
- 155 McCammon, J. A. and Harvey, S. C. (1987) *Dynamics of Proteins and Nucleic Acids.* Cambridge University Press, Cambridge, MA
- 156 McCammon, J. A., Gelin, B. R. and Karplus, M. (1977) Dynamics of Folded Proteins. *Nature.* **267**, 585-590
- 157 Essmann, U. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577-8593
- 158 Gota, K., Rossen, A., Narutoshi, K., Makoto, T., Ryutaro, H., Haruki, N. and Yasushige, Y. (2009) Application of MDGRAPE-3, a special purpose board for molecular dynamics simulations, to periodic biomolecular systems. *Journal of Computational Chemistry.* **30**, 110-118

- 159 Suenaga, A., Hatakeyama, M., Kiyatkin, A. B., Radhakrishnan, R., Taiji, M. and Kholodenko, B. N. (2009) Molecular Dynamics Simulations Reveal that Tyr-317 Phosphorylation Reduces Shc Binding Affinity for Phosphotyrosyl Residues of Epidermal Growth Factor Receptor. *96*, 2278-2288
- 160 Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., Young, C., Batson, B., Bowers, K. J., Chao, J. C., Eastwood, M. P., Gagliardo, J., Grossman, J. P., Ho, C. R., Ierardi, D. J., Kolossvary, I., Klepeis, J. L., Layman, T., McLeavey, C., Moraes, M. A., Mueller, R., Priest, E. C., Shan, Y. B., Spengler, J., Theobald, M., Towles, B. and Wang, S. C. (2008) Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM.* **51**, 91-97
- 161 Shan, Y., Seeliger, M. A., Eastwood, M. P., Frank, F., Xu, H., Jensen, M. O., Dror, R. O., Kuriyan, J. and Shaw, D. E. (2009) A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc Natl Acad Sci U S A.* **106**, 139-144
- 162 Friedrichs, M. S., Eastman, P., Vaidyanathan, V., Houston, M., Legrand, S., Beberg, A. L., Ensign, D. L., Bruns, C. M. and Pande, V. S. (2009) Accelerating molecular dynamic simulation on graphics processing units. *J Comput Chem.* **30**, 864-872
- 163 Stone, J. E., Phillips, J. C., Freddolino, P. L., Hardy, D. J., Trabuco, L. G. and Schulten, K. (2007) Accelerating molecular modeling applications with graphics processors. *J Comput Chem.* **28**, 2618-2640
- 164 Bajorath, J., Stenkamp, R. and Aruffo, A. (1993) Knowledge-based model building of proteins: concepts and examples. *Protein Sci.* **2**, 1798-1810
- 165 Baker, D. (2000) A surprising simplicity to protein folding. *Nature.* **405**, 39-42
- 166 Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F. and Sali, A. (2000) Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure.* **29**, 291-325
- 167 Kryzhafovych, A. and Fidelis, K. (2009) Protein structure prediction and model quality assessment. *Drug Discov Today.* **14**, 386-393
- 168 Dunbrack, R. L., Jr. (2006) Sequence comparison and protein structure prediction. *Curr Opin Struct Biol.* **16**, 374-384
- 169 Ginalski, K. (2006) Comparative modeling for protein structure prediction. *Curr Opin Struct Biol.* **16**, 172-177
- 170 Moulton, J. (2008) Comparative modeling in structural genomics. *Structure.* **16**, 14-16
- 171 Schueler-Furman, O., Wang, C., Bradley, P., Misura, K. and Baker, D. (2005) Progress in modeling of protein structures and interactions. *Science.* **310**, 638-642
- 172 Fiser, A. and Sali, A. (2003) Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* **374**, 461-491
- 173 Blake, J. D. and Cohen, F. E. (2001) Pairwise sequence alignment below the twilight zone. *J Mol Biol.* **307**, 721-735
- 174 Dewey, T. G. (2001) A sequence alignment algorithm with an arbitrary gap penalty function. *J Comput Biol.* **8**, 177-190
- 175 Sanchez, R. and Sali, A. (1998) Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. *Proc Natl Acad Sci U S A.* **95**, 13597-13602
- 176 Aszodi, A. and Taylor, W. R. (1996) Homology modelling by distance geometry. *Fold Des.* **1**, 325-334
- 177 Havel, T. F. and Snow, M. E. (1991) A new method for building protein conformations from sequence alignments with homologues of known structure. *J Mol Biol.* **217**, 1-7
- 178 Srinivasan, S., March, C. J. and Sudarsanam, S. (1993) An automated method for modeling proteins on known templates using distance geometry. *Protein Sci.* **2**, 277-289
- 179 Fiser, A., Do, R. K. and Sali, A. (2000) Modeling of loops in protein structures. *Protein Sci.* **9**, 1753-1773
- 180 Fidelis, K., Stern, P. S., Bacon, D. and Moulton, J. (1994) Comparison of systematic search and database methods for constructing segments of protein structure. *Protein Eng.* **7**, 953-960
- 181 Laskowski, R. A., M.W. MacArthur, D.S. Moss, and J.M. Thornton. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst.* **26**, 283-291

