Functional annotation of oncogenic mutations in epidermal growth factor receptor kinase

by

Zheng Ruan

(Under the Direction of Natarajan Kannan)

Abstract

The epidermal growth factor receptor (EGFR) kinase is activated by a variety of mutations in human cancer. Despite previous extensive research on EGFR, the structural and functional impact of the majority of cancer mutations remains poorly understood, hindered the effective treatment of patients harboring these mutations. In this dissertation, I establish a combined experimental and computational framework to investigate patient-derived mutations that map to the kinase domain of EGFR. First, I examine a recurrent R776H mutation and provide detailed mechanistic insight into EGFR signaling and kinase regulation. Next, I perform a mutation screening study focusing on the regulatory spine of EGFR. The study reveals a novel mode of kinase activation and drug resistance mechanism. Last, I characterize a class of less well-studied insertion mutations in the αC-β4 loop of the kinase domain. The findings presented in the dissertation not only shed light on mutation-mediated EGFR kinase activation, but also provide clues for patient treatment and future cancer drug design.

Index words: EGFR, protein kinase, cancer, mutation, allostery, MD simulation

Functional annotation of oncogenic mutations in epidermal growth factor receptor kinase

by

Zheng Ruan

B.S., Huazhong Agricultural University, 2012

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2018

Functional annotation of oncogenic mutations in epidermal growth factor receptor kinase

by

Zheng Ruan

Major Professor:	Natarajan Kannan
Committee:	Eileen J. Kennedy Robert J. Woods Shaying Zhao

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2018

Functional annotation of oncogenic mutations in epidermal growth factor receptor kinase

Zheng Ruan

August 2018

For Mom and Dad

Acknowledgments

I am grateful to all of those with whom I have had the pleasure to work during my Ph.D. study. I would like to express the deepest appreciation to my dissertation advisor, Dr. Natarajan Kannan, who has provided me extensive personal/professional guidance and financial support to my Ph.D. study, without which this dissertation would not have been possible. I would also like to thank all of my committee members, Eileen J. Kennedy, Robert J. Woods, and Shaying Zhao, who have provided invaluable advice to my research project. ESBG lab members Samiksha Katiyar, Krishnadev Oruganty, Daniel McSkimming, Eric W. Talevich, Smita Mohanty, Hyunjin Annie Kwon, Carlos E Sanz, Liang-Chin Huang, Rahil Taujale, Wayland Yeung, and Safal Shrestha are acknowledged for the helpful discussions in the lab. I must also express my gratitude to Mrs. Patty Grimes and her husband for their generous research fellowship and the UGA Graduate School for the Innovative and Interdisciplinary Research Grant 2017.

Contents

Acknowledgments **vi**

List of Figures **x**

List of Tables **xiii**

1 Introduction and literature review **1**

1.1 Motivation 1

1.2 Background 3

1.3 Major research questions addressed 10

Bibliography 13

2 Mechanistic Insights into R776H Mediated Activation of Epidermal Growth Factor Receptor (EGFR) Kinase **31**

2.1 Introduction 33

2.2 Results 35

2.3 Materials and Methods 46

2.4 Discussion 50

Bibliography 55

3 Computational and experimental characterization of patient derived mutations reveal an unusual mode of regualtory spine assembly and drug sensitivity in EGFR kinase 61

3.1 Introduction 63

3.2 Results 65

3.3 Materials and Methods 83

3.4 Discussion 88

Bibliography 94

4 Altered conformational landscape and dimerization dependency underpins the activation of EGFR by αC-β4 loop insertion mutations 104

4.1 Introduction 106

4.2 Result 108

4.3 Materials and Methods 124

4.4 Conclusion and Discussion 131

Bibliography 134

5 Discussion and conclusion remarks 145

5.1 Achievement of goals and broader impact 146

5.2 Future directions 148

Bibliography 154

Appendices 160

A Extended Results **160**

List of Figures

1.1 Crystal structures of EGFR in various conformational states 5

2.1 R776 mediated interactions in active/inactive state 34
2.2 R776H depends on the asymmetric dimer for activation 36
2.3 R776H is a "superacceptor" 38
2.4 Receiver kinase phosphorylation 40
2.5 Lateral phosphorylation of EGFR 41
2.6 Kinase activity of EGFR is responsible for lateral phosphorylation 42
2.7 Molecular dynamics simulation of EGFR 45
2.8 R776H loses the interaction with C-terminal tail of EGFR 47
2.9 Histogram of the capping interaction distance in EGFR crystal structures . . 52
2.10 Emerging acvitation scheme of oncogenic mutations in EGFR 54

3.1 R-spine assembly/disassembly 64
3.2 Western blot analyses and screening of RS3 mutations in EGFR 67
3.3 Western blot of H835L (RS1) and co-occurring L833V mutation 68
3.4 Phosphorylation of STAT3 by M766T EGFR 68

3.5 Dimerization dependency of M766T mutation 70
3.6 MD simulation of EGFR RS mutants in inactive state 72
3.7 MD simulation of EGFR RS mutants in inactive state 74
3.8 Root mean square fluctuation (RMSF) of the αC-helix in the inactive monomer simulation. 75
3.9 RMSF plot for receiver and activator kinase 77
3.10 Root mean square deviation (RMSD) plot of RS3 mutants in active (top panel) and inactive (bottom panel) states. 78
3.11 RMSF plot of C-helix residues in the active kinase monomer (red) and receiver kinase (cyan) in the asymmetric dimer 79
3.12 Drug sensitivity of WT and mutant (M766T) EGFR 81
3.13 Structural mapping of MM/PBSA energy decomposition analysis 84
3.14 Schematic activation mechanisms of M766T 88
3.15 Activity of EGFR L777F mutant 90
3.16 Comparison of the binding mode of WT EGFR and mutant (M766T) to the allosteric inhibitor EAI045. 91

4.1 EGFR insertion mutations 109
4.2 Cell-based screening of EGFR insertion mutations 111
4.3 Dimerization-dependency of EGFR insertion mutations 113
4.4 Densitometry analysis of the dimerization-dependency experiment 114
4.5 Activating insertion mutations are tolerant to C-lobe dimerization-deficient mutations 114
4.6 Drug inhibition assay of five EGFR inhibitors 116

4.7 MD simulation of EGFR and three activating insertion mutations in the active state 119
4.8 Modulation of auto-inhibitory capping interactions 121
4.9 Reaction coordinates for US simulation 124
4.10 Free energy landscape of EGFR and insertion mutations from umbrella sampling 125
4.11 The alternative inactive state stabilized by activating αC-β4 loop insertion mutations. 126

5.1 Alignment of EGFR exon 19 insertion mutations 149
5.2 β3-lysine mutation of exon 19 insertion mutation 150
5.3 Thermodynamic cycle of binding free energy calculation 153

List of Tables

1.1 FDA-approved EGFR kinase inhibitors 2

3.1 Binding free energy estimates of WT and mutant (M766T) EGFR for various inhibitors using MM/PBSA method. 83

3.2 Mutation impact prediction for EGFR RS3 mutants. 93

4.1 The statistics of $\Delta G_{\text{active-inactive}}$ of Figure 4.10b. 123

4.2 Information of all-atom MD simulations performed in the study 130

Chapter 1

Introduction and literature review

1.1 Motivation

Cancer is a disease driven by heritable or somatic mutations that result in uncontrolled cell division [1]. The rapid advancement of next generation sequencing technologies brings promises to diagnose cancer patients with the genetic alterations in a cost-effective manner, and to monitor disease progression at different stages of cancer development [2, 3]. While informative, the majority of mutations identified through genome sequencing may arise simply as a natural consequence of random DNA replication error. These genomic alterations, coined "passenger" mutations, are distinct from another group of mutations that truly contribute to the tumor development, i.e., "driver" mutations [4]. Until now, distinguishing genuine "driver" mutations against a background of "passenger" mutations is an unsolved problem [5].

Epidermal growth factor receptor (EGFR) is one of the most frequently mutated genes

Table 1.1: FDA-approved EGFR kinase inhibitors

Name	Year Approved	Targets	Indications
Gefitinib	2003-2005, 2015	EGFR	NSCLC
Erlotinib	2004	EGFR	NSCLC and pancreatic cancer
Lapatinib	2007	EGFR and ErbB2	Breast cancer
Vandetanib	2011	VEGFRs, EGFRs and Ret	Medullary thyroid cancer
Afatinib	2013 and 2016	EGFR, ErbB2, ErbB4	NSCLC and squamous NSCLC
Neratinib	2017	EGFR and ErbB2	breast cancer
Brigatinib	2017	ALK and EGFR	ALK^+ NSCLC after crizotinib
Osimertinib	2015 and 2017	EGFR	NSCLC

in human cancer [6, 7, 8, 9]. Mutational activation of EGFR results in constitutive cellular growth signaling and drives the development of many different human cancers [10, 11, 12]. EGFR is also a long-pursued drug target with 8 different FDA-approved kinase inhibitors until June 2018 (Table 1.1) and many in clinical trials [13, 14, 15, 16, 17]. Previous studies have structurally and functionally characterized a few frequently occurring mutations in EGFR and established their response profiles to EGFR-targeted inhibitors [18, 19, 20, 21, 22]. However, the vast majority of other patient-derived mutations in EGFR are poorly annotated in terms of their impact on kinase activity and drug sensitivity, significantly hindered the effective treatment of patients harboring these mutations [23, 24, 25, 26].

To overcome these challenges, I seek to experimentally and computationally characterize a subset of these mutations, with a special focus on those that are located at functionally important regions of the kinase domain. The established pepline not only allows me to provide phenotypic information regarding the activity and drug response of these mutations, but also detailed structural mechanisms underlie the experimental observations. The interdisci-

plinary strategies employed in this study provide fundamental insight into EGFR oncogenic activation and is broadly applicable to understand other mutations in protein kinases as well.

1.2 Background

1.2.1 EGFR Biology

The epidermal growth factor (EGF) and its receptor (EGFR) are first discovered by Stanley Cohen from Vanderbilt University, who was awarded the Nobel Prize in Medicine in 1986 because of his contribution in establishing ligand-induced activation of cellular growth signaling [27]. Human epidermal growth factor receptor, also known as ErbB1 or HER1, is an important receptor tyrosine kinases (RTKs) that regulates many biological processes [28, 13]. Genetic alterations of EGFR is commonly implicated in many diseases, in particular cancer [29, 30, 31]. EGFR belongs to the ErbB receptor family that also consists of ErbB2 (HER2), ErbB3 (HER3), and ErbB4 (HER4) [28]. All ErbB receptors are membrane proteins composed of a share domain organization, including a glycosylated extracellular ligand binding domain, a single pass transmembrane segment, and a cytoplasmic portion with a juxtamembrane segment, a protein kinase domain, and a carboxy-terminal tail [32]. Although the structure of full-length EGFR protein has yet to be determined, previous structural and functional analysis on fragments of EGFR has provided significant insight into the molecular details of EGFR activation and regulation (Figure 1.1) [33].

The protein kinase domain of EGFR contains a small flexible N-terminal lobe for ATP binding and a large stable C-terminal lobe for peptide substrate recognition [34]. The N-lobe and C-lobe are connected by a hinge linker that controls the interlobe movement critical for

the catalytic cycle of the enzyme [35]. The kinase domain of EGFR displays structural plasticity in that it can adopt distinct conformational states (Figure 1.1) [36]. In the catalytically competent "active" state, the important regulatory αC-helix in the kinase domain is placed in an "inward" orientation such that the critical ATP-coordinating K745-E762 salt bridge is formed [22]. In addition, the activation segment, defined as residues between D855 and E884, is in an extended conformation to allow substrate binding [22]. This is in contrast to the "inactive" state of the kinase domain, in which the activation segment folds into two short 3/10 helices and the αC-helix rotates to an "outward" conformation [34]. As a consequence, the K745-E762 salt bridge is broken and the kinase is catalytically incompetent [37]. The dynamic equilibrium of EGFR between the "active" and "inactive" states is the structural basis for it being a molecular switch of the growth signaling cascade upon various extracellular stimuli [38]. A diverse array of regulatory mechanisms are evolved to further fine-tune the normal signaling of EGFR and prevent uncontrolled kinase activation [22, 39, 25].

The catalytic activation of EGFR requires the presence of extracellular growth factors, such as epidermal growth factor (EGF) and transforming growth factor α (TGFα) [40]. Previous crystallographic studies on the extracellular domain of EGFR suggests that ligand binding induces the exposure of dimerization arm and facilitates the extracellular domain dimerization [41, 42, 43]. Extracellular domain dimerized EGFR brings the intracellular kinase domain into close proximity so that the kinase domain could also dimerize [28, 22, 38, 44]. In 2006, a seminal paper by John Kuriyan and colleagues discovered that EGFR kinase domain adheres to an unexpected asymmetric dimer organization in which the N-lobe of a EGFR ("donor") interacts with the C-lobe of another EGFR ("acceptor") (Figure 1.1) [22]. Such an intermolecular interaction allosterically activates the "acceptor" kinase by

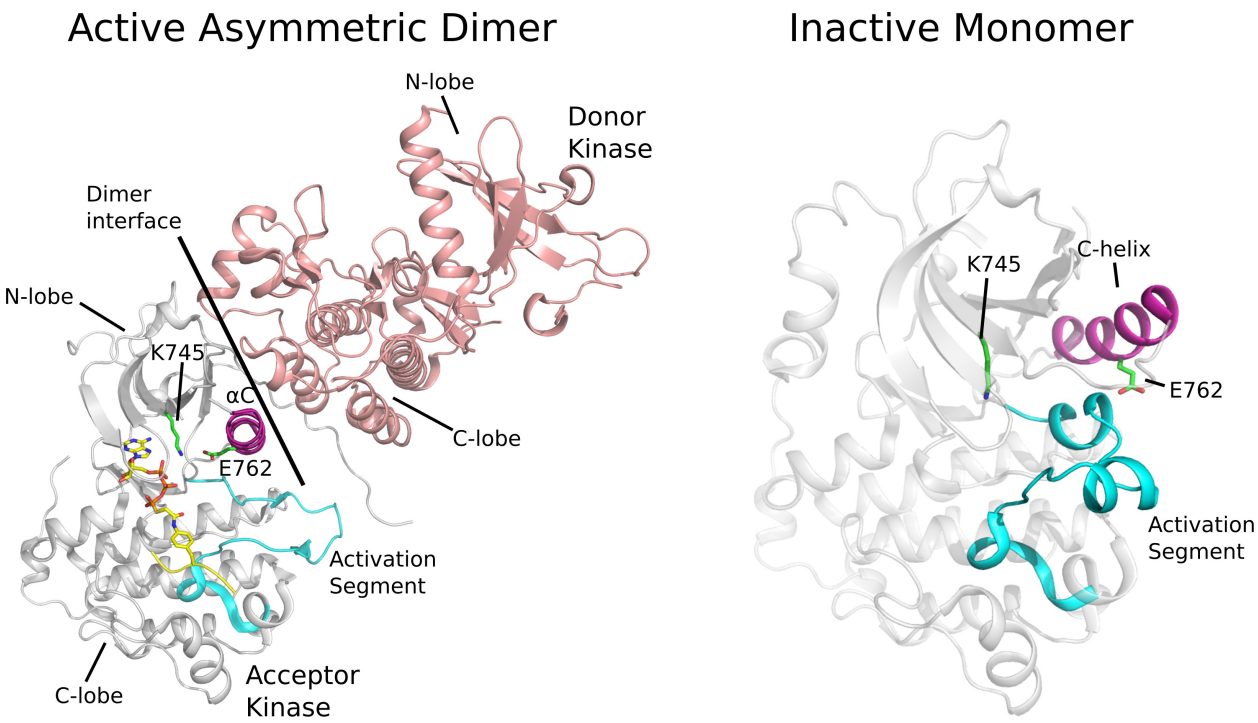

Figure 1.1: Crystal structures of EGFR in various conformational states. Left panel: the active asymmetric dimer of EGFR (PDBID: 2GS6). Right panel: the inactive monomeric form of EGFR (PDBID: 3W32).

positioning the catalytically crucial residues/motifs without the requirement of activation segment phosphorylation [22]. The C-terminal tail of EGFR will then be phosphorylated at several tyrosine residues to initiate downstream signaling [22, 45].

While the enzyme activity of WT EGFR is strictly controlled in normal state, in many cancer cells, such a regulation is lost due to various mutations in EGFR [28, 29]. Although patient-derived mutations are distributed throughout EGFR protein, more than 80% of the mutations map to the intracellular kinase domain [46]. Mutational activation of EGFR can be achieved through multiple different mechanisms. For example, the most frequently occurring L858R mutation in EGFR introduced a positively charged arginine residue at the activation segment, which is proposed to destabilize the characteristic 3/10 helix and destabilize the inactive state [28]. Further biochemical assays and computational simulations showed that the mutation activates EGFR by facilitating the dimerization process to achieve kinase activation [47, 20]. Another group of frequently occurring deletion mutations locate at the β3-αC loop of the kinase domain [48, 49]. Deletions at the loop is proposed to restrict the flexibility of αC-helix [39]. As a result, the deletion mutations are sensitive to EGFR inhibitor (erlotinib) that targets the active conformation of the kinase domain, but are resistant to the inhibitor (lapatinib) that targets the inactive kinase conformation [39].

The kinase domain of EGFR has been the target for many small-molecule inhibitor drugs (Table 1.1). Gefitinib, erlotinib and lapatinib are the three FDA-approved first-generation EGFR inhibitors for non-small cell lung cancer patients (NSCLC) [13]. Crystallographic study of EGFR with these inhibitors showed that they compete with ATP molecules to inhibit the enzyme activity of EGFR [21, 50, 51]. Although promising outcome has been reported for patients treated with these inhibitors, the responses are temporary in that the tu-

mor cells quickly develop secondary mutations and become resistant to the drugs [52, 53, 54]. The unfavorable prognosis is primarily due to the occurrence of the notorious "gatekeeper" mutation (T790M) in the kinase domain [55, 54]. To overcome the drug resistance, multiple strategies have been pursued in the past a few years. Since 2016, FDA has approved two novel inhibitors (afatinib and osimertinib) for NSCLC patients with specific mutations in EGFR kinase domain [56, 57]. The two inhibitors selectively target EGFR by covalently linking to the C797 residue and are effective for T790M-positive tumors [58, 59]. However, recent clinical report showed that the acquired C797S mutation is frequently observed in patients treated with afatinib and osimertinib, making the drug molecules ineffective [60, 61]. Novel allosteric inhibitors, such as EAI045, are currently under development for patients with both T790M and C797S mutations [18].

1.2.2 Molecular Dynamics

Molecular dynamics (MD) simulation is a widely used computer simulation technique to investigate the microscopic motions of molecules. MD simulation embraces a simple idea that the dynamic trajectory of an atomic system can be predicted by numerically solving classical equations of motion of interacting particles in a stepwise manner, where the potential energies between particles are determined by the force field [62]. While the first MD simulation was done by Alder and Wainwright in 1957 to study the interactions of hard spheres [63], the method was restricted to simple systems for several decades due to the limited computational power and the integration time step being in the range of several femtoseconds. The broad applicability of MD simulation is largely driven by the pervasiveness of the supercomputing environment and the development of fast and user-friendly simulation softwares [64, 65, 66,

67, 68, 69]. Community driven massive parallel computation and specialized computer such as Anton bring the modern MD simulation to millisecond regime for many biomolecules [70, 71, 72, 73]. The techniques is now being used regularly as a complement to lab experiments, allowing scientists to rationalize macroscopic observations with microscopic information.

While the traditional MD simulation is a good technique to study the intrinsic dynamics of biomolecules, it is inefficient to sample large conformational changes. This is primarily because the typical free energy landscape of biomolecules is rugged with many local minimums [74, 75]. Conventional MD simulations in the canonical ensemble sample the conformations according to the Boltzmann distribution, making it inefficient to jump across large free energy barrier and explore the full conformational space. Enhanced sampling algorithms, instead, sample from a modified/biased form of potential energy that favors regions of phase space that would be visited only infrequently in unbiased MD [76]. Once the MD snapshots drawn from this biased distribution are obtained, they can then be re-weighted to recover the unbiased free energy landscape [74, 76]. Over the past decades, many different flavors of biased/enhanced sampling methods have been proposed, including replica exchange MD [77], umbrella sampling [78], adaptive biasing methods [79], and metadynamics [80]. Many of these methods have been incorporated into the standard MD simulation software suite and allow researchers to apply these methods to reveal mechanistic insight into many interesting biological systems [81, 82, 83].

EGFR is one of the few receptor tyrosine kinases (RTKs) that have undergone extensive investigation through MD simulations. Early simulation of EGFR in various conformational states suggests that the interlobe movement of the kinase core is correlated with the dynamics of the juxtamembrane (JM) segment and the C-terminal tail [84]. In 2012, D.E.

Shaw Research performed more than 100 microseconds MD simulation on EGFR kinase domain and found that the N-lobe dimerization interface of EGFR is intrinsically disordered [20]. However, the disorder can be suppressed by kinase domain dimerization or the presence of activating oncogenic mutations [20]. A following MD study observed a spontaneous transition from the "active" state to the Src-like "inactive" state through two intermediate conformations [85]. In addition to the previously identified locally disordered state, the authors found an "extended" conformation, in which the two lobes of the kinase domain opens up to allow the folding and unfolding movement of the activation segment [85]. The entire EGFR protein in membrane environment has also been studied through coarse-grained MD simulation which yielded interesting insight into P-site recognition and self-phosphorylation events in the C-terminal tail [86].

MD simulation has also been used to study mutation mediated kinase activation and drug resistance mechanisms. For example, previous massive parallel tampering and metadynamics simulation quantified the equilibrium property and free energy landscape (FEL) of the kinase domain of WT EGFR and several clinically relevant point mutations (L858R, T790M, L858R/T790M) [87]. Later on, the FEL was combined with thermodynamic integration (TI) to understand the inhibitor binding to EGFR [88]. The analysis established that the origins of the difference in drug binding free energy between WT and mutants are due to the interplay between conformational dynamics and absolute binding affinity. In addition, The Molecular Mechanics Poisson-Boltzmann Surface Area (MM/PBSA) approach has been applied to predict the binding free energies with gefitinib and erlotinib to 113 EGFR mutations, providing a reference for clinical drug treatment [89].

1.3 Major research questions addressed

Despite previous intensive efforts on characterizing oncogenic mutations in EGFR, a significant knowledge gap of understanding the impact of patient-derived mutations is still present. The goal of the dissertation is to provide comprehensive and detailed mechanistic models of previously understudied mutations in EGFR which could also contribute to the uncontrolled growth signaling and tumor development. I prioritize EGFR mutations that map to functionally important regions of the kinase domain. Using a combination of experimental cell-based kinase assay and computational MD simulations, the dissertation identifies novel regulatory mechanisms underlie mutational activation of EGFR kinase. The three specific aspects investigated by the study are described below.

1.3.1 Mechanistic Insights into R776H Mediated Activation of Epidermal Growth Factor Receptor Kinase

R776H is a recurrent point mutation identified by iterative mining of protein kinase ontology (ProKinO) [24]. Previous study showed that EGFR bearing this mutation is constitutively active and are sensitive to gefitinib [24]. However, the mechanistic details of how R776H contributes to EGFR activation are not well understood. In this project, I use cell based assay and molecular dynamics (MD) simulations to understand the mutation impact of R776H in EGFR. Specifically, through detailed experimental design and mutagenesis, I uncover a previously underappreciated "lateral" phosphorylation mechanisms that are enhanced by oncogenic mutations. In addition, computational simulation identified a key auto-inhibitory interaction between R776 and regulatory αC-helix, which is disrupted by R776H mutant

and result in kinase activation. My results provide an emerging scheme of how mutations activate EGFR by modulating the critical structural element and provide clues for designing mutant specific kinase inhibitors.

1.3.2 Computational and Experimental Characterization of Patient Derived Mutations Reveal an Unusual Mode of Regulatory Spine Assembly and Drug Sensitivity in EGFR Kinase

The regulatory spine (RS) is a recently established structural motif present in all eukaryotic protein kinases [90, 91]. RS is consists of four non-consecutive hydrophobic residues that is spatially aligned in the active state of the kinase [92, 90, 91]. Previous mutagenesis studies on RS residues of PKA and B-Raf established that the integrity of RS assembly is critical for kinase activation and regulatory functions of the kinase [93, 94]. In particular, the RS of EGFR is also a mutation hotspot. The goal of this study is to characterize patient-derived RS mutations in EGFR and understand their mechanisms of action.

In this study, I establish a correlation between the side-chain size of the residue present at the RS3 position and EGFR autophosphorylation activity. Molecular modeling and molecular dynamics simulations of WT and mutant EGFR suggest a model in which RS3 mutations activate the kinase domain by disrupting the hydrophobic contact between RS3 residue and the activation segment. In addition, a buried water mediated hydrogen bond connecting T766 and catalytically important motifs of protein kinases is identified as a key determinant of M766T-mediated activation. M766T is resistant to FDA approved EGFR inhibitors such as gefitinib and erlotinib, and computational free energy analysis provides detailed mecha-

nisms associated with drug sensitivity. In sum, my studies suggest an unusual mode of RS assembly and oncogenic EGFR activation and provide new clues for the design of allosteric protein kinase inhibitors.

1.3.3 Altered conformational landscape and dimerization dependency underpins the activation of EGFR by αC-β4 loop insertion mutations

αC-β4 loop is an important structural motif conserved across all eukaryotic protein kinases (ePKs), including EGFR. Short in-frame insertion mutations at the αC-β4 loop of EGFR are frequently observed in many cancer patients and are often associated with drug resistance [95, 96, 97]. Despite the prevalence and clinical relevance of insertion mutations, the mechanisms by which they regulate EGFR activity and contribute to drug response is poorly understood. In this study, I demonstrate that insertion mutations in the αC-β4 loop alter kinase activity to various levels. I identify three mutations (N771_P772insN, D770_N771insG, and D770>GY) that constitutively activate EGFR and alter the dimerization dependency. The activating insertion mutations respond similarly to both reversible and irreversible EGFR inhibitors compared to WT EGFR. Based on structural modelling, molecular dynamics (MD) simulations and umbrella sampling techniques, I propose a model in which αC-β4 loop insertion mutations alter catalytic activity by controlling the conformational freedom of the regulatory αC-helix and lowering the transition free energy ($\Delta G_{\text{active-inactive}}$) between active and inactive states. My established pipeline not only provides detailed mechanisms associated with the αC-β4 loop mutation mediated kinase activation but also a framework to

understand αC-β4 loop mutation in other protein kinases.

Bibliography

[1] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144:646–74, 2011.

[2] Cliff Meldrum, Maria A Doyle, and Richard W Tothill. Next-generation sequencing for cancer diagnostics: a practical perspective. *The Clinical biochemist. Reviews*, 32:177–95, 2011.

[3] Shuyu Li and Mao Mao. Next generation sequencing reveals genetic landscape of hepatocellular carcinomas. *Cancer letters*, 340:247–53, 2013.

[4] Julia R Pon and Marco A Marra. Driver and passenger mutations in cancer. *Annual review of pathology*, 10:25–50, 2015.

[5] Simon Kebede Merid, Daria Goranskaya, and Andrey Alexeyenko. Distinguishing between driver and passenger mutations in individual cancer genomes by network enrichment analysis. *BMC bioinformatics*, 15:308, 2014.

[6] Ahmed El-Telbany and Patrick C Ma. Cancer genes in lung cancer: racial disparities: are there any? *Genes & cancer*, 3:467–80, 2012.

[7] Yong Chen, Jian-Xin Shi, Xu-Feng Pan, Jian Feng, and Heng Zhao. Identification of candidate genes for lung cancer somatic mutation test kits. *Genetics and molecular biology*, 36:455–64, 2013.

[8] Caterina Peraldo-Neia, Giorgia Migliardi, Maurizia Mello-Grand, Filippo Montemurro, Raffaella Segir, Ymera Pignochino, Giuliana Cavalloni, Bruno Torchio, Luciano Mosso, Giovanna Chiorino, and Massimo Aglietta. Epidermal growth factor receptor (egfr) mutation analysis, gene expression profiling and egfr protein expression in primary prostate cancer. *BMC cancer*, 11:31, 2011.

[9] T E Taylor, F B Furnari, and W K Cavenee. Targeting egfr for treatment of glioblastoma: molecular basis to overcome resistance. *Current cancer drug targets*, 12:197–209, 2012.

[10] A F Gazdar. Activating and resistance mutations of egfr in non-small-cell lung cancer: role in clinical response to egfr tyrosine kinase inhibitors. *Oncogene*, 28 Suppl 1:S24–31, 2009.

[11] Thomas J Lynch, Daphne W Bell, Raffaella Sordella, Sarada Gurubhagavatula, Ross A Okimoto, Brian W Brannigan, Patricia L Harris, Sara M Haserlat, Jeffrey G Supko, Frank G Haluska, David N Louis, David C Christiani, Jeff Settleman, and Daniel A Haber. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *The New England journal of medicine*, 350:2129–39, 2004.

[12] Giovanni Ciriello, Martin L Miller, BÃijlent Arman Aksoy, Yasin Senbabaoglu, Nikolaus Schultz, and Chris Sander. Emerging landscape of oncogenic signatures across human cancers. *Nature genetics*, 45:1127–33, 2013.

[13] Robert Roskoski. The erbb/her family of protein-tyrosine kinases and cancer. *Pharmacological research*, 79:34–74, 2014.

[14] Meredith K Bollinger, Amanda S Agnew, and Gerard P Mascara. Osimertinib: A third-generation tyrosine kinase inhibitor for treatment of epidermal growth factor receptor-mutated non-small cell lung cancer with the acquired thr790met mutation. *Journal of oncology pharmacy practice : official publication of the International Society of Oncology Pharmacy Practitioners*, page 1078155217712401, 2017.

[15] Siu-Fun Wong. Cetuximab: an epidermal growth factor receptor monoclonal antibody for the treatment of colorectal cancer. *Clinical therapeutics*, 27:684–94, 2005.

[16] Yi-Long Wu, Ying Cheng, Xiangdong Zhou, Ki Hyeong Lee, Kazuhiko Nakagawa, Seiji Niho, Fumito Tsuji, Rolf Linke, Rafael Rosell, Jesus Corral, Maria Rita Migliorino, Adam Pluzanski, Eric I Sbar, Tao Wang, Jane Liang White, Sashi Nadanaciva, Rickard Sandin, and Tony S Mok. Dacomitinib versus gefitinib as first-line treatment for patients with egfr-mutation-positive non-small-cell lung cancer (archer 1050): a randomised, open-label, phase 3 trial. *The Lancet. Oncology*, 18:1454–1466, 2017.

[17] Howard A Burris. Dual kinase inhibition in the treatment of breast cancer: initial experience with the egfr/erbb-2 inhibitor lapatinib. *The oncologist*, 9 Suppl 3:10–5, 2004.

[18] Yong Jia, Cai-Hong Yun, Eunyoung Park, Dalia Ercan, Mari Manuia, Jose Juarez, Chunxiao Xu, Kevin Rhee, Ting Chen, Haikuo Zhang, Sangeetha Palakurthi, Jaebong Jang, Gerald Lelais, Michael DiDonato, Badry Bursulaya, Pierre-Yves Michellys, Robert Epple, Thomas H Marsilje, Matthew McNeill, Wenshuo Lu, Jennifer Harris, Steven Bender, Kwok-Kin Wong, Pasi A JÃďnne, and Michael J Eck. Overcoming egfr(t790m)

and egfr(c797s) resistance with mutant-selective allosteric inhibitors. *Nature*, 534:129–32, 2016.

[19] Wenjun Zhou, Dalia Ercan, Liang Chen, Cai-Hong Yun, Danan Li, Marzia Capelletti, Alexis B Cortot, Lucian Chirieac, Roxana E Iacob, Robert Padera, John R Engen, Kwok-Kin Wong, Michael J Eck, Nathanael S Gray, and Pasi A JÃďnne. Novel mutant-selective egfr kinase inhibitors against egfr t790m. *Nature*, 462:1070–4, 2009.

[20] Yibing Shan, Michael P Eastwood, Xuewu Zhang, Eric T Kim, Anton Arkhipov, Ron O Dror, John Jumper, John Kuriyan, and David E Shaw. Oncogenic mutations counteract intrinsic disorder in the egfr kinase and promote receptor dimerization. *Cell*, 149:860–70, 2012.

[21] Cai-Hong Yun, Titus J Boggon, Yiqun Li, Michele S Woo, Heidi Greulich, Matthew Meyerson, and Michael J Eck. Structures of lung cancer-derived egfr mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer cell*, 11:217–27, 2007.

[22] Xuewu Zhang, Jodi Gureasko, Kui Shen, Philip A Cole, and John Kuriyan. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell*, 125:1137–49, 2006.

[23] Kaiyan Chen, Xiaoqing Yu, Haiyang Wang, Zhiyu Huang, Yanjun Xu, Lei Gong, and Yun Fan. Uncommon mutation types of epidermal growth factor receptor and response to egfr tyrosine kinase inhibitors in chinese non-small cell lung cancer patients. *Cancer chemotherapy and pharmacology*, 80:1179–1187, 2017.

[24] Daniel Ian McSkimming, Shima Dastgheib, Eric Talevich, Anish Narayanan, Samiksha Katiyar, Susan S Taylor, Krys Kochut, and Natarajan Kannan. Prokino: a unified resource for mining the cancer kinome. *Human mutation*, 36:175–86, 2015.

[25] Zheng Ruan and Natarajan Kannan. Mechanistic insights into r776h mediated activation of epidermal growth factor receptor kinase. *Biochemistry*, 54:4216–25, 2015.

[26] Shinji Kohsaka, Masaaki Nagano, Toshihide Ueno, Yoshiyuki Suehara, Takuo Hayashi, Naoko Shimada, Kazuhisa Takahashi, Kenji Suzuki, Kazuya Takamochi, Fumiyuki Takahashi, and Hiroyuki Mano. A method of high-throughput functional evaluation of egfr gene variants of unknown significance in cancer. *Science translational medicine*, 9, 2017.

[27] T N Raju. The nobel chronicles. 1986: Stanley cohen (b 1922); rita levi-montalcini (b 1909). *Lancet (London, England)*, 355:506, 2000.

[28] Ami Citri and Yosef Yarden. Egf-erbb signalling: towards the systems level. *Nature reviews. Molecular cell biology*, 7:505–16, 2006.

[29] Nicola Normanno, Antonella De Luca, Caterina Bianco, Luigi Strizzi, Mario Mancino, Monica R Maiello, Adele Carotenuto, Gianfranco De Feo, Francesco Caponigro, and David S Salomon. Epidermal growth factor receptor (egfr) signaling in cancer. *Gene*, 366:2–16, 2006.

[30] Sana Siddiqui, Meng Fang, Bin Ni, Daoyuan Lu, Bronwen Martin, and Stuart Maudsley. Central role of the egf receptor in neurometabolic aging. *International journal of endocrinology*, 2012:739428, 2012.

[31] Patrick Campbell, Penny E Morton, Takuya Takeichi, Amr Salam, Nerys Roberts, Laura E Proudfoot, Jemima E Mellerio, Kingi Aminu, Cheryl Wellington, Sachin N Patil, Masashi Akiyama, Lu Liu, James R McMillan, Sophia Aristodemou, Akemi Ishida-Yamamoto, Alya Abdul-Wahab, Gabriela Petrof, Kenneth Fong, Sarawin Harnchoowong, Kristina L Stone, John I Harper, W H Irwin McLean, Michael A Simpson, Maddy Parsons, and John A McGrath. Epithelial inflammation resulting from an inherited loss-of-function mutation in egfr. *The Journal of investigative dermatology*, 134:2570–2578, 2014.

[32] Kathryn M Ferguson. Structure-based view of epidermal growth factor receptor regulation. *Annual review of biophysics*, 37:353–73, 2008.

[33] Erika Kovacs, Julie Anne Zorn, Yongjian Huang, Tiago Barros, and John Kuriyan. A structural perspective on the regulation of the epidermal growth factor receptor. *Annual review of biochemistry*, 84:739–64, 2015.

[34] Jennifer Stamos, Mark X Sliwkowski, and Charles Eigenbrot. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *The Journal of biological chemistry*, 277:46265–72, 2002.

[35] Amar Mirza, Morad Mustafa, Eric Talevich, and Natarajan Kannan. Co-conserved features associated with cis regulation of erbb tyrosine kinases. *PloS one*, 5:e14310, 2010.

[36] Morgan Huse and John Kuriyan. The conformational plasticity of protein kinases. *Cell*, 109:275–82, 2002.

[37] Daniel Ian McSkimming, Khaled Rasheed, and Natarajan Kannan. Classifying kinase conformations using a machine learning approach. *BMC bioinformatics*, 18:86, 2017.

[38] Natalia Jura, Nicholas F Endres, Kate Engel, Sebastian Deindl, Rahul Das, Meindert H Lamers, David E Wemmer, Xuewu Zhang, and John Kuriyan. Mechanism for activation of the egf receptor catalytic domain by the juxtamembrane segment. *Cell*, 137:1293–307, 2009.

[39] Scott A Foster, Daniel M Whalen, Aysegul Ozen, Matthew J Wongchenko, JianPing Yin, Ivana Yen, Gabriele Schaefer, John D Mayfield, Juliann Chmielecki, Philip J Stephens, Lee A Albacker, Yibing Yan, Kyung Song, Georgia Hatzivassiliou, Charles Eigenbrot, Christine Yu, Andrey S Shaw, Gerard Manning, Nicholas J Skelton, Sarah G Hymowitz, and Shiva Malek. Activation mechanism of oncogenic deletion mutations in braf, egfr, and her2. *Cancer cell*, 29:477–493, 2016.

[40] Amy Doerner, Rebecca Scheck, and Alanna Schepartz. Growth factor identity is encoded by discrete coiled-coil rotamers in the egfr juxtamembrane region. *Chemistry & biology*, 22:776–84, 2015.

[41] Thomas P J Garrett, Neil M McKern, Meizhen Lou, Thomas C Elleman, Timothy E Adams, George O Lovrecz, Hong-Jian Zhu, Francesca Walker, Morry J Frenkel, Peter A Hoyne, Robert N Jorissen, Edouard C Nice, Antony W Burgess, and Colin W Ward. Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha. *Cell*, 110:763–73, 2002.

[42] Hideo Ogiso, Ryuichiro Ishitani, Osamu Nureki, Shuya Fukai, Mari Yamanaka, Jae-Hoon Kim, Kazuki Saito, Ayako Sakamoto, Mio Inoue, Mikako Shirouzu, and Shigeyuki

Yokoyama. Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell*, 110:775–87, 2002.

[43] Shiqing Li, Karl R Schmitz, Philip D Jeffrey, Jed J W Wiltzius, Paul Kussie, and Kathryn M Ferguson. Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer cell*, 7:301–11, 2005.

[44] Monica Red Brewer, Sung Hee Choi, Diego Alvarado, Katarina Moravcevic, Ambra Pozzi, Mark A Lemmon, and Graham Carpenter. The juxtamembrane region of the egf receptor functions as an activation domain. *Molecular cell*, 34:641–51, 2009.

[45] Yongjian Huang, Shashank Bharill, Deepti Karandur, Sean M Peterson, Morgan Marita, Xiaojun Shi, Megan J Kaliszewski, Adam W Smith, Ehud Y Isacoff, and John Kuriyan. Molecular basis for multimerization in the activation of the epidermal growth factor receptor. *eLife*, 5, 2016.

[46] S A Forbes, G Bhamra, S Bamford, E Dawson, C Kok, J Clements, A Menzies, J W Teague, P A Futreal, and M R Stratton. The catalogue of somatic mutations in cancer (cosmic). *Current protocols in human genetics*, Chapter 10:Unit 10.11, 2008.

[47] Zhihong Wang, Patti A Longo, Mary Katherine Tarrant, Kwangsoo Kim, Sarah Head, Daniel J Leahy, and Philip A Cole. Mechanistic insights into the activation of oncogenic forms of egf receptor. *Nature structural & molecular biology*, 18:1388–93, 2011.

[48] J Guillermo Paez, Pasi A JÃďnne, Jeffrey C Lee, Sean Tracy, Heidi Greulich, Stacey Gabriel, Paula Herman, Frederic J Kaye, Neal Lindeman, Titus J Boggon, Katsuhiko Naoki, Hidefumi Sasaki, Yoshitaka Fujii, Michael J Eck, William R Sellers, Bruce E

Johnson, and Matthew Meyerson. Egfr mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science (New York, N.Y.)*, 304:1497–500, 2004.

