All Theses and Dissertations (ETDs)

January 2009

Targeting The Dimerization Of ERBB Receptor Tyrosine Kinases

Rob Yang
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/etd

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences

Department of Biochemistry and Molecular Biophysics

Program in Computational Biology

Dissertation Examination Committee:

Linda Pike (Chair)
Garland Marshall (Advisor)
Nathan Baker
Tom Ellenberger
Jay Ponder
David Sept

TARGETING THE DIMERIZATION OF ERBB RECEPTOR TYROSINE KINASES

by

Robert Yunchuan Yang

A dissertation presented to the
Graduate School of Arts and
Sciences of Washington University
in St. Louis in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

August 2009

St. Louis, Missouri

ABSTRACT OF DISSERTATION

Targeting the dimerization of ErbB receptor tyrosine kinases

by

Robert Yunchuan Yang

Doctor of Philosophy in Biology and Biomedical Sciences

(Computational Biology)

Washington University in St. Louis, 2009

Professor Garland R. Marshall, Advisor

The epidermal growth factor receptor (EGFR) is a membrane receptor tyrosine kinase whose over-activation has been implicated to cause many human cancers. Novel strategies to inhibit the activation of EGF receptors other than the conventional antibody-based and tyrosine kinase inhibitors are virtually non-existent but could provide benefits both in the laboratory and clinical settings. In an effort to expand the current approaches, this thesis focused on targeting the homodimerization of the EGF receptors themselves and the heterodimerization of EGF receptors with the related ErbB2 receptor. Three sub-projects were completed in the process. The first project explored the feasibility of inhibiting the EGF receptor by targeting receptor dimerization with small molecules. Two lead compounds were initially predicted by virtual screening the NCI compound library, and were biochemically characterized. The benefit gained from the application of virtual screening in this project initiated another project to enhance the accessibility of virtual screening within the non-computational community. The OpenScreening project utilizes distributed computing resources and provides open-access screening server at: http://omg.phy.umassd.edu/xvhts. A final project identified the structural mechanism that

may explain the observed preference of EGFR-ErbB2 heterodimerization over EGFR homodimerization. Key residues were computationally predicted and biochemically tested to reveal critical dimerization interface.

Acknowledgement

Personal Dedications

I would like to open the Thank You's by thanking Garland for his guidance, patience, and most importantly, for granting and supporting me with complete research freedom.

I would also like to thank Linda for providing me with hands-on guidance in designing and carrying out experiments. To both Garland and Linda, it was truly a perfect combination of mentoring styles that made my Ph.D. training environment a successful and unique experience.

I would like to thank my colleagues J.W. Feng, Christy Taylor and Dan Kuster for great discussions and help in computational work, Jennifer Obermann for experimental procedures, and Charles Parnot for an unique "entirely over the internet" collaboration in the OpenScreening Project. I would also like to extend my thank-you's to the rest of the Marshall lab (Gregory Nikiforovich, Chris Ho, Sage Arbor, Jon Vabeno, Abby Fisher, Yaniv Barda, Xiaoming Zhang, Masayuki Hata) and the rest of the Pike lab (Yu Li, Sangeeta Adak).

I would like to thank my thesis committee, Professors Nathan Baker, Tom Ellenberger, Garland Marshall, Linda Pike (chair), Jay Ponder, and Dave Sept for great discussions and advices.

Finally, I will close out by thanking Katy, my colleague, collaborator, and wife for the wonderful Ph.D. chapter and many many more exciting chapters ahead of us.