- 182 Hooft, R. W., Vriend, G., Sander, C. and Abola, E. E. (1996) Errors in protein structures. *Nature*. **381**, 272
- 183 Luthy, R., Bowie, J. U. and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. *Nature*. **356**, 83-85
- 184 Xiang, Z. (2006) Advances in homology protein structure modeling. *Curr Protein Pept Sci*. **7**, 217-227
- 185 Ponder, J. W., Richard, F.M. (1987) Tertiary templates for protein use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **193**, 775-791
- 186 Xiang, Z. W. (2001) Extending the accuracy limit of side-chain prediction. *J. Mol. Biol.* **311**, 421-430
- 187 Zwanzig, R. W. (1954) High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics*. **22**, 1420-1426
- 188 Frenkel, D. and Smit, B. (1996) *Understanding Molecular Simulations. From Algorithms to Applications*. Academic Press, San Diego, CA
- 189 Beveridge, D. L. and DiCapua, F. M. (1989) Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu Rev Biophys Biophys Chem*. **18**, 431-492
- 190 Straatsma, T. P., Berendsen, H. J. C., Postma, J. P. M. (1986) *J. Chem. Phys.* **85**, 6720
- 191 Beutler, T. C., A.E. Mark, R.C. van Schaik, P.R. Gerber, and W.F. van Gunsteren. (1994) Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem Phys Lett*. **222**, 529-539
- 192 Pitera, J. W., and W.F. van Gunsteren. (2002) A comparison of non-bonded scaling approaches for free energy calculations. *Mol Sim*. **28**, 45-65
- 193 Florian, J., Goodman, M. F. and Warshel, A. (2003) Computer simulation of the chemical catalysis of DNA polymerases: Discriminating between alternative nucleotide insertion mechanisms for T7 DNA polymerase. *Journal of the American Chemical Society*. **125**, 8163-8177
- 194 Radmer, R. J. and Kollman, P. A. (1997) Free energy calculation methods: A theoretical and empirical comparison of numerical errors and a new method for qualitative estimates of free energy changes. *Journal of Computational Chemistry*. **18**, 902-919
- 195 Wang, J. Y., Deng, Y. Q. and Roux, B. (2006) Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophysical Journal*. **91**, 2798-2814
- 196 Zhou, R., Das, P. and Royyuru, A. K. (2008) Single Mutation Induced H3N2 Hemagglutinin Antibody Neutralization: A Free Energy Perturbation Study. *J Phys Chem B*, In Press: PMID 18975892.
- 197 Chandler, D. (1987) *Introduction to modern statistical mechanics*. Oxford University Press, New York, NY
- 198 Chandler, D. (1978) Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J. Chem. Phys.* **68**, 2959-2970
- 199 Bartels, C. and Karplus, M. (1998) Probability distribution for complex systems: Adaptive umbrella sampling of the potential energy. *J. Phys. Chem. B*. **102**, 865-880
- 200 Roux, B. (1995) The Calculation of the Potential of Mean Force Using Computer-Simulations. *Computer Physics Communications*. **91**, 275-282
- 201 Radhakrishnan, R. and Schlick, T. (2004) Biomolecular free energy profiles by a shooting/umbrella sampling protocol, "BOLAS". *J Chem Phys*. **121**, 2436-2444
- 202 Radhakrishnan, R. (2007) Coupling of Fast and Slow Modes in the Reaction Pathway of the Minimal Hammerhead Ribozyme Cleavage. *Biophys. J*. **93**, 2391-2399
- 203 Mezei, M. (1989) Evaluation of the Adaptive Umbrella Sampling Method. *Molecular Simulation*. **3**, 301 - 313
- 204 Szabo, A. and Ostlund, N. S. (1996) *Modern Quantum Chemistry*. Dover Publications, Mineola, New York
- 205 Parr, R. G. and Yang, W. (1989) *Density-functional theory of atoms and molecules*. Oxford University Press, Oxford
- 206 Jensen, F. (2007) *Introduction to Computational Chemistry*. John Wiley & Sons, Chichester

- 207 Warshel, A. and Parson, W. W. (2001) Dynamics of Biochemical and Biophysical Reactions: Insight from Computer Simulations. *Quart. Rev. Biophys.* **34**, 563-679
- 208 Warshel, A. (1989) Computer modeling of chemical reactions in enzymes and solution. John Wiley and Sons, New York
- 209 Shurki, A. and Warshel, A. (2003) Structure/function correlations of proteins using MM, QM/MM, and related approaches: Methods, concepts, pitfalls, and current progress. *Protein Simulations.* **66**, 249-313
- 210 Senn, H. M. and Thiel, W. (2007) QM/MM methods for biological systems. In *Atomistic Approaches in Modern Biology: from Quantum Chemistry to Molecular Simulations.* pp. 173-290, Springer-Verlag Berlin, Berlin
- 211 Das, D., Eurenus, K. P., Billings, E. M., Sherwood, P., Chatfield, D. C., Hodoscek, M. and Brooks, B. R. (2002) Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. *Journal of Chemical Physics.* **117**, 10534-10547
- 212 Reuter, N., Dejaegere, A., Maigret, B. and Karplus, M. (2000) Frontier bonds in QM/MM methods: A comparison of different approaches. *Journal of Physical Chemistry A.* **104**, 1720--1735
- 213 Field, M. J., Bash, P. A. and Karplus, M. (2002) A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **11**, 700-733
- 214 Zhang, Y. and Yang, W. (1999) A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.* **110**, 46-54
- 215 Garcia-Viloca, M. and Gao, J. (2004) Generalized Hybrid Orbital for the treatment of boundary atoms in combined quantum mechanical and molecular mechanical calculations using the semiempirical parameterized model 3 method. *Theoretical Chemistry Accounts.* **111**, 280-286
- 216 Pu, J., Truhlar, D. G. and Gao, J. (2004) The Generalized Hybrid Orbital (GHO) method for ab initio combined QM/MM calculations. *Journal of Physical Chemistry A.* **108**, 632-650
- 217 Friesner, R. A., Baik, M.-H., Guallar, V., Gherman, B. F., Wirstam, M., Murphy, R. B. and Lippard, S. J. (2003) How iron-containing proteins control dioxygen chemistry: a detailed atomic level description via accurate quantum chemical and mixed quantum mechanics/molecular mechanics calculations. *Coordination Chemistry Reviews.* **238-239**, 267-290
- 218 Rega, N., Iyengar, S. S., Voth, G. A., Schlegel, H. B., Vreven, T. and Frisch, M. J. (2004) Hybrid ab initio empirical molecular dynamics: combining the ONIOM scheme with the atom-centered density matrix propagation (ADMP) approach. *J. Phys. Chem. B.* **108**, 4210-4220
- 219 Zhou, R., Berne, B. J. and Germain, R. (2001) The free energy landscape for β hairpin folding in explicit water. *Proc. Natl. Acad. Sci. USA.* **98**, 14931-14936
- 220 Snow, C. D., Nguyen, H., Pande, V. S. and Gruebele, M. (2002) Absolute comparison of simulated and experimental protein-folding dynamics. *Nature.* **420**, 102-106
- 221 Zagrovic, B., Sorin, E. J. and Pande, V. (2001) beta-Hairpin Folding Simulations in Atomistic Detail Using an Implicit Solvent Model. *J. Mol. Biol.* **313**, 151-169
- 222 Duan, Y. and Kollman, P. A. (1998) Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution. *Science.* **282**, 740-744
- 223 Daggett, V. (2000) Long Timescale Simulations. *Curr. Opin. Struct. Biol.* **10**, 160-164
- 224 Elber, R., Meller, J. and Olender, R. (1999) Stochastic path approach to compute atomically detailed trajectories: Application to the folding of C peptide. *J. Phys. Chem. B.* **103**, 899-911
- 225 Zaloj, V. and Elber, R. (2000) Parallel Computations of Molecular Dynamics Trajectories Using The Stochastic Path Approach. *Comput. Phys. Comm.* **128**, 118-127
- 226 Daura, X., Jaun, B., Seebach, D., Gunsteren, W. F. V. and Mark, A. (1998) Reversible Peptide Folding in Solution by Molecular Dynamics Simulation. *J. Mol. Biol.* **280**, 925-932
- 227 Schlick, T. (2003) Engineering teams up with computer-simulation and visualization tools to probe biomolecular mechanisms. *Biophysical J.* **85**, 1-4
- 228 Isralewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D. and Schulten, K. (2001) Steered Molecular Dynamics Investigations of Protein Function. *J. Mol. Graph. Model.* **19**, 13-25