[49] David M Jackman, Beow Y Yeap, Lecia V Sequist, Neal Lindeman, Alison J Holmes, Victoria A Joshi, Daphne W Bell, Mark S Huberman, Balazs Halmos, Michael S Rabin, Daniel A Haber, Thomas J Lynch, Matthew Meyerson, Bruce E Johnson, and Pasi A Janne. Exon 19 deletion mutations of epidermal growth factor receptor are associated with prolonged survival in non-small cell lung cancer patients treated with gefitinib or erlotinib. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 12:3908–14, 2006.

[50] Jin H Park, Yingting Liu, Mark A Lemmon, and Ravi Radhakrishnan. Erlotinib binds both inactive and active conformations of the egfr tyrosine kinase domain. *The Biochemical journal*, 448:417–23, 2012.

[51] Edgar R Wood, Anne T Truesdale, Octerloney B McDonald, Derek Yuan, Anne Hassell, Scott H Dickerson, Byron Ellis, Christopher Pennisi, Earnest Horne, Karen Lackey, Krystal J Alligood, David W Rusnak, Tona M Gilmer, and Lisa Shewchuk. A unique structure for epidermal growth factor receptor bound to gw572016 (lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer research*, 64:6652–9, 2004.

[52] Erin L Stewart, Samuel Zhixing Tan, Geoffrey Liu, and Ming-Sound Tsao. Known and putative mechanisms of resistance to egfr targeted therapies in nsclc patients with egfr mutations-a review. *Translational lung cancer research*, 4:67–81, 2015.

[53] Lihua Huang and Liwu Fu. Mechanisms of resistance to egfr tyrosine kinase inhibitors. *Acta pharmaceutica Sinica. B*, 5:390–401, 2015.

[54] James Bean, Cameron Brennan, Jin-Yuan Shih, Gregory Riely, Agnes Viale, Lu Wang, Dhananjay Chitale, Noriko Motoi, Janos Szoke, Stephen Broderick, Marissa Balak, Wen-Cheng Chang, Chong-Jen Yu, Adi Gazdar, Harvey Pass, Valerie Rusch, William Gerald, Shiu-Feng Huang, Pan-Chyr Yang, Vincent Miller, Marc Ladanyi, Chih-Hsin Yang, and William Pao. Met amplification occurs with or without t790m mutations in egfr mutant lung tumors with acquired resistance to gefitinib or erlotinib. *Proceedings of the National Academy of Sciences of the United States of America*, 104:20932–7, 2007.

[55] Cai-Hong Yun, Kristen E Mengwasser, Angela V Toms, Michele S Woo, Heidi Greulich, Kwok-Kin Wong, Matthew Meyerson, and Michael J Eck. The t790m mutation in egfr kinase causes drug resistance by increasing the affinity for atp. *Proceedings of the National Academy of Sciences of the United States of America*, 105:2070–5, 2008.

[56] Natalie Minkovsky and Alan Berezov. Bibw-2992, a dual receptor tyrosine kinase inhibitor for the treatment of solid tumors. *Current opinion in investigational drugs (London, England : 2000)*, 9:1336–46, 2008.

[57] Sarah L Greig. Osimertinib: First global approval. *Drugs*, 76:263–73, 2016.

[58] Tristan A Barnes, Grainne M O'Kane, Mark David Vincent, and Natasha B Leighl. Third-generation tyrosine kinase inhibitors targeting epidermal growth factor receptor mutations in non-small cell lung cancer. *Frontiers in oncology*, 7:113, 2017.

[59] Roberta Minari, Paola Bordi, and Marcello Tiseo. Third-generation epidermal growth

factor receptor-tyrosine kinase inhibitors in t790m-positive non-small cell lung cancer: review on emerged mechanisms of resistance. *Translational lung cancer research*, 5:695–708, 2016.

[60] Shuhang Wang, Stella T Tsui, Christina Liu, Yongping Song, and Delong Liu. Egfr c797s mutation mediates resistance to third-generation inhibitors in t790m-positive non-small cell lung cancer. *Journal of hematology & oncology*, 9:59, 2016.

[61] Kenneth S Thress, Cloud P Paweletz, Enriqueta Felip, Byoung Chul Cho, Daniel Stetson, Brian Dougherty, Zhongwu Lai, Aleksandra Markovets, Ana Vivancos, Yanan Kuang, Dalia Ercan, Sarah E Matthews, Mireille Cantarini, J Carl Barrett, Pasi A JÃďnne, and Geoffrey R Oxnard. Acquired egfr c797s mutation mediates resistance to azd9291 in non-small cell lung cancer harboring egfr t790m. *Nature medicine*, 21:560–2, 2015.

[62] Ron O Dror, Robert M Dirks, J P Grossman, Huafeng Xu, and David E Shaw. Biomolecular simulation: a computational microscope for molecular biology. *Annual review of biophysics*, 41:429–52, 2012.

[63] B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *The Journal of Chemical Physics*, 27(5):1208–1209, 1957.

[64] David E. Shaw, Martin M. Deneroff, Ron O. Dror, Jeffrey S. Kuskin, Richard H. Larson, John K. Salmon, Cliff Young, Brannon Batson, Kevin J. Bowers, Jack C. Chao, Michael P. Eastwood, Joseph Gagliardo, J.P. Grossman, C. Richard Ho, Douglas J. Ierardi, Istvan Kolossvary, John L. Klepeis, Timothy Layman, Christine McLeavey,

Mark A. Moraes, Rolf Mueller, Edward C. Priest, Yibing Shan, Jochen Spengler, Michael Theobald, Brian Towles, and Stanley C. Wang. Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM*, 51(7):91–97, 2008.

[65] David E. Shaw, J.P. Grossman, Joseph A. Bank, Brannon Batson, J. Adam Butts, Jack C. Chao, Martin M. Deneroff, Ron O. Dror, Amos Even, Christopher H. Fenton, Anthony Forte, Joseph Gagliardo, Gennette Gill, Brian Greskamp, C. Richard Ho, Douglas J. Ierardi, Lev Iserovich, Jeffrey S. Kuskin, Richard H. Larson, Timothy Layman, Li-Siang Lee, Adam K. Lerer, Chester Li, Daniel Killebrew, Kenneth M. Mackenzie, Shark Yeuk-Hai Mok, Mark A. Moraes, Rolf Mueller, Lawrence J. Nociolo, Jon L. Peticolas, Terry Quan, Daniel Ramot, John K. Salmon, Daniele P. Scarpazza, U. Ben Schafer, Naseer Siddique, Christopher W. Snyder, Jochen Spengler, Ping Tak Peter Tang, Michael Theobald, Horia Toma, Brian Towles, Benjamin Vitale, Stanley C. Wang, and Cliff Young. Anton 2: Raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis(SC)*, volume 00, pages 41–53, Nov. 2014.

[66] David A Case, Thomas E Cheatham, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M Merz, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J Woods. The amber biomolecular simulation programs. *Journal of computational chemistry*, 26:1668–88, 2005.

[67] Mark James Abraham, Teemu Murtola, Roland Schulz, SzilÃąrd PÃąll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations

through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19 – 25, 2015.

[68] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant KalÃĺ, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26:1781–802, 2005.

[69] Peter Eastman, Jason Swails, John D Chodera, Robert T McGibbon, Yutong Zhao, Kyle A Beauchamp, Lee-Ping Wang, Andrew C Simmonett, Matthew P Harrigan, Chaya D Stern, Rafal P Wiewiora, Bernard R Brooks, and Vijay S Pande. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology*, 13:e1005659, 2017.

[70] Vincent A Voelz, Gregory R Bowman, Kyle Beauchamp, and Vijay S Pande. Molecular simulation of ab initio protein folding for a millisecond folder ntl9(1-39). *Journal of the American Chemical Society*, 132:1526–8, 2010.

[71] Mohammad M Sultan, Rajiah Aldrin Denny, Ray Unwalla, Frank Lovering, and Vijay S Pande. Millisecond dynamics of btk reveal kinome-wide conformational plasticity within the apo kinase domain. *Scientific reports*, 7:15604, 2017.

[72] David E. Shaw, Ron O. Dror, John K. Salmon, J. P. Grossman, Kenneth M. Mackenzie, Joseph A. Bank, Cliff Young, Martin M. Deneroff, Brannon Batson, Kevin J. Bowers, Edmond Chow, Michael P. Eastwood, Douglas J. Ierardi, John L. Klepeis, Jeffrey S. Kuskin, Richard H. Larson, Kresten Lindorff-Larsen, Paul Maragakis, Mark A. Moraes,

Stefano Piana, Yibing Shan, and Brian Towles. Millisecond-scale molecular dynamics simulations on anton. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, SC '09, pages 65:1–65:11, New York, NY, USA, 2009. ACM.

[73] Kresten Lindorff-Larsen, Paul Maragakis, Stefano Piana, and David E Shaw. Picosecond to millisecond structural dynamics in human ubiquitin. *The journal of physical chemistry. B*, 120:8313–20, 2016.

[74] Ayori Mitsutake, Yoshiharu Mori, and Yuko Okamoto. Enhanced sampling algorithms. *Methods in molecular biology (Clifton, N.J.)*, 924:153–95, 2013.

[75] Tatiana Maximova, Ryan Moffatt, Buyong Ma, Ruth Nussinov, and Amarda Shehu. Principles and overview of sampling methods for modeling macromolecular structure and dynamics. *PLoS computational biology*, 12:e1004619, 2016.

[76] Jeffrey Comer, James C Gumbart, JÃľrÃŧme HÃľnin, Tony LeliÃĺvre, Andrew Pohorille, and Christophe Chipot. The adaptive biasing force method: everything you always wanted to know but were afraid to ask. *The journal of physical chemistry. B*, 119:1129–51, 2015.

[77] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1):141 – 151, 1999.

[78] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for freeâĂŘenergy calculations

on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.

[79] Eric Darve, David Rodriguez-Gomez, and Andrew Pohorille. Adaptive biasing force method for scalar and vector free energy calculations. *The Journal of Chemical Physics*, 128(14):144120, 2008.

[80] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, 99:12562–6, 2002.

[81] Massimiliano Bonomi, Davide Branduardi, Giovanni Bussi, Carlo Camilloni, Davide Provasi, Paolo Raiteri, Davide Donadio, Fabrizio Marinelli, Fabio Pietrucci, Ricardo A. Broglia, and Michele Parrinello. Plumed: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications*, 180(10):1961 – 1972, 2009.

[82] Bradley M Dickson, Parker W de Waal, Zachary H Ramjan, H Eric Xu, and Scott B Rothbart. A fast, open source implementation of adaptive biasing potentials uncovers a ligand design strategy for the chromatin regulator brd4. *The Journal of chemical physics*, 145:154113, 2016.

[83] Rafael C Bernardi, Marcelo C R Melo, and Klaus Schulten. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et biophysica acta*, 1850:872–877, 2015.

[84] Morad Mustafa, Amar Mirza, and Natarajan Kannan. Conformational regulation of

the egfr kinase core by the juxtamembrane and c-terminal tail: a molecular dynamics study. *Proteins*, 79:99–114, 2011.

[85] Yibing Shan, Anton Arkhipov, Eric T Kim, Albert C Pan, and David E Shaw. Transitions to catalytically inactive conformations in egfr kinase. *Proceedings of the National Academy of Sciences of the United States of America*, 110:7270–5, 2013.

[86] John G Koland. Coarse-grained molecular simulation of epidermal growth factor receptor protein tyrosine kinase multi-site self-phosphorylation. *PLoS computational biology*, 10:e1003435, 2014.

[87] Ludovico Sutto and Francesco Luigi Gervasio. Effects of oncogenic mutations on the conformational free-energy landscape of egfr kinase. *Proceedings of the National Academy of Sciences of the United States of America*, 110:10616–21, 2013.

[88] Jiyong Park, Joseph J McDonald, Russell C Petter, and K N Houk. Molecular dynamics analysis of binding of kinase inhibitors to wt egfr and the t790m mutant. *Journal of chemical theory and computation*, 12:2066–78, 2016.

[89] Lichun Ma, Debby D Wang, Yiqing Huang, Hong Yan, Maria P Wong, and Victor H F Lee. Egfr mutant structural database: computationally predicted 3d structures and the corresponding binding free energies with gefitinib and erlotinib. *BMC bioinformatics*, 16:85, 2015.

[90] Susan S Taylor and Alexandr P Kornev. Protein kinases: evolution of dynamic regulatory proteins. *Trends in biochemical sciences*, 36:65–77, 2011.

[91] Alexandr P Kornev, Susan S Taylor, and Lynn F Ten Eyck. A helix scaffold for the assembly of active protein kinases. *Proceedings of the National Academy of Sciences of the United States of America*, 105:14377–82, 2008.

[92] Alexandr P Kornev, Nina M Haste, Susan S Taylor, and Lynn F Ten Eyck. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 103:17783–8, 2006.

[93] Jiancheng Hu, Lalima G Ahuja, Hiruy S Meharena, Natarajan Kannan, Alexandr P Kornev, Susan S Taylor, and Andrey S Shaw. Kinase regulation by hydrophobic spine assembly in cancer. *Molecular and cellular biology*, 35:264–76, 2015.

[94] Hiruy S Meharena, Philip Chang, Malik M Keshwani, Krishnadev Oruganty, Aishwarya K Nene, Natarajan Kannan, Susan S Taylor, and Alexandr P Kornev. Deciphering the structural basis of eukaryotic protein kinase regulation. *PLoS biology*, 11:e1001680, 2013.

[95] Hiroyuki Yasuda, Susumu Kobayashi, and Daniel B Costa. Egfr exon 20 insertion mutations in non-small-cell lung cancer: preclinical data and clinical implications. *The Lancet. Oncology*, 13:e23–31, 2012.

[96] Maria E Arcila, Khedoudja Nafa, Jamie E Chaft, Natasha Rekhtman, Christopher Lau, Boris A Reva, Maureen F Zakowski, Mark G Kris, and Marc Ladanyi. Egfr exon 20 insertion mutations in lung adenocarcinomas: prevalence, molecular heterogeneity, and clinicopathologic characteristics. *Molecular cancer therapeutics*, 12:220–9, 2013.

[97] Hiroyuki Yasuda, Eunyoung Park, Cai-Hong Yun, Natasha J Sng, Antonio R Lucena-Araujo, Wee-Lee Yeo, Mark S Huberman, David W Cohen, Sohei Nakayama, Kota Ishioka, Norihiro Yamaguchi, Megan Hanna, Geoffrey R Oxnard, Christopher S Lathan, Teresa Moran, Lecia V Sequist, Jamie E Chaft, Gregory J Riely, Maria E Arcila, Ross A Soo, Matthew Meyerson, Michael J Eck, Susumu S Kobayashi, and Daniel B Costa. Structural, biochemical, and clinical characterization of epidermal growth factor receptor (egfr) exon 20 insertion mutations in lung cancer. *Science translational medicine*, 5:216ra177, 2013.

Chapter 2

Mechanistic Insights into R776H Mediated Activation of Epidermal Growth Factor Receptor (EGFR) Kinase

Zheng Ruan and Natarajan Kannan (2015) *Biochemistry* 54(27):6216-25.
Reprinted here with permission from the publisher.

Abstract

Mutational activation of human Epidermal Growth Factor Receptor (EGFR) kinase is implicated in many different cancer patients. R776H is one recurrent mutation in the αC-β4 loop of the tyrosine kinase domain that activates EGFR in the absence of the activating EGF ligand. However, the mechanistic details of how R776H (R752H in another numbering system) contributes to kinase activation are not well understood. Here using cell-based co-transfection assays, we show that the R776H mutation activates EGFR in a dimerization dependent manner by preferentially adopting the acceptor position in the asymmetric dimer. The acceptor function, but not the donor function, is enhanced for the R776H mutant, supporting the "superacceptor" hypothesis proposed for oncogenic mutations in EGFR. We also find that phosphorylation of monomeric EGFR is increased by R776H mutation, providing insights into EGFR lateral phosphorylation and oligomerization. Based on molecular modeling and molecular dynamics simulation, we propose a model in which loss of key auto-inhibitory αC-helix capping interaction and alteration of co-conserved *cis* regulatory interactions between the kinase domain and the flanking regulatory segments contribute to mutational activation. Since the R776 equivalent position is mutated in ErbB2 and ErbB4, our studies have implications for understanding kinase mutational activation in other ErbB family members as well.

2.1 Introduction

Epidermal growth factor receptor (EGFR) associated pathways are critical for regulating cell growth, proliferation, differentiation, and survival [1, 2, 3]. While EGFR signaling is tightly controlled by a diverse array of regulatory mechanisms in normal cells, in many cancer cells, the regulatory constraint on EGFR signaling is lost, resulting in abnormal cell growth and proliferation [2, 3, 4, 5]. Cancer genome sequencing studies have revealed hundreds of mutations in EGFR, many of which map to the intracellular kinase domain. Although more than 800 unique mutations have been identified in EGFR, only a handful of recurrent mutations have been structurally and functionally characterized. The mechanisms of action of many less frequently occurring mutations are poorly understood. R776H/C/G is one such mutation in the αC-β4 loop region of EGFR kinase domain (Figure 2.1). A total of 22 unique samples (patients) have been reported to carry mutations at R776 position in COSMIC database, and 14 of them are R776H mutation. In one clinical report, R776H mutation is associated with lung cancer patients without smoking history and is found both in normal and tumor tissues [6]. In addition, R776H/C mutations are known to co-occur with other common oncogenic mutations, such as L858R, G719A and L861Q, and confer sensitivity to cancer drugs [6, 7, 8]. R776H/C mutations have been shown to activate EGFR kianse domain in the absence of the EGF ligand [9]. However, the mechanism by which R776H/C activates the kinase domain is not fully understood.

Here we employ a combination of computational and experimental approaches to characterize the R776H mutation in EGFR. First, we show that R776H activates EGFR in a ligand independent manner; however the mutant still relies on the asymmetric dimer for its activity. When co-expressed with WT EGFR, R776H mutant preferentially adopts the ac-

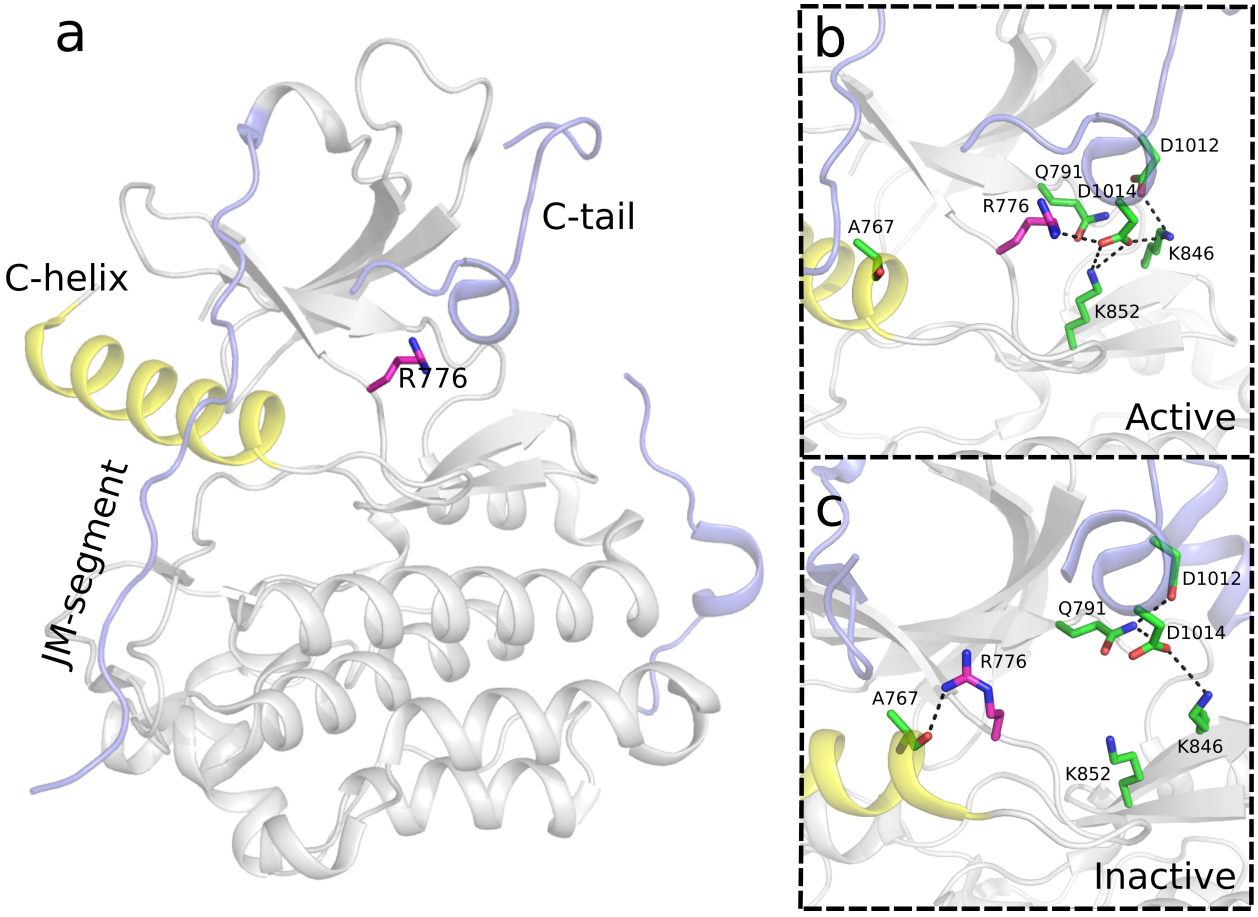

Figure 2.1: Structure of EGFR showing the position of R776 and associated interactions in active (2GS6) and inactive (3W32) states.

ceptor position to better exert its activity, i.e., it functions as a "superacceptor". We further demonstrate that the enzymatic activity of R776H is not restricted to the dimer itself in that the activated EGFR dimer can phosphorylate inactive monomeric EGFR *in vivo*. Lastly, we performed molecular dynamics to investigate the activation mechanism of R776H mutant. Our study suggests a model in which the R776H mutation activates EGFR by relieving auto-inhibitory interactions with the αC-helix as well as the auto-inhibitory C-terminal tail. Our results provide an emerging scheme of how mutations activate EGFR by regulation of the critical αC-helix, and provide clues for designing mutant specific kinase inhibitors.

2.2 Results

2.2.1 R776H mutation activates EGFR in a dimerization dependent manner

We previously demonstrated that the R776H mutant is constitutively active and displays catalytic activity in the absence of the activating EGF ligand [9](Figure 2.2a, lane 3-4); however, the role of dimerization in R776H-mediated EGFR activation was not studied. To test the dimerization dependency of R776H, we introduced an N-lobe dimerization deficient mutation L704N (L680N in another numbering system) and a C-lobe dimerization deficient mutation V948R (V924R in another numbering system) in the R776H background [10]. Western blot analysis indicates that dimerization deficient mutants, R776H/L704N and R776H/V948R, are inactive (Figure 2.2a, lane 5-8). However, Y1197 (Y1173 in another numbering system) phosphorylation can be restored when both constructs are co-expressed (Figure 2.2a, lane 9-10), indicating that the activation of R776H relies on the intact dimerization interface.

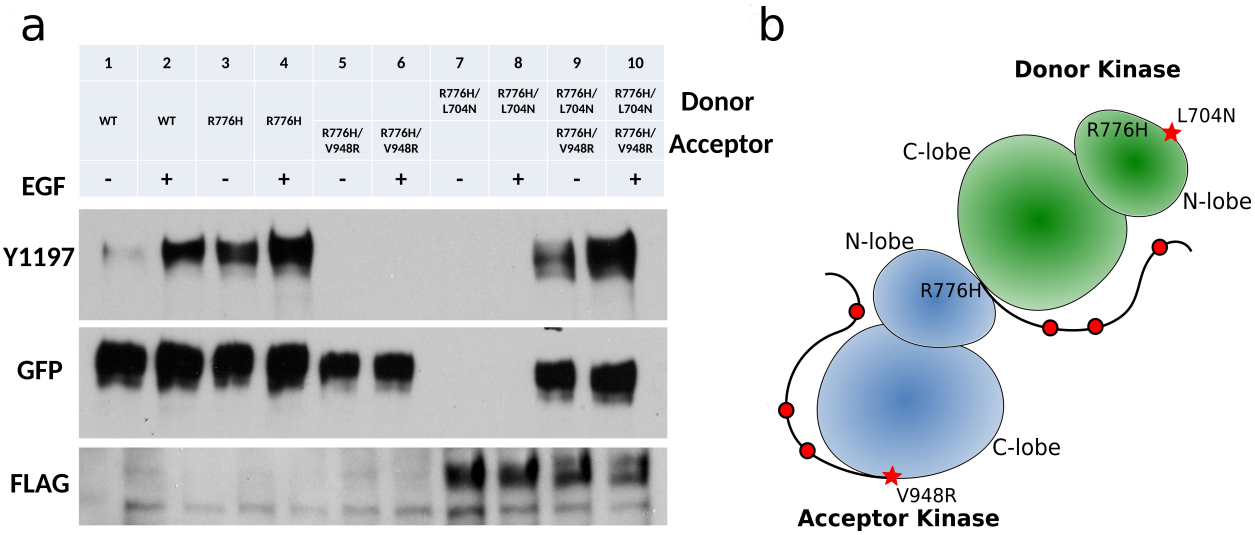

Figure 2.2: R776H depends on the asymmetric dimer for activation. a) Lanes from left to right, WT EGFR (-), WT EGFR (+), R776H (-), R776H (+), R776H/V948R (-), R776H/V948R (+), R776H/L704N (-), R776H/L704N (+), R776H/V948R & R776H/L704N (-), R776H/V948R & R776H/L704N (+). + and - indicate the presence and absence of EGF ligand, respectively. WT EGFR, R776H, R776H/V948R are GFP tagged, whereas R776H/L704N is FLAG tagged. b) Cartoon representation of the asymmetric dimer that is formed between R776H/V948R and R776H/L704N.

Thus, the asymmetric dimer is required for R776H mediated activation (Figure 2.2b). Ligand independent phosphorylation of R776H also depends on the asymmetric dimer (Figure 2.2a, lane 3, 9), because auto-phosphorylation is not detected when R776H is dimerization deficient (Figure 2.2a, lane 5, 7).

2.2.2 R776H mutant preferentially adopts the acceptor position in the asymmetric dimer

To test whether R776H mutant preferentially adopts the acceptor position in the asymmetric dimer, we used the complementation assay, as described in a recent study [11]. Specifically, we reconstituted the asymmetric dimer by co-expressing designated acceptor (V948R and R776H/V948R) (Figure 2.3, lane 4, 6, 14, 16) and donor (L704N and R776H/L704N) (Figure 2.3, lane 3, 5, 13, 15) constructs, and probed for C-terminal tail auto-phosphorylation (Y1197). Y1197 phosphorylation is greatly enhanced when R776H mutant is in the acceptor position (Figure 2.3, lane 7 vs 8, 9 vs 10 and Figure 2.3, lane 17 vs 18, lane 19 vs 20), suggesting that the R776H mutant has a higher intrinsic kinase activity compared to WT. Co-transfection of enforced donor and acceptor constructs shows that phosphorylation of Y1197 decreases when R776H is in the donor position (Figure 2.3, lane 7 vs 9, 8 vs 10, 17 vs 19, and 18 vs 20). Interestingly, the highest tyrosine phosphorylation is observed when WT donor (L704N) is reconstituted with R776H acceptor (R776H/V948R) (Figure 2.3, lane 8, 18). Taken together, these results indicate that the R776H mutant preferentially adopts the acceptor position when paired with WT EGFR, providing support for the "superacceptor" hypothesis proposed for other lung cancer mutations [11].

2.2.3 R776H mutant enhances lateral phosphorylation of monomeric EGFR

In the traditional view of EGFR activation, only the acceptor kinase is activated upon dimerization [10]. Once activated, the acceptor kinase phosphorylates tyrosine residues in

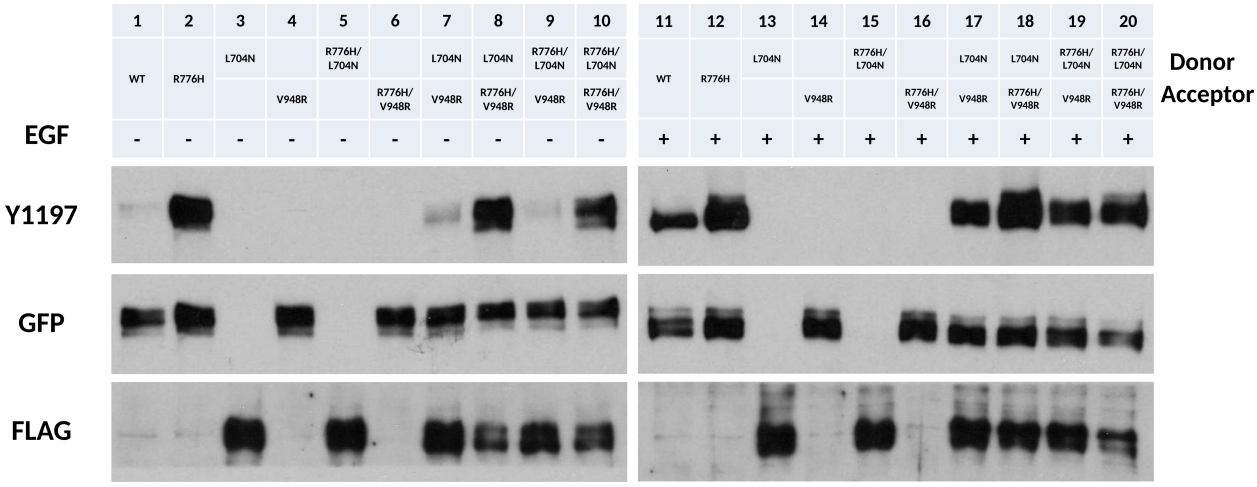

Figure 2.3: R776H is a "superacceptor". + and - indicate the presence and absence of EGF ligand, respectively. Lanes from left to right, WT EGFR (-), R776H (-), L704N (-), V948R (-), R776H/L704N (-), R776H/V948R (-), L704N & V948R (-), L704N & R776H/V948R (-), R776H/L704N & V948R (-), R776H/L704N & R776H/V948R (-), WT EGFR (+), R776H (+), L704N (+), V948R (+), R776H/L704N (+), R776H/V948R (+), L704N & V948R (+), L704N & R776H/V948R (+), R776H/L704N & V948R (+), R776H/L704N & R776H/V948R (+). + and - indicate the presence and absence of EGF ligand, respectively. WT EGFR, R776H, V948R, R776H/V948R are GFP tagged, whereas L704N and R776H/L704N are FLAG tagged.

the C-terminal tail of the donor in a *trans* manner, although recent coarse-grained MD simulation suggest an alternative hypothesis in which the acceptor kinase is phosphorylated in *cis* [12]. To test whether the C-terminal tail of the acceptor kinase is phosphorylated in the asymmetric dimer *in vivo*, we mutated Y1197 in the donor to a phenylalanine (Y1197F, L704N/Y1197F) (Figure 2.4, lane 3-6). When we co-expressed L704N/Y1197F with the enforced acceptor (V948R), Y1197 phosphorylation is detected in the presence of EGF (Figure 2.4, lane 9, 10). We also performed the same set of experiments in the background of R776H mutant (Figure 2.4, lane 11-20) and noticed that Y1197 phosphorylation is observed even in the absence of EGF (Figure 2.4, lane 19, 20) suggesting that the acceptor kinase gets phosphorylated in the asymmetric dimer and the R776H mutant enhances acceptor kinase phosphorylation.

Phosphorylation of the acceptor kinase in the asymmetric dimer can be explained by two competing hypotheses/models: 1) The acceptor kinase phosphorylates itself (*cis* phosphorylation) within the asymmetric dimer, because only the acceptor kinase is active when EGFR dimerizes [10]. 2) The acceptor kinase is phosphorylated by other EGFR dimers. To test the second hypothesis, we generated a monomeric form of EGFR that contains both the N-lobe dimer deficient mutation (L704N) and C-lobe dimer deficient mutation (V948R) [10]. As expected, the enforced kinase monomer is not active (Figure 2.5a, lane 7, 8; Figure 2.5c). However, when the enforced kinase monomer is co-transfected with WT EGFR, with Y1197 mutated to phenylalanine (Figure 2.5a, lane 3, 4, 2.5b), phosphorylation of Y1197 is detected in the presence of EGF (Figure 2.5a, lane 5, 6; Figure 2.5d). In addition, co-expression of R776H/Y1197F with the enforced kinase monomer reveals that Y1197 phosphorylation is enhanced even in the absence of EGF (Figure 2.5a, 13, 14; Figure 2.5d). Because Y1197

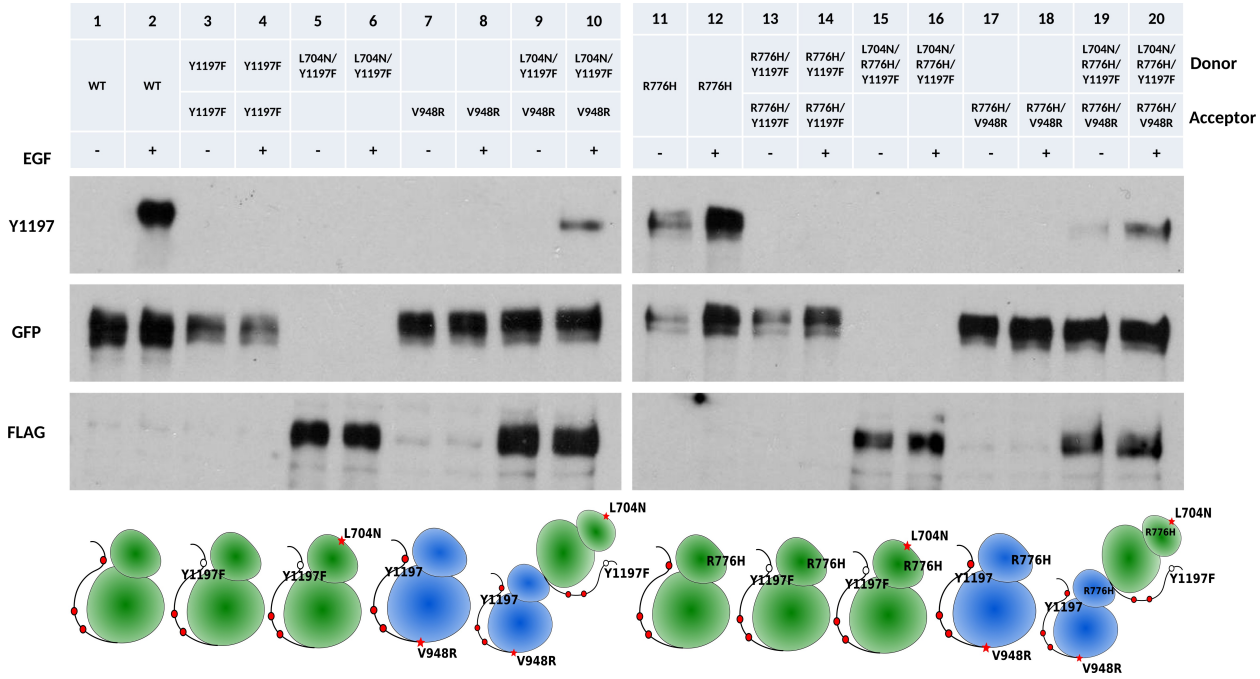

Figure 2.4: Receiver kinase phosphorylation. + and - indicate the presence and absence of EGF ligand, respectively. Cartoon scheme below shows the position of the mutation introduced to EGFR. Lanes from left to right, WT EGFR (-), WT EGFR (+), Y1197 (-), Y1197 (+), L704N/Y1197F (-), L704N/Y1197F (+), V948R (-), V948R (+), L704N/Y1197F & V948R (-), L704N/Y1197F & V948R (+), R776H (-), R776H (+), R776H/Y1197F (-), R776H/Y1197F (+), L704N/R776H/Y1197F (-), L704N/R776H/Y1197F (+), R776H/V948R & L704N/R776H/Y1197F (-), R776H/V948R & L704N/R776H/Y1197F (+). L704N/Y1197F and L704N/R776H/Y1197F are FLAG tagged. All other constructs are GFP tagged.

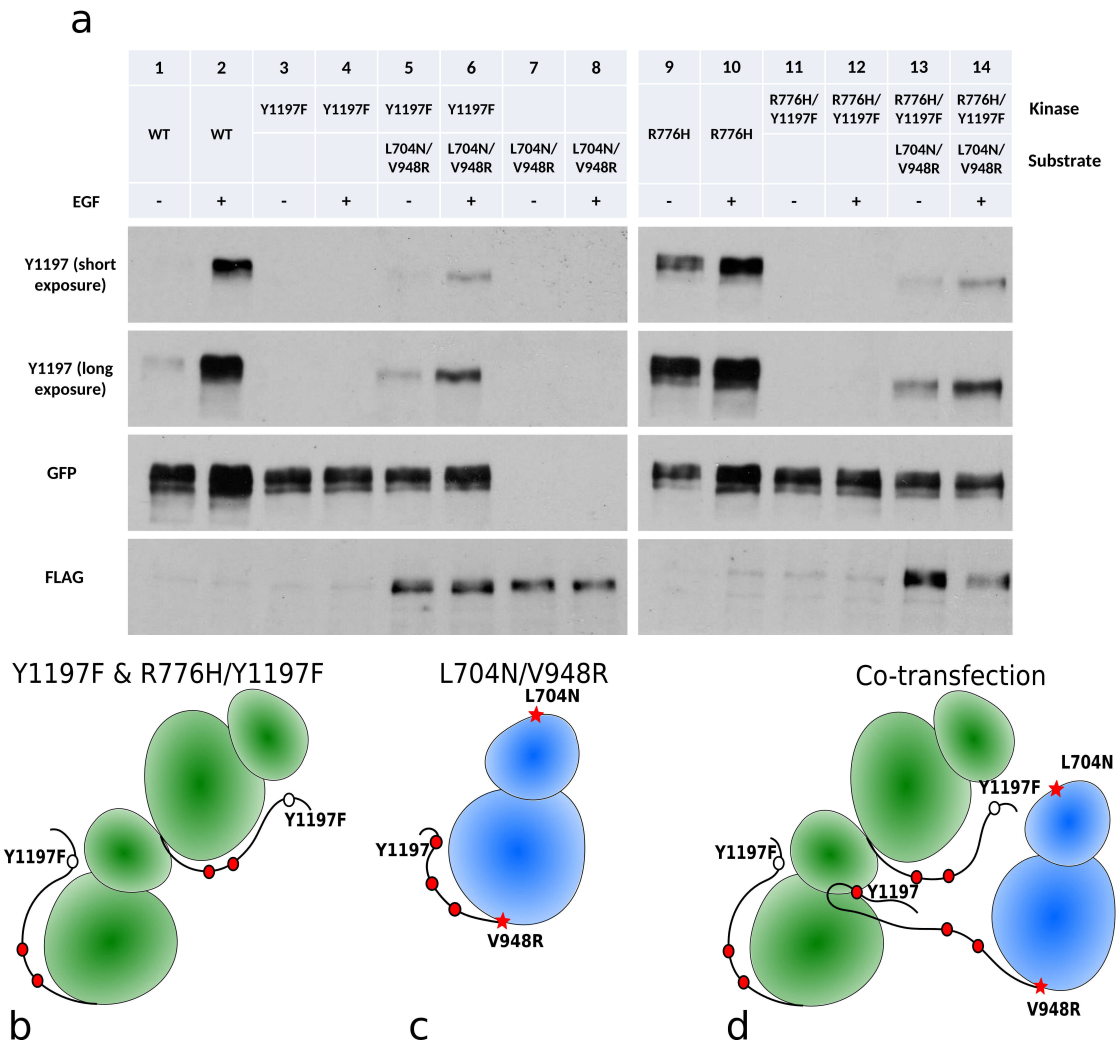

Figure 2.5: Lateral phosphorylation of EGFR. + and - indicate the presence and absence of EGF ligand, respectively. a) Lanes from left to right, WT EGFR (-), WT EGFR (+), Y1197F (-), Y1197F (+), Y1197F & L704N/V948R (-), Y1197F & L704N/V948R (+), L704N/V948R (-), L704N/V948R (+), R776H (-), R776H (+), R776H/Y1197F (-), R776H/Y1197F (+), R776H/Y1197F & L704N/V948R (-), R776H/Y1197F & L704N/V948R (+). L704N/V948R is FLAG tagged. All other constructs are GFP tagged. b) Cartoon scheme for lane 3, 4 and lane 11, 12. c) Cartoon scheme for enforced kinase monomer, lane 7, 8. d) Cartoon scheme for lane 5, 6 and lane 13, 14.