<div align="right">Robert Y.C. Yang</div>

Washington University in St. Louis School of Medicine

May 2009

Table of Contents

List of Tables

List of Figures

Chapter 5. Additional Hits and Future Directions

List of Abbreviations

EGFR   Epidermal Growth Factor Receptor

vHTS   Virtual High-throughput Screen

MM/PBSA   Molecular Mechanics Poisson Boltzmann Solvent Accessible method

SASA   Solvent-Accessible Surface Area

MD   Molecular Dynamics Simulation

CHAPTER 1. Introduction to the ErbB Receptors

Overview

The epidermal growth factor receptor, EGFR, is the founding member of the ErbB family of receptor tyrosine kinases that also includes ErbB2, ErbB3 and ErbB4[1]. Among the members, they share high sequence similarities and adopt a similar spatial domain arrangement that consists of an extracellular ligand-binding region, a single transmembrane helix, a juxtamembrane region, a tyrosine kinase domain and a carboxyl-terminal regulatory region[6]. Binding of stimulatory ligands to the extracellular domain induces receptor homo- and heterodimerization that leads to autophosphorylation of the tyrosine residues in the intracellular regulatory region. Phosphorylated tyrosine residues serve as docking sites for the recruitment of adaptor proteins which in turn stimulate downstream pathways leading to enhanced cell proliferation[1, 7]. Aberrant activation of the ErbB receptors have long been associated with a wide range of cancers and consequently established as validated cancer targets[8, 9]. Current therapeutics fall under two categories: extracellular-targeting monoclonal antibodies, and small-molecule tyrosine kinase inhibitors[8]. Translation from pre-clinical promises to clinical efficacies have been increasingly challenged by acquired resistance[8, 10, 11]. Ongoing studies have pinpointed escape routes through receptor heterodimerization leading to signal diversification and amplification as a main evasion mechanism[11, 12]. With crystal structures revealing the structural basis of the extracellular-mediated homodimerization mechanism, therapeutically targeting the interface with small molecules was one of the main goals put forth for this thesis. Heterodimeric crystal structures are challenging and perhaps, even improbable, given the heterogeneity of homo- and heterodimer populations.

Computational modeling of the interface, prediction of critical residues using the model and subsequent experimental testing of the predictions represented another main goal.

The rest of this chapter reviews the current state of ErbB receptors and is organized to reflect aspects with regards to history, biology, and therapeutic implications. Chapter 2 focuses on the discovery of the first class of small-molecule inhibitors of EGFR homodimerization. Chapter 3 reports the development of a distributed computing virtual screen application that initiated from the computational procedures used during the discovery of dimerization inhibitors. Chapter 4 summarizes EGFR-ErbB2 heterodimer modeling and its subsequent experimental testing of critical interface residues predicted by the model. Chapter 5 analyzes and elaborates on the additional data obtained from further development of the initial lead inhibitors

History

The growth factor research field originated here at Washington University in 1952 when Dr. Rita Levi-Montalcini, while working with Professor Viktor Hamburger, discovered secreted factor in mouse tumour cells that potently promoted neurite outgrowth in chicken embryos[13]. This factor was later identified as the nerve growth factor (NGF) after its purification from snake venom and mouse salivary glands by Dr. Levi-Montalcini and Professor Stanley Cohen, a postdoc at the time in 1957[14, 15]. Professor Cohen then carried on the research as faculty at Vanderbilt to understand why crude extracts containing NGF from the male mouse salivary gland, but not the purified nerve growth factor alone, induced early eyelid opening in mouse. Five years later in 1962 (would have been tenure-review time under the current system), Professor Cohen