- 229 Isralewitz, B., Gao, M. and Schulten, K. (2001) Steered Molecular Dynamics and Mechanical Functions of Proteins. *Curr. Opin. Struct. Biol.* **11**, 224-230
- 230 Izrailev, S., Crofts, A. R., Berry, E. A. and Schulten, K. (1999) Steered Molecular Dynamics Simulation of the Rieske Subunit Motion in the Cytochrome B Complex. *Biophys. J.* **77**, 1753-1768
- 231 Grayson, P., Tajkhorshid, E. and Schulten, K. (2003) Mechanisms of selectivity in channels and enzymes studied with interactive molecular dynamics. *Biophys. J.* **119**, 36-48
- 232 Yang, L., Beard, W. A., Wilson, S. H., Broyde, S. and Schlick, T. (2002) Polymerase beta simulations suggest that Arg258 rotation is a slow step rather than large subdomain motion per se. *J. Mol. Biol.* **317**, 651-671
- 233 Young, M. A., Gonfloni, S., Superti-Furga, G., Roux, B. and Kuriyan, J. (2001) Dynamic Coupling between the SH2 and SH3 Domains of c-Src and Hck Underlies Their Inactivation by C-Terminal Tyrosine Phosphorylation. *Cell.* **105**, 115-126
- 234 Ferrara, P., Apostolakis, J. and Caflisch, A. (2000) Targeted Molecular Dynamics Simulations of Protein Unfolding. *J. Phys. Chem. B.* **104**, 4511-4518
- 235 Yang, J. (2004) Development and evaluation of a generic evolutionary method for protein-ligand docking. *J. Comput. Chem.* **25**, 843-857
- 236 Kirkpatrick, S., Gelatt, C. and Vecchi, M. (1983) Optimization by simulated annealing. *Science.* **220**, 671-680
- 237 Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. (1998) Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **19**, 1639-1662
- 238 Lee, K., Czaplowski, C., Kim, S. Y. and Lee, J. (2005) An efficient molecular docking using conformational space annealing. *J. Comput. Chem.* **26**, 78-87
- 239 Camacho, J. C., W.Gatchell, D., Kimura, S. R. and Vajda, S. (2000) Scoring docked conformations generated by rigid body protein protein docking. *Proteins.* **40**, 525-537
- 240 Bindewald, E. and Skolnick, J. (2005) A scoring function for docking ligands to low-resolution protein structures. *J. Comput. Chem.* **26**, 374-383
- 241 Bursulaya, B. D., Totrov, M., Abagyan, R. and 3rd, C. L. B. (2003) Comparative study of several algorithms for flexible ligand docking. *Journal of Computer-Aided Molecular Design.* **17**, 755-763
- 242 Pearlman, D. A. (1999) Free energy grids: A practical qualitative application of free energy perturbation to ligand design using the OWFEG method. *J. Med. Chem.* **42**, 4313-4324
- 243 Pearlman, D. A. and Charifson, P. S. (2001) Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. *J. Med. Chem.* **44**, 3417-3423
- 244 Banerjee, A., Yang, W., Karplus, M. and Verdine, G. L. (2005) Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature.* **434**, 612-618
- 245 Zhang, Y., Kua, J. and McCammon, J. A. (2002) Role of the catalytic triad and oxyanion hole in acetylcholinesterase catalysis: an ab initio QM/MM study. *J. Amer. Chem. Soc.* **124**, 10572-10577
- 246 Guo, H., Cui, Q., Lipscomb, W. N. and Karplus, M. (2001) Substrate conformational transitions in the active site of chorismate mutase: Their role in the catalytic mechanism. *Proceedings of the National Academy of Sciences of the United States of America.* **98**, 9032-9037
- 247 Cui, Q. and Karplus, M. (2001) Triosephosphate isomerase: A theoretical comparison of alternative pathways. *Journal of the American Chemical Society.* **123**, 2284-2290
- 248 Mulholland, A. J. (2005) Modelling enzyme reaction mechanisms, specificity and catalysis. *Drug Discovery Today.* **10**, 1393-1402
- 249 Schmidt, M. W., Baldrige, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S. J., Windus, T. L., Dupuis, M. and Montgomery, J. A. (1993) General Atomic and Molecular Electronic-Structure System. *Journal of Computational Chemistry.* **14**, 1347-1363
- 250 Shimojo, F., Kalia, R. K., Nakano, A. and Vashishta, P. (2005) Embedded divide-and-conquer algorithm on hierarchical real-space grids: parallel molecular dynamics simulation based on linear-scaling density functional theory. *Computer Physics Communications.* **167**, 151-164