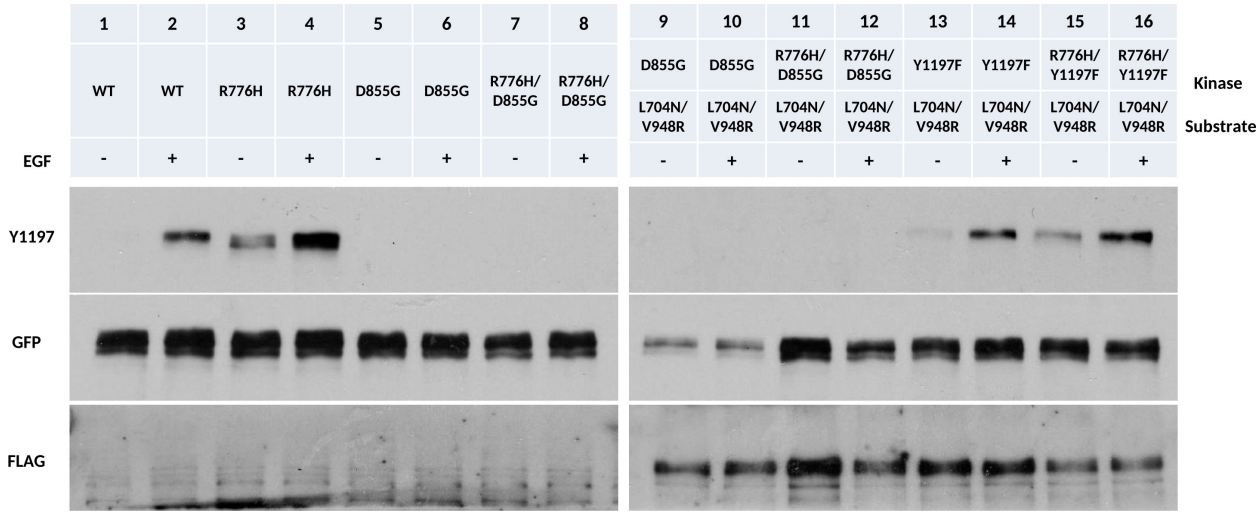

Figure 2.6: Kinase activity of EGFR is responsible for lateral phosphorylation. + and - indicate the presence and absence of EGF ligand. Lanes from left to right, WT EGFR (-), WT (+), R776H (-), R776H (+), D855G (-), D855G (+), R776H/D855G (-), R776H/D855G (+), D855G & L704N/V948R (-), D855G & L704N/V948R (+), R776H/D855G & L704N/V948R (-), R776H/D855G & L704N/V948R (+), Y1197F & R776H/V948R (-), Y1197F & L704N/V948R (+), R776H/Y1197F & R776H/V948R (-), R766H/Y1197F & L704N/V948R (+) L704N/V948R is FLAG tagged. All other constructs are GFP tagged.

is mutated to phenylalanine in the asymmetric dimer, the observed Y1197 phosphorylation is from the enforced kinase monomer (Figure 2.5d). CHO cells express low levels of endogenous ErbB2 [13, 14], which can potentially contribute to Y1197 phosphorylation by forming hetero-dimers with EGFR in our experiment. To rule out this possibility, we made a kinase dead construct, in which the catalytic DFG-Asp is mutated to a glycine (D855G). Co-expression of D855G or R776H/D855G with enforced kinase monomer (L704N/V948R) shows no Y1197 phosphorylation (Figure 2.6, lane 9-12 vs 13-16). Thus the enforced kinase monomer (L704N/V948R) phosphorylation is due to EGFR but not due to endogenous ErbB2. Although our experiments cannot rule out the possibility of *cis* phosphorylation, lateral phosphorylation of the enforced kinase monomer by EGFR dimers has implications for understanding substrate phosphorylation and EGFR signaling in the oligomeric state (See Discussion). In the R776H background, this lateral phosphorylation is enhanced even in the absence of EGF (Figure 2.5a, lane 13).

2.2.4 αC-helix conformational R776H is correlated with a capping interaction between R776 and A767

To investigate the atomic details of how R776H activates EGFR, we performed 1μs atomic molecular dynamics simulation of WT and mutant EGFR in the active state. Previous long time scale molecular dynamics simulation of EGFR showed that the regulatory αC-helix is intrinsically disordered and the salt bridge interaction between the conserved αC-helix glutamate (E762) and the ATP coordinating lysine (K745) breaks within a short period of time (less than 200 ns) during the simulation [15]. Consistent with these studies, we observe a state shift from active to inactive state during 240ns to 480ns in our simulation,

in which the K745-E762 salt bridge is lost and the αC-helix breaks/cracks into two parts (Figure 2.7a, Figure 2.7d). However, the K745-E762 salt bridge and αC-helix are stable in the R776H mutant (Figure 2.7a, 2.7c and 2.7d), indicating that the R776H mutant favors the active form. Furthermore, αC-helix cracking is strongly correlated with a capping interaction between R776 and A767 in WT simulations. Figure 2.7b plots the shortest distance between R776/(NH1,NH2,NE) group and A767/O (Figure 2.7a, 2.7b). At around 240ns to 480ns, R776 hydrogen bonds to the backbone oxygen of A767, which correlates with the loss of K745-E762 salt bridge and αC-helix breaking (Figure 2.7a and 2.7c). The breaking point of αC-helix is another alanine (A763), which is 4 residues N-terminus of A767 and forms the canonical i-i+4 hydrogen bond in the intact αC-helix conformation (Figure 2.7d). However, upon αC-helix breaking the canonical i-i+4 interaction between A763 and A767 is lost, and the unsatisfied backbone hydrogen bonds are partially stabilized by the capping interactions between R776 and A767 (Figure 2.7d). Notably, the hydrogen bond frequency between R776 and A767 is reduced in the R776H mutant (26.9% vs 9.5%), suggesting a loss of inhibitory capping interaction, which correlates with a stable αC-helix conformation (Figure 2.7d). However, it should be noted that the αC-helix capping interaction in itself does not fully explain αC-helix conformational transitions because at around 600ns the K745-E762 salt bridge is formed even though the αC-helix capping interaction is maintained (Figure 2.7a and 2.7b).

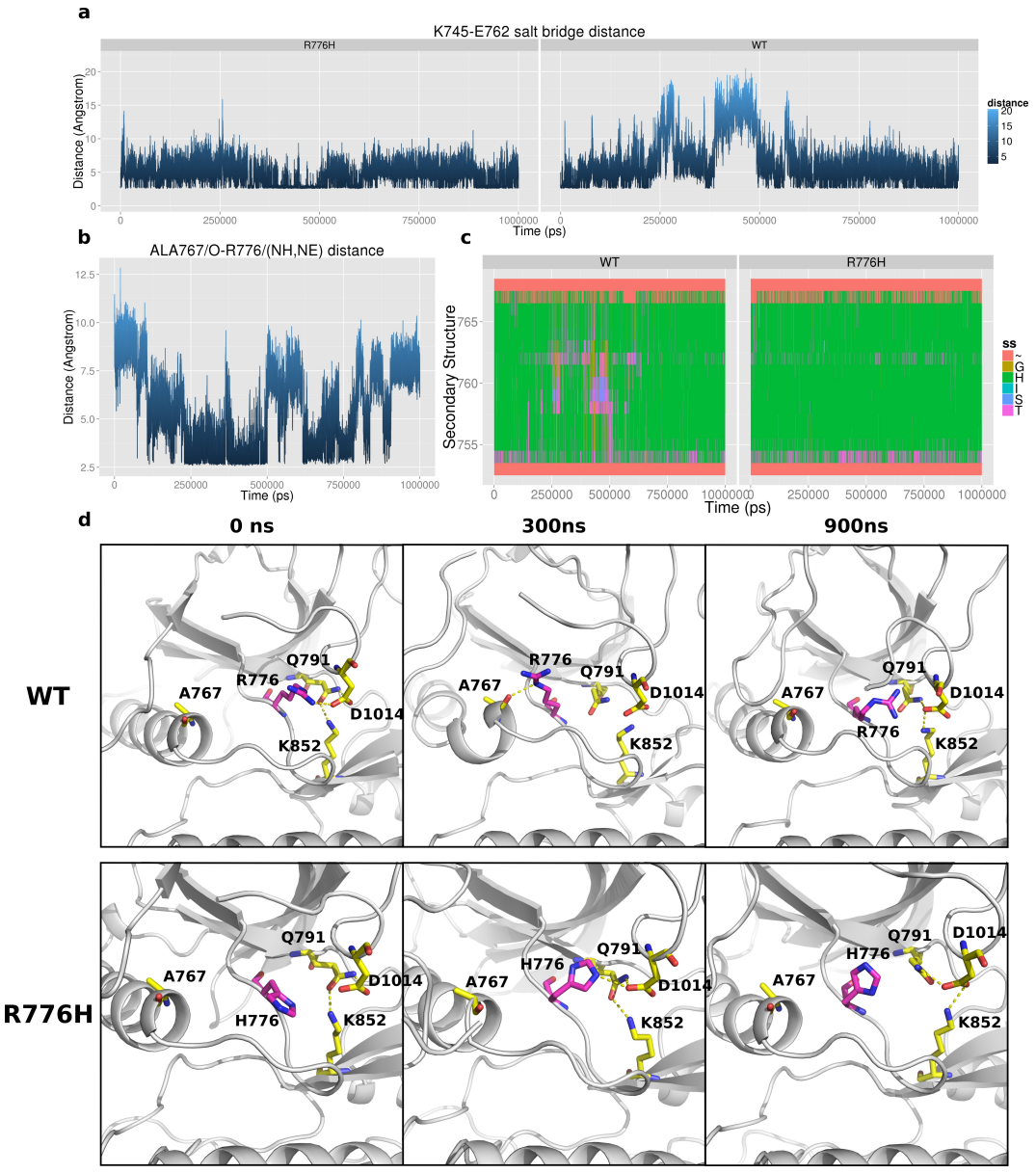

Figure 2.7: Molecular dynamics simulation of EGFR. a) K745-E762 salt bridge distance across 1μs active monomer simulation of WT and R776H. b) Distance plot between the side chain nitrogen of R767 and the backbone oxygen of A767 in WT simulation. c) Secondary structure assignment of αC-helix residues during the 1μs simulation. Coil: $\sim$, Bend: S, Turn: T, A-helix: H, 3-Helix: G. Y-axis spans residues 752-767 which correspond to αC-helix in EGFR. d) Representative snapshot of WT and R776H simulation at 0ns, 300ns and 900ns.

2.2.5 *cis* regulatory interactions between the kinase domain and the flanking JM and C-terminal tail contribute to kinase conformational transitions and R776H mediated activation

We previously demonstrated that the C-terminal tail and juxtamembrane (JM) segment are distinguishing features of the EGFR family that have co-evolved with the kinase core to uniquely regulate catalytic activity [16]. R776 associates with these conserved flanking segments in the crystal structures and MD trajectories. In particular, a segment of the C-tail (1011-1018) tethered to the kinase hinge forms extensive interaction with R776 during our simulation (Figure 2.8a, 2.8b). However, these interactions are not observed in the R776H mutant (Figure 2.8a, 2.8b). Per-residue interaction energy profiles reveal that C-tail residues: D1012, A1013, D1014, I1018 and P1019 form favorable interactions with R776 but not with the mutant (R776H). D1014 is one of the conserved C-tail residues that hydrogen bonds to the inter-lobe salt bridge (Q791 and K852) associated with inter-lobe movement [16]. The association of R776 with the C-tail and JM segment suggests that these *cis* regulatory interactions may also be altered in the R776H mutant in addition to the αC-helix capping interaction described above.

2.3 Materials and Methods

2.3.1 Antibodies and reagents

Anti-GFP, anti-p-Y1197, HRP conjugated mouse monoclonal and rabbit polyclonal antibodies were purchased from Cell signaling (Danvers, MA). Anti-FLAG and human recom-

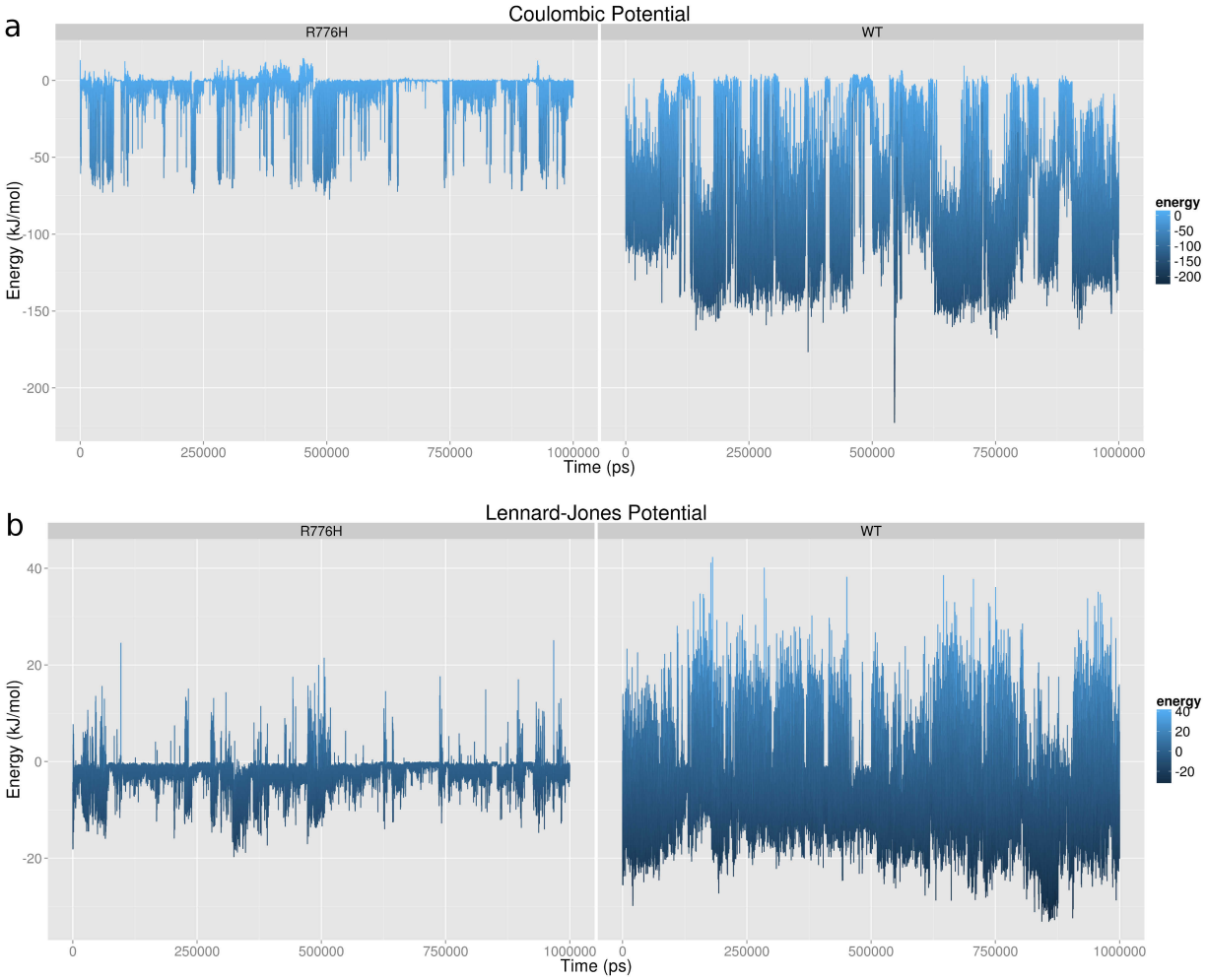

Figure 2.8: R776H loses the interaction with C-terminal tail of EGFR. a) Coulombic interaction energy profile between R776/H776 and the C-terminal tail tether. b) Van Der Waals interaction energy profile between R776/H776 and the C-terminal tail tether.

binant EGF was purchased from Sigma (St. Louis, MO). Lipofectamine was obtained from Invitrogen (Carlsbad, CA). Protease inhibitor cocktail were purchased from Calbiochem. Quick-Change site-directed mutagenesis kit was bought from Stratagene (San Diego, CA).

2.3.2 DNA constructs

pEGFP-N1-EGFR plasmid was a kind gift from Dr. Graham Carpenter (Vanderbilt University, Nashville, TN). Mutagenesis was performed using the QuikChange II kit and confirmed via DNA sequencing.

2.3.3 Cell culture and Transfections

CHO cells were grown in high-glucose Dulbecco's Modified Eagle Medium (DMEM) (Cellgro, Manassas, VA, USA) with 10% fetal bovine serum (Bioexpress, UT, USA) without antibiotics. Transfection in CHO cells was performed using lipfectamine-2000 according to manufacturer's protocol with GFP-pEGFP-N1-WT-EGFR (WT), GFP-pEGFP-N1-R776H-EGFR (R776H), FLAG-pEGFP-N1-L704N-EGFR (L704N), FLAG-pEGFP-N1-R776H/L704N-EGFR (R776H/L704N), GFP-pEGFP-N1-V948R-EGFR (V948R), GFP-pEGFP-N1-R776H/V948R-EGFR (R776H/V948R), GFP-pEGFP-N1-Y1197F-EGFR (Y1197F), GFP-pEGFP-N1-R776H/Y1197F-EGFR (R776H/Y1197F), FLAG-pEGFP-N1-L704N/R776H/Y1197F-EGFR (L704N/R776H/Y1197F), FLAG-pEGFP-N1-L704N/V948R-EGFR (L704N/V948R) DNA constructs. Transiently transfected cell population was pooled and protein expression was analyzed under fluorescent microscope for GFP tagged WT/mutant EGFR and on Western blot using target-specific antibodies.

2.3.4 EGF stimulation, cell lysis and immunoblotting

CHO cells transiently transfected with WT and mutant EGFR plasmids were cultured in DMEM containing 10% FBS on 60 mm plate. To detect autophosphorylation of WT and mutant EGFR, 30% confluent cells were serum-starved in Ham's F-12 media for 18 h . EGF stimulation was carried out using 50 ng/ml human EGF for 5 min . Cells were washed and immediately lysed with lysis buffer (50 mM Tris-HCL, pH 7.4, 150 mM NaCl, 10% glycerol, 1 mM EDTA, 10% Triton X-100, 1 mM PMSF, and 1X Protease Inhibitor Cocktail Set V, EDTA-free). Total cell lysate was spun at 15,000 rpm at 4 °C for 5 min. Samples for SDS-PAGE gel were prepared in 2x Laemelli buffer (25μg total protein). Proteins were resolved on 10% SDS-PAGE and transferred onto polyvinylidenedifluoride (PVDF) membrane using Trans-Blot SD semi-dry transfer cell (Bio-Rad). Western blotting was done using anti-GFP, anti-pY1197 and anti-FLAG antibodies. Proteins were detected by using chemiluminescent substrate (Western Blotting ECL substrate, Pierce, Rockford, IL).

2.3.5 Modeling and molecular dynamics simulations

The PDB structures of EGFR in active conformation (2GS6) were used to model the active state[10]. The two disordered regions (β3-αC loop and part of C-terminal tail) were modeled using modeller[17]. R776H mutation was then introduced in the modeled WT structure using the loop refine module. The backbone was nearly identical for both mutant and wild-type structures and no steric clashes were observed in the final structures.

Molecular dynamics (MD) simulations were done using GROMACS version 4.6.1[18]. All-atom modified AMBER ff99SB force field[19] was used with TIP3P water in a box that was at least 1 nm bigger than the protein on all sides. Steepest descent and conjugate-gradient

energy minimization was performed on the solvated protein for 10000 steps until the Fmax was less than 50 kcal/mol. NVT simulations were carried out by heating from 0 to 298.15 K by coupling it to a berendson thermostat for 200 ps. The restraints on the protein backbone atoms over multiple stages of equilibration under NPT ensemble (P = 1 atm, T = 298.15 K) were released to obtain a relaxed protein, and Parinello-Rahman barostat was used to maintain pressure and density. The unrestrained MD productions were run for 1 µs using a time step of 2 fs at the NPT ensemble. Root mean square deviation is checked to be stable before further analysis. Analysis of MD simulations was carried out using programs in the GROMACS suite. All protein visualization was done using PyMOL[20].

2.4 Discussion

The allosteric activation of EGFR kinase domain involves conformational transitions from an inactive αC-helix "out" conformation to an active αC-helix "in" conformation [21, 22]. Our studies are consistent with a model in which the R776H mutant enhances kinase activity by altering αC-helix conformational transition. In particular, the R776H mutantion increases affinity for dimerization by stabilizing the acceptor αC-helix in an "in" conformation, thereby priming the N-lobe interface for dimerization [15, 11]. The increased affinity for dimerization is the biochemical basis for "superacceptor" activity, and analogous to the recently described L858R/T790M mutation, R776H also appears to display "superacceptor" activity. Notably, both L858R/T790M and R776H display impaired donor activity compared to WT EGFR [11, 23], indicating that only acceptor functions are selectively enhanced by oncogenic mutations. Furthermore, the co-occurrence of R776H with L858R, L861Q, and G719A suggests that the

double mutants (R776H/L858R, R776H/L861Q, and R776H/G719A) may have a synergistic effect on EGFR activation.

Free energy landscape theory suggests that activating allosteric mutations shift the conformational ensemble of proteins towards an active state by destabilizing inactive conformations [24]. Our MD studies suggest a possible transition state intermediate in which the αC-helix is held in an inactive "broken" conformation by a capping interaction between R776 and A767. Additional support for the αC-helix capping interaction is provided by crystal structures of inactive symmetric dimer in which the R776 mediated capping interaction is stabilized through interactions with the juxtamembrane segment [25]. In fact, comparisons of all available crystal structures of EGFR indicates that the R776 to A767 capping interaction is correlated with αC-helix "out" conformation (p-value 6.077e-10) (Figure 2.9). Thus, disruption of auto-inhibitory αC-helix capping interaction and C-terminal tail interaction appears to be the most likely allosteric mechanism by which R776H mutation activates the kinase domain.

Because the αC-helix "in" conformation is critical for kinase activity, key auto-inhibitory mechanisms have evolved to prevent inadvertent EGFR activation [22]. Our studies indicate the αC-helix capping interaction mediated by R776 as a critical auto-inhibitory interaction that prevents αC-helix from adopting an active conformation in the monomeric form. The αC-helix cap may work in conjunction with other regions such as the β3-αC loop and the activation loop that are also associated with αC-helix movement. Oncogenic mutations appear to activate the kinase domain by overcoming these auto-inhibitory interactions (Figure 2.10). L858 and L861 in the activation loop, for example, pack against the αC-helix in the inactive conformation, and oncogenic mutations at these positions (L858R and L861Q)

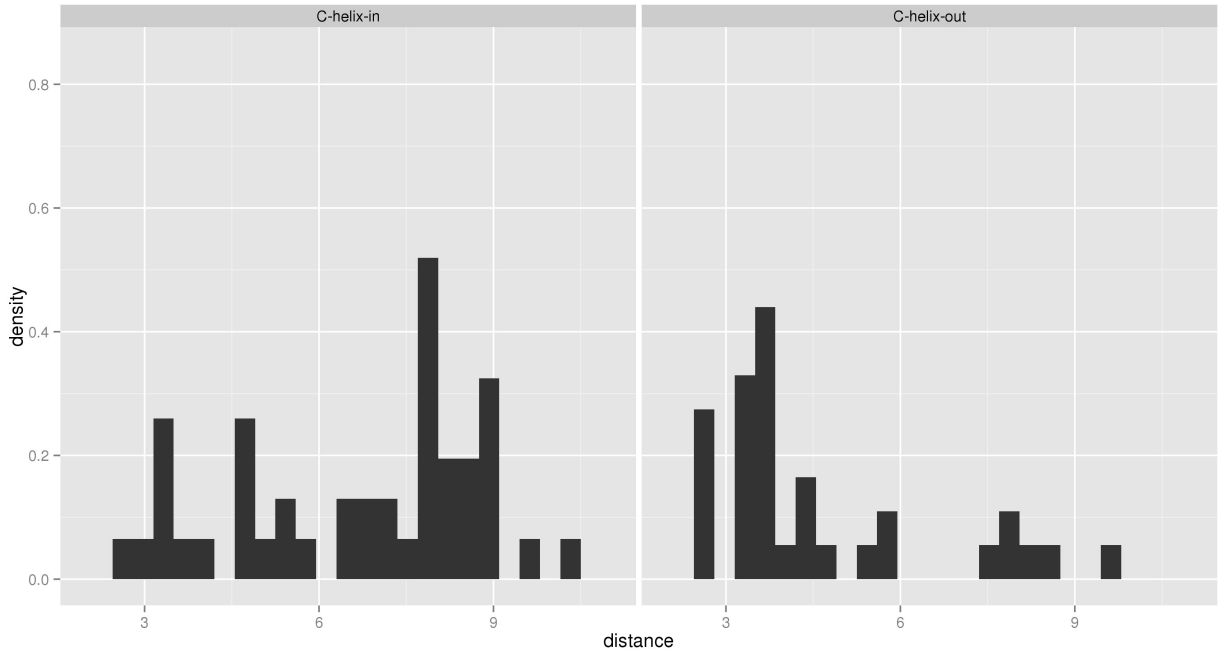

Figure 2.9: Histogram plot showing the shortest distances between R776(NE,NH1,NH2) to A767(O) between αC-helix "in" crystal structures and αC-helix "out" crystal structures.

potentially relieve auto-inhibitory interactions between the activation loop and αC-helix to activate the kinase domain [5, 15]. The β3-αC loop is rich in deletion mutation (account for 15% of EGFR cancer mutations in COSMIC database) and although these deletion mutants have not been structurally characterized, they are likely to favor an αC-helix "in" conformation, by restricting αC-helix conformational flexibility [15, 22]. The αC-β4 loop, on the other hand, is an insertion hotspot and insertions in the loop also favors αC-helix "in" conformation [26, 22]. Indeed, in one crystal structure of αC-β4 loop insertion mutant (D770_N771insNPG, PDBID: 4LRM), the inserted residues form another turn of the αC-helix that prohibit R776 from forming the αC-helix capping interaction [26].

R776 is a mutation hotspot in the kinase domain as the residue equivalent to R776 is mutated in multiple cancers and congenital disorders [9]. A total of 68 missense mutations have been found at the R776 equivalent position in 44 different kinases. Other than R776H, R776C/S/G are also observed in COSMIC database. These mutants are likely to be activating as well, since cysteine, serine and glycine are also predicted to destabilize the autoinhibitory αC-helix capping interaction. Indeed, we previously showed that R776C also activates EGFR in a ligand independent manner [9]. Notably, ErbB2 (R784C/L) and ErbB4 (R782Q) also harbor the arginine mutation at the equivalent position, but not ErbB3. Although the intrinsic disorder of the αC-helix is a unique feature of EGFR and not observed in closely related kinases such as ErbB2 and ErbB4, they may still share some aspect of EGFR conformational transition [15, 21]. Thus R776H equivalent mutations in ErbB2 and ErbB4 are predicted to be activating as well.

EGFR auto-phosphorylation is believed to occur in *trans* through the formation of the asymmetric dimer [10]. Our experimental results show that EGFR kinase activity is not re-

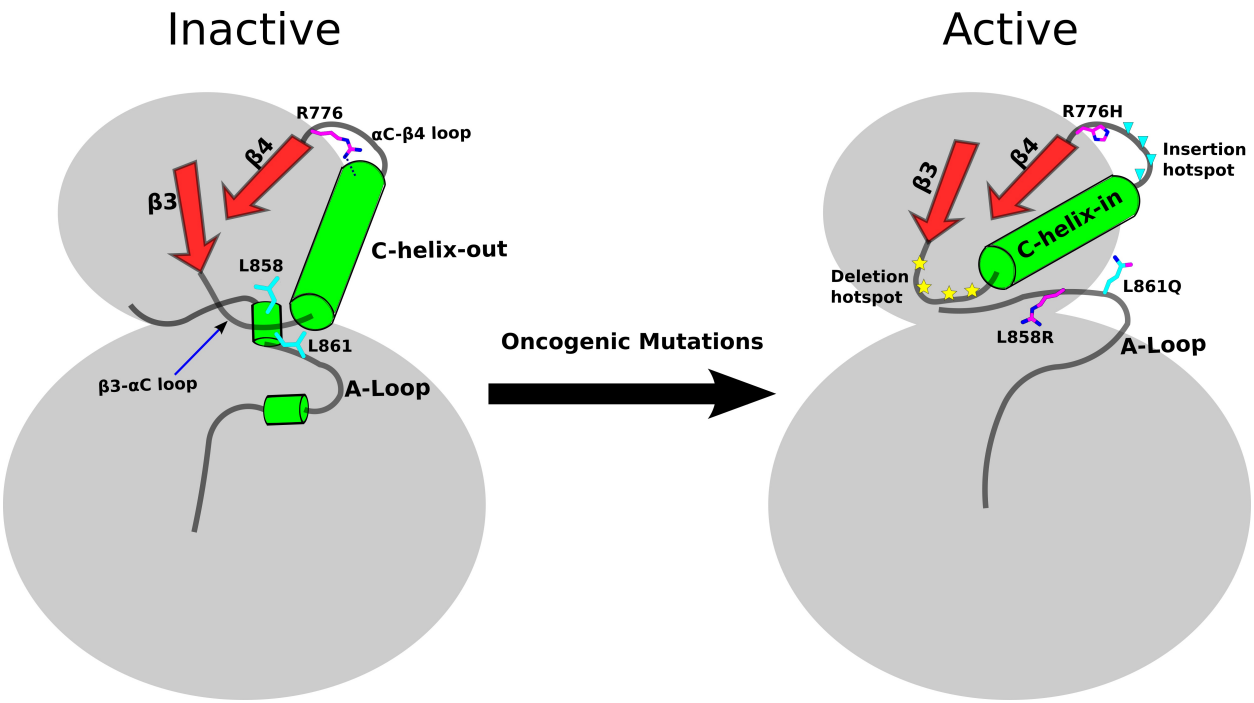

Figure 2.10: Emerging acvitation scheme of oncogenic mutations in EGFR. Right: autoinhibitory regions/interactions are highlighted rendering EGFR inactive. Left: oncogenic mutations overcome autoinhibitory mechanisms and actives EGFR

stricted to the asymmetric dimer itself. Upon ligand stimulation, the activated asymmetric dimer is able to phosphorylate monomeric inactive EGFR. This is an extension of our traditional view of EGFR signal transduction. The result also indirectly supports the observed oligmers reported by several studies [27, 28]. One previous report showed that EGFR lacking the extracellular domain is able to phosphorylate ErbB3 in the absence of ligand stimulation [29], similar to the lateral phosphorylation suggested by our studies. These data support the model that dimeric EGFR functions as a holoenzyme to phosphorylate other monomeric EGFR, and members of the ErbB family. An activating mutation in EGFR, such as R776H, increases such lateral phosphorylation even in the absence of ligand stimulation, leading to constitutive activation and downstream signaling. A logical extension of this model is that other members of the ErbB family (ErbB2, ErbB3, and ErbB4) can potentially be phosphorylated by the R776H dimer via lateral phosphorylation. This hypothesis, however, needs to be tested in future studies. The concept of lateral phosphorylation was first proposed by PJ. Verveer *et al.* in 2000 when lateral propagation of EGFR signals in the plasma membrane was observed using fluorescence imaging [30]. Our studies on the enforced kinase monomer are consistent with previous studies linking EGFR dimerization and lateral phosphorylation.

Bibliography

[1] Y Yarden and M X Sliwkowski. Untangling the erbb signalling network. *Nature reviews. Molecular cell biology*, 2:127–37, 2001.

[2] Robert Roskoski. The erbb/her family of protein-tyrosine kinases and cancer. *Pharmacological research*, 79:34–74, 2014.

[3] Chetan Yewale, Dipesh Baradia, Imran Vhora, Sushilkumar Patil, and Ambikanandan Misra. Epidermal growth factor receptor targeting in cancer: a review of trends and strategies. *Biomaterials*, 34:8690–707, 2013.

[4] Ami Citri and Yosef Yarden. Egf-erbb signalling: towards the systems level. *Nature reviews. Molecular cell biology*, 7:505–16, 2006.

[5] Ruth Nussinov and Chung-Jung Tsai. Allostery in disease and in drug discovery. *Cell*, 153:293–305, 2013.

[6] Jan van Noesel, Ward H van der Ven, Theo A M van Os, Peter W A Kunst, Jitske Weegenaar, Roy J A Reinten, Rama K Kancha, Justus Duyster, and Carel J M van Noesel. Activating germline r776h mutation in the epidermal growth factor receptor associated with lung cancer with squamous differentiation. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 31:e161–4, 2013.

[7] Chih-Hsin Yang, Chong-Jen Yu, Jin-Yuan Shih, Yeun-Chung Chang, Fu-Chang Hu, Meng-Chin Tsai, Kuan-Yu Chen, Zhong-Zhe Lin, Ching-Ju Huang, Chia-Tung Shun, Chin-Lun Huang, James Bean, Ann-Lii Cheng, William Pao, and Pan-Chyr Yang. Specific egfr mutations predict treatment outcome of stage iiib/iv patients with chemotherapy-naive non-small-cell lung cancer receiving first-line gefitinib monotherapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 26:2745–53, 2008.

[8] Jenn-Yu Wu, Shang-Gin Wu, Chih-Hsin Yang, Chien-Hung Gow, Yih-Leong Chang, Chong-Jen Yu, Jin-Yuan Shih, and Pan-Chyr Yang. Lung cancer with epidermal growth

factor receptor exon 20 mutations is associated with poor gefitinib treatment response. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 14:4877–82, 2008.

[9] Daniel Ian McSkimming, Shima Dastgheib, Eric Talevich, Anish Narayanan, Samiksha Katiyar, Susan S Taylor, Krys Kochut, and Natarajan Kannan. Prokino: a unified resource for mining the cancer kinome. *Human mutation*, 36:175–86, 2015.

[10] Xuewu Zhang, Jodi Gureasko, Kui Shen, Philip A Cole, and John Kuriyan. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell*, 125:1137–49, 2006.

[11] Monica Red Brewer, Cai-Hong Yun, Darson Lai, Mark A Lemmon, Michael J Eck, and William Pao. Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 110:E3595–604, 2013.

[12] John G Koland. Coarse-grained molecular simulation of epidermal growth factor receptor protein tyrosine kinase multi-site self-phosphorylation. *PLoS computational biology*, 10:e1003435, 2014.

[13] E Tzahar, H Waterman, X Chen, G Levkowitz, D Karunagaran, S Lavi, B J Ratzkin, and Y Yarden. A hierarchical network of interreceptor interactions determines signal transduction by neu differentiation factor/neuregulin and epidermal growth factor. *Molecular and cellular biology*, 16:5276–87, 1996.

[14] Abasalt Hosseinzadeh Colagar, Omolbanin Amjadi, Reza Valadan, and Alireza Rafiei.

A flexible algorithm for calculating pair interactions on simd architectures. *Res. Mol. Med.*, 1:6–12, 2013.

[15] Yibing Shan, Michael P Eastwood, Xuewu Zhang, Eric T Kim, Anton Arkhipov, Ron O Dror, John Jumper, John Kuriyan, and David E Shaw. Oncogenic mutations counteract intrinsic disorder in the egfr kinase and promote receptor dimerization. *Cell*, 149:860–70, 2012.

[16] Amar Mirza, Morad Mustafa, Eric Talevich, and Natarajan Kannan. Co-conserved features associated with cis regulation of erbb tyrosine kinases. *PloS one*, 5:e14310, 2010.

[17] A Fiser, R K Do, and A Sali. Modeling of loops in protein structures. *Protein science : a publication of the Protein Society*, 9:1753–73, 2000.

[18] Mark James Abraham, Teemu Murtola, Roland Schulz, SzilÃąrd PÃąll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19 – 25, 2015.

[19] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins*, 78:1950–8, 2010.

[20] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

[21] Yibing Shan, Anton Arkhipov, Eric T Kim, Albert C Pan, and David E Shaw. Transitions to catalytically inactive conformations in egfr kinase. *Proceedings of the National Academy of Sciences of the United States of America*, 110:7270–5, 2013.

[22] Chung-Jung Tsai and Ruth Nussinov. The molecular basis of targeting protein kinases in cancer therapeutics. *Seminars in cancer biology*, 23:235–42, 2013.

[23] Peter Littlefield and Natalia Jura. Egfr lung cancer mutants get specialized. *Proceedings of the National Academy of Sciences of the United States of America*, 110:15169–70, 2013.

[24] Chung-Jung Tsai and Ruth Nussinov. The free energy landscape in translational science: how can somatic mutations result in constitutive oncogenic activation? *Physical chemistry chemical physics : PCCP*, 16:6332–41, 2014.

[25] Natalia Jura, Nicholas F Endres, Kate Engel, Sebastian Deindl, Rahul Das, Meindert H Lamers, David E Wemmer, Xuewu Zhang, and John Kuriyan. Mechanism for activation of the egf receptor catalytic domain by the juxtamembrane segment. *Cell*, 137:1293–307, 2009.

[26] Hiroyuki Yasuda, Eunyoung Park, Cai-Hong Yun, Natasha J Sng, Antonio R Lucena-Araujo, Wee-Lee Yeo, Mark S Huberman, David W Cohen, Sohei Nakayama, Kota Ishioka, Norihiro Yamaguchi, Megan Hanna, Geoffrey R Oxnard, Christopher S Lathan, Teresa Moran, Lecia V Sequist, Jamie E Chaft, Gregory J Riely, Maria E Arcila, Ross A Soo, Matthew Meyerson, Michael J Eck, Susumu S Kobayashi, and Daniel B Costa. Structural, biochemical, and clinical characterization of epidermal growth factor recep-

tor (egfr) exon 20 insertion mutations in lung cancer. *Science translational medicine*, 5:216ra177, 2013.

[27] Erik G Hofman, Arjen N Bader, Jarno Voortman, Dave J van den Heuvel, Sara Sigismund, Arie J Verkleij, Hans C Gerritsen, and Paul M P van Bergen en Henegouwen. Ligand-induced egf receptor oligomerization is kinase-dependent and enhances internalization. *The Journal of biological chemistry*, 285:39481–9, 2010.

[28] Andrew H A Clayton, Francesca Walker, Suzanne G Orchard, Christine Henderson, Dominik Fuchs, Julie Rothacker, Edouard C Nice, and Antony W Burgess. Ligand-induced dimer-tetramer transition during the activation of the cell surface epidermal growth factor receptor-a multidimensional microscopy analysis. *The Journal of biological chemistry*, 280:30392–9, 2005.

[29] Rama Krishna Kancha, Nikolas von Bubnoff, and Justus Duyster. Asymmetric kinase dimer formation is crucial for the activation of oncogenic egfrviii but not for erbb3 phosphorylation. *Cell communication and signaling : CCS*, 11:39, 2013.

[30] P J Verveer, F S Wouters, A R Reynolds, and P I Bastiaens. Quantitative imaging of lateral erbb1 receptor signal propagation in the plasma membrane. *Science (New York, N.Y.)*, 290:1567–70, 2000.

Chapter 3

Computational and experimental characterization of patient derived mutations reveal an unusual mode of regualtory spine assembly and drug sensitivity in EGFR kinase

Zheng Ruan, Samiksha Katiyar and Natarajan Kannan (2017) *Biochemistry* 56(1):22-32.
Reprinted here with permission from the publisher.

Abstract

The catalytic activation of protein kinases requires precise positioning of key conserved catalytic and regulatory motifs in the kinase core. The Regulatory Spine (RS) is one such structural motif that is dynamically assembled upon kinase activation. The RS is also a mutational hotspot in cancers; however, the mechanisms by which cancer mutations impact RS assembly and kinase activity are not fully understood. In this study, through mutational analysis of patient derived mutations in the RS of EGFR kinase, we identify an activating mutation, M766T, at the RS3 position. RS3 is located in the regulatory αC-helix, and a series of mutations at the RS3 position suggest a strong correlation between the amino-acid type present at the RS3 position and ligand (EGF) independent EGFR activation. Small polar amino-acids increase ligand independent activity, while large aromatic amino-acids decrease kinase activity. M766T relies on the canonical asymmetric dimer for full activation. Molecular modeling and molecular dynamics simulations of WT and mutant EGFR suggest a model in which M766T activates the kinase domain by disrupting conserved autoinhibitory interactions between M766 and hydrophobic residues in the activation segment. In addition, a water mediated hydrogen bond network between T766, the conserved K745-E762 salt bridge, and the backbone amide of the DFG motif is identified as a key determinant of M766T-mediated activation. M766T is resistant to FDA approved EGFR inhibitors such as gefitinib and erlotinib, and computational estimation of ligand binding free energy identifies key residues associated with drug sensitivity. In sum, our studies suggest an unusual mode of RS assembly and oncogenic EGFR activation, and provide new clues for the design of allosteric protein kinase inhibitors.