2

described the purification of a 53-amino acid "tooth-lid factor" capable of inducing "precocious eruption" of teeth and eyelids in newborn mice in 1962[16]. Upon subsequent studies where he showed that this factor stimulated the growth of epidermis on eyelids[17] and the proliferation of epidermal cells[18], Professor Cohen renamed this factor as the epidermal growth factor (EGF)[17]. These initial discoveries ignited a rapid growth in the field that led to the identification of the human homolog EGF[19-21]. Although widely-accepted that EGF acted by binding to specific receptors[22, 23], the actual identification of its specific receptor, the 170-kDa membrane protein EGF receptor (EGFR), was not until 1978 by Graham Carpenter and Professor Cohen[24]. Subsequent identification of the mechanism by which EGFR was phosphorylated upon EGF-stimulation established EGFR as the first receptor tyrosine kinase (RTK)[25-27]. In 1985, ErbB2 became the second member of the ErbB family[28, 29], and inadvertently popularized the name "ErbB". A year later in 1986, Professor Stanley Cohen and Dr. Rita Levi-Montalcini were awarded a Nobel Prize to recognize their initial discovery of the growth factors NGF and EGF in 1986. ErbB3[30] and ErbB4[31] were eventually discovered in 1989 and 1993, respectively. Figure 1.1 outlines a timeline of significant events in the field of ErbB receptors.
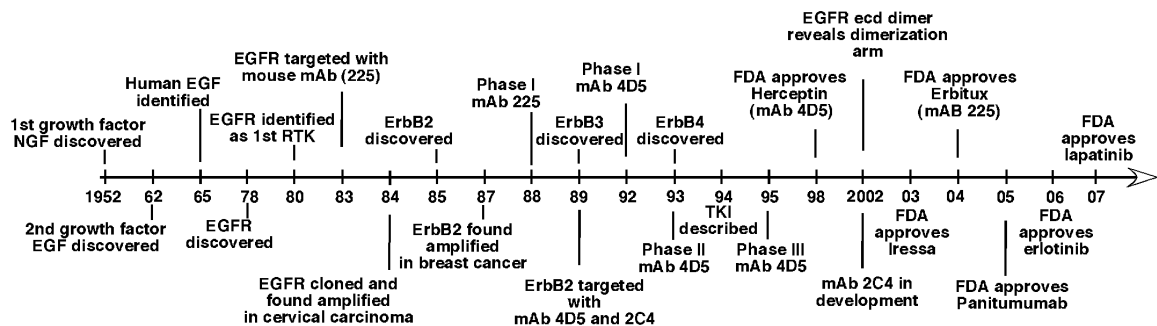
Figure 1.1. Historic timeline of the ErbB receptors. Adapted from Yarden *et al*[5].

3

Another important and often confusing historic aspect of the ErbB field is the redundancies in nomenclature. Given that the complexity of jargons in a given field is almost always positively correlated with the length of its history, it is not surprising that the field of ErbB/EGF/HER/p185, owing to its impact in both pre-clinical and clinical settings, has accumulated redundancy in describing only four receptors. I gave up tracing the exact documented origin of the switch from its original intuitive "EGF" to the present "ErbB" family name. However, based on the fact that ErbB2 was discovered as the gene "v-erb-b2 erythroblastic leukemia viral oncogene homolog 2", and that GenBank had just celebrated its 25th birthday in 2008, I could not help but wonder if this nomenclature showdown between a gene name and a protein name was an early hint of the budding impact of bioinformatics. As the references cited in this thesis and in literature span a broad disciplinary, I am including a jargon look-up table (Table 1.1) for convenience.

Table 1.1. Common names to describe EGFR and ErbB2 receptors

Name	ID
EGFR, ErbB1, Her1	EGFR
ErbB2, Her2, p185neu, NGL, NEU proto-oncogene, TKR1, c-erbB-2	ErbB2

Structure, function and mechanism

Under normal physiological conditions, activation of the ErbB receptors is controlled by spatial and temporal stimulations with a wide range of ligands. The evolutionary complexity of the ErbB receptors can be visibly traced to a one-to-one ligand-to-receptor paradigm in nematode *Caenorhabditis elegans* where EGFR homolog LET-23 is stimulated and activated by LIN-3 during vulva development[32]. In fruitfly *Drosophila melanogaster*, the system has evolved into five ligands (Spitz, Gurken, Keren, Vein, and Argo) and one receptor (DER)[33-35]. Interestingly, Argo is an antagonist

that inhibits DER activation, a unique mechanism not found in higher organisms. In mammalians, the four ErbB receptors are activated by a number of growth factors. ErbB2 currently has no known ligand but instead is thought to be the preferred heterodimer partner for the members[6].