- 251 Zhou, R. and Berne, B. J. (1995) A New Molecular Dynamics Method Combining the Reference System Propagator Algorithm with a Fast Multipole Method for Simulating Proteins and Other Complex Systems. *J. Chem. Phys.* **103**, 9444-9459
- 252 Berne, B. J. and Straub, J. E. (1997) Novel method of sampling phase space in the simulation of biological systems. *Curr. Opin. Struct. Biol.* **7**, 181-189
- 253 Berneche, S. and Roux, B. (2001) Energetics of Ion Conduction Through the K⁺ Channel. *Nature*. **414**, 73-77
- 254 Simonson, T., Archontis, G. and Karplus, M. (2002) Free energy simulations come of age: protein-ligand recognition. *Acc. Chem. Res.* **35**, 430-437
- 255 Boczko, E. M. and Brooks, C. L. (1995) First principles calculation of the folding free energy of a three helix bundle protein. *Science*. **269**, 393-396
- 256 Schlitter, J., Engels, M. and Kruger, P. (1994) Targeted Molecular-Dynamics - a New Approach for Searching Pathways of Conformational Transitions. *Journal of Molecular Graphics*. **12**, 84-89
- 257 Laio, A. and Parrinello, M. (2002) Escaping free energy minima. *Proc. Natl. Acad. Sci. USA*. **99**, 12562-12566
- 258 Ensing, B. and Klein, M. (2005) Perspective on reactions between F⁻ and CH₃CH₂F: free energy landscape of E₂ and S_N2 reaction channels. *Proc. Natl. Acad. Sci. USA*. **102**, 6755-6759
- 259 Barth, E. J., Laird, B. B. and Leimkuhler, B. J. (2003) Generating generalized distributions from dynamical simulation. *J. Chem. Phys.* **118**, 5759-5768
- 260 Tsallis, C. (1988) Possible generalization of the Boltzmann-Gibbs statistics. *J. Stat. Phys.* **542**, 479-487
- 261 E, W. N., Ren, W. and Vanden-Eijnden, E. (2005) Finite temperature string method for the study of rare events. *J. Phys. Chem. B*. **109**, 6688-6693
- 262 Wang, F. and Landau, D. P. (2001) Efficient, multiple-range random walk algorithm to calculate density of states. *Phys. Rev. Lett.* **86**, 2050-2053
- 263 Yan, Q. and Pablo, J. J. d. (2003) Fast calculation of density of states of a fluid by Monte Carlo simulations. *Phys. Rev. Lett.* **90**, 035701
- 264 Weinan, E., Ren, W. Q. and Vanden-Eijnden, E. (2005) Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes. *Chemical Physics Letters*. **413**, 242-247
- 265 Elber, R., Ghosh, A. and Cardenas, A. (2002) Long time dynamics of complex systems. *Acc. Chem. Res.* **35**, 396-403
- 266 Elber, R., Cárdenas, A., Ghosh, A. and Stern, H. (2003) Bridging the gap between long time trajectories and reaction pathways. *Adv. Chem. Phys.* **126**, 93-129
- 267 Henkelman, G. and Jonsson, H. (2000) Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **113**, 9978
- 268 Henkelman, G., Uberuaga, B. P. and Jonsson, H. (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **113**, 9901
- 269 Jonsson, H., Mills, G. and Jacobsen, K. W. (1998) Nudged elastic band method for finding minimum energy paths of transitions. *World Scientific*
- 270 Bolhuis, P. G., Chandler, D., Dellago, C. and Geissler, P. L. (2002) Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu Rev Phys Chem.* **53**, 291-318
- 271 Dellago, C., Bolhuis, P. G. and Geissler, P. L. (2002) Transition Path Sampling. *Adv. Chem. Phys.* **123**, 1-81
- 272 Bolhuis, P. G., Dellago, C. and Chandler, D. (1998) Sampling ensembles of deterministic transition pathways. *Faraday Discuss.* **110**, 421-436
- 273 Radhakrishnan, R. and Schlick, T. (2004) Orchestration of cooperative events in DNA synthesis and repair mechanism unraveled by transition path sampling of DNA polymerase beta's closing. *P Natl Acad Sci USA*. **101**, 5970-5975
- 274 Bolhuis, P. G. (2003) Transition path sampling of beta-hairpin folding. *Proc. Natl. Acad. Sci. USA*. **100**, 12129-12134

- 275 Radhakrishnan, R. and Schlick, T. (2005) Fidelity discrimination in DNA polymerase beta: differing closing profiles for a mismatched (G:A) versus matched (G:C) base pair. *J. Am. Chem. Soc.* **127**, 13245-13253
- 276 Bustamante, C., Chemla, Y. R., Forde, N. R. and Izhaky, D. (2004) Mechanical processes in biochemistry. *Annual Review of Biochemistry.* **73**, 705-748
- 277 Bustamante, C. (2005) Recent advances in single molecule biophysics. *Febs J.* **272**, 337-337
- 278 Venkatramani, R. and Radhakrishnan, R. (2008) A computational study of the force dependence of phosphoryl transfer during DNA synthesis by a high fidelity polymerase. *Phys Rev Lett.* **100**, 088102
- 279 Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A. and T. E. Cheatham, I. (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889-897
- 280 Balsera, M. A., Wriggers, W., Oono, Y. and Schulten, K. (1996) Principal component analysis and long time protein dynamics. **100**, 2567--2572
- 281 Arora, K. and Schlick, T. (2004) In Silico evidence for DNA Polymerase beta's Substrate-Induced Conformational Change. *Biophysical J.* **87**, 3088-3099
- 282 Shih, A., Purvis, J. and Radhakrishnan, R. (2008) Molecular systems biology of ErbB1 signaling: bridging the gap through multiscale modeling and high-performance computing. *Mol Biosyst.* **4**, 1151-1159
- 283 Yarden, Y. and Sliwkowski, M. X. (2001) Untangling the ErbB Signaling Network. *Nature Reviews Molecular Cell Biology.* **2**, 127--137
- 284 Schoeberl, B., Jonsson, C. E., Gilles, E. D. and Muller, G. (2002) Computational modeling of the dynamics of MAPKinase cascade activated by surface and internalized receptors. *Nature Biotech.* **20**, 370-375
- 285 Hatakeyama, M., Kimura, S., Yumoto, N., Ichikawa, M., Kawasaki, T., Naka, T. and Konagaya, A. (2003) A computational model on the modulation of {MAPK} and {Akt} pathways in Heregulin-induced ERB signaling. *Biochem. J.* **373**, 451--463
- 286 Kholodenko, B. N., Demin, O. V., Moehren, G. and Hoek, J. B. (1999) Quantification of short term signaling by the epidermal growth factor receptor. *Journal of Biological Chemistry.* **274**, 30169-30181
- 287 Kholodenko, B. N. (2006) Cell-signalling dynamics in time and space. *Nat Rev Mol Cell Biol.* **7**, 165-176
- 288 Schlessinger, J. (2000) Cell signaling by receptor tyrosine kinases. *Cell.* **103**, 211-225
- 289 Kloth, M. T., Laughlin, K. K., Biscardi, J. S., Boerner, J. L., Parsons, S. J. and M.Silva, C. (2003) STAT5b, a Mediator of Synergism between c-Src and the Epidermal Growth Factor Receptor. *J. Biol. Chem.* **278**, 1671-1679
- 290 Vieira, A. V., Lamaze, C. and Schmid, S. L. (1996) Control of {EGF} Receptor signaling by clathrin-mediated endocytosis. *Science.* **274**, 2066--2089
- 291 Mendelsohn, J. and Baselga, J. (2000) The EGF receptor family as targets for cancer therapy. [Review] [155 refs]. *Oncogene.* **19**, 6550-6565
- 292 Linggi, B. and Carpenter, G. (2006) ErbB receptors: new insights on mechanisms and biology. *Trends in Cell Biology.* **16**, 649-656
- 293 Liu, Y., Purvis, J., Shih, A., Weinstein, J., Agrawal, N. and Radhakrishnan, R. (2007) A Multiscale Computational Approach to Dissect Early Events in the Erb Family Receptor Mediated Activation, Differential Signaling, and Relevance to Oncogenic Transformations. *Annals of Biomedical Engineering.* **35**, 1012-1025
- 294 A. Shih, S-H. Choi, M. A. Lemmon and Radhakrishnan, R. (2008) Network of Stabilizing Hydrogen Bonds is Integral to the Dimer-Mediated Allosteric Activation Mechanism in the ErbB1 and ErbB4 Kinase. Submitted for publication
- 295 Sordella, R., Bell, D. W., Haber, D. A. and Settleman, J. (2004) Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science.* **305**, 1163-1167
- 296 Carey, K. D., Garton, A. J., Romero, M. S., Kahler, J., Thomson, S., Ross, S., Park, F., Haley, J. D., Gibson, N. and Sliwkowski, M. X. (2006) Kinetic analysis of epidermal growth factor receptor somatic