3.1 Introduction

Eukaryotic protein kinases (EPKs) are a large family of signaling proteins that propagate cellular signals through the controlled phosphorylation of serine, threonine and tyrosine residues on protein substrates [1, 2]. EPKs share a conserved structural fold consisting of an N-terminal ATP binding lobe (N-lobe) and a C-terminal substrate binding lobe (C-lobe)[2]. The N-lobe of the kinase domain contains five β-strands and the regulatory αC-helix [1, 2, 3], while the C-lobe of the kinase domain is mainly helical and serves as a scaffold for protein docking and substrate recognition[1, 2, 3, 4]. ATP binding and phosphoryl-transfer occur at the catalytic cleft located between the N and C-lobes[2, 3]. The wealth of crystal structure data available on protein kinases has also enabled structural bioinformatics approaches in the study of kinase activation and regulation. In particular, comparisons of the active and inactive conformations of various kinases using spatial pattern matching techniques identified the Regulatory Spine (RS), a set of 4 non-consecutive residues spanning the ATP and substrate binding regions, as a key spatial motif for kinase regulation[5, 6, 7]. The RS is formed by a contiguous network of hydrophobic interactions (RS1, RS2, RS3 and RS4) that are structurally assembled in the active state, but disassembled in the inactive state (Figure 3.1)[5, 7, 8], though some exceptions to this rule have been noted[9, 10]. Large-scale statistical comparisons of protein kinase sequences and crystal structures have also identified additional structural and conformational features associated with RS assembly and kinase activity. For example, a conformational "strain" switch in the catalytic site was shown to be correlated with RS assembly and kinase activity[9]. Likewise, extension of the RS through family and group-specific residues have been suggested to contribute to unique modes of allosteric regulation in some kinases[11, 12]. A molecular dynamics (MD) based structural

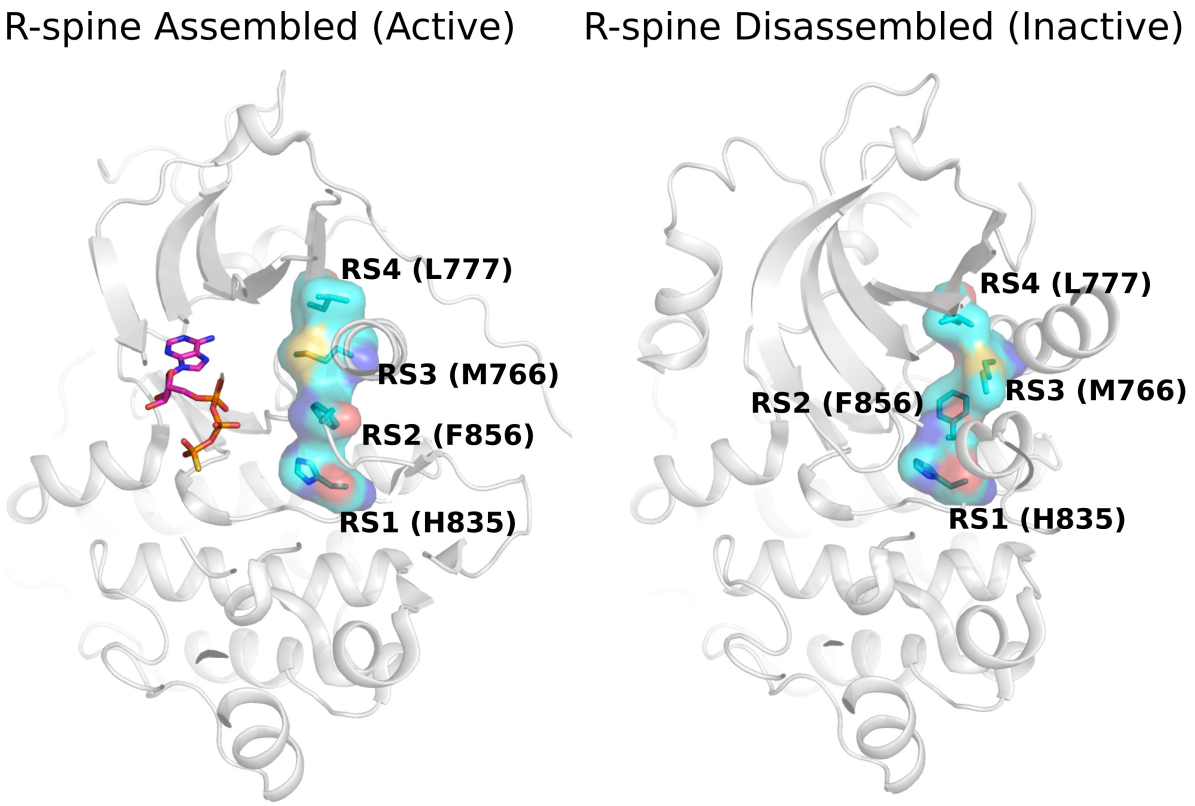

Figure 3.1: R-spine assembly/disassembly in active (PDB: 2GS6) and inactive (PDB: 3W32) state of EGFR.

network modelling approach identified the RS as a central hub for structural stability and allosteric communication[13].

The RS residues are also frequently mutated in human cancers. Mining of cancer mutations in the context of kinase sequence and structural motifs conceptualized in the Protein Kinase Ontology (ProKinO) revealed the RS as a mutational hotspot in cancers [14, 15]. Experimental characterization of some of these mutations in the oncogenic kinase B-Raf revealed that oncogenic variants that introduce additional aromatic interactions in the RS

result in constitutive B-Raf activity [16, 17]. Here, we computationally and experimentally screen patient-derived RS mutation in EGFR and identify M766T, a mutation at the RS3 position, as a novel activating mutation. We find that the nature and size of amino acids present at the RS3 position strongly correlate with ligand independent EGFR activation. Based on molecular modelling and molecular dynamics simulation studies, we propose a model in which M766T activates the kinase domain by both destabilizing the inactive hydrophobic packing between αC-helix and A-loop, and stabilizing the active state through a novel water mediated interaction in the dynamic core. We show that EGFR harboring the M766T mutation is resistant to gefitinib and erlotinib treatment, but not to lapatinib treatment. Since M766 is part of the binding site for some allosteric protein kinase inhibitors [18], our findings are relevant for the design of structure guided protein kinase inhibitors as well.

3.2 Results

3.2.1 Identification and mutational analysis of patient derived EGFR R-spine mutations

To investigate the impact of mutations mapping to the RS of EGFR, we queried the protein kinase ontology, ProKinO, which integrates and conceptualizes the relationships connecting protein kinase sequence, structure, function, evolution and disease in a human and machine readable format[14, 15]. Queries requesting mutations mapping to the RS in EGFR revealed a total of 18 distinct samples containing 7 unique mutations. To understand how these mutations alter kinase activity, we generated all 7 RS mutations in EGFR and monitored

the extent of Y1197 auto-phosphorylation in the absence/presence of EGF as a measure of kinase activation (see methods). Ligand dependent and independent activity of H835L (RS1 mutation) is less than WT (Figure 3.2a, lane 7-8; Figure 3.3, lane 3-4). H835L tends to co-occur with the L833V mutation (9 out of 11 patient samples). Western blot analysis shows that L833V itself is highly active, but the double mutant H835L/L833V is less active than H835L (Figure 3.3). The DFG-Phe (RS2) is mutated to a leucine (F856L) or serine (F856S) in cancer samples. F856L shows reduced Y1197 auto-phosphorylation (Figure 3.2a, lane 4) whereas F856S completely abrogates Y1197 auto-phosphorylation (Figure 3.2a, lane 6). M766T at the RS3 position is the only mutation more active than WT EGFR in the absence of EGF (Figure 3.2a, lane 11), whereas M766V is less active (Figure 3.2a, lane 13). Mutations at the RS4 position, L777Q and L777P, reduced Y1197 phosphorylation in the presence and absence of EGF (Figure 3.2a, lane 16, 18). Thus, our initial screening identified M766T as an activating mutation in the RS of EGFR. Downstream signaling of EGFR is also altered by the M766T mutation, as the phosphorylation status of STAT3 is increased in M766T relative to WT (Figure 3.4).

3.2.2 Activity of EGFR is correlated with side-chain size at the RS3 position

To understand the activation mechanism of M766T in EGFR, we made a series of mutations (M766A/T/S/V/F) at the RS3 position (Figure 3.2b). The phosphorylation levels of WT and mutant EGFR are comparable in the presence of EGF. However, ligand independent activity of EGFR differs for the various RS3 mutants. In general, when the RS3 methionine is replaced by a smaller side-chain residue (M766T/S/A), the mutants display higher ligand

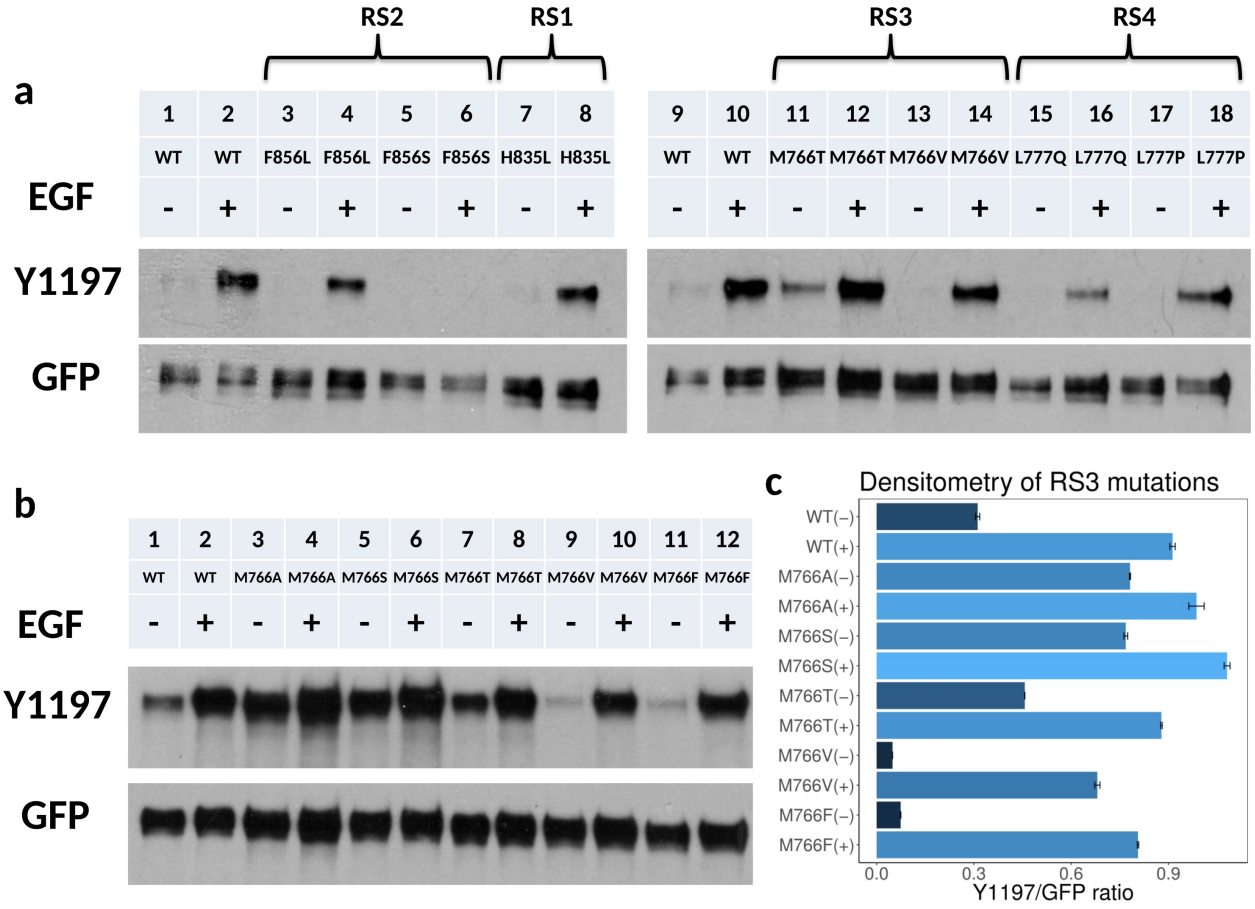

Figure 3.2: Western blot analyses and screening of RS3 mutations in EGFR. (a) Lanes from left to right: WT EGFR (−), WT EGFR (+), F856L (−), F856L (+), F856S (−), F856S (+), H835L (−), H835L (+), WT EGFR (−), WT EGFR (+), M766T (−), M766T (+), M766V (−), M766V (+), L777Q (−), L777Q (+), L777P (−), L777P (+). − and + indicate the absence and presence of EGF stimulation. (b) Series of mutations at RS3 residue. Lanes from left to right: WT EGFR (−), WT EGFR (+), M766A (−), M766A (+), M766S (−), M766S (+), M766T (−), M766T (+), M766V (−), M766V (+), M766F (−), M766F (+). − and + indicate the absence and presence of EGF stimulation. (c) Densitometry of three independent experiments of RS3 residue mutations.

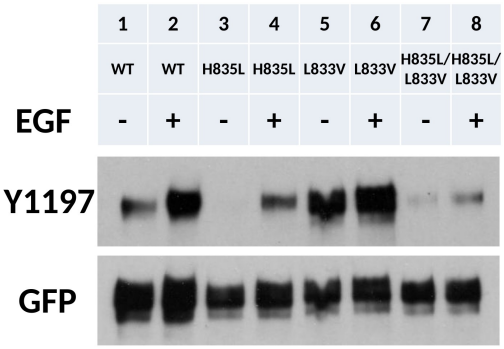

Figure 3.3: Western blot of H835L (RS1) and co-occurring L833V mutation. Lanes from left to right: WT EGFR (−), WT EGFR (+), H835L (−), H835L (+), L833V (−), L833V (+), H835L/L833V (−), H835L/L833V (+). − and + indicate the absence and presence of EGF stimulation.

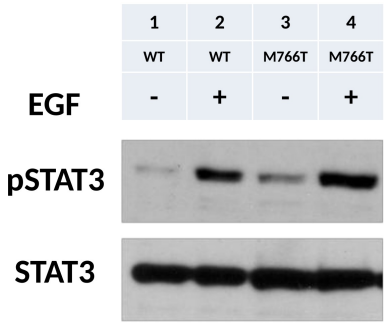

Figure 3.4: Phosphorylation of STAT3 by M766T EGFR. Lanes from left to right: WT EGFR (−), WT EGFR (+), M766T (−), M766T (+). − and + indicate the absence and presence of EGF stimulation.

independent activity compared to WT (Figure 3.2b, lane 3-8), except for M766V, which displays lower activity in comparison to WT (Figure 3.2b lane 9-10). M766T is 46% more active than WT EGFR in the absence of EGF, as indicated by densitometry quantification (Figure 3.2c). However, when RS3 methionine is replaced by a phenylalanine, activity of M766F is significantly reduced in comparison to WT (Figure 3.2b, lane 11-12), suggesting that the nature and size of amino acid side chain at the RS3 position is critical for normal EGFR functions.

3.2.3 M766T mediated EGFR activation is dimerization dependent

It is well established that WT EGFR requires the formation of the asymmetric dimer to be active[19]. To test if M766T mediated activation of EGFR also requires the asymmetric dimer, we performed complementation assays, as described in a previous study[19]. Specifically, we introduced kinase C-lobe dimerization deficient mutation (V948R) and N-lobe dimerization deficient mutation (L760R) [19] in the M766T background and probed for Y1197 phosphorylation. Y1197 phosphorylation cannot be detected once the dimerization deficient mutation is introduced (M766T/V948R and M766T/L760R) (Figure 3.5, lane 5-8). However, Y1197 phosphorylation can be rescued by co-transfection of M766T/V948R (receiver kinase) and M766T/L760R (activator kinase) in the presence of EGF (Figure 3.5, lane 9-10), since the asymmetric dimer can still form when both constructs are present. These data demonstrate that M766T mediated EGFR activation requires the asymmetric dimer.

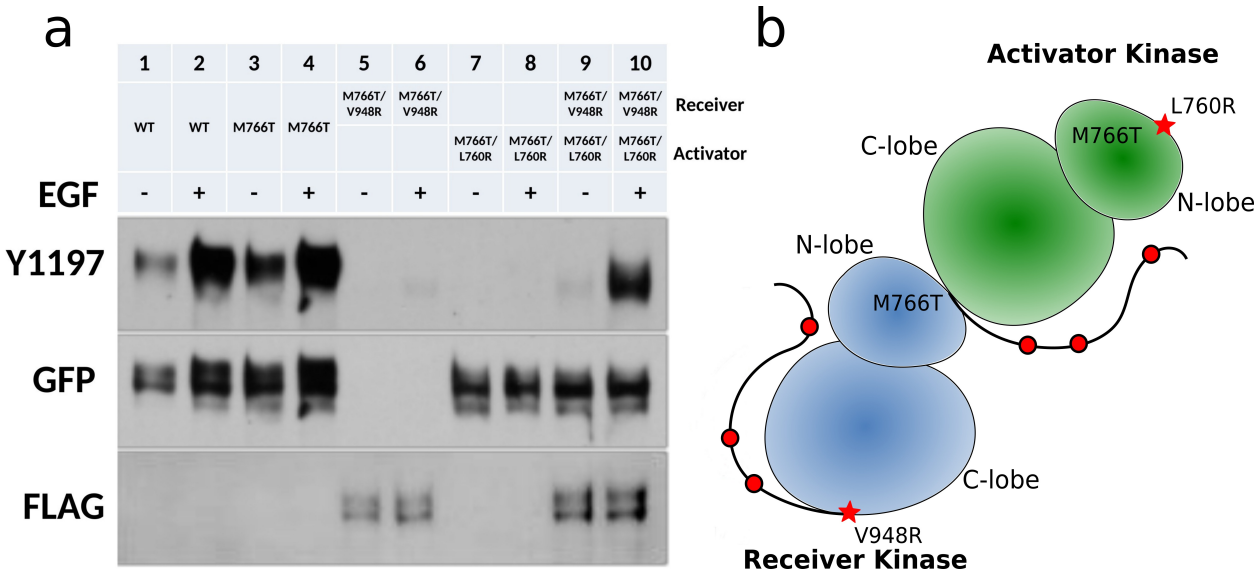

Figure 3.5: Dimerization dependency of M766T mutation. Right panel: Lanes from left to right: WT EGFR (−), WT EGFR (+), M766T (−), M766T (+), M766T/V948R (−), M766T/V948R (+), M766T/L760R (−), M766T/L760R (+), M766T/V948R+M766T/L760R (−), M766T/V948R+M766T/L760R (+). − and + indicate the absence and presence of EGF stimulation. Left panel: a cartoon representation of EGFR asymmetric dimer with N-lobe dimerization deficient mutation (L760R) and C-lobe dimerization deficient mutation (V948R).

3.2.4 Molecular dynamics simulations of WT and mutant (M766T) EGFR

To investigate the activation mechanism of M766T mediated activation, we performed MD simulations of WT and mutant (M766T/S/A/V/F) EGFR in the active, inactive and asymmetric dimer forms with two independent replicates, accumulating a total of 7.5 μs MD data (see method). Although we performed Kullback Leibler (KL) divergence comparison for all RS3 mutants, we primarily focus our discussions on M766T since it is a novel activating mutation.

M766T relieves auto-inhibitory packing interactions between αC-helix and the activation segment Previous studies have shown that the auto-inhibitory hydrophobic packing interactions between the regulatory αC-helix and the activation segment are critical for EGFR mutational activation[19, 20]. M766 is part of the hydrophobic network connecting the αC-helix and activation segment in the inactive state and makes van der Waals interactions with residues F856, L858 and L861 in the activation segment (Figure 3.6a). In our inactive MD simulation, the interaction between the two turn helix in the A-loop and the RS3 residue is weaker in M766T/S/A/V compared to WT EGFR (Figure 3.6b). In addition, the sulfur group of M766 is optimally positioned to form an S-pi interaction with the DFG-Phe (F856) (Figure 3.6c)[21]. Quantification of the contact area between the A-loop and the RS3 residue also suggests that M766T/S/A/V decreases hydrophobic packing between these two regions (Figure 3.6d) and can potentially activate EGFR by destabilizing the inactive conformation. M766F, on the other hand, makes a strong van der Waals interaction with the A-loop, and increases the contact area between A-loop and αC-helix (Figure 3.6b,

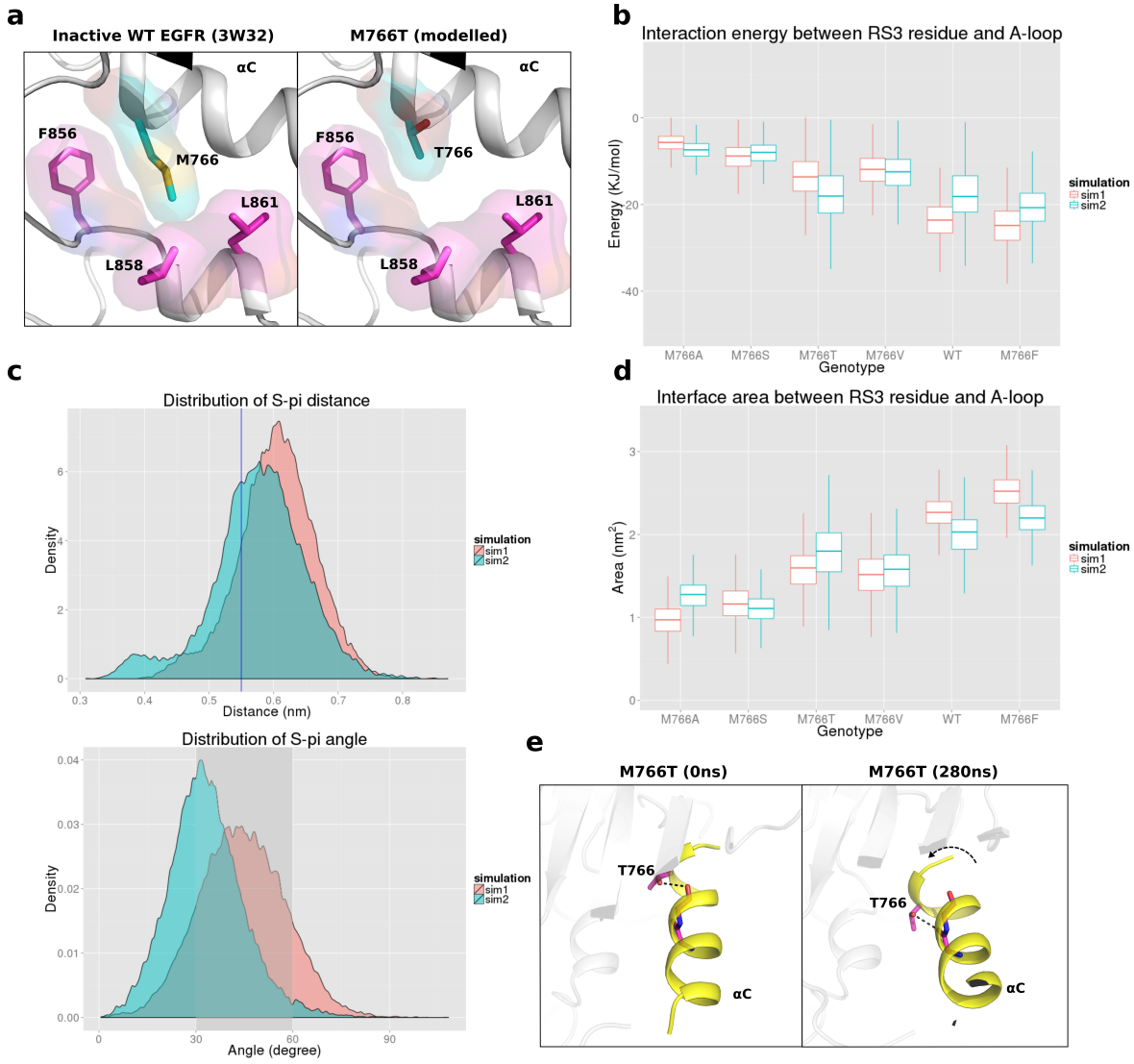

Figure 3.6: (a) Packing interaction between M766 in the αC-helix and hydrophobic residues in the activation loop. (b) Boxplot representing potential (VDW and couloumbic) interaction energy distributionbetween RS3 residue and A-loop helix. Distribution quantile (5%, 25%, 50%, 75% and 95%) is shown in each vertical box plot. (c) Distance distribution of of M766 sulfur atom and the center of F856 benzyl group (upper panel) and angle distribution of M766 sulfur atom and the normal vector of F856 benzyl group (bottom panel). Blue line or shaded area indicate optimal interaction region. (d) Boxplot representing the surface area between RS3 residue and A-loop in different RS3 mutations. (e) MD snapshots showing the intra helix hydrogen bonding interactions between the hydroxyl group of T766 and carbonyl oxygen of E762 or A763, resulting in a bent αC-helix conformation.

3.6d). Notably, M766F shows less activity in comparison to WT, suggesting mutational stabilization of the inactive conformation. M766T also displays the highest RMSF value at the C-terminal region of the αC-helix among all the mutants investigated (Figure 3.8). This is likely due to an intra helical hydrogen bond between the hydroxyl group of T766 and the carbonyl oxygen of E762 or A763 (92.6% occupancy), which introduces a bend in the αC-helix conformation (Figure 3.6e). Although S766, in theory, can also initiate such αC-helix bend, our MD data suggests that S766 is very flexible and does not form stable intra helical hydrogen bonds (27.7% occupancy).

Conserved water mediated interactions may contribute to M766T mediated activation of EGFR To further investigate the structural basis of M766T mediated activation, we performed two independent 300 ns MD simulations of WT and mutant (M766T/A/S/V/F) EGFR in the active monomeric conformation. Active monomeric EGFR is relatively unstable as the average RMSD value of the MD simulation is around 4 angstroms, whereas the RMSD of inactive monomeric EGFR is less than 3 angstroms (Figure 3.10). In addition, a previous MD study noted that the αC-helix is intrinsically disordered in the active monomeric state, and oncogenic mutations promote receptor dimerization by quenching intrinsic disorder [22]. Consistent with this, our active monomeric simulations identified different degrees of αC-helix disorder for the various RS3 mutants (Figure 3.7a). Notably, the αC-helix is relatively stable in both simulation replicates for the M766T mutant (Figure 3.7a). We also quantified the K745-E762 salt bridge distance of M766T and other RS3 mutants (Figure 3.7b). Surprisingly, the largest variance in K745-E762 salt bridge distance is observed for M766V (Figure 3.7b). Although threonine and valine are similar in terms of their side-chain size, their physicochemical properties are different. Unlike valine, threonine is capable of forming

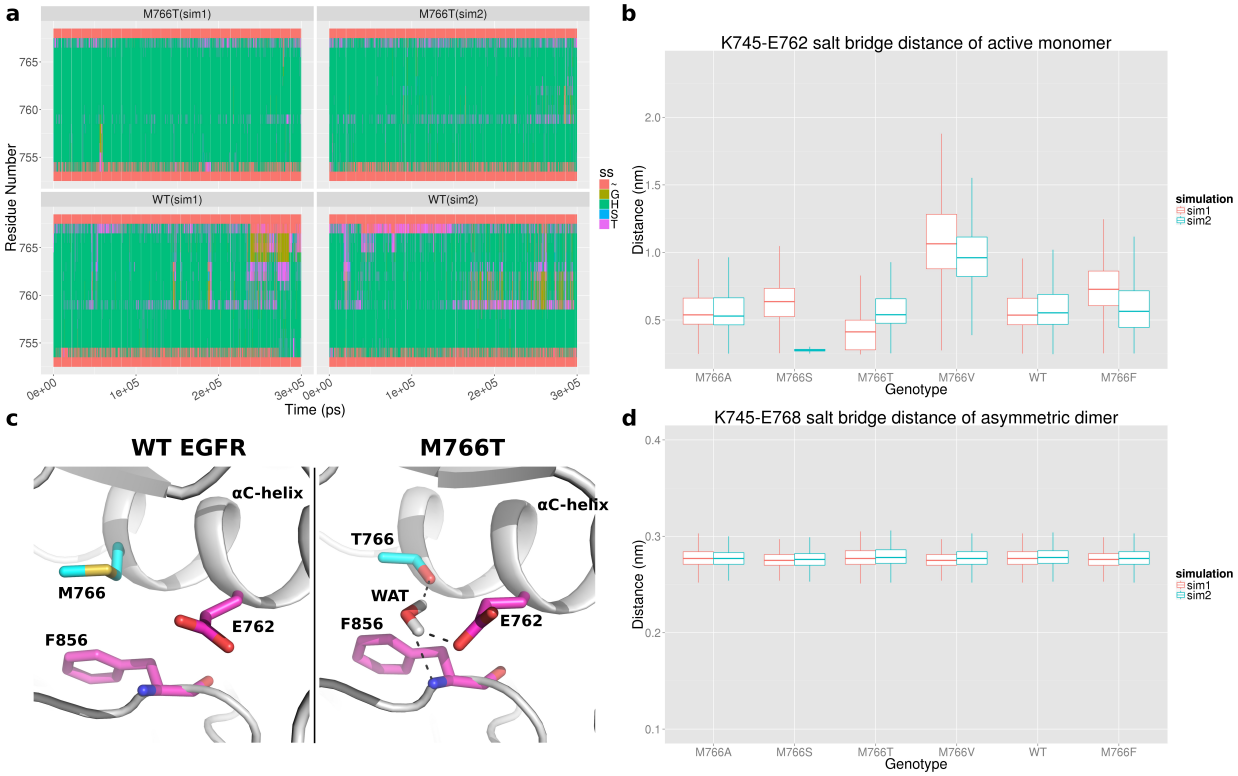

Figure 3.7: (a) Secondary structure assignment of the αC-helix region in WT nd M766T EGFR simulation. Coil: sin, 3-Helix: G, A-helix: H, 5-Helix: I, Bend: S, Turn: T. (b) Boxplot showing the K745-E768 salt bridge distance in the active monomer simulation. (c) Representative MD snapshots of WT EGFR and M766T mutation in the active state. (d) Boxplot showing the K745-E766 salt bridge distance in the active asymmetric dimer simulation.

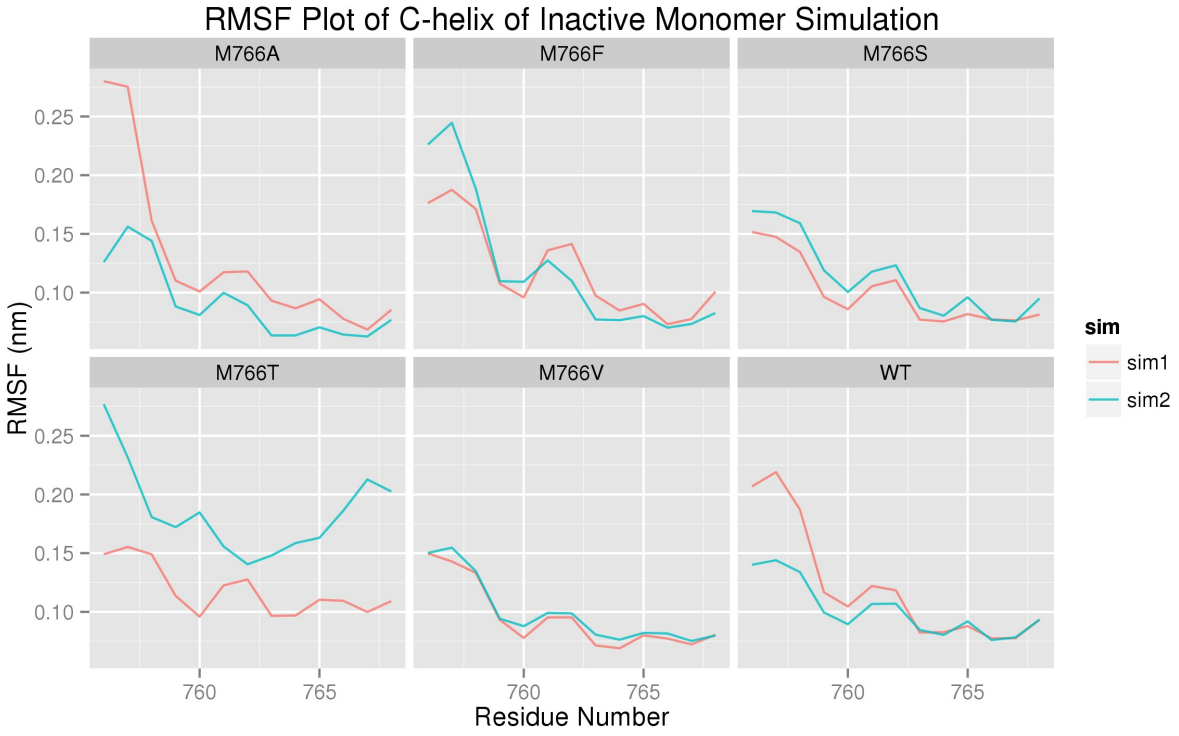

Figure 3.8: Root mean square fluctuation (RMSF) of the αC-helix in the inactive monomer simulation.

hydrogen bonds. Indeed, in the M766T simulation, the hydroxyl group of T766 hydrogen bonds to the E762 side-chain and the DFG-Phe backbone through a water molecule, which is absent in other RS3 mutant simulations (Figure 3.7c). Such an interaction might help maintain the catalytically important K745-E762 salt bridge and rigidify αC-helix in M766T, but not in M766V. Furthermore, comparison of the torsion angle dynamics between M766T and WT EGFR identified other residues (L833, R836, and H893) that are allosterically perturbed by M766T.

Our experimental studies show that M766T mediated EGFR activation requires the asymmetric dimer (Figure 3.5). To determine how RS3 mutations are coupled to dimer assembly, we performed 150 ns MD simulations for each mutant in the active asymmetric dimer and compared the RMSF values of key residues in the receiver and activator. As observed in Figure 3.9, the RMSF values for the αC-helix residues in the receiver and activator kinase are comparable for the activating mutations (M766T/A/S). In contrast, for the inactivating mutants (M766F and M766V), αC-helix residue fluctuations in the activator are significantly higher than the receiver. A similar trend is observed when the active monomer is compared against the receiver in the asymmetric dimer (Figure 3.11). Analysis of the K745-E762 salt bridge distance indicates that the salt bridge is well maintained in the activating M766T mutant (Figure 3.7d) and the conserved water molecule that mediates hydrogen bonding network with T766 and E762 is present in 49.2% of the MD trajectories, suggesting that the identified water-mediated network is functional and relevant for M766T mediated activation. Together, these data support the hypothesis that M766T reduces intrinsic flexibility of the αC-helix to promote dimerization and kinase activation.

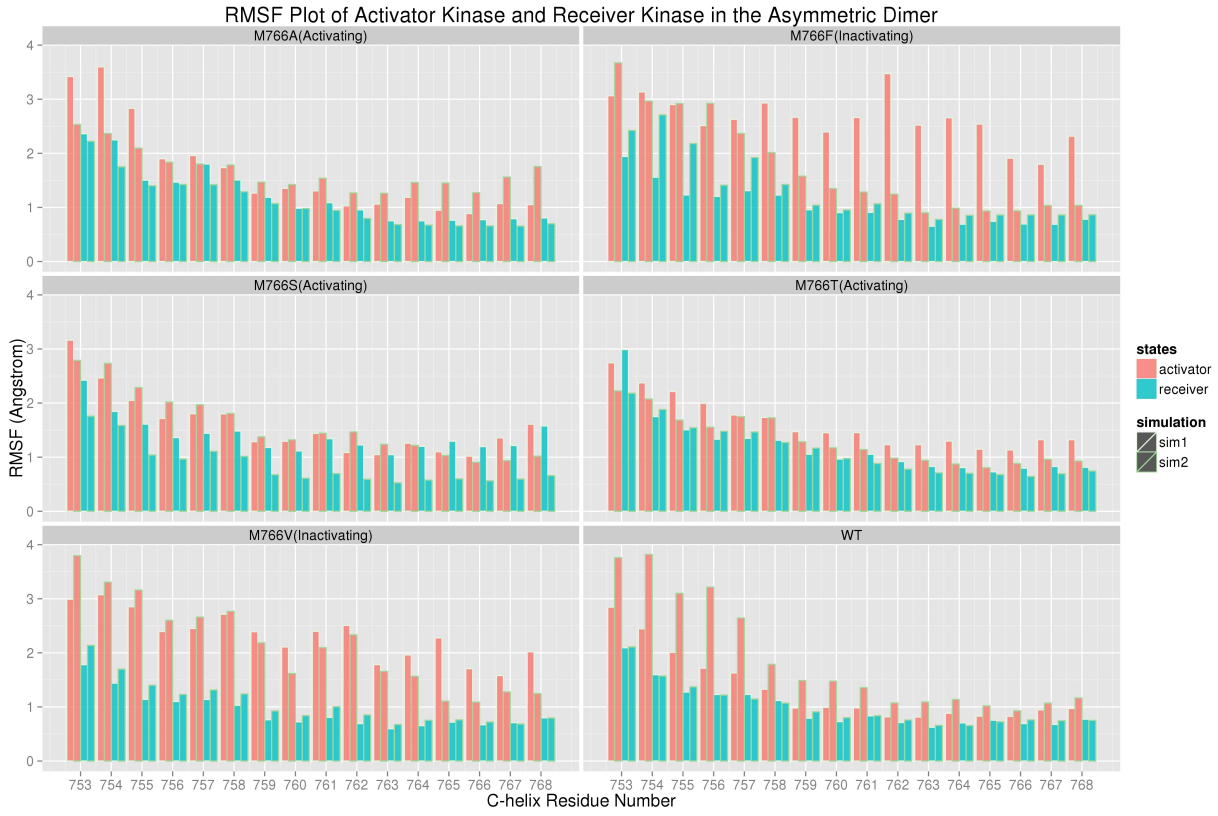

Figure 3.9: Root mean square fluctuation (RMSF) plot of αC-helix residues of activator kinase (red) and receiver kinase (receiver). For each RS3 mutants, two independent replicates are shown.

Figure 3.10: Root mean square deviation (RMSD) plot of RS3 mutants in active (top panel) and inactive (bottom panel) states.