Like all ErbB receptors, EGFR is a receptor tyrosine kinase that is composed of: a ligand-binding extracellular domain (~620 residues); a single transmembrane alpha helix (23 residues); a juxamembrane domain (~40 residues); an intracellular kinase domain (~260 residues); and, a C-terminal tail (~232 residues)[5] (Figure 1.2). Binding of stimulating ligands induces homo- and heterodimerization through a ligand-induced receptor-mediated interaction of the extracellular domains[36]. This leads to the activation of tyrosine kinase domains through the formation of an asymmetric kinase-dimer[37] (Figure 1.3). Stimulated kinases phosphorylate specific tyrosine residues in the C-terminal tail of the receptor[36] (Figure 1.4). These phosphotyrosines serve as docking sites for SH2 and PTB domain-containing proteins that promote the activation of multiple downstream pathways and ultimately regulate proliferation[38], differentiation[39], apoptosis[40], and angiogenesis[41] (Figure 1.5).
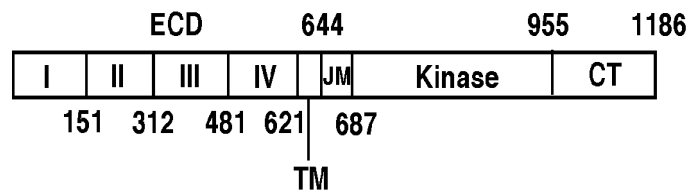
Figure 1.2. Schematic arrangement of EGFR. The extracellular domain (ECD) has 4 subdomains, followed by the transmembrane domain (TM), the juxtamembrane domain (JM), the kinase domain, and the C-terminal tail (CT). Adapted from Jorissen *et al*[2].