mutant proteins shows increased sensitivity to the epidermal growth factor receptor tyrosine kinase inhibitor, erlotinib. *Cancer Research*. **66**, 8163-8171

297 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. and Leahy, D. J. (2008) Mechanism of activation and inhibition of the HER4/ErbB4 kinase. *Structure*. **16**, 460-467

298 Shih, A. J., Telesco, S. E., Choi, S. H., Lemmon, M. A. and Radhakrishnan, R. Molecular Dynamics Analysis of Conserved Hydrophobic and Hydrophilic Bond Interaction Networks in ErbB Family Kinases In Review

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

300 Salomon, D. S., Brandt, R., Ciardiello, F. and Normanno, N. (1995) Epidermal growth factor-related peptides and their receptors in human malignancies. *Critical Reviews in Oncology/Hematology*. **19**, 183--232

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

302 Slamon, D. J., Godolphin, W., Jones, L. A., Holt, J. A., Wong, S. G. and Keith, D. E. (1989) Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science*. **244**, 707-712

303 Gassmann, M., Casagrande, F., Orioli, D., Simon, H., Lai, C., Klein, R. and Lemke, G. (1995) Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor. *Nature*. **378**, 390-394

304 Tidcombe, H., Jackson-Fisher, A., Mathers, K., Stern, D. F., Gassmann, M. and Golding, J. P. (2003) Neural and mammary gland defects in ErbB4 knockout mice genetically rescued from embryonic lethality. *Proceedings of the National Academy of Sciences*. **100**, 8281-8286

305 Linggi, B. and Carpenter, G. (2006) ErbB-4 s80 Intracellular Domain Abrogates ETO2-dependent Transcriptional Repression. *J. Biol. Chem*. **281**, 25373-25380

306 Gotoh, N., Tojo, A., Hino, M., Yazaki, Y., Shibuya, M., Gotoh, N., Tojo, A., Muroya, K., Hashimoto, Y., Hattori, S., Nakamura, S., Takenawa, T., Yazaki, Y. and Shibuya, M. (1992) EGFR mutant lacking the auto-phosphorylation sites induces phosphorylation of Shc protein and Shc-Grb2/Ash association and retains mitogenic activity. *Biochem. Biophys. Res. Commun*. **186**, 768--774

307 Biscardi, J. S., Maa, M.-C., Tice, D. A., Cox, M. E., Leu, T.-H. and Parsons, S. J. (1999) c-Src-mediated Phosphorylation of the Epidermal Growth Factor Receptor on Tyr845 and Tyr1101 Is Associated with Modulation of Receptor Function. *J. Biol. Chem*. **274**, 8335-8343

308 Ozkirimli, E. and Post, C. B. (2006) Src kinase activation: A switched electrostatic network. *Protein Science*. **15**, 1051-1062

309 Ozkirimli, E., Yadav, S. S., Miller, W. T. and Post, C. B. (2008) An electrostatic network and long-range regulation of Src kinases. *Protein Science*. **17**, 1871-1880

310 Dixit, A. and Verkhivker, G. M. (2009) Hierarchical Modeling of Activation Mechanisms in the ABL and EGFR Kinase Domains: Thermodynamic and Mechanistic Catalysts of Kinase Activation by Cancer Mutations. *PLoS Comput Biol*. **5**, e1000487

311 Papakyriakou, A., Vourloumis, D., Tzortzatou-Stathopoulou, F. and Karpusas, M. (2009) Conformational dynamics of the EGFR kinase domain reveals structural features involved in activation. *Proteins: Structure, Function, and Bioinformatics*. **76**, 375-386

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

313 Boggon, T. J. and Eck, M. J. (2004) Structure and regulation of Src family kinases. *Oncogene*. **23**, 7918-7927

314 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*. **79**, 926--935

315 Grubmüller, H. Solvate 1.0. Theoretische Biophysik, Institut für Medizinische Optik, Universität München, Theresienstr.