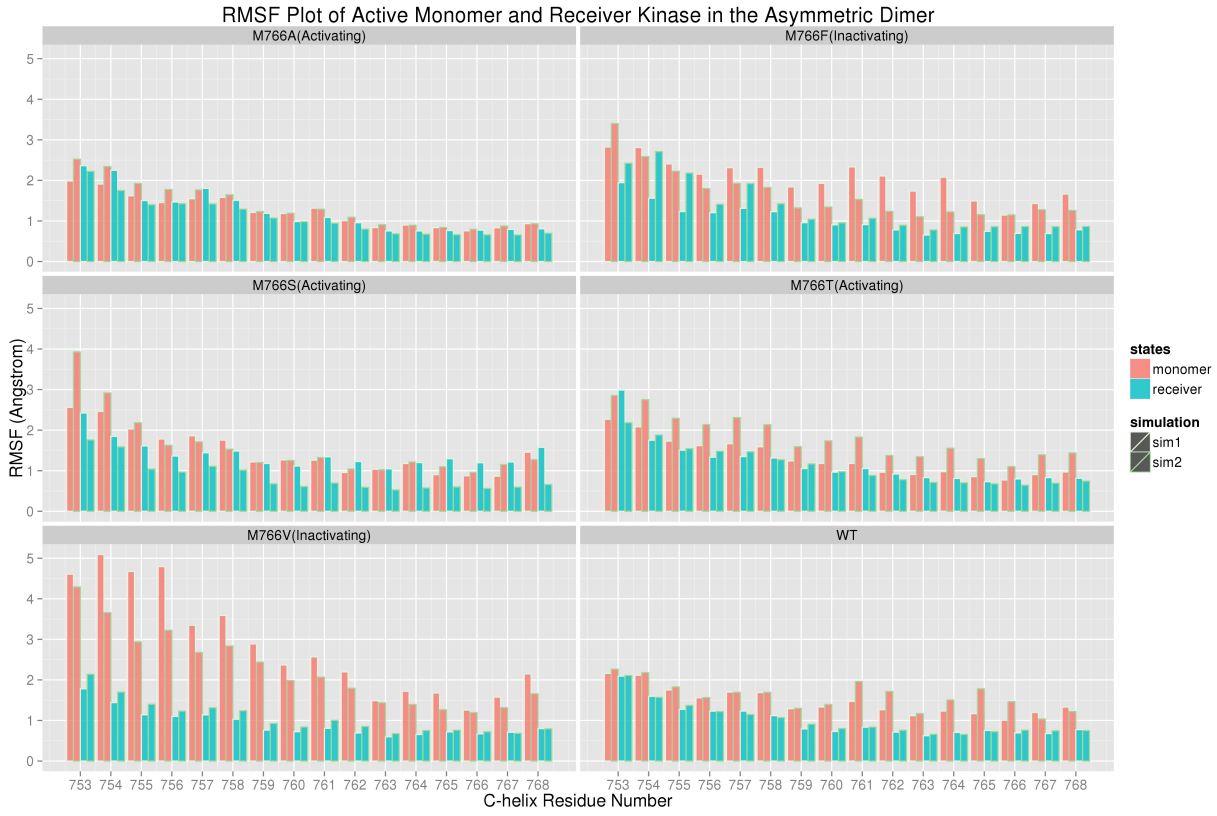

Figure 3.11: RMSF plot of C-helix residues in the active kinase monomer (red) and receiver kinase (cyan) in the asymmetric dimer

3.2.5 M766T is resistant to gefitinib/erlotinib treatment, but not lapatinib treatment

Because M766 is part of the inhibitor binding site in the lapatinib/gefitinib/erlotinib bound structure of EGFR [23, 24], we next wanted to explore the efficacy of clinically approved EGFR inhibitors on WT and mutant EGFR activity. To this end, we performed drug binding assays using inhibitors that target the inactive (lapatinib)[23], active (gefitinib) [25], and active/inactive conformations (erlotinib) [24, 26] of the kinase domain (Figure 3.12b). CHO cells were treated with increasing doses of EGFR inhibitors, and Y1197 phosohorylation was probed as readout for kinase inhibition. As observed in Figure 3.12a, M766T is slightly more sensitive to lapatinib in comparison to WT EGFR (Figure 3.12a, upper panel). The half inhibition concentration (IC50) of gefitinib (Figure 3.12a, middle panel) and erlotinib (Figure 3.12a, bottom panel) for WT EGFR is in the 0.01-0.1 μM range, whereas the IC50 for M766T is in the range of 1.0-10.0 μM, suggesting that M766T is more resistant to gefitinib and erlotinib compared to WT. These findings are consistent with previous IC50 estimates for various EGFR drug resistance mutations[27].

To understand the mechanisms by which M766T contributes to drug resistance, we performed modelling and 300 ns MD simulations of WT and mutant EGFR in the presence of EGFR inhibitors (lapatinib/gefitinib and erlotinib). Binding free energy estimates indicate a more favorable binding of lapatinib to M766T versus WT EGFR (Table 3.1). However, M766T binds less stably to gefitinib/erlotinib compared to WT EGFR, as indicated by positive $\Delta\Delta G$ values (Table 3.1). To pinpoint residues that confer the binding free energy difference, we performed energy decomposition analysis on both WT EGFR and M766T mutant and mapped the energy difference of each residue in available crystal structures (Figure

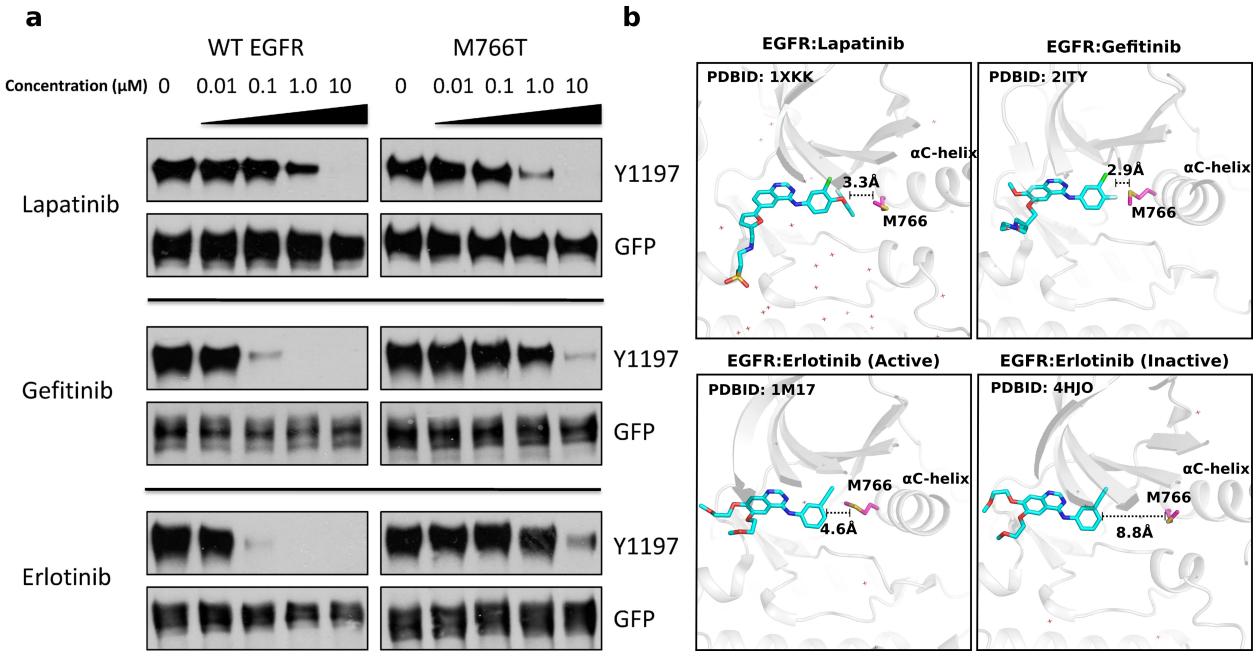

Figure 3.12: Drug sensitivity of WT and mutant (M766T) EGFR. Drug concentration from left to right: 0 μM, 0.01 μM, 0.1 μM, 1.0 μM, and 10.0 μM. Upper panel (lapatinib) [23], middle panel (gefitinib) [25], lower panel (erlotinib) [24, 26].

3.13). The analysis picked up residue clusters in the ligand binding pocket (Figure 3.13). Residues are colored based on energy contribution difference ($\Delta G_{M766T} - \Delta G_{WT} < 0$: red, $\Delta G_{M766T} - \Delta G_{WT} > 0$: blue) and a stick representation is shown if the energy difference is above a threshold. Residues that contribute most to the binding free energy difference in the lapatinib bound MD simulation includes M766 (RS3), K745, V762, T790, D800, D855, T854, and L1001 (Figure 3.13). Although M766, T854, T790, V726, and L1001 positively affect the binding free energy, this is compensated by several other charged residues (K745, D855 and D800) in the ligand binding pocket. The overall binding free energy for M766T:lapatinib is more negative than WT:lapatinib (Table 3.1). The gefitinib bound simulation identified the K745-E762 salt bridge residues as contributing most to the binding free energy (Figure 3.13). This difference might relate to the water mediated interaction between M766T and E762, as the hydrophobic local environment created by M766 is replaced with a water molecule, destabilizing the protein ligand interaction. Although MD simulation of erlotinib in active and inactive states both suggest a destabilizing effect of M766T (Table 3.1), the residues contributing to binding free energy differences are different (Figure 3.13). In the active state, the salt bridge residue E762 carries the most destabilizing effect, similar to gefitinib simulation (Figure 3.13). In the inactive state, the binding mode of lapatinib is distal from the RS3 residue (Figure 3.12b). Our analysis suggest that residues in the ligand binding pocket (K728, L792, T790, K745, D855, N842, R841, D800, R803) are allosterically perturbed by the M766T mutation (Figure 3.13). D855 and D800, in particular, exert the most positive impact on the binding free energy (Figure 3.13). Thus, RS3 is not the only residue that affects drug binding. M766T communicates with other residues in the ligand binding pocket to allosterically perturb the specific mode of binding, leading to the observed

resistance/sensitivity.

Table 3.1: Binding free energy estimates of WT and mutant (M766T) EGFR for various inhibitors using MM/PBSA method.

Drug	$\Delta\Delta G_{M766T-WT}$
Lapatinib	-4.60 kJ/mol
Gefitinib	17.67 kJ/mol
Erlotinib (Active)	4.24 kJ/mol
Erlotinib (Inactive)	7.81 kJ/mol

3.3 Materials and Methods

3.3.1 EGF stimulation, cell lysis and immunoblotting:

The pEGFP-N1-EGFR plasmid from our previous study[14, 28] was used to generate point mutations. Before transfection, Chinese Hamster Ovary (CHO) cells were grown in high-glucose Dubecco's Modified Eagle Medium (DMEM) (Cellgro, Manassas, VA, USA) with 10% fetal bovine serum (Bioexpress, UT, USA) without antibiotics at 30% confluency. Transient transfection was then performed using lipofectamine-2000 (Invitrogen, Carlsbad, CA) according to manufacturer's protocol with WT and mutant EGFR. To detect autophosphorylation, transfected cells were serum-starved in Ham's F-12 media for 18 h followed by ligand stimulation with EGF (100 ng/ml) (Sigma, St. Louis, MO) for 5 min. Cells were washed with PBS, and lysed with lysis buffer (50 mM Tris-HCL, pH 7.4, 150 mM NaCl, 10% glycerol, 1 mM EDTA,1 mM sodium orthovanadate, 1% Triton X-100, 1 mM PMSF, and 1X Protease Inhibitor Cocktail Set V, EDTA-free). Total protein was resolved on 10% SDS-PAGE and transferred onto polyvinylidenedifluoride (PVDF) membrane. Western blotting

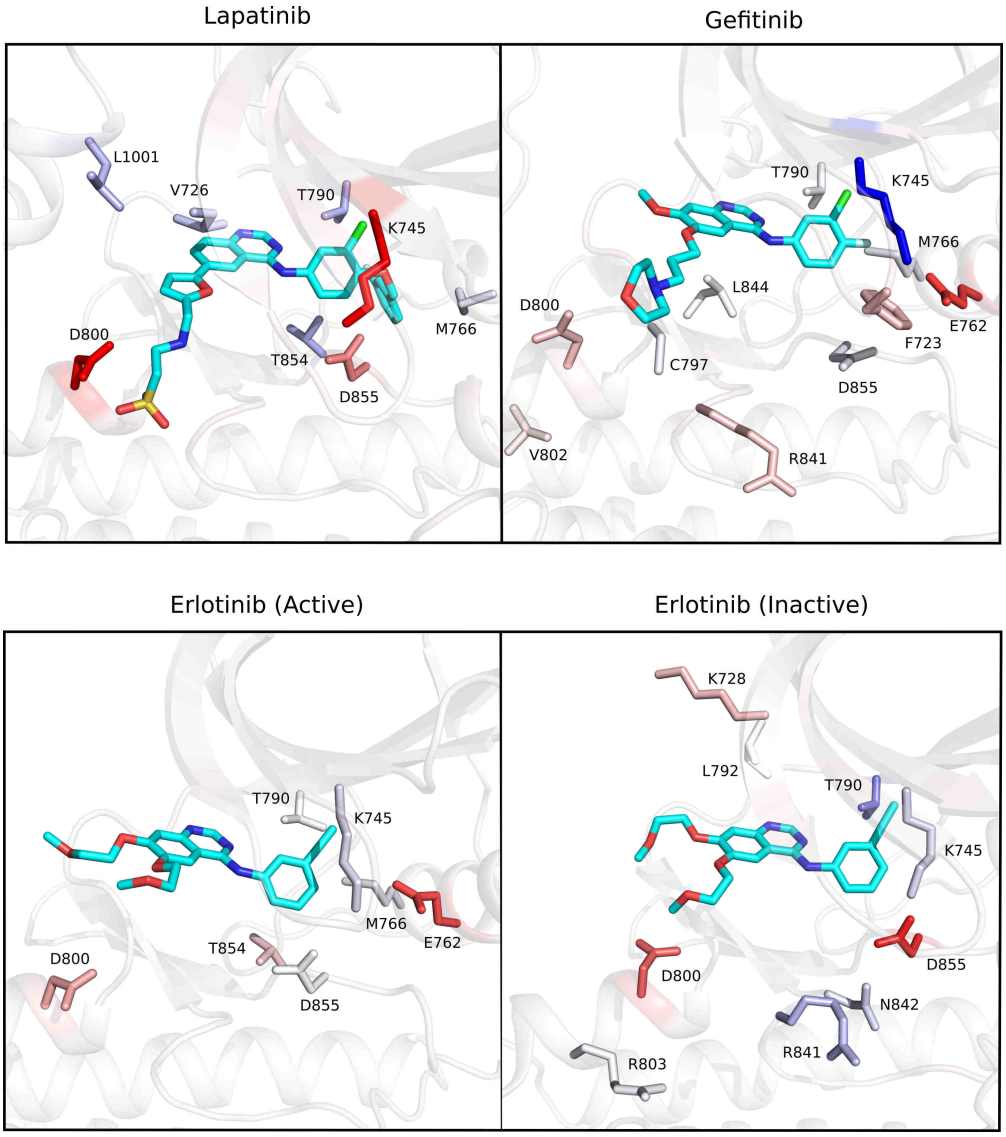

Figure 3.13: Structural mapping of MM/PBSA energy decomposition analysis. Ligand binding free energy difference between M766T and WT EGFR ($\Delta\Delta G = \Delta G_{M766T} - \Delta G_{WT}$) is mapped to each residue ($\Delta\Delta G < 0$: red; $\Delta\Delta G > 0$: blue). Stick representation is shown if the absolute $\Delta\Delta$G for the residue is above 2kJ/mol for lapatinib and 1kJ/mol for gefitinib and erlotinib.

was done using anti-GFP, anti-pY1197-EGFR, anti-FLAG, anti-STAT3, anti-pY705-STAT3 antibodies (Cell signaling, Danvers, MA). Proteins were detected using chemiluminescent substrate (ECL substrate, Pierce, Rockford, IL).

3.3.2 Drug sensitivity assays

For drug sensitivity assays, CHO cells were cultured, transfected, and starved as described above. After 18h of starvation, cells were treated with 0, 0.01, 0.1, 1.0 and 10μM lapatinib, gefitinib, or erlotinib for 2 h in Ham's 12 media without fetal bovine serum followed by EGF (100 ng/ml) stimulation for 5 min. Immunoblotting was performed as described above.

3.3.3 Structural Modelling

Inactive and active states of EGFR were modeled based on the crystal structures of EGFR kinase domain in the inactive (3W32)[29] and active conformations (2GS6)[19], respectively. Gefitinib, lapatinib and erlotinib bound structure of EGFR were modeled based on PDB entries 4WKQ (gefitinib), 1XKK (lapatinib), 1M17 (erlotinib active) and 4HJO (erlotinib inactive), respectively [23, 26, 24]. PDB entry 5D41 was used to model EGFR bound to the allosteric inhibitor EAI045 [18]. Disordered regions were modeled using the MODELLER program (version 9.12)[30]. Mutations introduced during crystallization were reverted to WT EGFR sequence during the modelling step. RS mutations were introduced in the modeled WT structure using the loop refine module and all modeled structures were manually inspected for steric clashes.

3.3.4 Molecular dynamics simulation

Molecular dynamics (MD) simulations were done using GROMACS version 5.0.4 [31] with AMBER ff99SB-ildn force field[32] and TIP3P water model. Force field parameter for various ligands were generated using 22824207 [33] based on General Amber Force Field (15116359)[34]. Long-range electrostatics were computed with PME method and a cutoff of 0.9 nm. Steepest descent and conjugate-gradient energy minimization was performed on the solvated protein for 20000 steps until Fmax was less than 100 kcal/mol. NVT simulations were performed to heat the system from 0 to 310 K by coupling it to berendsen thermostat for 200 ps with a restraint on the non-hydrogen atoms of the protein and the ligand, if present. Then, constant pressure dynamics (P = 1 atm, T = 310 K) were carried out under Parrinello-Rahman barostat to maintain pressure and density with the same restraint. The unrestrained MD productions were run for 300 ns for EGFR monomer and 150 ns for EGFR dimer with a timestep of 2 fs. Trajectories were determined to be stable based on root mean square deviation (RMSD) and root mean square fluctuation (RMSF) measurements. Analysis of MD trajectories was carried out using programs in the GROMACS suite[31]. All structure visualization was done using PyMOL 1.7.6[35].

3.3.5 MM/PBSA calculations

Molecular mechanics poisson boltzmann surface area (MM/PBSA) method has been widely used to estimate ligand-binding free energy[36, 37, 38]. In this study, MM/PBSA calculation was performed using the `g_mmpbsa` program[39]. Specifically, MD simulations with various ligands were performed for 300 ns following the protocol described above. To ensure equilibrium, the first 50 ns trajectory was omitted and a total of 210 snapshots were ex-

tracted from the remaining 250 ns MD data. Molecular mechanics potential energy (G_{mm}) for protein and ligand was calculated based on the force field parameters. Polar solvation energy (G_{polar}) was calculated using the linear poisson boltzmann equation solver in APBS[40]. Non-polar solvation energy ($G_{nonpolar}$) was generated using SASA-only model ($G_{nonpolar} = 0.0226778A + 3.84928$). Entropic contribution was not considered, as previous studies showed that it does not improve prediction accuracy [41, 42]. The energy decomposition was analyzed by summing the energy component (E_{mm}, G_{polar}, $G_{nonpolar}$) of each residue under consideration.

3.3.6 Kullback-Leibler (KL) divergence analysis

Kullback-Leibler divergence expansion has recently been proposed as a powerful approach to compare conformational ensembles of biomolecules [43, 44]. The MutInf package was used to perform KL divergence analysis on MD trajectories [43, 45]. In this study, we primarily focus on the local KL divergence to identify residues perturbed by a given mutation. Specifically, we split the 2 independent 300 ns MD trajectories (150 ns for EGFR dimer) into 4 equal sized blocks (ignoring the first 50 ns to ensure equilibrium) and quantify the residue torsion angles for each snapshot. The distribution of each torsion angle is approximated using bins with a size of 15 degrees. The KL divergence was used to identify residues that show divergence in their torsional angle distributions. The identified residues were mapped to structures and manually inspected for divergence in dynamics.

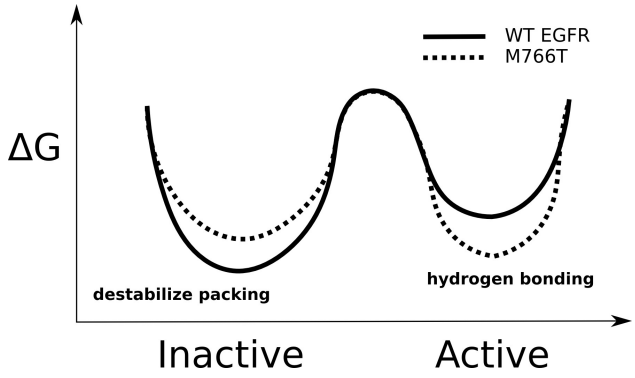

Figure 3.14: M766T activates EGFR by lowering the energy barrier for transition between active and inactive conformations.

3.4 Discussion

RS assembly is considered a key event in kinase activation [5, 7, 16, 17]. Although most RS mutations inactivate EGFR, our screening identified an activating mutation, M766T, at the RS3 position. Our studies demonstrate that M766T mediated allosteric activation involves multiple aspects (Figure 3.14). First, M766T activates EGFR by destabilizing the inactive conformation, the conformation that EGFR predominantly adopts in the absence of EGF. Second, M766T can stabilize the active conformation through water mediated interactions with the conserved K745-E762 salt bridge and backbone of the DFG motif, the precise positioning of which is critical for kinase activation[2, 3]. Third, as suggested by other MD studies, M766T can promote dimerization by rigidifying αC-helix conformation. It is worth mentioning that L833V also significantly increases the auto-phosphorylation of EGFR (Figure 3.3). L833 resides beneath the RS3 residue and packs against the A-loop, which is also part of the auto-inhibitory network. The fact that L833V is activating further enforces

the importance of RS3 associated interactions in the inactive state.

Although the RS is conserved across diverse kinases, the impact of RS mutations is different in each kinase. For example, M766F inactivates EGFR by stabilizing the hydrophobic packing in the inactive conformation. This is in contrast to B-Raf, where introducing a phenylalanine residue at the RS3 position activates the kinase[16, 17]. Comparing the local structural environment of RS3 in the active and inactive states of EGFR and B-Raf provides some clues. First, the DFG-Phe adopts a canonical "out" conformation in the inactive state of B-Raf while in EGFR the DFG-Phe still adopts an "in" conformation. Second, the nature of residues conserved at the RS4 position is different in EGFR and B-Raf. In B-Raf, the RS4 position is an aromatic (phenylalanine) residue, which can potentially stabilize the active conformation through $\pi - \pi$ aromatic stacking interactions with the phenylalanine at the RS3 position[46]. Such aromatic stabilization is not possible in EGFR since a leucine is present at the RS4 position. Replacing RS4 leucine in EGFR by phenylalanine (L777F) activates EGFR (Figure 3.15), supporting the critical role of aromatic interactions in RS assembly and kinase activation.

Our study also suggests that the hydrophobic nature of RS residues is not a strict requirement for RS assembly and kinase activity. The active state of a kinase can be achieved in multiple ways. Most protein kinases conserve a hydrophobic residue at the RS3 position that mediates van der Waals interaction with the DFG-Phe (RS2 position). However, there are many exceptions to this rule. For example, Aurora kinase A and CHK1 conserve an RS3 glutamine, which coordinates with the backbone of the DFG motif via a water molecule [47, 48]. In addition, substitution of RS3 (L95) to hydrophilic residues (asparagine) in PKA only has a small effect on catalytic activity [7]. Together, these data suggest that hydrophilic

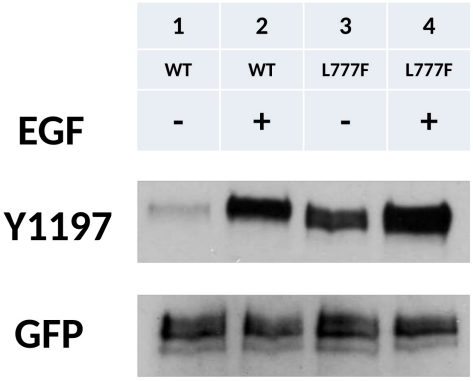

Figure 3.15: Activity of EGFR L777F mutant. Lanes from left to right: WT EGFR (−), WT EGFR (+), L777F (−), L777F (+). − and + indicate the absence and presence of EGF stimulation.

residues at the RS3 position can also facilitate RS assembly with the help of water mediated interactions.

The pocket between the αC-helix and activation segment is emerging as an important site for the design of allosteric protein kinase inhibitors [18]. The RS3 methionine is part of the drug binding pocket in several inhibitor bound crystal structures of EGFR [18, 23, 25, 24]. In fact, in one of the recently solved crystal structures of EGFR bound to the allosteric inhibitor, EAI045, M766 is part of the allosteric binding pocket (Figure 3.16). We also modeled the binding mode of the allosteric inhibitor in M766T, which drastically reduced the contact between αC-helix and EAI045 (Figure 3.16). Based on our findings, we hypothesize that patients harboring M766T mutation would be resistant to EAI045. This hypothesis, however, needs to be tested in future studies.

The RS3 position is a mutational hotspot in the kinase domain. A total of 70 different mutations spanning 55 different kinases have been identified at the RS3 position. EphA2,

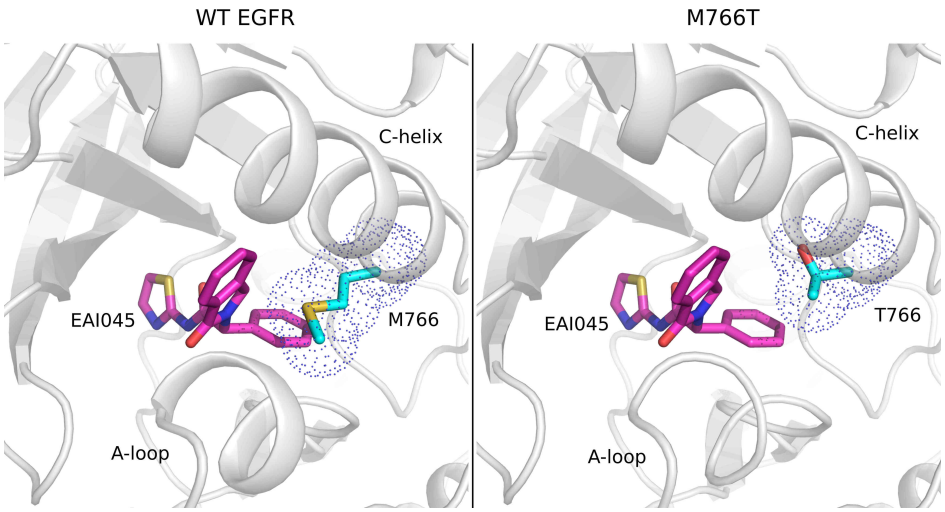

Figure 3.16: Comparison of the binding mode of WT EGFR and mutant (M766T) to the allosteric inhibitor EAI045.

EphA7, FRK, MET and ZAP70 harbor the exact same methionine to threonine mutation as seen in EGFR. MET and ZAP70 also adopt an inactive conformation similar to EGFR[49, 50]. Thus, although it is tempting to speculate that methionine to threonine substitution in these kinases will be activating, a detailed analysis of the local structural environment and family-specific variations in the hydrophobic core will be essential to accurately predict the structural and functional impact of RS3 mutations. We also note the preponderance of hydrophilic mutations at the RS3 position in cancer samples. Analogous to the M766T mutation studied here, hydrophilic residues at the RS3 position may involve a water molecule to stabilize the regulatory spine. Additional support for this notion comes from haspin kinase, where the RS3 residue is naturally conserved as a serine, which makes water mediated hydrogen bonds with the K511-E535 salt bridge and backbone amide of DYG-Y688 in a

manner similar to the model proposed for T766 in EGFR [51]. Thus, a deeper understanding of how naturally occurring variations contribute to protein functions and dynamics will be critical in accurately predicting the structural and functional impact of oncogenic mutations.

Mutation impact prediction is an important bioinformatics problem and several specialized web servers are available online for predicting mutation impact on protein structure and function. Our experimental studies on EGFR RS3 mutations provide a good dataset to test the performance of existing algorithms. We employed a total of 5 different mutation impact prediction algorithms (PANTHER-PSEP[52], PolyPhen2[53], I-Mutant2.0 [54], PROVEAN[55, 56], MutPred[57]) to predict the impact of EGFR RS3 mutations. The result is summarized in Table 3.2. PANTHER-PSEP, PolyPhen2 and PROVEAN predict all 5 mutants (M766A/S/T/V/F) as "probably damaging" or "deleterious". I-Mutant2.0 predicts all 5 mutants to "decrease stability". MutPred predicted all five RS3 mutations to be deleterious with a probability more than 0.8. In addition, several predefined features in the MutPred algorithm annotate RS3 mutations as "Loss of stability", "Gain of disorder", "Loss of catalytic residue at M766", and "Loss of phosphorylation at Y764". However, none of the prediction methods annotate M766T as potentially activating or gain of function. Thus, improved methods based on a deeper understanding of kinase structure, function and evolution will be essential to accurately predict the structural and functional impact of somatic variants identified in cancer genome sequencing studies.

Table 3.2: Mutation impact prediction for EGFR RS3 mutants.

	PANTHER	PolyPhen2	I-Mutant2.0	PROVEAN	MutPred
M766A	probably damaging	PROBABLY DAMAGING	decrease stability	Deleterious	Mutation Probability of deleterious mutation: 0.846
M766S	probably damaging	PROBABLY DAMAGING	decrease stability	Deleterious	Mutation Probability of deleterious mutation: 0.856
M766T	probably damaging	PROBABLY DAMAGING	decrease stability	Deleterious	Mutation Probability of deleterious mutation: 0.839
M766V	probably damaging	PROBABLY DAMAGING	decrease stability	Deleterious	Mutation Probability of deleterious mutation: 0.802
M766F	probably damaging	PROBABLY DAMAGING	decrease stability	Deleterious	Mutation Probability of deleterious mutation: 0.856

Bibliography

[1] Robert Roskoski. The erbb/her family of protein-tyrosine kinases and cancer. *Pharmacological research*, 79:34–74, 2014.

[2] Jane A Endicott, Martin E M Noble, and Louise N Johnson. The structural basis for control of eukaryotic protein kinases. *Annual review of biochemistry*, 81:587–613, 2012.

[3] S K Hanks and T Hunter. Protein kinases 6. the eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 9:576–96, 1995.

[4] Arvin C Dar, Thomas E Dever, and Frank Sicheri. Higher-order substrate recognition of eif2alpha by the rna-dependent protein kinase pkr. *Cell*, 122:887–900, 2005.

[5] Susan S Taylor and Alexandr P Kornev. Protein kinases: evolution of dynamic regulatory proteins. *Trends in biochemical sciences*, 36:65–77, 2011.

[6] Alexandr P Kornev, Nina M Haste, Susan S Taylor, and Lynn F Ten Eyck. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 103:17783–8, 2006.

[7] Hiruy S Meharena, Philip Chang, Malik M Keshwani, Krishnadev Oruganty, Aishwarya K Nene, Natarajan Kannan, Susan S Taylor, and Alexandr P Kornev. Deciphering the structural basis of eukaryotic protein kinase regulation. *PLoS biology*, 11:e1001680, 2013.

[8] Alexandr P Kornev, Susan S Taylor, and Lynn F Ten Eyck. A helix scaffold for the assembly of active protein kinases. *Proceedings of the National Academy of Sciences of the United States of America*, 105:14377–82, 2008.

[9] Krishnadev Oruganty, Nakul Suhas Talathi, Zachary A Wood, and Natarajan Kannan. Identification of a hidden strain switch provides clues to an ancient structural mechanism in protein kinases. *Proceedings of the National Academy of Sciences of the United States of America*, 110:924–9, 2013.

[10] Henrik Mobitz. The abc of protein kinase conformations. *Biochimica et biophysica acta*, 1854:1555–66, 2015.

[11] Krishnadev Oruganty and Natarajan Kannan. Evolutionary variation and adaptation in a conserved protein kinase allosteric network: implications for inhibitor design. *Biochimica et biophysica acta*, 1834:1322–9, 2013.

[12] Smita Mohanty, Krishnadev Oruganty, Annie Kwon, Dominic P Byrne, Samantha Ferries, Zheng Ruan, Laura E Hanold, Samiksha Katiyar, Eileen J Kennedy, Patrick A Eyers, and Natarajan Kannan. Hydrophobic core variations provide a structural framework for tyrosine kinase evolution and functional specialization. *PLoS genetics*, 12:e1005885, 2016.

[13] Kevin A James and Gennady M Verkhivker. Structure-based network analysis of activation mechanisms in the erbb family of receptor tyrosine kinases: the regulatory spine residues are global mediators of structural stability and allosteric interactions. *PloS one*, 9:e113488, 2014.

[14] Daniel Ian McSkimming, Shima Dastgheib, Eric Talevich, Anish Narayanan, Samiksha Katiyar, Susan S Taylor, Krys Kochut, and Natarajan Kannan. Prokino: a unified resource for mining the cancer kinome. *Human mutation*, 36:175–86, 2015.

[15] Gurinder Gosal, Krys J Kochut, and Natarajan Kannan. Prokino: an ontology for integrative analysis of protein kinases in cancer. *PloS one*, 6:e28782, 2011.

[16] Jiancheng Hu, Lalima G Ahuja, Hiruy S Meharena, Natarajan Kannan, Alexandr P Kornev, Susan S Taylor, and Andrey S Shaw. Kinase regulation by hydrophobic spine assembly in cancer. *Molecular and cellular biology*, 35:264–76, 2015.

[17] Jiancheng Hu, Edward C Stites, Haiyang Yu, Elizabeth A Germino, Hiruy S Meharena, Philip J S Stork, Alexandr P Kornev, Susan S Taylor, and Andrey S Shaw. Allosteric activation of functionally asymmetric raf kinase dimers. *Cell*, 154:1036–1046, 2013.

[18] Yong Jia, Cai-Hong Yun, Eunyoung Park, Dalia Ercan, Mari Manuia, Jose Juarez, Chunxiao Xu, Kevin Rhee, Ting Chen, Haikuo Zhang, Sangeetha Palakurthi, Jaebong Jang, Gerald Lelais, Michael DiDonato, Badry Bursulaya, Pierre-Yves Michellys, Robert Epple, Thomas H Marsilje, Matthew McNeill, Wenshuo Lu, Jennifer Harris, Steven Bender, Kwok-Kin Wong, Pasi A JÃďnne, and Michael J Eck. Overcoming egfr(t790m) and egfr(c797s) resistance with mutant-selective allosteric inhibitors. *Nature*, 534:129–32, 2016.

[19] Xuewu Zhang, Jodi Gureasko, Kui Shen, Philip A Cole, and John Kuriyan. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell*, 125:1137–49, 2006.

[20] S H Choi, J M Mendrola, and M A Lemmon. Egf-independent activation of cell-surface egf receptors harboring mutations found in gefitinib-sensitive lung cancer. *Oncogene*, 26:1567–76, 2007.

[21] Christopher C Valley, Alessandro Cembran, Jason D Perlmutter, Andrew K Lewis, Nicholas P Labello, Jiali Gao, and Jonathan N Sachs. The methionine-aromatic motif plays a unique role in stabilizing protein structure. *The Journal of biological chemistry*, 287:34979–91, 2012.

[22] Yibing Shan, Michael P Eastwood, Xuewu Zhang, Eric T Kim, Anton Arkhipov, Ron O Dror, John Jumper, John Kuriyan, and David E Shaw. Oncogenic mutations counteract intrinsic disorder in the egfr kinase and promote receptor dimerization. *Cell*, 149:860–70, 2012.

[23] Edgar R Wood, Anne T Truesdale, Octerloney B McDonald, Derek Yuan, Anne Hassell, Scott H Dickerson, Byron Ellis, Christopher Pennisi, Earnest Horne, Karen Lackey, Krystal J Alligood, David W Rusnak, Tona M Gilmer, and Lisa Shewchuk. A unique structure for epidermal growth factor receptor bound to gw572016 (lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer research*, 64:6652–9, 2004.

[24] Jin H Park, Yingting Liu, Mark A Lemmon, and Ravi Radhakrishnan. Erlotinib binds both inactive and active conformations of the egfr tyrosine kinase domain. *The Biochemical journal*, 448:417–23, 2012.

[25] Cai-Hong Yun, Titus J Boggon, Yiqun Li, Michele S Woo, Heidi Greulich, Matthew Meyerson, and Michael J Eck. Structures of lung cancer-derived egfr mutants and

inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer cell*, 11:217–27, 2007.

[26] Jennifer Stamos, Mark X Sliwkowski, and Charles Eigenbrot. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *The Journal of biological chemistry*, 277:46265–72, 2002.

[27] Egle Avizienyte, Richard A Ward, and Andrew P Garner. Comparison of the egfr resistance mutation profiles generated by egfr-targeted tyrosine kinase inhibitors and the impact of drug combinations. *The Biochemical journal*, 415:197–206, 2008.

[28] Zheng Ruan and Natarajan Kannan. Mechanistic insights into r776h mediated activation of epidermal growth factor receptor kinase. *Biochemistry*, 54:4216–25, 2015.

[29] Youichi Kawakita, Masaki Seto, Tomohiro Ohashi, Toshiya Tamura, Tadashi Yusa, Hiroshi Miki, Hidehisa Iwata, Hidenori Kamiguchi, Toshimasa Tanaka, Satoshi Sogabe, Yoshikazu Ohta, and Tomoyasu Ishikawa. Design and synthesis of novel pyrimido[4,5-b]azepine derivatives as her2/egfr dual inhibitors. *Bioorganic & medicinal chemistry*, 21:2250–2261, 2013.

[30] A Fiser, R K Do, and A Sali. Modeling of loops in protein structures. *Protein science : a publication of the Protein Society*, 9:1753–73, 2000.

[31] Mark James Abraham, Teemu Murtola, Roland Schulz, SzilÃąrd PÃąll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19 – 25, 2015.

[32] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins*, 78:1950–8, 2010.

[33] Alan W Sousa da Silva and Wim F Vranken. Acpype - antechamber python parser interface. *BMC research notes*, 5:367, 2012.

[34] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25:1157–74, 2004.

[35] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

[36] Hector A Velazquez and Donald Hamelberg. Conformation-directed catalysis and coupled enzyme-substrate dynamics in pin1 phosphorylation-dependent cis-trans isomerase. *The journal of physical chemistry. B*, 117:11509–17, 2013.

[37] Arghya Barman and Donald Hamelberg. Coupled dynamics and entropic contribution to the allosteric mechanism of pin1. *The journal of physical chemistry. B*, 120:8405–15, 2016.

[38] Minghui Li, Marharyta Petukh, Emil Alexov, and Anna R Panchenko. Predicting the impact of missense mutations on protein-protein binding affinity. *Journal of chemical theory and computation*, 10:1770–1780, 2014.

[39] Samuel Genheden and Ulf Ryde. The mm/pbsa and mm/gbsa methods to estimate ligand-binding affinities. *Expert opinion on drug discovery*, 10:449–61, 2015.

[40] N A Baker, D Sept, S Joseph, M J Holst, and J A McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, 98:10037–41, 2001.

[41] Nad'a SpackovÃą, Thomas E Cheatham, Filip RyjÃącek, Filip Lankas, Luc Van Meervelt, Pavel Hobza, and JirÃŋ Sponer. Molecular dynamics simulations and thermodynamics analysis of dna-drug complexes. minor groove binding between 4',6-diamidino-2-phenylindole and dna duplexes in solution. *Journal of the American Chemical Society*, 125:1759–69, 2003.

[42] Tianyi Yang, Johnny C Wu, Chunli Yan, Yuanfeng Wang, Ray Luo, Michael B Gonzales, Kevin N Dalby, and Pengyu Ren. Virtual screening using molecular simulations. *Proteins*, 79:1940–51, 2011.

[43] Christopher L McClendon, Lan Hua, Abriela Barreiro, and Matthew P Jacobson. Comparing conformational ensembles using the kullback-leibler divergence expansion. *Journal of chemical theory and computation*, 8:2115–2126, 2012.

[44] Jonathan A Stefely, Floriana Licitra, Leila Laredj, Andrew G Reidenbach, Zachary A Kemmerer, Anais Grangeray, Tiphaine Jaeg-Ehret, Catherine E Minogue, Arne Ulbrich, Paul D Hutchins, Emily M Wilkerson, Zheng Ruan, Deniz Aydin, Alexander S Hebert, Xiao Guo, Elyse C Freiberger, Laurence Reutenauer, Adam Jochem, Maya Chergova, Isabel E Johnson, Danielle C Lohman, Matthew J P Rush, Nicholas W Kwiecien, Pankaj K Singh, Anna I Schlagowski, Brendan J Floyd, Ulrika Forsman, Pavel J Sindelar, Michael S Westphall, Fabien Pierrel, Joffrey Zoll, Matteo Dal Peraro, Natarajan Kannan, Craig A Bingman, Joshua J Coon, Philippe Isope, HÃĪlÃĺne Puccio,

and David J Pagliarini. Cerebellar ataxia and coenzyme q deficiency through loss of unorthodox kinase activity. *Molecular cell*, 63:608–620, 2016.

[45] Christopher L McClendon, Gregory Friedland, David L Mobley, Homeira Amirkhani, and Matthew P Jacobson. Quantifying correlations between allosteric sites in thermodynamic ensembles. *Journal of chemical theory and computation*, 5:2486–2502, 2009.

[46] N Kannan and S Vishveshwara. Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein engineering*, 13:753–61, 2000.

[47] Jacek Nowakowski, CiarÃąn N Cronin, Duncan E McRee, Mark W Knuth, Christian G Nelson, Nikola P Pavletich, Joe Rogers, Bi-Ching Sang, Daniel N Scheibe, Ronald V Swanson, and Devon A Thompson. Structures of the cancer-related aurora-a, fak, and epha2 protein kinases from nanovolume crystallography. *Structure (London, England : 1993)*, 10:1659–67, 2002.

[48] Vibha Oza, Susan Ashwell, Patrick Brassil, Jason Breed, Jaychandran Ezhuthachan, Chun Deng, Michael Grondine, Candice Horn, DongFang Liu, Paul Lyne, Nicholas Newcombe, Martin Pass, Jon Read, Mei Su, Dorin Toader, Dingwei Yu, Yan Yu, and Sonya Zabludoff. Synthesis and evaluation of triazolones as checkpoint kinase 1 inhibitors. *Bioorganic & medicinal chemistry letters*, 22:2330–7, 2012.