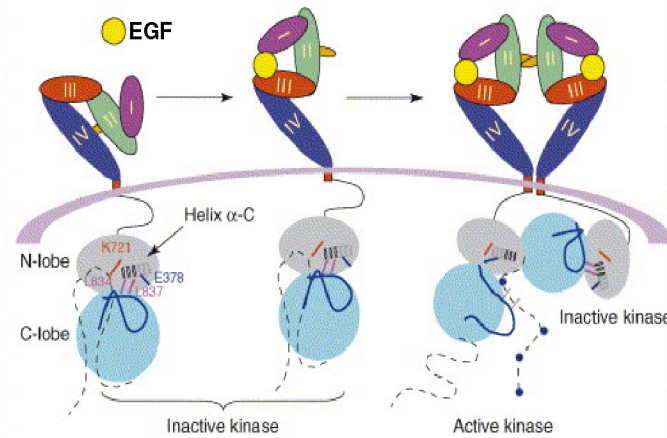
Figure 1.3. Mechanism of extracellular-driven dimerization leading to receptor phosphorylation. Upon EGF binding to the extracellular domain, the receptor adopts an open conformation exposing the dimerization arm. Receptor dimerization mediated by the arm induces the formation of asymmetric kinase dimer and subsequent kinase activation. Adapted from Bublil, *et al*[3].
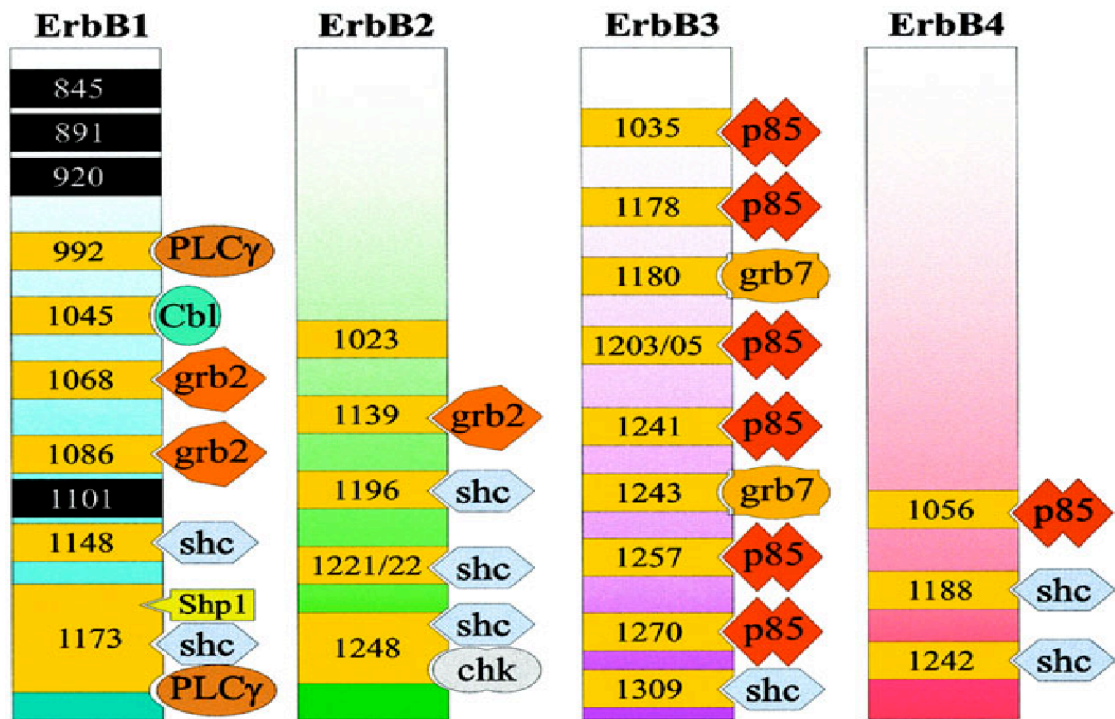
Figure 1.4. Signaling phosphotyrosines and the adaptor proteins. Tyrosine residues identified as autophosphorylation sites are colored yellow. Sites for the Src kinase are colored in black. Adapted from Olayioye *el al*[4].
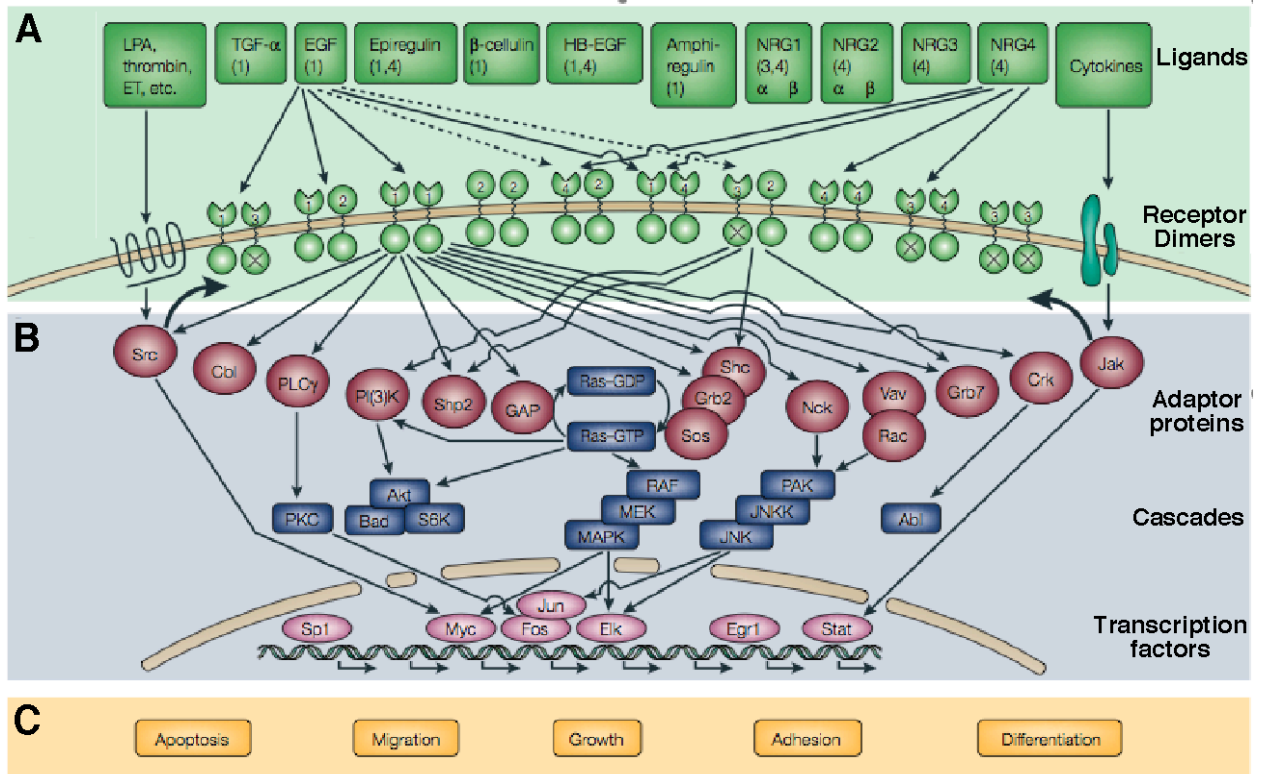
Figure 1.5. ErbB receptor overview. A) Stimulation by a wide range of ligands induces different combinations of dimerization. ErbB2 and ErbB3 homodimers are rarely found. ErbB2 lacks any known ligands while ErbB3 has non-functional kinase. B) Downstream signaling cascade propogates signals to control protein expressions. C) Ultimate cell fate decisions. Adapted from Yarden, *et al*[1].