- 316 Andersen, H. C. (1983) Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*. **52**, 24-34
- 317 Scott, E. F., Yuhong, Z., Richard, W. P. and Bernard, R. B. (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics*. **103**, 4613-4621
- 318 Glykos, N. M. and Kokkinidis, M. (2004) Structural polymorphism of a marginally stable 4-alpha-helical bundle. Images of a trapped molten globule? *Proteins*. **56**, 420-425
- 319 Buck, M. and Karplus, M. (2001) Hydrogen Bond Energetics: A Simulation and Statistical Analysis of N-Methyl Acetamide (NMA), Water, and Human Lysozyme. *The Journal of Physical Chemistry B*. **105**, 11000-11015
- 320 Godawat, R., Jamadagni, S. N. and Garde, S. (2009) Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations. *Proceedings of the National Academy of Sciences*. **106**, 15119-15124
- 321 Acharya, H., Vembanur, S., Jamadagni, S. N. and Garde, S. (2010) Mapping hydrophobicity at the nanoscale: Applications to heterogeneous surfaces and proteins. *Faraday Discuss.* **In Press** DOI: **10.1039/b927019a**
- 322 Tsai, C.-J., Sol, A. d. and Nussinov, R. (2009) Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Molecular BioSystems*. **5**, 207-216
- 323 Yang, S., Banavali, N. K. and Roux, B. t. (2009) Mapping the conformational transition in Src activation by cumulating the information from multiple molecular dynamics trajectories. *Proceedings of the National Academy of Sciences*. **106**, 3776-3781
- 324 Banavali, N. K. and Roux, B. (2007) Anatomy of a structural pathway for activation of the catalytic domain of Src kinase Hck. *Proteins: Structure, Function, and Bioinformatics*. **67**, 1096-1112
- 325 Gan, W., Yang, S. and Roux, B. (2009) Atomistic View of the Conformational Activation of Src Kinase Using the String Method with Swarms-of-Trajectories. *Biophysical Journal*. **97**, L8-L10
- 326 Wang, S. E., Narasanna, A., Perez-Torres, M., Xiang, B., Wu, F. Y., Yang, S., Carpenter, G., Gazdar, A. F., Muthuswamy, S. K. and Arteaga, C. L. (2006) HER2 kinase domain mutation results in constitutive phosphorylation and activation of HER2 and EGFR and resistance to EGFR tyrosine kinase inhibitors. *Cancer Cell*. **10**, 25-38
- 327 Liu, Y. and Radhakrishnan, R. (2010) Computational Delineation of the Tyrosyl-Substrate Binding, Pre-Catalytic, and Catalytic Landscapes in the Epidermal Growth Factor Receptor (EGFR) Tyrosine Kinase Domain. To be submitted
- 328 Feng, M. H., Philippopoulos, M., A. D. MacKerell, J. and Lim, C. (1996) Structural Characterization of the Phosphotyrosine Binding Region of a High-Affinity SH2 Domain-Phosphopeptide Complex by Molecular Dynamics Simulation and Chemical Shift Calculations. *Journal of the American Chemical Society*. **118**, 11265--11277
- 329 Brooks, B. R. (1983) CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem*. **4**, 187-217
- 330 Telesco, S. E., Shih, A. J., Jia, F. and Radhakrishnan, R. A multiscale modeling approach to investigate molecular mechanisms of pseudokinase activation and drug resistance in the HER3/ErbB3 receptor tyrosine kinase signaling network. In Review
- 331 Manning, G., Whyte, D. B., Martinez, R., Hunter, T. and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science*. **298**, 1912-1934
- 332 Min, X., Lee, B. H., Cobb, M. H. and Goldsmith, E. J. (2004) Crystal structure of the kinase domain of WNK1, a kinase that causes a hereditary form of hypertension. *Structure*. **12**, 1303-1311
- 333 Mukherjee, K., Sharma, M., Urlaub, H., Bourenkov, G. P., Jahn, R., Sudhof, T. C. and Wahl, M. C. (2008) CASK Functions as a Mg²⁺-independent neurexin kinase. *Cell*. **133**, 328-339
- 334 Yarden, Y. and Sliwkowski, M. X. (2001) Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol*. **2**, 127-137
- 335 Linggi, B. and Carpenter, G. (2006) ErbB receptors: new insights on mechanisms and biology. *Trends Cell Biol*. **16**, 649-656