[49] Sean G Buchanan, Jorg Hendle, Patrick S Lee, Christopher R Smith, Pierre-Yves Bounaud, Katti A Jessen, Crystal M Tang, Nanni H Huser, Jeremy D Felce, Karen J Froning, Marshall C Peterman, Brandon E Aubol, Steve F Gessert, J Michael Sauder, Kenneth D Schwinn, Marijane Russell, Isabelle A Rooney, Jason Adams, Barbara C

Leon, Tuan H Do, Jeff M Blaney, Paul A Sprengeler, Devon A Thompson, Lydia Smyth, Laura A Pelletier, Shane Atwell, Kevin Holme, Stephen R Wasserman, Spencer Emtage, Stephen K Burley, and Siegfried H Reich. Sgx523 is an exquisitely selective, atp-competitive inhibitor of the met receptor tyrosine kinase with antitumor activity in vivo. *Molecular cancer therapeutics*, 8:3181–90, 2009.

[50] Sebastian Deindl, Theresa A Kadlecek, Tomas Brdicka, Xiaoxian Cao, Arthur Weiss, and John Kuriyan. Structural basis for the inhibition of tyrosine kinase activity of zap-70. *Cell*, 129:735–46, 2007.

[51] Jeyanthy Eswaran, Debasis Patnaik, Panagis Filippakopoulos, Fangwei Wang, Ross L Stein, James W Murray, Jonathan M G Higgins, and Stefan Knapp. Structure and functional characterization of the atypical human kinase haspin. *Proceedings of the National Academy of Sciences of the United States of America*, 106:20198–203, 2009.

[52] Haiming Tang and Paul D Thomas. Panther-psep: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics (Oxford, England)*, 32:2230–2, 2016.

[53] Ivan Adzhubei, Daniel M Jordan, and Shamil R Sunyaev. Predicting functional effect of human missense mutations using polyphen-2. *Current protocols in human genetics*, Chapter 7:Unit7.20, 2013.

[54] Emidio Capriotti, Piero Fariselli, and Rita Casadio. I-mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic acids research*, 33:W306–10, 2005.

[55] Yongwook Choi, Gregory E Sims, Sean Murphy, Jason R Miller, and Agnes P Chan. Predicting the functional effect of amino acid substitutions and indels. *PloS one*, 7:e46688, 2012.

[56] Yongwook Choi and Agnes P Chan. Provean web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics (Oxford, England)*, 31:2745–7, 2015.

[57] Biao Li, Vidhya G Krishnan, Matthew E Mort, Fuxiao Xin, Kishore K Kamati, David N Cooper, Sean D Mooney, and Predrag Radivojac. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics (Oxford, England)*, 25:2744–50, 2009.

Chapter 4

Altered conformational landscape and dimerization dependency underpins the activation of EGFR by αC-β4 loop insertion mutations

Submitted to *Proceedings of the National Academy of Sciences of USA*, Feb 26, 2018

Abstract

Mutational activation of EGFR in human cancers involves both point mutations and complex mutations (insertions and deletions). In particular, short in-frame insertion mutations within a conserved αC-β4 loop in the EGFR kinase domain are frequently observed in tumor samples and patients harboring these mutations are insensitive to first-generation EGFR inhibitors. Despite the prevalence and clinical relevance of insertion mutations, the mechanisms by which these mutations regulate EGFR activity and contribute to drug sensitivity are poorly understood. Using cell-based mutation screening, we find that the precise location, length, and sequence of the inserted segment are critical for ligand-independent EGFR activation and downstream signaling. We identify three insertion mutations (N771_P772insN, D770_N771insG, and D770>GY) that activate EGFR in a unique way by relying more on the "acceptor" interface for kinase activation. Our drug inhibition studies indicate that these activating insertion mutations respond more favorably to osimertinib, a recently FDA-approved EGFR inhibitors for T790M positive lung cancer patients. Molecular dynamics simulations and umbrella sampling of WT and mutant EGFR suggest a model in which activating insertion mutations increase catalytic activity by relieving key auto-inhibitory interactions associated with αC-helix movement and by lowering the transition free energy ($\Delta G_{\text{active-inactive}}$) between active and inactive states. Our studies also identify a novel transition state sampled by activating insertion mutations that can be exploited in the design of mutant-selective EGFR inhibitors.

4.1 Introduction

Many human cancers are caused by the accumulation of somatic mutations in oncogenes that confer selective growth advantage. The epidermal growth factor receptor (EGFR) kinase is one such oncogene that is mutationally activated in many human cancers, in particular lung cancer [1, 2, 3, 4]. Cancer genome sequencing studies have revealed numerous genomic alterations in EGFR, including point mutations, deletions and insertions [5, 6, 7]. A detailed understanding of these mutations is essential for developing effective therapeutic strategies [8]. While some frequently occurring point mutations and deletion mutations, such as L858R, G719S, T790M, and E746_A750del, have been well studied [9, 10, 11, 12, 13], the structural and functional impact of many other recurrent mutations still remain largely unknown. In-frame insertions in exon 20 are a subcategory of understudied mutations that are frequently observed in cancer patients [14, 15, 16, 17]. Most of the exon 20 insertions map to a short loop connecting the αC-helix and β4 strand in the protein kinase domain, i.e. αC-β4 loop [15]. Exon 20 insertions confer resistance to first generation EGFR inhibitors and are associated with poor clinical outcomes [6, 18]. Recent studies suggest that the drug response profile of exon 20 insertions are heterogeneous and depends on the nature of the inserted segment [15, 16, 19]. However, an incomplete understanding of how exon 20 insertion mutations regulate EGFR activity has hindered the development of therapies for patients harboring these mutations.

In this study, we systematically analyzed exon 20 insertion mutations in EGFR using a combination of experimental and computational approaches. We identified the αC-β4 loop, which is involved in kinase regulation [20, 21, 22], evolution [23] and chaperone-assisted folding [24, 25], as a hotspot for insertion mutations. Using a cell-based screen of twelve most

commonly occurring insertions in the αC-β4 loop, we identified three activating mutations (N771_P772insN, D770_N771insG, and D770>GY) that activate EGFR in a dimerization-dependent manner. These activating mutations rely more on the canonical "acceptor" interface than the "donor" interface for kinase activation. Drug response studies indicate that the three insertion mutations are insensitive towards reversible EGFR inhibitors (gefitinib, erlotinib, and lapatinib), but display a heterogeneous response towards irreversible inhibitors. In particular, all three insertion mutations respond to osimertinib more favorably. Likewise, D770>GY is more sensitive to dacomitinib treatment compared to WT EGFR. To investigate the structural basis for mutational activation, we generated representative structural models for each of the insertion mutations and performed a total of 35 µs MD simulations in various conformational states. MD simulations suggest that insertion mutations restrict the αC-helix conformational freedom in the active state by hindering the formation of a critical auto-inhibitory capping interaction associated with αC-helix movement. Steered MD and umbrella sampling suggest that activating insertion mutations not only lower the free energy difference between active and inactive states ($\Delta G_{active-inactive}$) but also stabilize key intermediate states between the two states. Taken together, these results suggest a unique mode of EGFR activation by insertion mutations and provide new clues for EGFR-targeted cancer therapies.

4.2 Result

4.2.1 αC-β4 loop is a hotspot for insertion mutations in EGFR

To comprehensively identify EGFR insertion mutations, we mined the Catalogue of Somatic Mutations in Cancer (COSMIC v83) database and the Protein Kinase Ontology (ProKinO v2.0) [26, 27]. We identified a total of 97 distinct insertion mutations mapping to the EGFR kinase domain (Figure 4.1a). The identified insertion mutations are heterogeneous in that they are variable in length and map to different regions of the protein. In particular, the αC-β4 loop, which connects the αC-helix and β4 strand in the N-terminal ATP binding lobe of the kinase domain, is an insertion hotspot accounting for 86.9% of unique insertion mutations in EGFR (Figure 4.1a). Some of the frequently occurring insertion mutations in this loop include V769_D770insASV, D770_N771insSVD, H773_V774insNPH, H773_V774insH, D770_N771insG, D770>GY, V774_C775insHV, H773_V774insPH, P772_H773insPR, N771_P772insN, H773_V774insAH, and N771_P772>SVDNR (See Figure 4.1 legend for insertion mutation nomenclature). These mutations are also referred to as exon 20 insertion mutations in the literature [6, 15]. In addition to the αC-β4 loop, insertion mutations are also identified in the β3 strand (also referred to as exon 19 insertion mutations), the juxtamembrane segment, and N-terminus of the αC-helix (Figure 4.1a) [17, 28]. These mutations, however, are not as frequent as the αC-β4 loop insertion mutations.

4.2.2 Identification of activating insertion mutations

We next investigated the impact of αC-β4 loop insertion mutations on EGFR autophosphorylation activity and downstream signaling using cell-based kinase assays (see Materials and

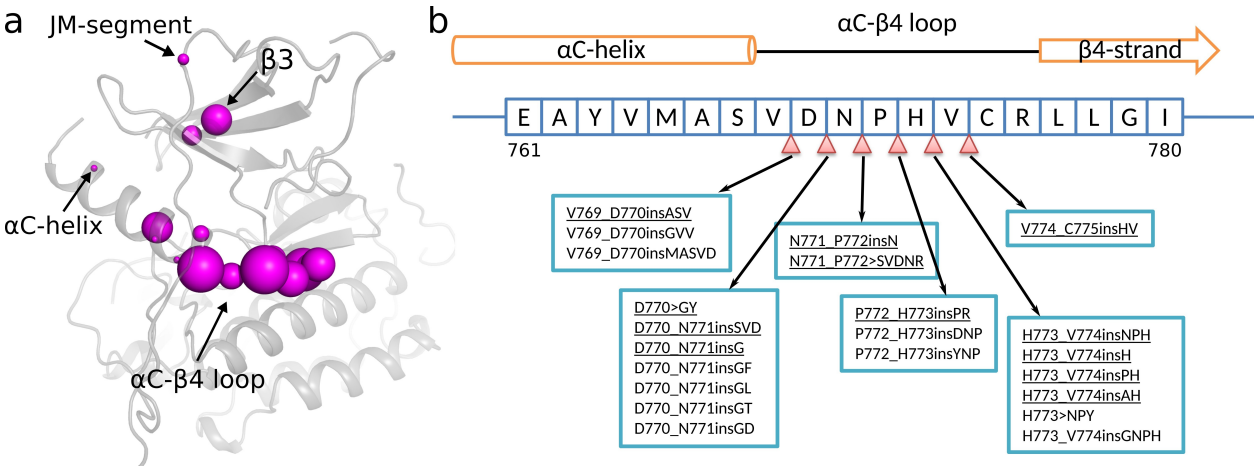

Figure 4.1: EGFR insertion mutations. (a) Structural location of insertion mutations in the EGFR kinase domain crystal structure (PDBID: 2GS6). The size of the magenta sphere is proportional to the log of the number of patient samples containing the insertion mutation at the particular residue position. (b) Sequence location of insertion mutations within the EGFR αC-β4 loop. The nomenclature of insertion mutations follows the guidelines established previously [29]. For example, V769_D770insASV indicates an Ala-Ser-Val insertion between residues V769 and D770, and D770>GY indicates a complex insertion mutation in which D770 is replaced by Gly-Tyr sequence. Insertion mutations characterized in this study are underlined.

Methods). The insertion mutations exhibit varying levels of C-terminal tail tyrosine (Y1197) phosphorylation, activation loop (Y869) phosphorylation and downstream protein phosphorylation in comparison to WT (Figure 4.2a, 4.2b). For most mutations, including the frequently occurring V769_D770insASV, D770_N771insSVD and H773_V774insNPH mutations, the activity of the mutant form is comparable, or slightly reduced, relative to WT (Figure 4.2a, 4.2b). N771_P772insN, D770_N771insG, D770_N771insSVD, V769_D770insASV, D770>GY, and N771_P772>SVDNR exhibit similar Y1197 phosphorylation in the presence and absence of EGF ligand, suggesting that these insertion mutations are insensitive to EGF stimulation (Figure 4.2a). This is different from the patient-derived point mutation (D770N) in the αC-β4 loop, which still shows ligand-dependent activation (Figure 4.2a). Interestingly, three mutations (D770_N771insG, and D770>GY, N771_P772insN) display increased autophosphorylation activity both in the presence and absence of EGF (Figure 4.2a, 4.2b). Notably, downstream STAT3 phosphorylation is also enhanced for these mutants relative to WT (Figure 4.2a), suggesting that N771_P772insN, D770_N771insG, and D770>GY are activating gain-of-function mutations.

4.2.3 Differential dimerization dependency of activating αC-β4 loop insertion mutations

Previous studies have shown that activating point mutations in EGFR such as L858R/T790M and R776H enhance the "acceptor" activity in the asymmetric dimer [30, 20]. To investigate the dimerization dependency of the activating αC-β4 loop insertion mutations (D770_N771insG, D770>GY, and N771_P772insN), we introduced a C-lobe dimerization-deficient mutation (V948R or M952R) or an N-lobe dimerization-deficient mutation (L760R

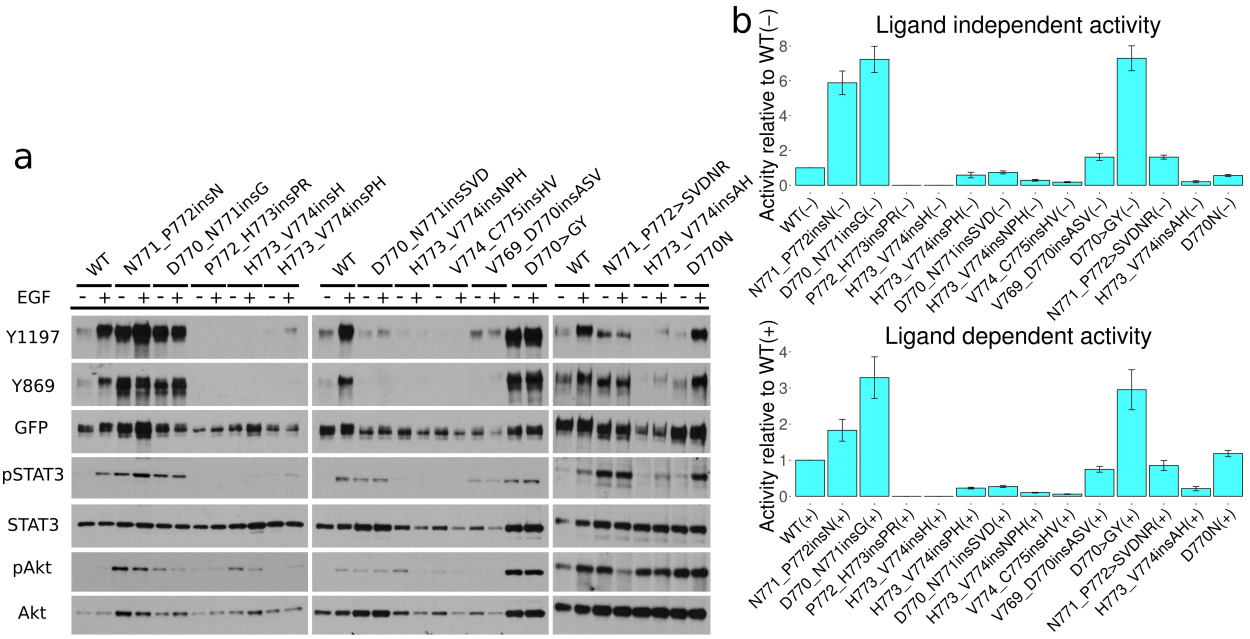

Figure 4.2: Cell-based screening of EGFR insertion mutations. (a) Western blot of 13 different EGFR mutations. − and + indicate the absence and presence of EGF ligand. Autophosphorylation activity is indicated by Y1197 and Y869 phosphorylation. GFP signal indicates the total expression level of EGFR. Downstream signaling of STAT3 and Akt is also shown. (b) Densitometry of relative phosphorylation activity of various EGFR insertion mutations compared to WT. The ratio between Y1197 and GFP signal is used as a measure of activity. The score for each insertion mutation is normalized to WT EGFR. Both ligand-independent activity (left panel) and ligand-dependent activity (right panel) are shown. Standard error (SE) represents 3-5 independent experiments.

or L704N) in the insertion mutation plasmids. We selected two mutations in each interface to rule out the possibility that the mutations might alter the kinase domain itself rather than the dimer interaction. Western blot analysis indicates that insertion mutations are still active in the presence of C-lobe dimerization-deficient mutation (V948R or M952R) (Figure 4.3a, lanes 11-12, 17-18, 23-24, 39-40, 45-46, 51-52, and Figure 4.4). In contrast, the presence of the N-lobe dimerization-deficient mutation (L760R or L704N) results in significantly reduced Y1197 phosphorylation (Figure 4.3a, lanes 13-14, 19-20, 25-26, 37-38, 43-44, 49-50, and Figure 4.4). As a control, Y1197 phosphorylation activity is not observed for WT EGFR when dimerization-deficient mutations are present (Figure 4.3a, lanes 3-6, 29-33, and Figure 4.4); however, Y1197 phosphorylation can be rescued by co-transfecting the two mutants (V948R+L760R or M952R+L704N) (Figure 4.3a, lanes 7-8, 33-34, and Figure 4.4). These data indicate that activating αC-β4 loop insertion mutations rely more on the N-lobe interface than the C-lobe interface for EGF-dependent and EGF-independent activation. In addition, we also tested the effect of introducing both V948R and M952R mutations in the insertion mutation background and observed comparable Y1197 autophosphorylation (Figure 4.4, 4.5), suggesting that the αC-β4 loop insertion mutations could bypass the canonical C-lobe asymmetric dimer interface to achieve kinase activation.

4.2.4 Drug sensitivity of activating αC-β4 loop insertion mutations to reversible and irreversible EGFR inhibitors

Exon 20 insertion mutations are generally associated with poor clinical response to first-generation EGFR inhibitors, such as gefitinib, erlotinib, and lapatinib [6, 14]. We tested the dose-dependent response of the three activating αC-β4 loop insertion mutations against

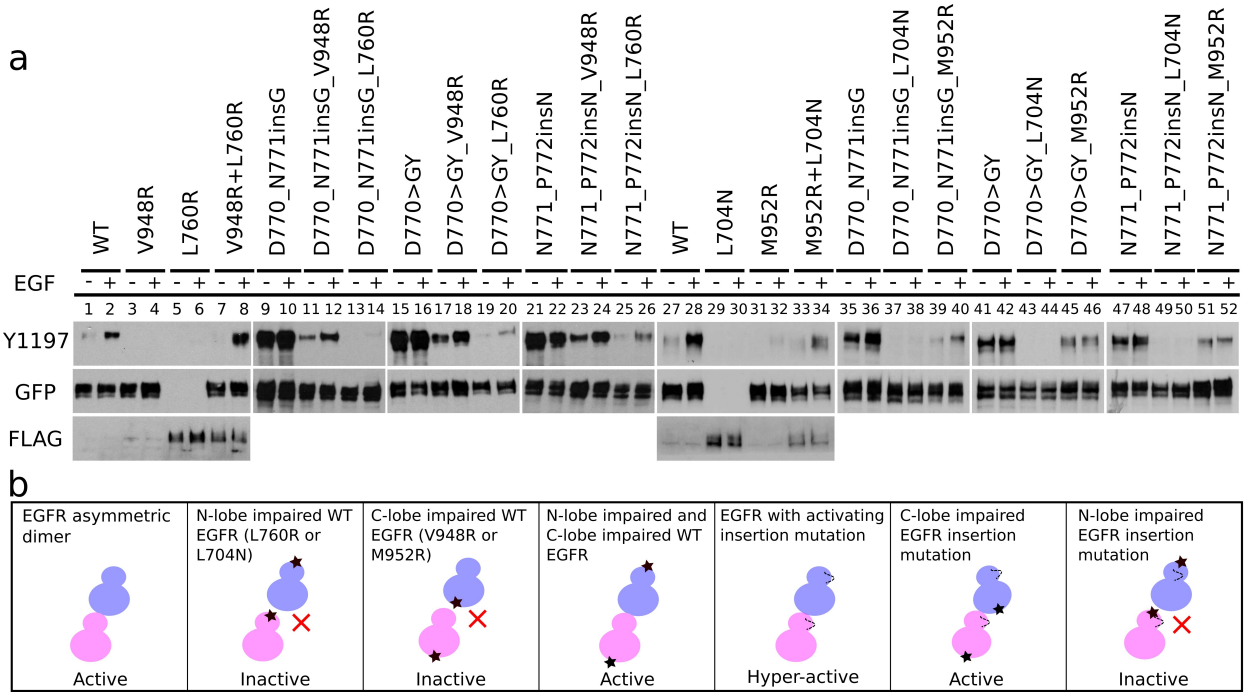

Figure 4.3: Dimerization-dependency of EGFR insertion mutations. (a) The impact of N-lobe dimerization-deficient mutation (L760R or L704N) and the C-lobe dimerization-deficient mutation (V948R or M952R) on Y1197 autophosphorylation for WT and mutant (D770_N771insG, N771_P772insN, and D770<GY) EGFR. All the constructs used are tagged with GFP except for L760R and L704N, which is tagged with FLAG. Co-transfection of V948R and L760R is indicated by V948R+L760R. Co-transfection of M952R and L704N is indicated by M952R+L704N. Densitometry analysis of 3-5 independent experiments is shown in Figure 4.4. (b) A schematic figure summarizing the observed differences in WT and mutant EGFR in the different dimerization deficient mutant background.

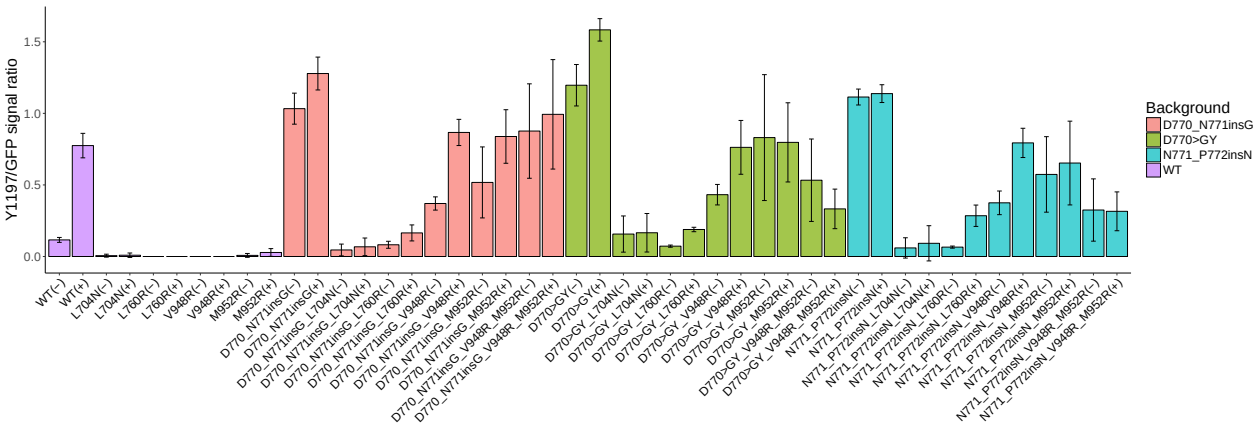

Figure 4.4: Densitometry analysis of the Western blot result in Figure 4.3. Error bar represents standard error (SE) of three independent experiments.

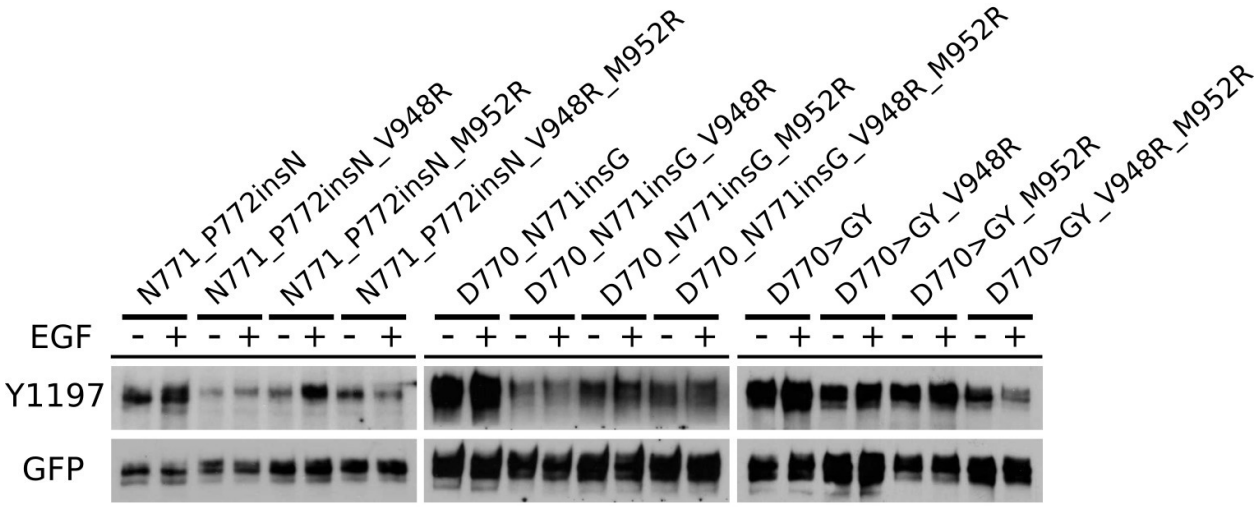

Figure 4.5: Activating insertion mutations are tolerant to C-lobe dimerization-deficient mutations. EGFR C-lobe dimerization- deficient mutations, V948R or M952R or V948R+M952R, are introduced to N771_P772insN, D770_N771insG, and D770>GY mutants background. Auto-phosphorylation of Y1197 is probed through phospho-specific antibody. The total expression of the designated constructs is detected using GFP antibody.

these inhibitors. Our data indicate that the activating insertion mutants (D770_N771insG, D770>GY, and N771_P772insN) are insensitive to these inhibitors as the inhibition profile is either unaltered or resistant with increasing drug concentration (Figure 4.6a), suggesting that patients harboring these mutations might not benefit from first-generation EGFR inhibitors. Recently, second-generation non-reversible covalent bond inhibitors, such as osimertinib, dacomitinib, and afatinib, have been developed to overcome drug resistance conferred by the gatekeeper (T790M) mutation [31, 32, 33]. We tested the responsiveness of activating αC-β4 loop insertion mutations to these non-reversible EGFR inhibitors (osimertinib and dacomitinib). Interestingly, we find that the three insertion mutations respond to osimertinib more favorably, as the Y1197 autophosphorylation is inhibited to various degrees at 0.1 µM drug concentration for insertion mutations but not for WT EGFR (Figure 4.6a, 4.6b). In addition, we find that D770>GY mutant displays a sensitive profile towards dacomitinib (Figure 4.6a, 4.6b). In conclusion, our drug sensitivity assays demonstrate a heterogeneous response to irreversible EGFR inhibitors by activating αC-β4 loop insertion mutations.

4.2.5 Activating insertion mutations restrict kinase conformational freedom

To investigate the structural basis for kinase domain activation by αC-β4 loop insertion mutations, we employed homology modeling and loop refinement strategies to generate an ensemble of structural models for each activating insertion mutation (see Materials and Methods). The generated models were clustered based on the root mean square deviation (RMSD) values. Representative models from each cluster were used for MD simulations. Comparisons of the MD trajectories indicate that the overall thermal fluctuations of WT and

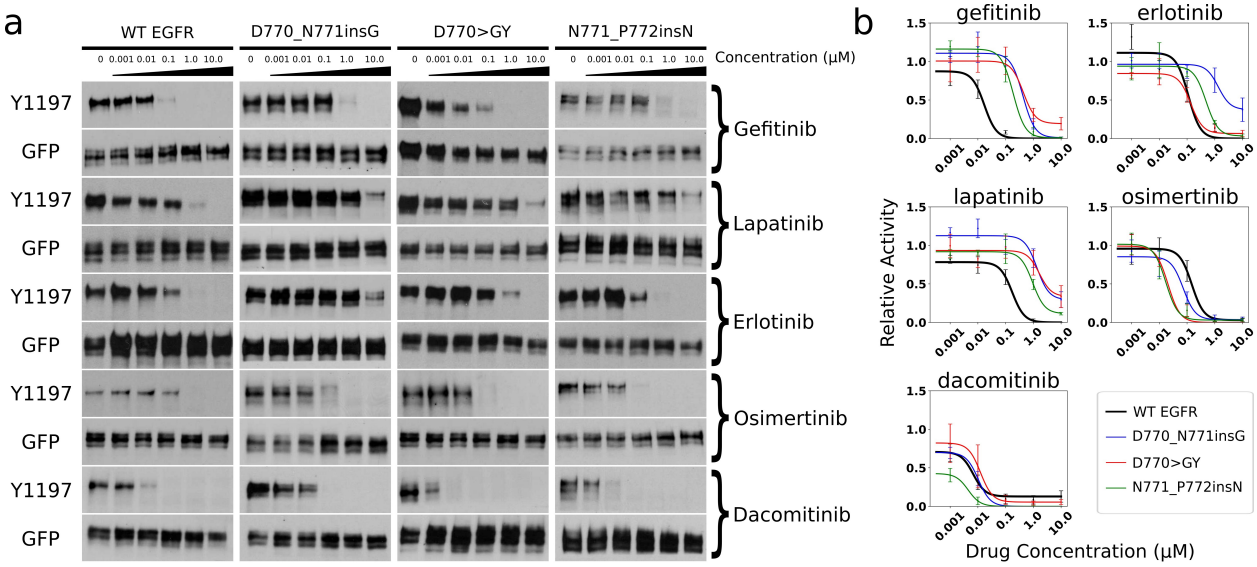

Figure 4.6: Drug inhibition assay of five EGFR inhibitors (gefitinib, lapatinib, erlotinib, osimertinib and dacomitinib). (a) The amount of Y1197 phosphorylation at drug concentration (0, 0.001 μM, 0.01 μM, 0.1 μM, 1.0 μM and 10.0 μM) is tested for WT and three activating insertion mutations. (b) Densitometry quantification of the drug treatment experiments with 3-5 replicates.

mutant EGFR are comparable during the course of the simulations. However, the intrinsic dynamics of the kinase domain, particularly for regions involved in kinase regulation, are significantly different, as described below.

Restricted conformational sampling by activating αC-β4 loop insertions in the active state

We used the root mean square deviation (RMSD) of the regulatory αC-helix and the K745-E762 salt bridge distance as proxies (variables) to compare the conformational freedom of WT and mutant EGFR in the active state. A joint probability distribution plot of these two variables in energy units is shown in Figure 4.7a, where each point represents the Boltzmann-weighted frequency of MD snapshots adopting the corresponding αC-helix RMSD and K745-E762 distance. Based on the plot, distinct clusters/states can be identified for WT and mutant EGFR. In general, WT is more flexible and samples four distinct states/clusters (as indicated by the arrows in Figure 4.7a), while the three activating insertion mutations only explore a subset of these states. State 1 corresponds to a native-like "active" conformation shared by WT and mutant EGFR with an intact αC-helix and a well-established K745-E762 salt bridge (Figure 4.7a, 4.7b). State 2 is a meta state, also shared by WT EGFR and all three insertion mutations (Figure 4.7a). The K745-E762 distance in state 2 centers around 5 Å and the αC-helix is in an "inward" conformation (Figure 4.7b). State 3 represents a cluster of conformations in which the N-terminus of the αC-helix is unfolded/disordered (Figure 4.7b). State 3 is sampled by N771_P772insN, but not by D770>GY and D770_N771insG. In state 3, a truncated αC-helix is only formed after residue K757, which makes the αC-helix four residues shorter (Figure 4.7b). Notably, such helix truncation is also observed in some of the

recently solved EGFR crystal structures (PDBID: 4WRG and 4ZAU) [34], indicating that our simulations are sampling biologically relevant conformations. State 4 corresponds to a highly destabilized αC-helix and a broken K745-E762 salt bridge, as observed in previous microsecond time-scale MD simulations [11]. In state 4, E762 in the C-helix makes a charged interaction with K860 from the activation loop, an interaction only seen in the inactive crystal structure of EGFR (Figure 4.7b). None of the mutants sample state 4, suggesting that the conformational freedom of the activating mutants are restricted to the active state (state 1) or meta states close to the active state (states 2-3), consistent with our cell-based assay in which these three insertion mutations are more active than WT (Figure 4.2, Figure 4.7a). Interestingly, the conformation heterogeneity of the mutants in the active state is also correlated with kinase activity in that the conformationally flexible N771_P772insN mutant is less active compared to the conformationally restricted D770>GY and D770_N771insG mutants (Figure 4.2b, Figure 4.7a).

Activating insertion mutations occlude auto-inhibitory interactions associated with αC-helix movement

We previously reported that stabilization of αC-helix in the inactive state is correlated with an auto-inhibitory capping interaction formed between R776/(NE,NH1,NH2) in the αC-β4 loop and A767/O in the αC-helix [20]. We hypothesize that activating insertion mutations in the αC-β4 loop could also modulate the αC-helix movement by altering the αC-helix capping interaction. To test this hypothesis, we analyzed our MD trajectories to correlate changes in the R776/(NE,NH1,NH2)-A767/O distance with the K745-E762 salt bridge, which serves as a proxy for kinase conformational transition (Figure 4.8a). The

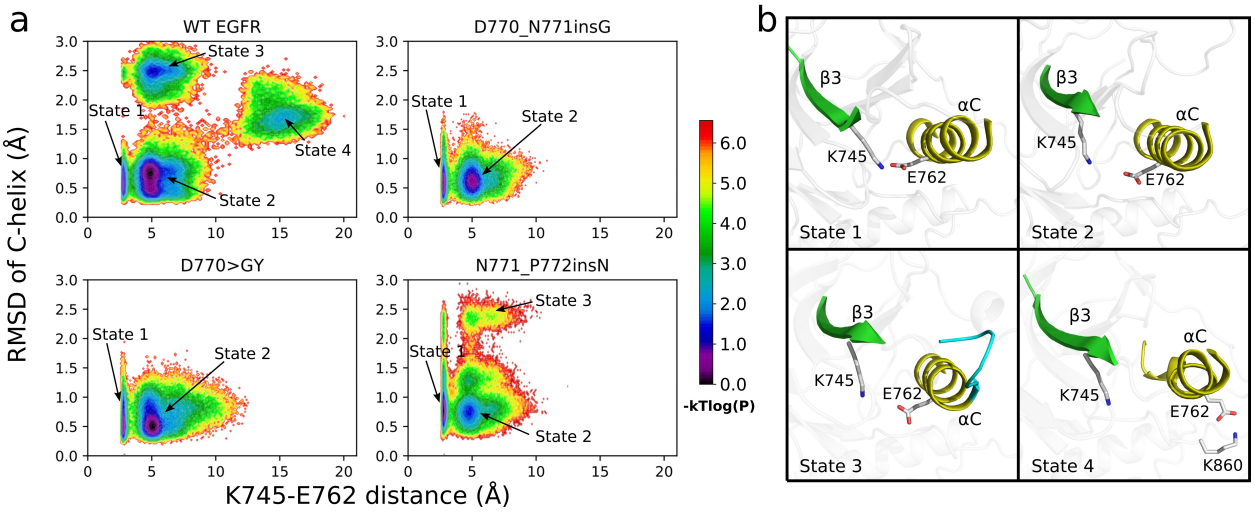

Figure 4.7: MD simulation of EGFR and three activating insertion mutations in the active state. (a) The joint distribution of the K745/NZ-E762/(OE1,OE2) distance and the αC-helix RMSD are shown in logarithmic scale. The plot is constructed using 100 different bins on both axes. The probability of each bin is estimated using the frequency of the 3 µs MD snapshots that fall into each bin. The distribution plot is transformed into an energy unit by $-kTlog(P)$, where kT is set to 1.0 in generating the contour image. Four distinct state clusters are labeled. (b) Representative snapshots from each conformational state labeled in (a) are shown.

Boltzmann weighted probability distribution plot of these two distance variables indicates an energy well at a K745-E762 distance of 5 Å and R776-A767 distance of 2.5 Å for WT (Figure 4.8a). However, for the mutants, the energy well is distributed across various R776-A767 distance values in the vertical axis (Figure 4.8a). For the D770>GY mutant, the energy well is maintained at 10 Å whereas for the D770_N771insG mutant, it is mostly populated at 7.5 Å (Figure 4.8a), indicating that the capping interaction is not frequently sampled by these mutants. By visualizing the MD trajectories of WT and mutant EGFR, we identify key atomic interactions that provide plausible explanations for the observed differences in the conformational ensemble (Figure 4.8b). In the D770>GY mutant, for example, the inserted tyrosine residue orients in a way that prevents R776 from interacting with A767 (Figure 4.8b). Likewise, in the D770_N771insG mutant, the inserted segment introduces an additional turn in the αC-helix that prevents R776 from forming the canonical auto-inhibitory capping interaction with A767 (Figure 4.8b). Although no clear interactions are identified for N771_P772insN, the flat energy distribution observed for this mutant suggests that the R776-A767 interaction is destabilized relative to WT EGFR (Figure 4.8a).

Stabilized inactive state simulation

We also performed MD simulations of WT and mutant forms in the inactive state. Unlike the active state, the overall thermal fluctuations of WT and mutant forms are similar, with only minor differences in the C-terminus of the αC-helix. This may be due to the overall global stabilization of the inactive state relative to the active state, as suggested by our free energy landscape analysis described below.

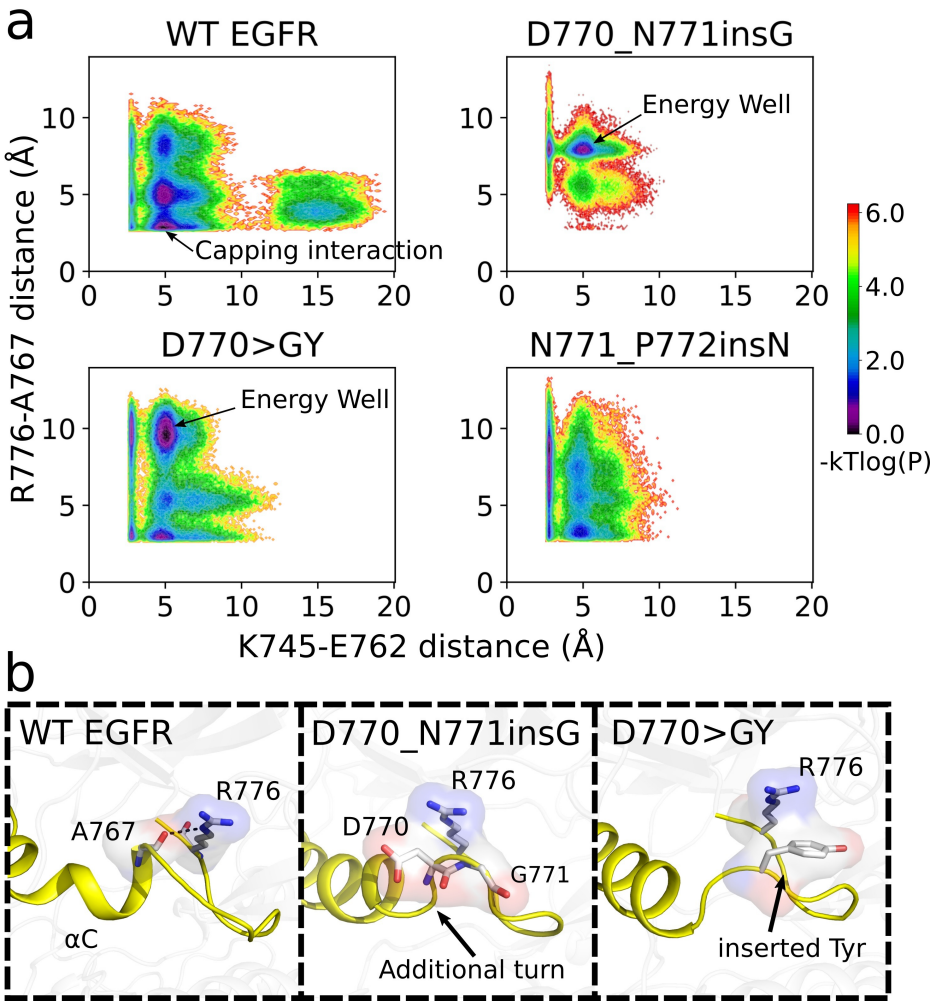

Figure 4.8: Modulation of auto-inhibitory capping interactions (R776/(NE,NH1,NH2)-A767/O hydrogen bonds) by activating insertion mutations. (a) The joint distribution of K745-E762 distance and R776-A767 distance is shown in logarithmic scale. The same binning and log-transformation procedure used in Figure 4.7a was followed to generate the figure. (b) Representative snapshots from the energy wells (indicated by arrows in (a).