Receptor homo- and heterodimerization

ErbB receptor dimerization is a pre-requisite for receptor activation[1]. ErbB receptor dimerization is driven by interactions between the extracellular domains of the two partners[42-45]. Recent crystal structures of the EGFR extracellular domain homodimer show that the most extensive part of the dimer interface centers on the intermolecular interaction between the ß-hairpin dimerization arm (residues 242-259) of one EGF receptor monomer and a corresponding docking pocket (the "armpit") on the second receptor monomer (Figure 1.6). Tyr-246 and Tyr-251 on the dimerization arm appear to be key residues involved in the stabilization of homodimers as mutation of either of these residues, even to a conservative Phe ablates dimer formation[42, 46]. These and the other residues located on the tip of the dimerization arm of monomer A interacts with residues that form the "armpit" of monomer B. As with mutation of Tyr-246 or Tyr-251 in the dimerization arm, mutation of certain residues in the armpit blocks receptor activation. For example, the R285S/F263A EGF receptor is defective in dimerization and hence its kinase activity cannot be stimulated by addition of EGF[45]. These data underscore the central role played by the dimerization arm in mediating EGFR homodimerization. Given the high sequence conservation and structural homology of the ErbB receptors, it is likely that this mode of interaction is conserved within the entire family.
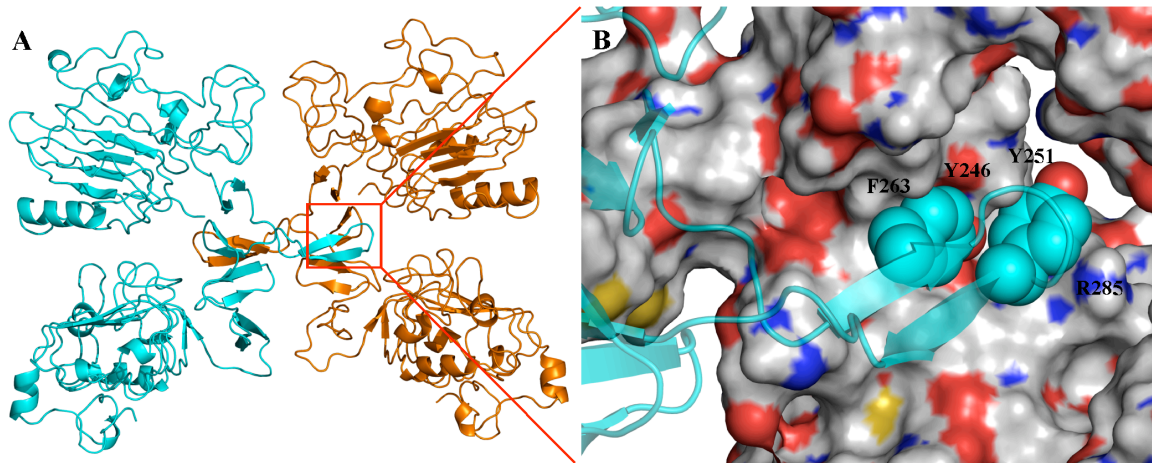
Figure 1.6. Crystal structure of the extracellular EGFR homodimer (PDB: 1MOX). A) dimer in ribbon form. B) Zoomed-in at the dimerization arm-armpit interface. Y246/Y251 are critical residues on the arm, F263/R285 are critical residues on the armpit.