- 336 Plowman, G. D., Whitney, G. S., Neubauer, M. G., Green, J. M., McDonald, V. L., Todaro, G. J. and Shoyab, M. (1990) Molecular cloning and expression of an additional epidermal growth factor receptor-related gene. *Proc Natl Acad Sci U S A*. **87**, 4905-4909
- 337 Shi, F., Telesco, S. E., Liu, Y., Radhakrishnan, R. and Lemmon, M. A. (2010) ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation. *Proceedings of the National Academy of Sciences*. **107**, 7692-7697
- 338 Choi, S. H., Mendrola, J. M. and Lemmon, M. A. (2007) EGF-independent activation of cell-surface EGF receptors harboring mutations found in gefitinib-sensitive lung cancer. *Oncogene*. **26**, 1567-1576
- 339 Ji, H., Li, D., Chen, L., Shimamura, T., Kobayashi, S., McNamara, K., Mahmood, U., Mitchell, A., Sun, Y., Al-Hashem, R., Chirieac, L. R., Padera, R., Bronson, R. T., Kim, W., Janne, P. A., Shapiro, G. I., Tenen, D., Johnson, B. E., Weissleder, R., Sharpless, N. E. and Wong, K. K. (2006) The impact of human EGFR kinase domain mutations on lung tumorigenesis and in vivo sensitivity to EGFR-targeted therapies. *Cancer Cell*. **9**, 485-495
- 340 Lee, J. W., Soung, Y. H., Kim, S. Y., Nam, S. W., Park, W. S., Wang, Y. P., Jo, K. H., Moon, S. W., Song, S. Y., Lee, J. Y., Yoo, N. J. and Lee, S. H. (2006) ERBB2 kinase domain mutation in the lung squamous cell carcinoma. *Cancer Lett*. **237**, 89-94
- 341 Minami, Y., Shimamura, T., Shah, K., LaFramboise, T., Glatt, K. A., Liniker, E., Borgman, C. L., Haringsma, H. J., Feng, W., Weir, B. A., Lowell, A. M., Lee, J. C., Wolf, J., Shapiro, G. I., Wong, K. K., Meyerson, M. and Thomas, R. K. (2007) The major lung cancer-derived mutants of ERBB2 are oncogenic and are associated with sensitivity to the irreversible EGFR/ERBB2 inhibitor HKI-272. *Oncogene*. **26**, 5023-5027
- 342 Shigematsu, H., Takahashi, T., Nomura, M., Majmudar, K., Suzuki, M., Lee, H., Wistuba, II, Fong, K. M., Toyooka, S., Shimizu, N., Fujisawa, T., Minna, J. D. and Gazdar, A. F. (2005) Somatic mutations of the HER2 kinase domain in lung adenocarcinomas. *Cancer Res*. **65**, 1642-1646
- 343 Sergina, N. V., Rausch, M., Wang, D., Blair, J., Hann, B., Shokat, K. M. and Moasser, M. M. (2007) Escape from HER-family tyrosine kinase inhibitor therapy by the kinase-inactive HER3. *Nature*. **445**, 437-441
- 344 Engelman, J. A., Zejnullahu, K., Mitsudomi, T., Song, Y., Hyland, C., Park, J. O., Lindeman, N., Gale, C. M., Zhao, X., Christensen, J., Kosaka, T., Holmes, A. J., Rogers, A. M., Cappuzzo, F., Mok, T., Lee, C., Johnson, B. E., Cantley, L. C. and Janne, P. A. (2007) MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science*. **316**, 1039-1043
- 345 Hamburger, A. W. (2008) The role of ErbB3 and its binding partners in breast cancer progression and resistance to hormone and tyrosine kinase directed therapies. *J Mammary Gland Biol Neoplasia*. **13**, 225-233
- 346 Liu, B., Ordonez-Ercan, D., Fan, Z., Edgerton, S. M., Yang, X. and Thor, A. D. (2007) Downregulation of erbB3 abrogates erbB2-mediated tamoxifen resistance in breast cancer cells. *Int J Cancer*. **120**, 1874-1882
- 347 Liu, Y., Purvis, J., Shih, A., Weinstein, J., Agrawal, N. and Radhakrishnan, R. (2007) 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*. **35**, 1012-1025
- 348 Shih, A. J., Purvis, J. and Radhakrishnan, R. (2008) Molecular systems biology of ErbB1 signaling: bridging the gap through multiscale modeling and high-performance computing. *Mol Biosyst*. **4**, 1151-1159
- 349 Rausenberger, J., Hussong, A., Kircher, S., Kirchenbauer, D., Timmer, J., Nagy, F., Schafer, E. and Fleck, C. (2010) An integrative model for phytochrome B mediated photomorphogenesis: from protein dynamics to physiology. *PLoS One*. **5**, e10721
- 350 Fedosov, D. A., Caswell, B. and Karniadakis, G. E. (2010) A multiscale red blood cell model with accurate mechanics, rheology, and dynamics. *Biophys J*. **98**, 2215-2225
- 351 Vasalou, C. and Henson, M. A. (2010) A multiscale model to investigate circadian rhythmicity of pacemaker neurons in the suprachiasmatic nucleus. *PLoS Comput Biol*. **6**, e1000706

- 352 Telesco, S. E. and Radhakrishnan, R. (2009) Atomistic insights into regulatory mechanisms of the HER2 tyrosine kinase domain: a molecular dynamics study. *Biophys J.* **96**, 2321-2334
- 353 Zhang, X., Gureasko, J., Shen, K., Cole, P. A. and Kuriyan, J. (2006) An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell.* **125**, 1137-1149
- 354 Shih, A. J., S.E. Telesco, S.H. Choi, M.A. Lemmon, and R. Radhakrishnan. Molecular dynamics analysis of conserved hydrophobic and hydrophilic bond interaction networks in ErbB family kinases. In review.
- 355 Taylor, S. S. and Kornev, A. P. (2010) Yet another "active" pseudokinase, Erb3. *Proc Natl Acad Sci U S A.* **107**, 8047-8048
- 356 Shi, F., Telesco, S. E., Liu, Y., Radhakrishnan, R. and Lemmon, M. A. (2010) ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation. *Proc Natl Acad Sci U S A.* **107**, 7692-7697
- 357 Jorgensen, W. L., J. Chandrasekar, J. Madura, R. Impey, and M. Klein. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935
- 358 Grubmuller, H., Heymann, B. and Tavan, P. (1996) Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science.* **271**, 997-999
- 359 Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L. and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem.* **26**, 1781-1802
- 360 Feller, S. E., Y. Zhang, R.W. Pastor, and B.R. Brooks. (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys.* **103**, 4613-4621
- 361 Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J Mol Graph.* **14**, 33-38, 27-38
- 362 Glykos, N. M. (2006) Software news and updates. Carma: a molecular dynamics analysis program. *J Comput Chem.* **27**, 1765-1768
- 363 Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. and de Vries, A. H. (2007) The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B.* **111**, 7812-7824
- 364 Shih, A., Telesco, S. E., Choi, S. H., Lemmon, M. A. and Radhakrishnan, R. (2010) Molecular Dynamics Analysis of Conserved Hydrophobic and Hydrophilic Bond Interaction Networks in ErbB Family Kinases Submitted to *Biochemical Journal*
- 365 Nagar, B., Hantschel, O., Young, M. A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G. and Kuriyan, J. (2003) Structural Basis for the Autoinhibition of c-Abl Tyrosine Kinase. *Cell.* **112**, 859-871
- 366 Superti-Furga, G. and Courtneidge, S. A. (1995) Structure-function relationships in Src family and related protein tyrosine kinases. *BioEssays.* **17**, 321-330
- 367 Furdui, C. M., Lew, E. D., Schlessinger, J. and Anderson, K. S. (2006) Autophosphorylation of FGFR1 Kinase Is Mediated by a Sequential and Precisely Ordered Reaction. *Molecular Cell.* **21**, 711-717
- 368 Bae, J. and Schlessinger, J. (2010) Asymmetric tyrosine kinase arrangements in activation or autophosphorylation of receptor tyrosine kinases. *Molecules and Cells.* **29**, 443-448
- 369 Landau, M., Fleishman, S. J. and Ben-Tal, N. (2004) A Putative Mechanism for Downregulation of the Catalytic Activity of the EGF Receptor via Direct Contact between its Kinase and C-Terminal Domains. *Structure.* **12**, 2265-2275
- 370 Yang, K. S., Macdonald-Obermann, J. L., Piwnica-Worms, D. and Pike, L. J. (2010) Asp-960/Glu-961 Controls the Movement of the C-terminal Tail of the Epidermal Growth Factor Receptor to Regulate Asymmetric Dimer Formation. *Journal of Biological Chemistry.* **285**, 24014-24022
- 371 Shen, M. Y. and Sali, A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507-2524
- 372 Stein, R. A. and Staros, J. V. (2006) Insights into the evolution of the ErbB receptor family and their ligands from sequence analysis. *BMC Evol Biol.* **6**, 79
- 373 Larsson, P., Wallner, B., Lindahl, E. and Elofsson, A. (2008) Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Sci.* **17**, 990-1002