4.2.6 Insertion mutations alter the free energy landscape of EGFR conformational transitions

We next wanted to investigate how insertion mutations in the αC-β4 loop alter the free energy landscape between active and inactive states. To this end, we employed an umbrella sampling technique, which allows efficient sampling of the entire transition event by restraining the simulation at various intermediate states. We used two collective variables (CVs) to describe the transition between active and inactive states (Figure 4.9a). CV1 is the pseudo-dihedral angle formed by the Cα atom of the DFG motif and the DFG+1 residue (D855, F856, G857, L858). The pseudo-dihedral angle describes the conformation of the activation loop and serves as an important conformational feature for distinguishing kinase active and inactive conformations [35]. CV2 is the difference in the distance between two salt bridge forming pairs, i.e., K745-E762 (d_1) and K860-E762 (d_2) (see Materials and Methods). The K745-E762 salt bridge is the catalytically important interaction formed in the active state of EGFR [9], whereas the K860-E762 salt bridge is only observed in the inactive state of EGFR [36]. The difference between d_1 and d_2 (CV2) describes the αC-helix movement during the conformational transition and is orthogonal to the DFG motif conformation captured by CV1. The initial transition pathway is generated by 4 independent steered molecular dynamics (SMD) simulations (Figure 4.9b). A total of 178 umbrella windows are chosen to cover the CV space explored by the SMD simulations. The 2D potential of mean force (PMF) profiles for WT and mutant EGFR are shown in (Figure 4.10a). The lower left corner of the PMF profile corresponds to the inactive conformation of EGFR, whereas the upper right corner of the 2D PMF profile corresponds to the active state. Umbrella sampling identified the inactive state to be the global energy minimum for WT EGFR and all three insertion

Table 4.1: The statistics of $\Delta G_{\text{active-inactive}}$ of Figure 4.10b.

Phenotype	$\Delta G_{\text{active-inactive}}$ (kJ mol^{-1})
WT EGFR	6.02±0.18
D770_N771insG	4.11±0.41
D770>GY	2.45±0.44
N771_P772insN	2.83±0.72

mutations, consistent with our unbiased MD simulations in which the inactive state was found to be more stable than the active state. Quantification of the free energy difference between active and inactive states along the 2D PMF shows that insertion mutations lower the $\Delta G_{active-inactive}$ compared to WT (Figure 4.10b and Table 4.1). We also identified the lowest free energy pathway (LFEP) connecting the active and inactive states in the free energy landscape (Figure 4.10a). The LFEP suggests stabilization of intermediate states along the transition pathway for the N771_P772insN and D770>GY mutants (Figure 4.10c). The three activating mutants also explore a unique inactive-like conformation with high frequencies in the 2D PMF (Figure 4.10a, red dashed frame). Structurally, this inactive state corresponds to a conformation in which the DFG-Phe (F856) is in an inward conformation, analogous to the inactive state of EGFR. However, the characteristic 3/10 helix right after the DFG motif is not formed (Figure 4.11). This inactive-like conformation appears to be readily accessed by the insertion mutations (D770_N771insG, D770>GY, and N771_P772insN) due to the interaction between the extended αC-β4 loop and the DFG motif in the activation segment (Figure 4.11). Notably, this inactive-like conformation is less predominant in WT EGFR, suggesting that it is uniquely explored by the activating insertion mutations.

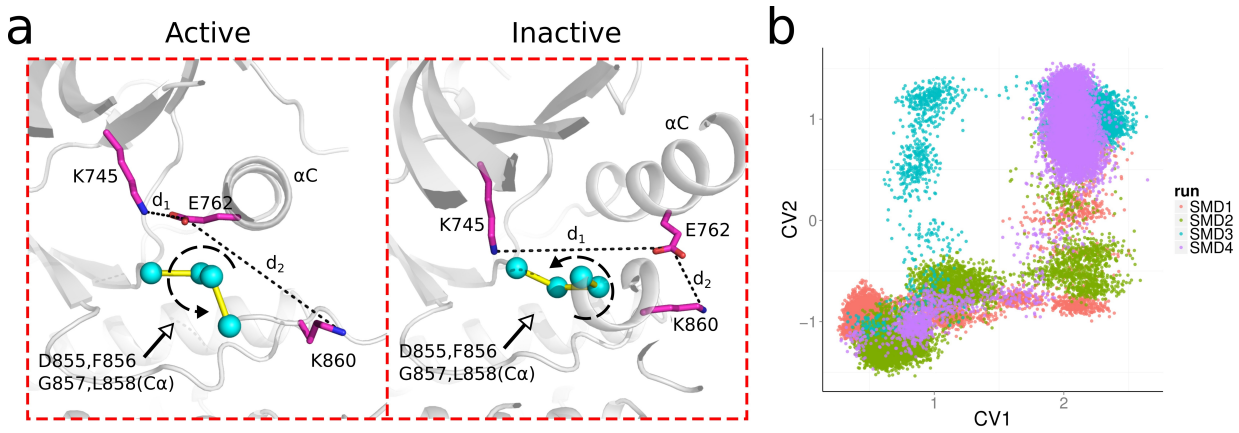

Figure 4.9: Reaction coordinates for US simulation. (a) The two collective variables used in our US simulation are shown in the active and inactive structures of EGFR. CV1: the dihedral angle of Cα[D855]-Cα[F856]-Cα[G857]-Cα[L858]. CV2: d_2[E762-K860]-d_1[K745-E762]. (b) Visualization of four independent steered MD simulation through CV1 and CV2.

4.3 Materials and Methods

4.3.1 Materials and Reagents

Q5 site-directed mutagenesis kit for generating mutant EGFR and DH5α competent cells were purchased from New England Biolabs (Ipswich, MA). Qiaprep spin miniprep/midiprep kits were obtained from Qiagen (Hilden, Germany). 1X Dulbecco's Modification of Eagle's Medium (DMEM), 1X Ham's F12 and Phosphate-Buffered Saline (PBS) were purchased from Mediatech (Manassas, VA). Lipofectamine-2000 was obtained from Invitrogen (Carlsbad, CA). Anti-pY1197-EGFR, Anti-pY869-EGFR, anti-GFP, anti-STAT3, anti-pY705-STAT3, anti-Akt, anti-pS473-Akt, and horseradish peroxidase (HRP) conjugated mouse monoclonal, and rabbit polyclonal antibodies were obtained from Cell Signaling (Danvers, MA). Anti-

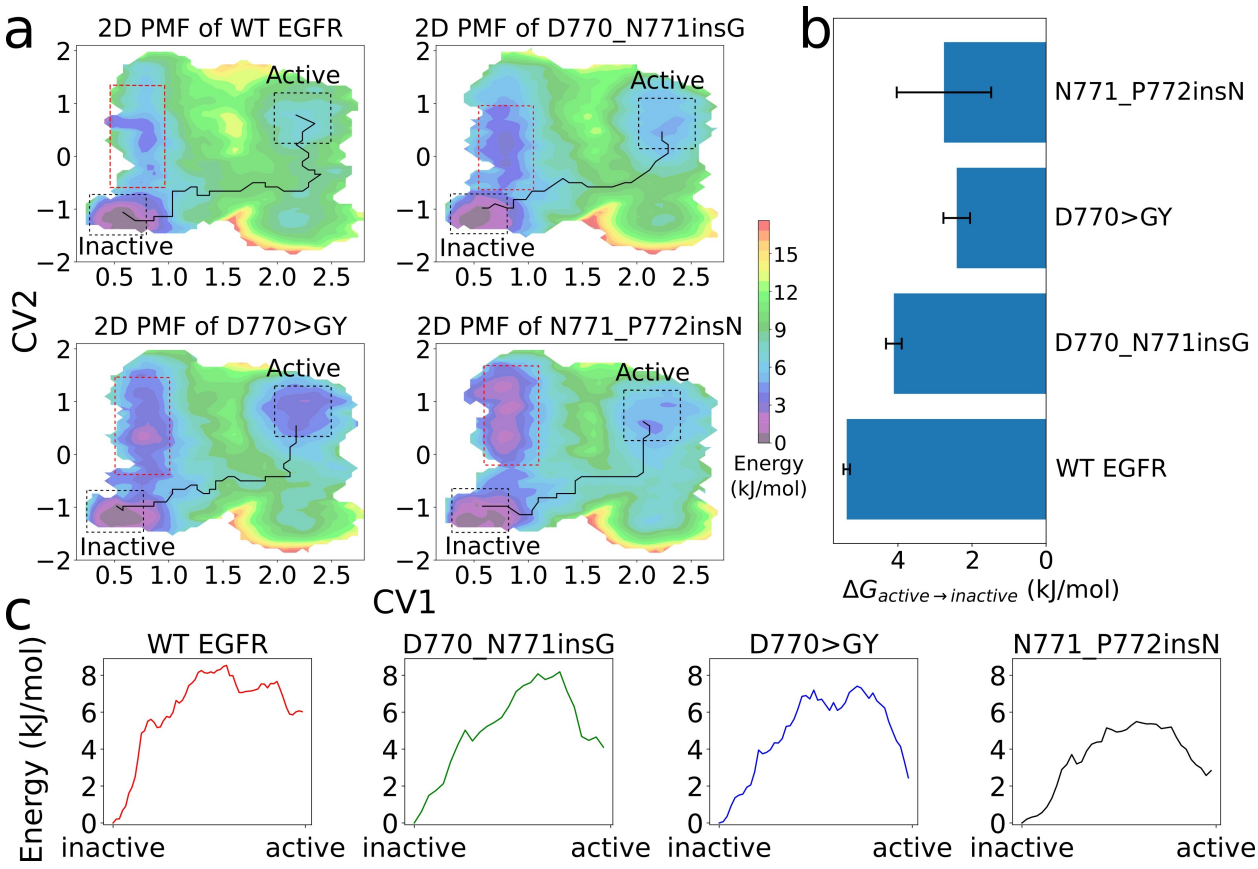

Figure 4.10: Free energy landscape of EGFR and insertion mutations from umbrella sampling. (a) 2D PMF of CV1 and CV2. The lowest free energy transition pathway (LFEP) connecting the active and inactive state is highlighted. Red dashed frames show the CV space that is stabilized by insertion mutations. (b) The $\Delta G_{active-inactive}$ between active and inactive states. The error bars are estimated from 3 independent US simulations. (c) The free energy associated with the lowest free energy pathway (LFEP) identified from 2D-PMF from (a) of various mutants.

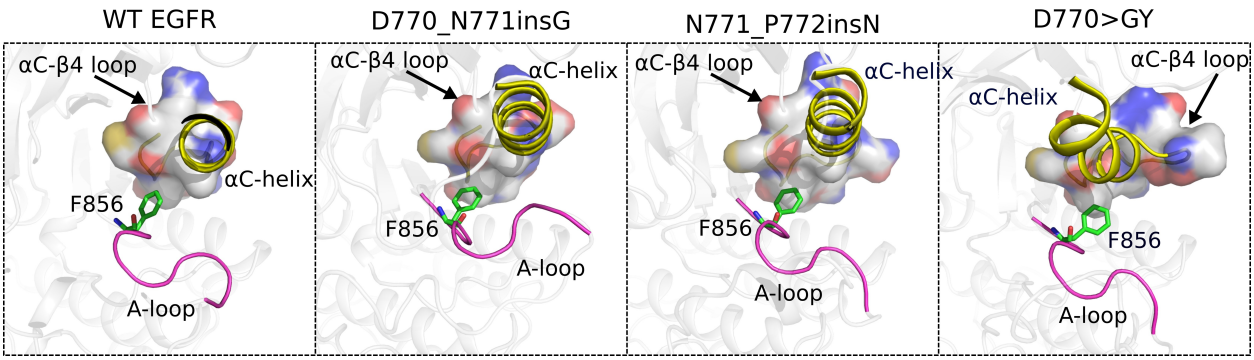

Figure 4.11: The alternative inactive state stabilized by activating αC-β4 loop insertion mutations.

FLAG human recombinant epidermal growth factor (EGF) and dacomitinib were bought from Sigma-Aldrich (St. Louis, MO). Bovine Serum Albumin (BSA) was purchased from Rockland Immunochemicals (Limerick, PA). Pierce ECL Western Blotting Substrate was bought from Thermo Scientific (Waltham, MA). Gefitinib, erlotinib, and lapatinib were kind gifts from Dr. Kojo Mensa-Wilmot at UGA Department of Cellular Biology. Osimertinib was purchased from LC Laboratories (Woburn, MA).

The pEGFP-N1-EGFR plasmid from our previous studies was used for mutagenesis [20, 37]. The PCR primers for site-directed mutagenesis are designed using the NEBaseChanger v1.2.6 web server from New England Biolabs (Ipswich, MA). All mutated plasmids were confirmed via Sanger sequencing through Eurofins Genomics (Louisville, KY).

4.3.2 Cell Culture, Transfection and Starvation

Chinese hamster ovary (CHO) cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum (FBS). DNA transfection was performed using Lipofectamine-

2000 by following manufacturer's protocol. After 12-16 h, the transfection efficacy was inspected under the microscope for GFP signal. Transfected CHO cells were washed with PBS and starved in Ham's 12 media for more than 12 h before EGF stimulation/drug treatment.

4.3.3 EGF Stimulation, Cell Lysis, and Immunoblotting

EGF stimulation was carried out by incubating serum starved CHO cells in 50 $\mathrm{ng\,ml^{-1}}$ human EGF for 5 min. Cells were washed and immediately lysed with lysis buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 10% glycerol, 1 mM EDTA, 10% Triton X-100, 1 mM PMSF, and 1xprotease inhibitor cocktail Set V, EDTA-free). The total cell lysate was spun at 15000 rpm at 4 °C for 5 min. Samples for SDS-PAGE gel were prepared in 4xLaemelli buffer. Proteins were resolved on 8% SDS-PAGE and transferred onto polyvinylidenedifluoride (PVDF) membrane using Trans-Blot Turbo Transfer System (Bio-Rad, Hercules, CA). Western blotting was done using target specific primary antibodies, followed by the rabbit/mouse secondary antibodies. Protein luminesce were detected by using Pierce Western Blotting ECL substrate (Rockford, IL).

4.3.4 Drug inhibition assay

Cells were transfected and then incubated with drugs at 5 different concentrations (0.001 µM, 0.01 µM, 0.1 µM, 1.0 µM, and 10.0 µM) in Ham's 12 media without serum. After 2 h drug treatment, the cells were processed as described above.

4.3.5 Image digitization and densitometry analysis

The exposed Western Blotting films were scanned using EPSON perfection 4490 photo scanner at 300 dpi resolution and saved in Portable Document Format (PDF). The resulting images were converted to black and white using Inkscape software without any additional means of post-processing. Densitometry analysis was performed using ImageJ software.

4.3.6 Structural Modeling

The PDB structures 2GS6 (residue 696-1019) and 3W32 (residue 701-1017) were used to model the active and inactive state of EGFR kinase domain, respectively [9, 36]. All crystallographic ligands, ions and water molecules were removed from the original structures. The two missing loops of 2GS6 (β3-αC loop and part of C-terminal tail) were fixed using the automodel class of MODELLER (version 9.12) [38]. We then introduced the αC-β4 loop insertions in the WT EGFR structure using the same protocol. The insertion loop of various mutants was then refined by robotics-inspired kinematic closure (KIC) algorithm implemented in Rosetta, in which 1000 loop structures were generated [39]. We clustered the structure ensemble of the insertion mutations based on the root mean square (RMS) similarity of the modeled αC-β4 loop. The analysis revealed 2 clusters for D770_N771insG and 3 clusters for D770>GY and N771_P772insN in the active state, and 4 clusters for D770_N771insG and 5 clusters for D770>GY and N771_P772insN in the inactive state. The cluster center of each cluster was selected for subsequent molecular dynamics simulations.

4.3.7 Molecular Dynamics (MD) Simulations

Unbiased MD simulation

Full-atom unbiased MD simulations was performed for each insertion mutation using GROMACS version 5.0.14 patched with plumed version 2.3.3 [40, 41]. Representative structural models from each cluster center were used as the starting conformation for MD. All hydrogen atoms were converted to virtual sites to remove the fastest vibrational freedom. The protein was parameterized with amber99sb-ildn force field and solvated with TIP3P water model in a dodecahedron box that was 1 nm larger than the protein in all directions [42]. Sodium and chloride ions were added to the system to neutralize the charge of the protein. The Verlet cutoff scheme was used to maintain the neighbor list for calculating non-bonded interactions [43]. Long-range electrostatics was calculated using particle mesh Ewald (PME) algorithms. Energy minimization was performed with the steepest descent algorithm followed by the conjugate-gradient algorithm until the $F_{\max}$ was less than 100 $\mathrm{kJ\,mol^{-1}}$. Canonical ensemble was carried out by heating the system from 0 to 310 K using velocity rescaling for 200 ps [44]. Isothermal-isobaric ensemble (P = 1 bar, T = 310 K) were carried out using the Berendsen barostat for 200 ps [45]. Position restraint was applied to non-hydrogen atoms of the protein during the equilibration steps. The unrestrained MD productions were collected using a time step of 5 fs after the isothermal-isobaric ensemble. The detailed simulation information can be found in Table 4.2. The trajectories were processed and analyzed using the built-in tools of Gromacs and plumed. Structural visualization was performed using PyMOL [46].

Steered MD (SMD) Simulation

Steered MD (SMD) is a technique to transform the system from an initial configuration to a target configuration by applying an external potential biased towards the target state. In our simulation, the external potential takes the form of a harmonic function of the instantaneous atom coordinates ($X[t]$) to the corresponding target atoms coordinates (X'): $V[t] = \frac{k}{2}(X[t] - X')^2$, where $V[t]$ is the external potential being added to the system Hamiltonian at each time step and k is the spring constant controlling the strength of the pulling force. Because the crystal structure of the active state of EGFR is several residues longer than the inactive state, we trimmed the active model of EGFR to match the inactive model so that the two structures contain identical sequences (residue 701-1017). The system preparation protocol for SMD simulation is similar to the MD simulation described in the previous subsection.

Table 4.2: Information of all-atom MD simulations performed in the study

Phenotype	State	Simulation Length
WT EGFR	Active	1μs×3
	Inactive	1μs×3
D770_N771insG	Active	1.5μs×2
	Inactive	1μs×4
D770>GY	Active	1μs×3
	Inactive	1μs×5
N771_P772insN	Active	1μs×3
	Inactive	1μs×5
L704N	Active	1μs×3
L760R	Active	1μs×3

In the production run, the plumed engine was enabled to calculate the collective variables (CVs) and bias the system potential [41]. A carefully chosen CV is important to obtain the reliable intermediate transition states connecting the two conformations [47]. In this study, we used two different CVs to perform SMD simulation, i.e. 1) the Cα atoms RMSD of the whole protein (residue 701-1017) and 2) the Cα atoms RMSD of αC-helix and activation loop (residue 756-769 and 855-863). A total of 4 independent SMD simulations were run by performing bidirectional transformations (active to inactive or inactive to active). The convergence of conformational transition was determined by monitoring the endpoint RMSD value to be less than 0.05 nm to the target state.

4.4 Conclusion and Discussion

Using a combination of computational and experimental approaches we find that patient-derived insertion mutations in the EGFR αC-β4 loop impact kinase autophosphorylation and drug sensitivity to varying degrees. We identify three gain-of-function mutations (D770_N771insG, D770>GY and N771_P772insN) that activate EGFR in a ligand-independent manner. Interestingly, these activating mutations rely more on the N-lobe asymmetric dimer interface than the C-lobe interface for kinase activation. Consistent with our findings, a previous study focusing on the dimerization dependence of the D770_N771insNPG mutant found that the N-lobe dimerization deficient mutation (L704N) had a more dramatic effect on the transforming activity of EGFR compared to C-lobe dimerization deficient mutation (I941R) [48]. These findings, however, cannot be reconciled based on the only available low resolution crystal structure of an exon 20 insertion mutation (D770_N771insNPG mutation; PDBID:

4LRM) in which the C-lobe of the "donor" kinase docks to the N-lobe of the "acceptor" in a manner similar to WT EGFR [15]. Thus, based on the asymmetric dimer model, it is unclear how activating insertion mutations in the αC-β4 loop can maintain EGFR activity in the presence of the C-lobe dimerization deficient mutant. Based on our cell-based mutagenesis experiment, we conclude that the asymmetric dimer is important for the full activation of EGFR insertion mutations. However, with a compromised C-lobe dimer interface, the mutants can still achieve activation through alternative configurations (Figure 4.3b).

The αC-β4 loop is a conserved regulatory segment present in all eukaryotic protein kinases (ePKs) [23]. Through quantitative comparisons of diverse eukaryotic protein kinases and distantly related small molecule kinases, we previously proposed that the HxN motif in the αC-β4 loop (NPH motif in EGFR) serves as a hinge point for αC-helix and interlobe movement [23]. While the length of the αC-β4 loop is typically 7-8 residues long in most protein kinases, some kinases such as CK2α, VRK1, VRK2, and SRPK1 [49, 50, 51] naturally contain an extended αC-β4 loop due to sequence insertions. Interestingly, the αC-helix in these kinases are stabilized in an active conformation [49, 50, 51], though the precise mechanisms by which such stabilization occurs are unclear. Our studies suggest that modulation of key auto-inhibitory interactions associated with the αC-helix movement, such as the R776-A767 capping interaction in EGFR, is one possible mechanism by which sequence and conformational variation in the αC-β4 loop may impact αC-helix conformational freedom and, consequently, kinase activity. The variable impact on EGFR auto-phosphorylation activity based on the nature, location and size of the inserted segment further supports the hypothesis that the αC-β4 loop is a rheostat that can be fine-tuned to regulate kinase activity [52]. For example, the two insertion mutations (N771_P772insN and N771_P772insH)

that only differ by a single residue at the NPH-Asn position have a dramatically different effect on kinase autophosphorylation. N771_P772insH abolishes Y1197 autophosphorylation, whereas N771_P772insN is constitutively active (Figure 4.2). Thus, subtle sequence and conformational variations in the αC-β4 loop may have long-range allosteric effects on ATP binding and activation loop conformation as proposed previously for other tyrosine kinases [21, 22, 53].

Our drug response assays indicate that activating insertion mutations are more sensitive to osimertinib compared to WT EGFR. This is consistent with a recently established high throughput functional evaluation assay in which the relative viability of Ba/F3 cells expressing N771_P772insN is reported to be sensitive to osimertinib treatment [19]. Another recently published study showed that D770>GY is partially responsive to dacomitinib compared to other insertion mutations in the αC-β4 loop [16]. This is in agreement with our experimental result in that the other two activation insertion mutations (D771_N772insG and N771_P772insN) show similar inhibition profile compared to WT EGFR. Our finding also suggests that non-reversible EGFR inhibitors, originally designed to overcome drug-resistant T790M mutant, may also be repurposed to treat patients with activating αC-β4 loop insertion mutations. Moreover, establishing a comprehensive drug response profile for αC-β4 loop insertion mutations to irreversible EGFR inhibitors is critical to guide therapeutical strategies for patient treatment.

The αC-β4 loop has recently been appreciated as an important region for Hsp90 association and co-chaperone cdc37 binding [24]. While EGFR is not a typical client protein for Hsp90, insertion mutations in the αC-β4 loop could alter this preference [54]. A previous study has shown that D770_N771insNPG and D770_N771insNPH mutants are sensitive

to Hsp90 inhibitors, opening another therapeutic window for clinical treatment [54]. In addition, our umbrella sampling analysis of WT and mutant EGFR identified alternative inactive states that are uniquely sampled by αC-β4 loop insertion mutants (Figure 4.10 and Figure 4.11). Thus selectively targeting these inactive states can be beneficial for treating patients harboring αC-β4 loop insertion mutations. Finally, by mechanistically characterizing oncogenic mutations in the αC-β4, our study provides a framework for investigating the underappreciated role of the αC-β4 loop in kinase structure, function, regulation and disease.

Bibliography

[1] Sreenath V Sharma, Daphne W Bell, Jeffrey Settleman, and Daniel A Haber. Epidermal growth factor receptor mutations in lung cancer. *Nature reviews. Cancer*, 7:169–81, 2007.

[2] Veronique Frattini, Vladimir Trifonov, Joseph Minhow Chan, Angelica Castano, Marie Lia, Francesco Abate, Stephen T Keir, Alan X Ji, Pietro Zoppoli, Francesco Niola, Carla Danussi, Igor Dolgalev, Paola Porrati, Serena Pellegatta, Adriana Heguy, Gaurav Gupta, David J Pisapia, Peter Canoll, Jeffrey N Bruce, Roger E McLendon, Hai Yan, Ken Aldape, Gaetano Finocchiaro, Tom Mikkelsen, Gilbert G PrivÃĬ, Darell D Bigner, Anna Lasorella, Raul Rabadan, and Antonio Iavarone. The integrated landscape of driver genomic alterations in glioblastoma. *Nature genetics*, 45:1141–9, 2013.

[3] Kathleen C Day, Guadalupe Lorenzatti Hiles, Molly Kozminsky, Scott J Dawsey, Alyssa Paul, Luke J Broses, Rajal Shah, Lakshmi P Kunja, Christopher Hall, Nallasivam

Palanisamy, Stephanie Daignault-Newton, Layla El-Sawy, Steven James Wilson, Andrew Chou, Kathleen Woods Ignatoski, Evan Keller, Dafydd Thomas, Sunitha Nagrath, Todd Morgan, and Mark L Day. Her2 and egfr overexpression support metastatic progression of prostate cancer to bone. *Cancer research*, 77:74–85, 2017.

[4] Hiroko Masuda, Dongwei Zhang, Chandra Bartholomeusz, Hiroyoshi Doihara, Gabriel N Hortobagyi, and Naoto T Ueno. Role of epidermal growth factor receptor in breast cancer. *Breast cancer research and treatment*, 136:331–45, 2012.

[5] A F Gazdar. Activating and resistance mutations of egfr in non-small-cell lung cancer: role in clinical response to egfr tyrosine kinase inhibitors. *Oncogene*, 28 Suppl 1:S24–31, 2009.

[6] Maria E Arcila, Khedoudja Nafa, Jamie E Chaft, Natasha Rekhtman, Christopher Lau, Boris A Reva, Maureen F Zakowski, Mark G Kris, and Marc Ladanyi. Egfr exon 20 insertion mutations in lung adenocarcinomas: prevalence, molecular heterogeneity, and clinicopathologic characteristics. *Molecular cancer therapeutics*, 12:220–9, 2013.

[7] David M Jackman, Beow Y Yeap, Lecia V Sequist, Neal Lindeman, Alison J Holmes, Victoria A Joshi, Daphne W Bell, Mark S Huberman, Balazs Halmos, Michael S Rabin, Daniel A Haber, Thomas J Lynch, Matthew Meyerson, Bruce E Johnson, and Pasi A Janne. Exon 19 deletion mutations of epidermal growth factor receptor are associated with prolonged survival in non-small cell lung cancer patients treated with gefitinib or erlotinib. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 12:3908–14, 2006.

[8] Chung-Jung Tsai and Ruth Nussinov. The molecular basis of targeting protein kinases in cancer therapeutics. *Seminars in cancer biology*, 23:235–42, 2013.

[9] Xuewu Zhang, Jodi Gureasko, Kui Shen, Philip A Cole, and John Kuriyan. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell*, 125:1137–49, 2006.

[10] Cai-Hong Yun, Titus J Boggon, Yiqun Li, Michele S Woo, Heidi Greulich, Matthew Meyerson, and Michael J Eck. Structures of lung cancer-derived egfr mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer cell*, 11:217–27, 2007.

[11] Yibing Shan, Michael P Eastwood, Xuewu Zhang, Eric T Kim, Anton Arkhipov, Ron O Dror, John Jumper, John Kuriyan, and David E Shaw. Oncogenic mutations counteract intrinsic disorder in the egfr kinase and promote receptor dimerization. *Cell*, 149:860–70, 2012.

[12] Cai-Hong Yun, Kristen E Mengwasser, Angela V Toms, Michele S Woo, Heidi Greulich, Kwok-Kin Wong, Matthew Meyerson, and Michael J Eck. The t790m mutation in egfr kinase causes drug resistance by increasing the affinity for atp. *Proceedings of the National Academy of Sciences of the United States of America*, 105:2070–5, 2008.

[13] Scott A Foster, Daniel M Whalen, Aysegul Ozen, Matthew J Wongchenko, JianPing Yin, Ivana Yen, Gabriele Schaefer, John D Mayfield, Juliann Chmielecki, Philip J Stephens, Lee A Albacker, Yibing Yan, Kyung Song, Georgia Hatzivassiliou, Charles Eigenbrot, Christine Yu, Andrey S Shaw, Gerard Manning, Nicholas J Skelton, Sarah G Hymowitz,

and Shiva Malek. Activation mechanism of oncogenic deletion mutations in braf, egfr, and her2. *Cancer cell*, 29:477–493, 2016.

[14] Hiroyuki Yasuda, Susumu Kobayashi, and Daniel B Costa. Egfr exon 20 insertion mutations in non-small-cell lung cancer: preclinical data and clinical implications. *The Lancet. Oncology*, 13:e23–31, 2012.

[15] Hiroyuki Yasuda, Eunyoung Park, Cai-Hong Yun, Natasha J Sng, Antonio R Lucena-Araujo, Wee-Lee Yeo, Mark S Huberman, David W Cohen, Sohei Nakayama, Kota Ishioka, Norihiro Yamaguchi, Megan Hanna, Geoffrey R Oxnard, Christopher S Lathan, Teresa Moran, Lecia V Sequist, Jamie E Chaft, Gregory J Riely, Maria E Arcila, Ross A Soo, Matthew Meyerson, Michael J Eck, Susumu S Kobayashi, and Daniel B Costa. Structural, biochemical, and clinical characterization of epidermal growth factor receptor (egfr) exon 20 insertion mutations in lung cancer. *Science translational medicine*, 5:216ra177, 2013.

[16] Takayuki Kosaka, Junko Tanizaki, Raymond M Paranal, Hideki Endoh, Christine Lydon, Marzia Capelletti, Claire E Repellin, Jihyun Choi, Atsuko Ogino, Antonio Calles, Dalia Ercan, Amanda J Redig, Magda Bahcall, Geoffrey R Oxnard, Michael J Eck, and Pasi A JÃd̈nne. Response heterogeneity of egfr and her2 exon 20 insertions to covalent egfr and her2 inhibitors. *Cancer research*, 77:2712–2721, 2017.

[17] Mai He, Marzia Capelletti, Khedoudja Nafa, Cai-Hong Yun, Maria E Arcila, Vincent A Miller, Michelle S Ginsberg, Binsheng Zhao, Mark G Kris, Michael J Eck, Pasi A Janne, Marc Ladanyi, and Geoffrey R Oxnard. Egfr exon 19 insertions: a new family of

sensitizing egfr mutations in lung adenocarcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 18:1790–7, 2012.

[18] J Naidoo, C S Sima, K Rodriguez, N Busby, K Nafa, M Ladanyi, G J Riely, M G Kris, M E Arcila, and H A Yu. Epidermal growth factor receptor exon 20 insertions in advanced lung adenocarcinomas: Clinical outcomes and response to erlotinib. *Cancer*, 121:3212–3220, 2015.

[19] Shinji Kohsaka, Masaaki Nagano, Toshihide Ueno, Yoshiyuki Suehara, Takuo Hayashi, Naoko Shimada, Kazuhisa Takahashi, Kenji Suzuki, Kazuya Takamochi, Fumiyuki Takahashi, and Hiroyuki Mano. A method of high-throughput functional evaluation of egfr gene variants of unknown significance in cancer. *Science translational medicine*, 9, 2017.

[20] Zheng Ruan and Natarajan Kannan. Mechanistic insights into r776h mediated activation of epidermal growth factor receptor kinase. *Biochemistry*, 54:4216–25, 2015.

[21] Huaibin Chen, Jinghong Ma, Wanqing Li, Anna V Eliseenkova, Chongfeng Xu, Thomas A Neubert, W Todd Miller, and Moosa Mohammadi. A molecular brake in the kinase hinge region regulates the activity of receptor tyrosine kinases. *Molecular cell*, 27:717–30, 2007.

[22] Tobias Klein, Navratna Vajpai, Jonathan J Phillips, Gareth Davies, Geoffrey A Holdgate, Chris Phillips, Julie A Tucker, Richard A Norman, Andrew D Scott, Daniel R Higazi, David Lowe, Gary S Thompson, and Alexander L Breeze. Structural and dynamic insights into the energetics of activation loop rearrangement in fgfr1 kinase. *Nature communications*, 6:7877, 2015.

[23] Natarajan Kannan and Andrew F Neuwald. Did protein kinase regulatory mechanisms evolve through elaboration of a simple structural component? *Journal of molecular biology*, 351:956–72, 2005.

[24] Kliment A Verba, Ray Yu-Ruei Wang, Akihiko Arakawa, Yanxin Liu, Mikako Shirouzu, Shigeyuki Yokoyama, and David A Agard. Atomic structure of hsp90-cdc37-cdk4 reveals that hsp90 traps and stabilizes an unfolded kinase. *Science (New York, N.Y.)*, 352:1542–7, 2016.

[25] Wanping Xu, Xitong Yuan, Zhexin Xiang, Edward Mimnaugh, Monica Marcu, and Len Neckers. Surface charge and hydrophobicity determine erbb2 binding to the hsp90 chaperone complex. *Nature structural & molecular biology*, 12:120–6, 2005.

[26] Gurinder Gosal, Krys J Kochut, and Natarajan Kannan. Prokino: an ontology for integrative analysis of protein kinases in cancer. *PloS one*, 6:e28782, 2011.

[27] Daniel Ian McSkimming, Shima Dastgheib, Eric Talevich, Anish Narayanan, Samiksha Katiyar, Susan S Taylor, Krys Kochut, and Natarajan Kannan. Prokino: a unified resource for mining the cancer kinome. *Human mutation*, 36:175–86, 2015.

[28] Katerina Politi and Thomas J Lynch. Two sides of the same coin: Egfr exon 19 deletions and insertions in lung cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 18:1490–2, 2012.

[29] J T den Dunnen and S E Antonarakis. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum. Mutat.*, 15:7–12, 2000.

[30] Monica Red Brewer, Cai-Hong Yun, Darson Lai, Mark A Lemmon, Michael J Eck, and William Pao. Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 110:E3595–604, 2013.

[31] Darren A E Cross, Susan E Ashton, Serban Ghiorghiu, Cath Eberlein, Caroline A Nebhan, Paula J Spitzler, Jonathon P Orme, M Raymond V Finlay, Richard A Ward, Martine J Mellor, Gareth Hughes, Amar Rahi, Vivien N Jacobs, Monica Red Brewer, Eiki Ichihara, Jing Sun, Hailing Jin, Peter Ballard, Katherine Al-Kadhimi, Rachel Rowlinson, Teresa Klinowska, Graham H P Richmond, Mireille Cantarini, Dong-Wan Kim, Malcolm R Ranson, and William Pao. Azd9291, an irreversible egfr tki, overcomes t790m-mediated resistance to egfr inhibitors in lung cancer. *Cancer discovery*, 4:1046–61, 2014.

[32] Andrea J Gonzales, Kenneth E Hook, Irene W Althaus, Paul A Ellis, Erin Trachet, Amy M Delaney, Patricia J Harvey, Teresa A Ellis, Danielle M Amato, James M Nelson, David W Fry, Tong Zhu, Cho-Ming Loi, Stephen A Fakhoury, Kevin M Schlosser, Karen E Sexton, R Thomas Winters, Jessica E Reed, Alex J Bridges, Daniel J Lettiere, Deborah A Baker, Jianxin Yang, Helen T Lee, Haile Tecle, and Patrick W Vincent. Antitumor activity and pharmacokinetic properties of pf-00299804, a second-generation irreversible pan-erbb receptor tyrosine kinase inhibitor. *Molecular cancer therapeutics*, 7:1880–9, 2008.

[33] Natalie Minkovsky and Alan Berezov. Bibw-2992, a dual receptor tyrosine kinase in-

hibitor for the treatment of solid tumors. *Current opinion in investigational drugs (London, England : 2000)*, 9:1336–46, 2008.

[34] Yuliana Yosaatmadja, Shevan Silva, James M Dickson, Adam V Patterson, Jeff B Smaill, Jack U Flanagan, Mark J McKeage, and Christopher J Squire. Binding mode of the breakthrough inhibitor azd9291 to epidermal growth factor receptor revealed. *Journal of structural biology*, 192:539–544, 2015.

[35] Daniel Ian McSkimming, Khaled Rasheed, and Natarajan Kannan. Classifying kinase conformations using a machine learning approach. *BMC bioinformatics*, 18:86, 2017.

[36] Youichi Kawakita, Masaki Seto, Tomohiro Ohashi, Toshiya Tamura, Tadashi Yusa, Hiroshi Miki, Hidehisa Iwata, Hidenori Kamiguchi, Toshimasa Tanaka, Satoshi Sogabe, Yoshikazu Ohta, and Tomoyasu Ishikawa. Design and synthesis of novel pyrimido[4,5-b]azepine derivatives as her2/egfr dual inhibitors. *Bioorganic & medicinal chemistry*, 21:2250–2261, 2013.

[37] Zheng Ruan, Samiksha Katiyar, and Natarajan Kannan. Computational and experimental characterization of patient derived mutations reveal an unusual mode of regulatory spine assembly and drug sensitivity in egfr kinase. *Biochemistry*, 56:22–32, 2017.

[38] Narayanan Eswar, Ben Webb, Marc A Marti-Renom, M S Madhusudhan, David Eramian, Min-Yi Shen, Ursula Pieper, and Andrej Sali. Comparative protein structure modeling using modeller. *Current protocols in bioinformatics*, Chapter 5:Unit–5.6, 2006.

[39] Daniel J Mandell, Evangelos A Coutsias, and Tanja Kortemme. Sub-angstrom accuracy

in protein loop reconstruction by robotics-inspired conformational sampling. *Nature methods*, 6:551–2, 2009.

[40] Mark James Abraham, Teemu Murtola, Roland Schulz, SzilÃąrd PÃąll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19 – 25, 2015.

[41] Massimiliano Bonomi, Davide Branduardi, Giovanni Bussi, Carlo Camilloni, Davide Provasi, Paolo Raiteri, Davide Donadio, Fabrizio Marinelli, Fabio Pietrucci, Ricardo A. Broglia, and Michele Parrinello. Plumed: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications*, 180(10):1961 – 1972, 2009.

[42] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins*, 78:1950–8, 2010.

[43] Szilard Pall and Berk Hess. A flexible algorithm for calculating pair interactions on simd architectures. *Comput. Phys. Commun.*, 184:2641–2650, 2013.

[44] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126:014101, 2007.

[45] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, DiNola A., and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.

[46] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

[47] Albert C Pan, Thomas M Weinreich, Yibing Shan, Daniele P Scarpazza, and David E Shaw. Assessing the accuracy of two enhanced sampling methods using egfr kinase transition pathways: The influence of collective variable choice. *Journal of chemical theory and computation*, 10:2860–5, 2014.

[48] Jeonghee Cho, Liang Chen, Naveen Sangji, Takafumi Okabe, Kimio Yonesaka, Joshua M Francis, Richard J Flavin, William Johnson, Jihyun Kwon, Soyoung Yu, Heidi Greulich, Bruce E Johnson, Michael J Eck, Pasi A JÃďnne, Kwok-Kin Wong, and Matthew Meyerson. Cetuximab response of lung cancer-derived egf receptor mutants is associated with asymmetric dimerization. *Cancer research*, 73:6770–9, 2013.

[49] Ewa Sajnaga, Konrad KubiÅĎski, and Ryszard Szyszka. Catalytic activity of mutants of yeast protein kinase ck2alpha. *Acta biochimica Polonica*, 55:767–76, 2008.

[50] Eric D Scheeff, Jeyanthy Eswaran, Gabor Bunkoczi, Stefan Knapp, and Gerard Manning. Structure of the pseudokinase vrk3 reveals a degraded catalytic site, a highly conserved kinase fold, and a putative regulatory binding site. *Structure (London, England : 1993)*, 17:128–38, 2009.

[51] Jacky Chi Ki Ngo, Justin Gullingsrud, Kayla Giang, Melinda Jean Yeh, Xiang-Dong Fu, Joseph A Adams, J Andrew McCammon, and Gourisankar Ghosh. Sr protein kinase 1 is resilient to inactivation. *Structure (London, England : 1993)*, 15:123–33, 2007.

[52] Sarah Meinhardt, Michael W Manley, Daniel J Parente, and Liskin Swint-Kruse.