EGFR is able to heterodimerize with other members of the ErbB family. Essentially all combinations of ErbB receptor heterodimers have been shown to exist[47, 48] with ErbB2 as the preferred hetero-dimerization partner in all cases[48]. In fact, ErbB2 is activated almost exclusively through heterodimerization as it has no known activating ligand[49] and ErbB2 homodimers have not been reported except in highly-overexpressed states[1]. ErbB3 is an obligate heterodimerization partner since it has very weak, tyrosine kinase activity and must utilize the kinase activity of its partner to signal[50]. All of these observations lead to the recognition that central to the function of ErbB receptors is their ability to form dimeric complexes that amplify and diversify the signals generated

Role in cancer and current therapy

The tight association between ErbB receptors and cancer was evident from the very beginning when ErbB2 was initially discovered because of its oncogenic potentials[29, 51]. It is therefore not surprising that they are among the most studied targets in the paradigm of molecular-targeted cancer therapies[8, 52, 53]. Perturbation of the EGFR

system leads to a variety of tumors in organs including breast[54], lung[55], ovary[56], pancreas[57] and prostate[58] (Table 1.2).

Current EGF receptor-directed strategies include monoclonal antibodies that target the extracellular domain[59-61], and small molecule tyrosine kinase inhibitors that compete with ATP at the nucleotide-binding site of the kinase domain[62-64] (Table 1.3). These drugs are promising, but show highly variable efficacy among patients in clinical applications[38, 65]. Ongoing researches have pointed resistance as a major and an increasingly substantial challenge. Two most observed resistance are in the forms of mutations within the ATP-binding site of the kinase domain[66, 67] and through homo- and heterodimerizations leading to signal diversification/amplification[12]. This is consistent with the reports that tumors characterized by abnormal levels of EGFR homodimerization[68, 69] and EGFR-ErbB2 heterodimerization[70] often display increased resistance to current treatments. In this setting, inhibitors that block signaling by multiple ErbB receptors would likely represent more effective chemotherapeutic agents than drugs specifically targeting one of the ErbB receptors. Based on the initial development of the heterodimer-inhibiting pertuzumab[71], this strategy must have been conceived by many at the time when I started this thesis work— additional multi-targeting agents are now being reported[63, 72]. Nevertheless, there has not yet been any report of targeting the dimerization arm with small-molecules. This is likely because that although the concept is fairly obvious with the availability of the crystal structures, the lack of an appropriate dimerization assay represents a challenge compared to the already established protocols for antibody generation and kinase inibition. As described in the next chapter, I embarked on what was in hindsight a high-risk project to identify small-molecule dimerization

inhibitors that, with some luck, led to the identification of two lead inhibitors.