- 374 Sharma, S. V., Gajowniczek, P., Way, I. P., Lee, D. Y., Jiang, J., Yuza, Y., Classon, M., Haber, D. A. and Settleman, J. (2006) A common signaling cascade may underlie "addiction" to the Src, BCR-ABL, and EGF receptor oncogenes. *Cancer Cell*. **10**, 425-435
- 375 Purvis, J., Ilango, V. and Radhakrishnan, R. (2008) Role of Network Branching in Eliciting Differential Short-Term Signaling Responses in the Hyper-Sensitive Epidermal Growth Factor Receptor Mutants Implicated in Lung Cancer. *Biotechnol Prog*. **24**, 540-553
- 376 Purvis, J., Liu, Y., Ilango, V. and Radhakrishnan, R. (2007) Efficacy of tyrosine kinase inhibitors in the mutants of the epidermal growth factor receptor: A multiscale molecular/ systems model for phosphorylation and inhibition. *Proc. Foundations in Systems Biology II*. **IRB Verlag Stuttgart**, 289-294
- 377 Duesbery, N. and Vande Woude, G. (2006) BRAF and MEK mutations make a late entrance. *Sci STKE*. **2006**, pe15
- 378 Bentires-Alj, M., Gil, S. G., Chan, R., Wang, Z. C., Wang, Y., Imanaka, N., Harris, L. N., Richardson, A., Neel, B. G. and Gu, H. (2006) A role for the scaffolding adapter GAB2 in breast cancer. *Nat Med*. **12**, 114-121
- 379 Masciaux, C., Iannino, N., Martin, B., Paesmans, M., Berghmans, T., Dusart, M., Haller, A., Lothaire, P., Meert, A. P., Noel, S., Lafitte, J. J. and Sculier, J. P. (2005) The role of RAS oncogene in survival of patients with lung cancer: a systematic review of the literature with meta-analysis. *Br J Cancer*. **92**, 131-139
- 380 Samuels, Y., Wang, Z., Bardelli, A., Silliman, N., Ptak, J., Szabo, S., Yan, H., Gazdar, A., Powell, S. M., Riggins, G. J., Willson, J. K. V., Markowitz, S., Kinzler, K. W., Vogelstein, B. and Velculescu, V. E. (2004) High Frequency of Mutations of the PIK3CA Gene in Human Cancers. *Science*. **304**, 554-
- 381 Sjoblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S. D., Willis, J., Dawson, D., Willson, J. K. V., Gazdar, A. F., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B. H., Bachman, K. E., Papadopoulos, N., Vogelstein, B., Kinzler, K. W. and Velculescu, V. E. (2006) The Consensus Coding Sequences of Human Breast and Colorectal Cancers. *Science*. **314**, 268-274
- 382 Schulze, W. X., Deng, L. and Mann, M. (2005) Phosphotyrosine interactome of the ErbB-receptor kinase family. *Molecular Systems Biology*
- 383 Bing Liu, B. B. J. H. W. (2006) Impact of EGFR point mutations on the sensitivity to gefitinib: Insights from comparative structural analyses and molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*. **65**, 331-346
- 384 Mulloy, R., Ferrand, A., Kim, Y., Sordella, R., Bell, D. W., Haber, D. A., Anderson, K. S. and Settleman, J. (2007) Epidermal growth factor receptor mutants from human lung cancers exhibit enhanced catalytic activity and increased sensitivity to gefitinib. *Cancer Research*. **67**, 2325-2330
- 385 Baselga, J. and Swain, S. M. (2009) Novel anticancer targets: revisiting ERBB2 and discovering ERBB3. *Nat Rev Cancer*. **9**, 463-475
- 386 Hutcheson, I. R., Knowlden, J. M., Hiscox, S. E., Barrow, D., Gee, J. M., Robertson, J. F., Ellis, I. O. and Nicholson, R. I. (2007) Heregulin beta1 drives gefitinib-resistant growth and invasion in tamoxifen-resistant MCF-7 breast cancer cells. *Breast Cancer Res*. **9**, R50
- 387 Menendez, J. A. and Lupu, R. (2007) Transphosphorylation of kinase-dead HER3 and breast cancer progression: a new standpoint or an old concept revisited? *Breast Cancer Res*. **9**, 111
- 388 Hsieh, M. H., Fang, Y. F., Chang, W. C., Kuo, H. P., Lin, S. Y., Liu, H. P., Liu, C. L., Chen, H. C., Ku, Y. C., Chen, Y. T., Chang, Y. H., Chen, Y. T., Hsi, B. L., Tsai, S. F. and Huang, S. F. (2006) Complex mutation patterns of epidermal growth factor receptor gene associated with variable responses to gefitinib treatment in patients with non-small cell lung cancer. *Lung Cancer*. **53**, 311-322
- 389 Offtenderinger, M., Schofer, C., Weipoltshammer, K. and Grunt, T. W. (2002) c-erbB-3: a nuclear protein in mammary epithelial cells. *J Cell Biol*. **157**, 929-939
- 390 Schoeberl, B., Pace, E. A., Fitzgerald, J. B., Harms, B. D., Xu, L., Nie, L., Linggi, B., Kalra, A., Paragas, V., Bukhalid, R., Grantcharova, V., Kohli, N., West, K. A., Leszczyniecka, M., Feldhaus, M. J., Kudla, A. J. and Nielsen, U. B. (2009) Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis. *Sci Signal*. **2**, ra31

- 391 Ferrer-Soler, L., Vazquez-Martin, A., Brunet, J., Menendez, J. A., De Llorens, R. and Colomer, R. (2007) An update of the mechanisms of resistance to EGFR-tyrosine kinase inhibitors in breast cancer: Gefitinib (Iressa) -induced changes in the expression and nucleo-cytoplasmic trafficking of HER-ligands (Review). *Int J Mol Med.* **20**, 3-10
- 392 Motoyama, A. B. and Hynes, N. E. (2003) {BAD}: a good therapeutic target? *Breast Cancer Research.* **5**, 27--30