Rheostats and toggle switches for modulating protein function. *PloS one*, 8:e83502, 2013.

[53] Chung-Jung Tsai and Ruth Nussinov. The free energy landscape in translational science: how can somatic mutations result in constitutive oncogenic activation? *Physical chemistry chemical physics : PCCP*, 16:6332–41, 2014.

[54] W Xu, S Soga, K Beebe, M-J Lee, Y S Kim, J Trepel, and L Neckers. Sensitivity of epidermal growth factor receptor and erbb2 exon 20 insertion mutants to hsp90 inhibition. *British journal of cancer*, 97:741–4, 2007.

Chapter 5

Discussion and conclusion remarks

I have carried out both experimental and computational experiments to address each of the research questions stated at the beginning of this dissertation. The novel contribution from this study is the systematic characterization of patient-derived cancer mutations located in the regulatory structural motifs of EGFR kinase domain. By productive integration of computational and experimental approaches, I establish the dimerization dependency, drug response, and structural implications of the activating EGFR mutations. The mechanistic insight emerged from this study deepens our understanding of EGFR biology and provides a framework to understand other cancer mutations as well.

5.1 Achievement of goals and broader impact

5.1.1 R776H mediated activation of EGFR

The characterization of R776H mutation reveals a previously overlooked "lateral" phosphorylation mechanism that separates the kinase function of EGFR asymmetric dimer from the substrate function of monomeric EGFR. The "lateral" phosphorylation is significantly enhanced by oncogenic mutations, highlighting the importance of pathogenic EGFR in the oligomeric state. The mutagenesis strategy of my study has been adopted by another research group to investigate the mechanisms of EGFR multimerization [1]. My experimental result also suggests that R776H mutant of EGFR prefers to adopt the "acceptor" position when paired with WT EGFR in the asymmetric dimer, supporting the previously established "superacceptor" model for oncogenic EGFR mutations. MD simulations of WT and R776H uncover several auto-inhibitory interactions that are disrupted by oncogenic mutations. After the publication of the result, the capping interaction between R776 and αC-helix of EGFR has also been hypothesized to be important in understanding mutation impact and drug response of exon 20 insertion mutations [2].

5.1.2 Novel mode of R-spine assembly

The regulatory spine (RS) of protein kinases was initially identified as a hydrophobic organizing center for the active state transition of protein kinases. My mutational analysis of patient-derived R-spine mutations in EGFR finds that M766T at the RS3 position unexpectedly elevates the autophosphorylation activity. Detailed experimental and computational analysis of mutations at RS3 position shows that T766 could form hydrogen bonds

with E762 and the backbone of DFG-motif through a conserved water-mediated interaction. The result is surprising because it shows a novel mode of R-spine assembly in which the hydrophobic packing seen in the WT EGFR is replaced with hydrogen bonding interactions in pathologic EGFR. Such a novel RS arrangement is the underlying basis for the drug resistance effect conferred by the M766T mutant. Soon after this work got published in *Biochemistry*, another group of scientists proposed a similar water-mediated allosteric network in Aurora A kinase [3], suggesting that nature has already explored the alternative R-spine assembly for kinase regulation and activation.

5.1.3 Mechanistic insight into αC-β4 loop insertion mutations of EGFR

My work on characterizing EGFR αC-β4 loop insertion mutations shows that the precise location, sequence, and size of the inserted segment is critical for kinase activity. This is important because previous studies don't explicitly separate activating insertion mutations from other non-activating insertion mutations. In addition, I find that insertion mutations alter the dimerization dependency of EGFR, which is distinct from WT EGFR and many previously characterized point mutations. Through microsecond time-scale MD simulation, I show that insertion mutations activate EGFR by preventing the R776-A767 capping interaction identified in the second chapter. I also establish the free energy landscape (FEL) of activating insertion mutations through umbrella sampling. The FEL clearly indicates that insertion mutations change the relative probability of adopting various conformations of EGFR kinase domain. In particular, the analysis finds novel states that are uniquely stabilized by insertion mutations, but not by WT EGFR. Such states could be exploited for future drug discovery efforts to design selective EGFR inhibitors.

My analysis on both R776H point mutation and EGFR insertion mutations highlight that fact that the conserved αC-β4 loop plays an important role in EGFR kinase function and disease. This is in line with several recent studies of the αC-β4 loop on other kinases, which revealed remarkable insight of kinase conformational regulation and chaperone-assisted folding of the protein [4, 5, 6]. The mechanistic insight emerged from one kinase might shed light on understanding the function of other kinases due to the conservation of their sequence and structure. To condense the knowledge of these recent discoveries, I'm preparing a review manuscript focusing on the various mode of kinase regulation by the αC-β4 loop.

5.2 Future directions

The work and pipeline established in this dissertation provide the basis for future investigation of mutation-mediated kinase activation and conformational landscape of protein kinases. A few possible future directions are discussed below.

5.2.1 Characterization of exon 19 insertion mutations of EGFR

EGFR exon 19 insertions are in-frame duplications/insertions of 5-6 residues occurring within exon 19 of EGFR [7]. Structurally, the insertions map to the β3 strand of the protein kinase domain. These mutations have been shown to transform Ba/F3 cells and are sensitive to EGFR-TKIs, such as gefitinib and afatinib [7]. However, the structural mechanisms associated with the mutation are poorly understood. The β3 strand insertion mutations have been identified in chapter 4, but I decide to focus on the more prevalent αC-β4 mutations in the dissertation. Detailed biochemical characterization of exon 19 insertion mutations is

also important for understanding the inherent malleability of the kinase domain as well as clinical treatment.

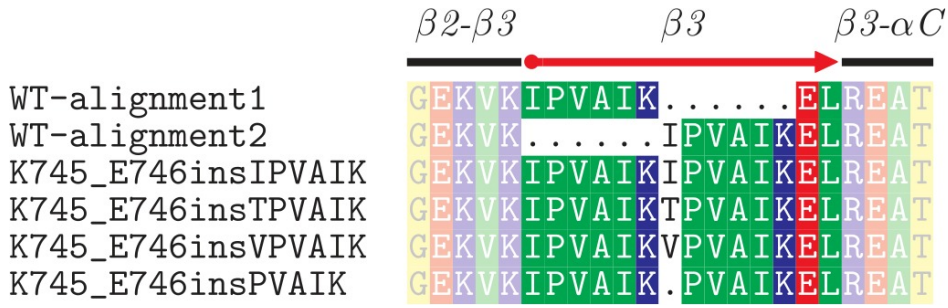

Figure 5.1: Alignment of EGFR exon 19 insertion mutations. WT and four exon 19 insertion mutations in the COSMIC database are aligned. Two possible ways of WT sequence alignment are shown at the top.

The β3 strand of protein kinases contains a highly conserved ATP-coordinating lysine residue (K745 in EGFR). The formation of the K745-E762 salt bridge is necessary for EGFR kinase in the active state [8, 9]. All of the exon 19 insertion mutations result in a duplication of the conserved lysine residue in EGFR (K745_E746insIPVAIK, K745_E746insTPVAIK, or K745_E746insPVAIK). Therefore, depending on which lysine residue makes the salt bridge interaction with αC-helix-Glu (E762), the kinase domain will either has an extended β2-β3 loop or αC-β3 loop in the structure. Previous studies have shown that mutations breaking the formation of the K745-E762 salt bridge will inactivate the kinase [10, 11]. Based on this knowledge, we can probe the structural alterations conferred by exon 19 insertion mutations by mutating each lysine residue of the exon 19 insertion mutations and detect the autophosphorylation activity of the mutants. My preliminary result has shown that when the first lysine is mutated to methionine, the kinase is still active (Figure 5.2). However, if the second lysine is mutated to methionine, the kinase activity is abrogated (Figure 5.2),

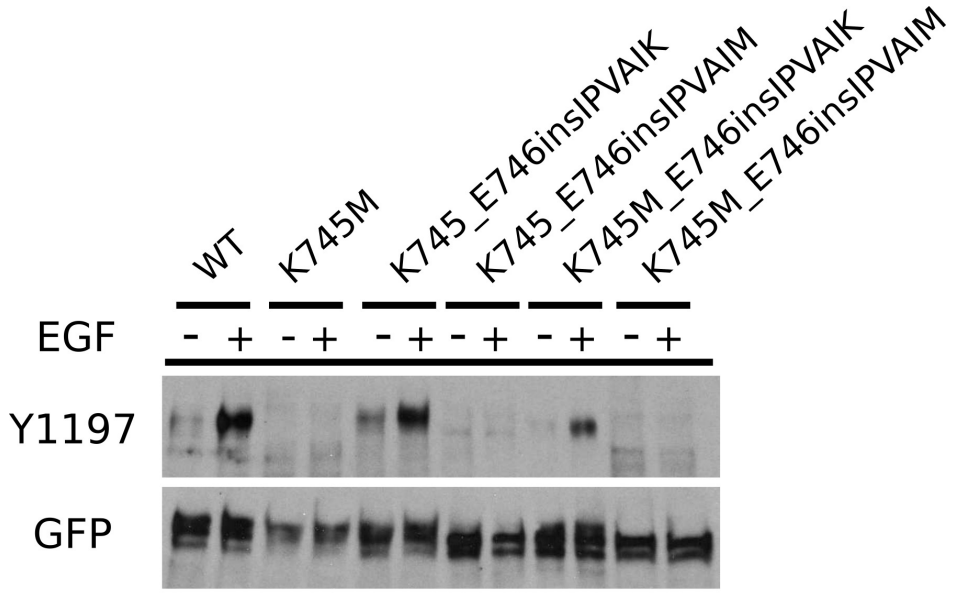

Figure 5.2: β3-lysine mutation of exon 19 insertion mutation. Lanes from left to right: WT, K745_E746insIPVAIK, K745_E746insIPVAIM, K745M_E746insIPVAIK, K745M_E746insIPVAIM. − and + indicate the absence or presence of EGF treatment. The autophosphorylation of Y1197 and total expression (indicated by GFP) are probed by target specific antibodies.

suggesting that the second lysine residue of the β3 strand in exon 19 insertion mutations forms the catalytically important salt bridge. Based on this information, I can infer that exon 19 insertion mutations will result in an extended β2-β3 loop.

Although the structural alterations mediated by exon 19 insertion mutations have been established by my preliminary data, the functional significance of the recurrent mutations is still less well understood. In particular, exon 19 insertion mutations don't seem to significantly elevate the kinase activity of EGFR (Figure 5.2). Therefore, the selection advantage of fixing such complex insertion mutations during tumor progression remain elusive. Fu-

ture work in establishing the dimerization dependency, drug response, and conformational landscape profile of these group of mutations could provide novel insight into EGFR in the disease state. Moreover, the extended β2-β3 loop lies at the interface of the inactive dimer of EGFR [12]. How exon 19 insertion mutations modulate the inactive dimer formation can be investigated through structural modeling, computational docking, and MD simulations.

5.2.2 Establishing the free energy landscape (FEL) by incorporating other conformations seen in the crystal structures of EGFR

The structural study of EGFR in this dissertation assumes a two state transition model, in which the kinase domain of EGFR can only adopt the "active" state or another Src-like "inactive" state. Although the Src-like "inactive" state is the most prevalent conformation of EGFR kinase domain seen in PDB database, recent structural studies have uncovered several novel "inactive" states that are also accessible by EGFR. Specifically, several DFG-"out" inactive states of EGFR has been crystallized with various kinase inhibitors [13, 14, 15]. Therefore, an obvious extension to the current research is to include additional conformational states of EGFR and perform a more exhaustive conformational sampling. The umbrella sampling technique employed in chapter 4 is well suited for this purpose.

To clearly delineating different conformational states of EGFR kinase domain, a set of carefully chosen collective variables (CVs) is important in obtaining meaningful result [16]. Knowledge-based variables such as the αC-helix or activation loop RMSD could be used for this purpose, but these CVs only describe the local dynamics of the kinase domain and may omit important degrees of freedom during the conformational transition. Another more objective option is to derive the CVs based on the principal component (PC) or time-lagged

independent component (tiCA) [17, 18]. These methods aim to identify a linear combination of the coordinates/torsion angle of the MD trajectory that captures the largest conformational variation or slow reaction coordinates. Since I have already performed extensive MD simulation on WT EGFR, the information can be utilized to extract CVs and guide future sampling of the mutant form [19]. Once the CVs and intermediate states along the CVs are selected, the FEL incorporating the other "inactive" structures of EGFR kinase domain can be established by standard umbrella sampling. Moreover, the FEL for mutant EGFR can also be easily obtained by modeling the intermediate states with the mutation and perform US simulations by following the identical protocol.

5.2.3 Accurate prediction of the drug response of EGFR oncogenic mutations

MD-based free energy prediction has been widely used in the virtual screening and drug discovery effort [20, 21, 22, 23]. However, the prediction of accurate drug binding free energy in protein kinases is especially challenging due to several factors. First, many kinase inhibitors dedicate to occupy the ATP-binding pocket and are thus ATP-competitive inhibitors [24]. Therefore, the drug response profile of a mutant depends on the binding affinity of both the kinase inhibitor and ATP molecule [24, 25], which is not explicitly considered in many previous studies [26, 27]. Second, the remarkable conformational plasticity of protein kinase domain introduces another layer of complexity [28]. In WT EGFR, the Src-like "inactive" state is the global free energy minimum and the majority of the kinase population will adopt such a state in the monomeric form (Figure 4.10). If an inhibitor selectively targets the "active" state of EGFR, then there is an energy penalty associated with the conformational

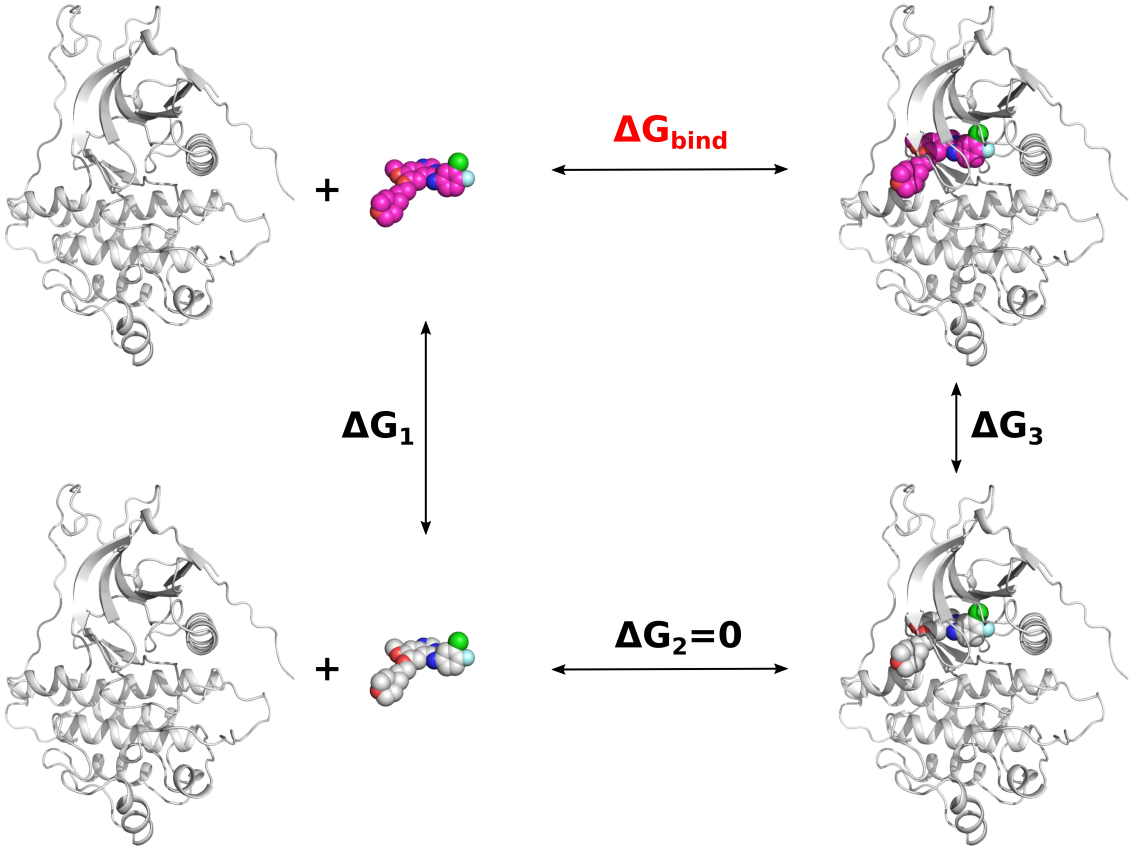

Figure 5.3: Thermodynamic cycle of binding free energy calculation. A meganta ligand interacts with the environment whereas a gray ligand is not. The binding free energy $\Delta G_{bind} = \Delta G_3 - \Delta G_1$.

transition. Consequently, mutations in the kinase domain may alter the free energy landscape and indirectly affects the drug binding [29, 30].

The first challenge can be overcome by performing thermodynamic integration (TI) on both kinase inhibitor and ATP. By decoupling the ligand from the kinase through a thermodynamic cycle (Figure 5.3), the relative binding free energy can be obtained through a series of λ-coupled simulations, where the free energy estimate is $\Delta G = \int_0^1 \langle \frac{dU(\lambda,\vec{q})}{d\lambda} \rangle_\lambda d\lambda$ [31, 32].

The difference between the inhibitor and ATP binding free energy determines the relative drug sensitivity of the mutant. To address the second challenge, the established FEL in the previous section could be utilized to analyze the FEL of EGFR mutants. Therefore, the conformational energy difference between the mutant and WT EGFR is also taken into consideration. This strategy will not only allow me to achieve accurate drug response prediction but also track the origin of the changes in drug binding energy.

Bibliography

[1] Yongjian Huang, Shashank Bharill, Deepti Karandur, Sean M Peterson, Morgan Marita, Xiaojun Shi, Megan J Kaliszewski, Adam W Smith, Ehud Y Isacoff, and John Kuriyan. Molecular basis for multimerization in the activation of the epidermal growth factor receptor. *eLife*, 5, 2016.

[2] Takayuki Kosaka, Junko Tanizaki, Raymond M Paranal, Hideki Endoh, Christine Lydon, Marzia Capelletti, Claire E Repellin, Jihyun Choi, Atsuko Ogino, Antonio Calles, Dalia Ercan, Amanda J Redig, Magda Bahcall, Geoffrey R Oxnard, Michael J Eck, and Pasi A JÃďnne. Response heterogeneity of egfr and her2 exon 20 insertions to covalent egfr and her2 inhibitors. *Cancer research*, 77:2712–2721, 2017.

[3] Soreen Cyphers, Emily F Ruff, Julie M Behr, John D Chodera, and Nicholas M Levinson. A water-mediated allosteric network governs activation of aurora kinase a. *Nature chemical biology*, 13:402–408, 2017.

[4] Huaibin Chen, Jinghong Ma, Wanqing Li, Anna V Eliseenkova, Chongfeng Xu,

Thomas A Neubert, W Todd Miller, and Moosa Mohammadi. A molecular brake in the kinase hinge region regulates the activity of receptor tyrosine kinases. *Molecular cell*, 27:717–30, 2007.

[5] Kliment A Verba, Ray Yu-Ruei Wang, Akihiko Arakawa, Yanxin Liu, Mikako Shirouzu, Shigeyuki Yokoyama, and David A Agard. Atomic structure of hsp90-cdc37-cdk4 reveals that hsp90 traps and stabilizes an unfolded kinase. *Science (New York, N.Y.)*, 352:1542–7, 2016.

[6] Tobias Klein, Navratna Vajpai, Jonathan J Phillips, Gareth Davies, Geoffrey A Holdgate, Chris Phillips, Julie A Tucker, Richard A Norman, Andrew D Scott, Daniel R Higazi, David Lowe, Gary S Thompson, and Alexander L Breeze. Structural and dynamic insights into the energetics of activation loop rearrangement in fgfr1 kinase. *Nature communications*, 6:7877, 2015.

[7] Mai He, Marzia Capelletti, Khedoudja Nafa, Cai-Hong Yun, Maria E Arcila, Vincent A Miller, Michelle S Ginsberg, Binsheng Zhao, Mark G Kris, Michael J Eck, Pasi A Janne, Marc Ladanyi, and Geoffrey R Oxnard. Egfr exon 19 insertions: a new family of sensitizing egfr mutations in lung adenocarcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 18:1790–7, 2012.

[8] Brad Nolen, Susan Taylor, and Gourisankar Ghosh. Regulation of protein kinases; controlling activity through activation segment conformation. *Molecular cell*, 15:661–75, 2004.

[9] J A Adams. Kinetic and catalytic mechanisms of protein kinases. *Chemical reviews*, 101:2271–90, 2001.

[10] Monica Red Brewer, Sung Hee Choi, Diego Alvarado, Katarina Moravcevic, Ambra Pozzi, Mark A Lemmon, and Graham Carpenter. The juxtamembrane region of the egf receptor functions as an activation domain. *Molecular cell*, 34:641–51, 2009.

[11] T B Deb, L Su, L Wong, E Bonvini, A Wells, M David, and G R Johnson. Epidermal growth factor (egf) receptor kinase-independent signaling by egf. *The Journal of biological chemistry*, 276:15554–60, 2001.

[12] Natalia Jura, Nicholas F Endres, Kate Engel, Sebastian Deindl, Rahul Das, Meindert H Lamers, David E Wemmer, Xuewu Zhang, and John Kuriyan. Mechanism for activation of the egf receptor catalytic domain by the juxtamembrane segment. *Cell*, 137:1293–307, 2009.

[13] Simon Planken, Douglas C Behenna, Sajiv K Nair, Theodore O Johnson, Asako Nagata, Chau Almaden, Simon Bailey, T Eric Ballard, Louise Bernier, Hengmiao Cheng, Sujin Cho-Schultz, Deepak Dalvie, Judith G Deal, Dac M Dinh, Martin P Edwards, Rose Ann Ferre, Ketan S Gajiwala, Michelle Hemkens, Robert S Kania, John C Kath, Jean Matthews, Brion W Murray, Sherry Niessen, Suvi T M Orr, Mason Pairish, Neal W Sach, Hong Shen, Manli Shi, James Solowiej, Khanh Tran, Elaine Tseng, Paolo Vicini, Yuli Wang, Scott L Weinrich, Ru Zhou, Michael Zientek, Longqing Liu, Yiqin Luo, Shuibo Xin, Chengyi Zhang, and Jennifer Lafontaine. Discovery of n-((3r,4r)-4-fluoro-1-(6-((3-methoxy-1-methyl-1h-pyrazol-4-yl)amino)-9-methyl-9h-purin-2-yl)pyrrolidine-3-yl)acrylamide (pf-06747775) through structure-based drug design: A high affinity irreversible inhibitor targeting oncogenic egfr mutants with selectivity over wild-type egfr. *Journal of medicinal chemistry*, 60:3002–3019, 2017.

[14] Xuewu Zhang, Kerry A Pickin, Ron Bose, Natalia Jura, Philip A Cole, and John Kuriyan. Inhibition of the egf receptor by binding of mig6 to an activating kinase domain interface. *Nature*, 450:741–4, 2007.

[15] Ketan S Gajiwala, Junli Feng, Roseann Ferre, Kevin Ryan, Oleg Brodsky, Scott Weinrich, John C Kath, and Al Stewart. Insights into the aberrant activity of mutant egfr kinase domain and drug recognition. *Structure (London, England : 1993)*, 21:209–19, 2013.

[16] Albert C Pan, Thomas M Weinreich, Yibing Shan, Daniele P Scarpazza, and David E Shaw. Assessing the accuracy of two enhanced sampling methods using egfr kinase transition pathways: The influence of collective variable choice. *Journal of chemical theory and computation*, 10:2860–5, 2014.

[17] Christian R Schwantes and Vijay S Pande. Improvements in markov state model construction reveal many non-native interactions in the folding of ntl9. *Journal of chemical theory and computation*, 9:2000–2009, 2013.

[18] Mohammad M Sultan and Vijay S Pande. tica-metadynamics: Accelerating metadynamics by using kinetically selected collective variables. *Journal of chemical theory and computation*, 13:2440–2447, 2017.

[19] Mohammad M Sultan and Vijay S Pande. Transfer learning from markov models leads to efficient sampling of related systems. *The journal of physical chemistry. B*, 122:5291–5299, 2018.

[20] Niu Huang, Chakrapani Kalyanaraman, Katarzyna Bernacki, and Matthew P Jacobson.

Molecular mechanics methods for predicting protein-ligand binding. *Physical chemistry chemical physics : PCCP*, 8:5166–77, 2006.

[21] Baofeng Zhang, Michael P D'Erasmo, Ryan P Murelli, and Emilio Gallicchio. Free energy-based virtual screening and optimization of rnase h inhibitors of hiv-1 reverse transcriptase. *ACS omega*, 1:435–447, 2016.

[22] David L Mobley and Pavel V Klimovich. Perspective: Alchemical free energy calculations for drug discovery. *The Journal of chemical physics*, 137:230901, 2012.

[23] Agastya P Bhati, Shunzhou Wan, David W Wright, and Peter V Coveney. Rapid, accurate, precise, and reliable relative free energy prediction using ensemble based thermodynamic integration. *Journal of chemical theory and computation*, 13:210–222, 2017.

[24] Zachary A Knight and Kevan M Shokat. Features of selective kinase inhibitors. *Chemistry & biology*, 12:621–37, 2005.

[25] Cai-Hong Yun, Kristen E Mengwasser, Angela V Toms, Michele S Woo, Heidi Greulich, Kwok-Kin Wong, Matthew Meyerson, and Michael J Eck. The t790m mutation in egfr kinase causes drug resistance by increasing the affinity for atp. *Proceedings of the National Academy of Sciences of the United States of America*, 105:2070–5, 2008.

[26] Lichun Ma, Debby D Wang, Yiqing Huang, Hong Yan, Maria P Wong, and Victor H F Lee. Egfr mutant structural database: computationally predicted 3d structures and the corresponding binding free energies with gefitinib and erlotinib. *BMC bioinformatics*, 16:85, 2015.

[27] Jin H Park, Yingting Liu, Mark A Lemmon, and Ravi Radhakrishnan. Erlotinib binds both inactive and active conformations of the egfr tyrosine kinase domain. *The Biochemical journal*, 448:417–23, 2012.

[28] Morgan Huse and John Kuriyan. The conformational plasticity of protein kinases. *Cell*, 109:275–82, 2002.

[29] Tamjeed Saleh, Paolo Rossi, and Charalampos G Kalodimos. Atomic view of the energy landscape in the allosteric regulation of abl kinase. *Nature structural & molecular biology*, 24:893–901, 2017.

[30] Jiyong Park, Joseph J McDonald, Russell C Petter, and K N Houk. Molecular dynamics analysis of binding of kinase inhibitors to wt egfr and the t790m mutant. *Journal of chemical theory and computation*, 12:2066–78, 2016.

[31] Michael R Shirts. Best practices in free energy calculations for drug design. *Methods in molecular biology (Clifton, N.J.)*, 819:425–67, 2012.

[32] Andrew Pohorille, Christopher Jarzynski, and Christophe Chipot. Good practices in free-energy calculations. *The journal of physical chemistry. B*, 114:10235–53, 2010.

Appendix A

Extended Results

In addition to the results presented in the main dissertation, I contribute to multiple collaborative projects during my Ph.D. study and result in eight co-authored journal publications. The abstract of these publications is listed below.

STRAW: species TRee analysis web server

Timothy I. Shaw, Zheng Ruan, Travis C. Glenn, Liang Liu

Abstract

The coalescent methods for species tree reconstruction are increasingly popular because they can accommodate coalescence and multilocus data sets. Herein, we present STRAW, a web server that offers workflows for reconstruction of phylogenies of species using three species tree methods-MP-EST, STAR and NJst. The input data are a collection of rooted gene trees (for STAR and MP-EST methods) or unrooted gene trees (for NJst). The output includes the estimated species tree, modified Robinson-Foulds distances between gene trees and the estimated species tree and visualization of trees to compare gene trees with the estimated species tree. The web sever is available at `http://bioinformatics.publichealth.uga.edu/SpeciesTreeAnalysis/`.

Timothy I. Shaw, Zheng Ruan, Travis C. Glenn, Liang Liu (2013) *Nucleic Acids Res* 41(Web Server issue):W238-41.

Co-Conserved MAPK Features Couple D-Domain Docking Groove to Distal Allosteric Sites via the C-Terminal Flanking Tail

Tuan Nguyen, Zheng Ruan, Krishnadev Oruganty, Natarajan Kannan

Abstract

Mitogen activated protein kinases (MAPKs) form a closely related family of kinases that control critical pathways associated with cell growth and survival. Although MAPKs have been extensively characterized at the biochemical, cellular, and structural level, an integrated evolutionary understanding of how MAPKs differ from other closely related protein kinases is currently lacking. Here, we perform statistical sequence comparisons of MAPKs and related protein kinases to identify sequence and structural features associated with MAPK functional divergence. We show, for the first time, that virtually all MAPK-distinguishing sequence features, including an unappreciated short insert segment in the β4-β5 loop, physically couple distal functional sites in the kinase domain to the D-domain peptide docking groove via the C-terminal flanking tail (C-tail). The coupling mediated by MAPK-specific residues confers an allosteric regulatory mechanism unique to MAPKs. In particular, the regulatory ÎśC-helix conformation is controlled by a MAPK-conserved salt bridge interaction between an arginine in the αC-helix and an acidic residue in the C-tail. The salt-bridge interaction is modulated in unique ways in individual sub-families to achieve regulatory specificity. Our study is consistent with a model in which the C-tail co-evolved with the D-domain docking site to allosterically control MAPK activity. Our study provides testable mechanistic hypotheses for biochemical characterization of MAPK-conserved residues and new avenues for the design of allosteric MAPK inhibitors.

Tuan Nguyen, Zheng Ruan, Krishnadev Oruganty, Natarajan Kannan (2015) *PLoS One* 10(3):e0119636. Reprinted here with permission from the publisher.

Hydrophobic Core Variations Provide a Structural Framework for Tyrosine Kinase Evolution and Functional Specialization

Smita Mohanty, Krishnadev Oruganty, Annie Kwon, Dominic P. Byrne, Samantha Ferries, Zheng Ruan, Laura E. Hanold, Samiksha Katiyar, Eileen J. Kennedy, Patrick A. Eyers, Natarajan Kannan

Abstract

Protein tyrosine kinases (PTKs) are a group of closely related enzymes that have evolutionarily diverged from serine/threonine kinases (STKs) to regulate pathways associated with multi-cellularity. Evolutionary divergence of PTKs from STKs has occurred through accumulation of mutations in the active site as well as in the commonly conserved hydrophobic core. While the functional significance of active site variations is well understood, relatively little is known about how hydrophobic core variations contribute to PTK evolutionary divergence. Here, using a combination of statistical sequence comparisons, molecular dynamics simulations, mutational analysis and in vitro thermostability and kinase assays, we investigate the structural and functional significance of key PTK-specific variations in the kinase core. We find that the nature of residues and interactions in the hydrophobic core of PTKs is strikingly different from other protein kinases, and PTK-specific variations in the core contribute to functional divergence by altering the stability and dynamics of the kinase domain. In particular, a functionally critical STK-conserved histidine that stabilizes the regulatory spine in STKs is selectively mutated to an alanine, serine or glutamate in PTKs, and this loss-of-function mutation is accommodated, in part, through compensatory PTK-specific

Smita Mohanty *et al* (2016) *PLoS Genet* 12(2):e1005885.
Reprinted here with permission from the publisher.

interactions in the core. In particular, a PTK-conserved phenylalanine in the I-helix appears to structurally and functionally compensate for the loss of STK-histidine by interacting with the regulatory spine, which has far-reaching effects on enzyme activity, inhibitor sensing, and stability. We propose that hydrophobic core variations provide a selective advantage during PTK evolution by increasing the conformational flexibility, and therefore the allosteric potential of the kinase domain. Our studies also suggest that Tyrosine Kinase Like kinases such as RAF are intermediates in PTK evolutionary divergence inasmuch as they share features of both PTKs and STKs in the core. Finally, our studies provide an evolutionary framework for identifying and characterizing disease and drug resistance mutations in the kinase core.

Cerebellar ataxia and coenzyme Q deficiency through loss of unorthodox kinase activity

Jonathan A. Stefely, Floriana Licitra, Leila Laredj, Andrew G. Reidenbach, Zachary A. Kemmerer, Anais Grangeray, Tiphaine Jaeg-Ehret, Catherine E. Minogue, Arne Ulbrich, Paul D. Hutchins, Emily M. Wilkerson, Zheng Ruan, Deniz Aydin, Alexander S. Hebert, Xiao Guo, Elyse C. Freiberger, Laurence Reutenauer, Adam Jochem, Maya Chergova, Isabel E. Johnson, Danielle C. Lohman, Matthew J.P. Rush, Nicholas W. Kwiecien, Pankaj K. Singh, Anna I. Schlagowski, Brendan J. Floyd, Ulrika Forsman, Pavel J. Sindelar, Michael S. Westphall, Fabien Pierrel, Joffrey Zoll, Matteo Dal Peraro, Natarajan Kannan, Craig A. Bingman, Joshua J. Coon, Philippe Isope, Hlene Puccio, David J. Pagliarini

Abstract

The UbiB protein kinase-like (PKL) family is widespread, comprising one-quarter of microbial PKLs and five human homologs, yet its biochemical activities remain obscure. COQ8A (ADCK3) is a mammalian UbiB protein associated with ubiquinone (CoQ) biosynthesis and an ataxia (ARCA2) through unclear means. We show that mice lacking COQ8A develop a slowly progressive cerebellar ataxia linked to Purkinje cell dysfunction and mild exercise intolerance, recapitulating ARCA2. Interspecies biochemical analyses show that COQ8A and yeast Coq8p specifically stabilize a CoQ biosynthesis complex through unorthodox PKL functions. Although COQ8 was predicted to be a protein kinase, we demonstrate that it lacks canonical protein kinase activity in trans. Instead, COQ8 has ATPase activity and interacts with lipid CoQ intermediates, functions that are likely conserved across all domains of life. Collectively,

Jonathan A. Stefely *et al* (2016) *Mol Cell* 63(4):608-620.
Reprinted here with permission from the publisher.

our results lend insight into the molecular activities of the ancient UbiB family and elucidate the biochemical underpinnings of a human disease.

The Hippo Pathway Maintains the Equatorial Division Plane in the Ciliate *Tetrahymena*

Yu-Yang Jiang, Wolfgang Maier, Ralf Baumeister, Gregory Minevich, Ewa Joachimiak, Zheng Ruan, Natarajan Kannan, Diamond Clarke, Joseph Frankel and Jacek Gaertig

Abstract

The mechanisms that govern pattern formation within the cell are poorly understood. Ciliates carry on their surface an elaborate pattern of cortical organelles that are arranged along the anteroposterior and circumferential axes by largely unknown mechanisms. Ciliates divide by tandem duplication: the cortex of the predivision cell is remodeled into two similarly sized and complete daughters. In the conditional *cdaI-1* mutant of *Tetrahymena thermophila*, the division plane migrates from its initially correct equatorial position toward the cell's anterior, resulting in unequal cell division, and defects in nuclear divisions and cytokinesis. We used comparative whole genome sequencing to identify the cause of *cdaI-1* as a mutation in a Hippo/Mst kinase. CdaI is a cortical protein with a cell cycle-dependent, highly polarized localization. Early in cell division, CdaI marks the anterior half of the cell, and later concentrates at the posterior end of the emerging anterior daughter. Despite the strong association of CdaI with the new posterior cell end, the *cdaI-1* mutation does not affect the patterning of the new posterior cortical organelles. We conclude that, in *Tetrahymena*, the Hippo pathway maintains an equatorial position of the fission zone, and, by this activity, specifies the relative dimensions of the anterior and posterior daughter cell.

Yu-Yang Jiang *et al* (2017) *Genetics* 206(2):873-888.
Reprinted here with permission from the publisher.

Conformationally constrained peptides target the allosteric kinase dimer interface and inhibit EGFR activation

Melody D. Fulton, Laura E. Hanold, Zheng Ruan, Sneha Patel, Aaron M. Beedle, Natarajan Kannan, Eileen J. Kennedy

Abstract

Although EGFR is a highly sought-after drug target, inhibitor resistance remains a challenge. As an alternative strategy for kinase inhibition, we sought to explore whether allosteric activation mechanisms could effectively be disrupted. The kinase domain of EGFR forms an atypical asymmetric dimer via head-to-tail interactions and serves as a requisite for kinase activation. The kinase dimer interface is primarily formed by the H-helix derived from one kinase monomer and the small lobe of the second monomer. We hypothesized that a peptide designed to resemble the binding surface of the H-helix may serve as an effective disruptor of EGFR dimerization and activation. A library of constrained peptides was designed to mimic the H-helix of the kinase domain and interface side chains were optimized using molecular modeling. Peptides were constrained using peptide "stapling" to structurally reinforce an alpha-helical conformation. Peptide stapling was demonstrated to notably enhance cell permeation of an H-helix derived peptide termed EHBI2. Using cell-based assays, EHBI2 was further shown to significantly reduce EGFR activity as measured by EGFR phosphorylation and phosphorylation of the downstream signaling substrate Akt. To our knowledge, this is the first H-helix-based compound targeting the asymmetric interface of the kinase domain that can successfully inhibit EGFR activation and signaling. This study presents a novel, alternative targeting site for allosteric inhibition of EGFR.

Melody D. Fulton, *et al* (2018) *Bioorg Med Chem.* 26(6):1167-1173.
Reprinted here with permission from the publisher.

Coupled regulation by the juxtamembrane and sterile α motif (SAM) linker is a hallmark of Ephrin tyrosine kinase evolution

Annie Kwon, Mihir John, Zheng Ruan, Natarajan Kannan

Abstract

Ephrin (Eph) receptor tyrosine kinases have evolutionarily diverged from other tyrosine kinases to respond to specific activation and regulatory signals that require close coupling of kinase catalytic and regulatory functions. However, the evolutionary basis for such functional coupling is not fully understood. We employed an evolutionary systems approach involving statistical mining of large sequence and structural data sets to define the hallmarks of Eph kinase evolution and functional specialization. We found that some of the most distinguishing Eph-specific residues structurally tether the flanking juxtamembrane and sterile α motif (SAM) linker regions to the kinase domain, and substitutions of these residues in EphA3 resulted in faster kinase activation. We report for the first time that the SAM domain linker is functionally coupled to the juxtamembrane through co-conserved residues in the kinase domain and that together these residues provide a structural framework for coupling catalytic and regulatory functions. The unique organization of Eph-specific tethering networks and the identification of other Eph-specific sequence features of unknown functions provide new hypotheses for future functional studies and new clues to disease mutations altering Eph kinase-specific functions.

Annie Kwon, Mihir John, Zheng Ruan, Natarajan Kannan (2018) *J Biol Chem* 293(14):5102-5116.
Reprinted here with permission from the publisher.

Integrative annotation and knowledge discovery of kinase post-translational modifications and cancer-associated mutations through federated protein ontologies and resources

Liang-Chin Huang, Karen E. Ross, Timothy R. Baffi, Harold Drabkin, Krzysztof J. Kochut, Zheng Ruan, Peter D' Eustachio, Daniel McSkimming, Cecilia Arighi, Chuming Chen, Darren A. Natale, Cynthia Smith, Pascale Gaudet, Alexandra C. Newton, Cathy Wu, Natarajan Kannan

Abstract

Many bioinformatics resources with unique perspectives on the protein landscape are currently available. However, generating new knowledge from these resources requires interoperable workflows that support cross-resource queries. In this study, we employ federated queries linking information from the Protein Kinase Ontology, iPTMnet, Protein Ontology, neXtProt, and the Mouse Genome Informatics to identify key knowledge gaps in the functional coverage of the human kinome and prioritize understudied kinases, cancer variants and post-translational modifications (PTMs) for functional studies. We identify 32 functional domains enriched in cancer variants and PTMs and generate mechanistic hypotheses on overlapping variant and PTM sites by aggregating information at the residue, protein, pathway and species level from these resources. We experimentally test the hypothesis that S768 phosphorylation in the C-helix of EGFR is inhibitory by showing that oncogenic variants altering S768 phosphorylation increase basal EGFR activity. In contrast, oncogenic variants altering conserved phosphorylation sites in the 'hydrophobic motif' of PKCβII (S660F and

Liang-Chin Huang *et al* (2018) *Sci Rep* 8(1):6518.
Reprinted here with permission from the publisher.

S660C) are loss-of-function in that they reduce kinase activity and enhance membrane translocation. Our studies provide a framework for integrative, consistent, and reproducible annotation of the cancer kinomes.