Table 1.2. ErbB receptors in cancer (reviewed in [1, 73])

Receptor	Activation	Cancer type	Notes
EGFR	Mutation (type III)	Glioma, Breast, Ovarian	Deletion of part of the extracellular domain yields a constitutively active receptor
EGFR	Overexpression	SCCHN, Breast, Ovarian, Prostate, Kidney, NSCLC, glioma	Amplification in 40% gliomas, correlates with reduced survival
ErbB2	Overexpression	Breast, Ovarian, Stomach, Bladder, Salivary, Lung	Gene amplification in 15-30% breast cancers
ErbB3	Expression	Breast, Colon gastric, Prostate	Co-expression of ErbB2 with EGFR or ErbB3 in breast cancer
	Overexpression	Oral squamous cell cancer	Overexpression reduces survival
ErbB4	Expression	Breast, Prostate, Childhood medullo-blastoma	Co-expression with ErbB2

HNSCC, head and neck squamous-cell cancer; NSCLC, non-small-cell lung cancer.

Table 1.3. ErbB-targeted therapeutics in clinical use (reviewed in [8, 73, 74])

Compound	Type	Target	Company	Status and comments
Trastuzumab (Herceptin)	Humanized mAb	ErbB2	Genetech/ Roche	Approved for the treatment of ErbB2-overexpressing breast cancer
Pertuzumab (Omnitarg)	Humanized mAb	ErbB2	Genetech	Phase III trials to treat ovarian cancer, breast cancer, prostate cancer and NSCLC; based on its ability to block ErbB2 dimerization, trials are ongoing in cancer that express low ErbB2 levels
Cetuximab (Erbitux)	Chimeric mAb	EGFR	ImClone/ Merck KGaA	Approved for the treatment of CRC; ongoing trials in combination with various drugs for treatment of pancreatic cancer, HNSCC and NSCLC
Matuzumab (EMD 72000)	Humanized mAb	EGFR	Merck KGaA	Phase II trials for NSCLC, gynaecological cancer, pancreatic cancer and oesophageal cancer
Panitumumab (Vecitibix, ABX-EGF)	Fully humanized mAb	EGFR	Abgenix	Approved for CRC, RCC and NSCLC
Nimotuzumab (Thera CIM)	Humanized mAb	EGFR	YM	Approved to treat HNSCC, NPC, glioma; reported reduced toxicity and immunogenicity
Gefitinib (Iressa, ZD1839)	Quinazoline TKI	EGFR	AstraZeneca	Approved for the treatment of NSCLC; ongoing trials in HNSCC, gastrointestinal cancer and breast cancer
Erlotinib (Tarceva, OSI-774)	Quinazoline TKI	EGFR	Genetech/ Roche/OSI Pharmaceuticals	Approved for NSCLC; ongoing trials in many cancer types
Lapatinib (GW2016)	TKI	EGFR/ ErbB2	GlaxoSmithKline	Phase III trial on breast cancer patients who are refractory to trastuzumab and chemotherapy
AEE788	TKI	EGFR/ ErbB2/ VEGFR	Novartis	Phase I trials underway—first multifunction EGFR/ErbB2/VEGFR inhibitor
PKI166	TKI	EGFR	Novartis	Development of PKI166 was interrupted due to high incidence of liver toxicity, 17% of patients showed Grade 3 elevated liver transaminases
CI-1033	4-anilinoquinazoline irreversible TKI	EGFR/ ErbB2	Pfizer	Phase II trials underway in breast and NSCLC
EKB-569	4-Anilinoquinoline-3-carbonitrile irreversible TKI	EGFR/ ErbB2	Wyeth-Ayerst	Phase II trials underway in NSCLC
EXEL 7647/EXEL 0999	TKI	EGFR/ ErbB2/ VEGFR	EXELIXIS	Phase I trials underway

CRC, colorectal cancer; HNSCC, head and neck squamous-cell cancer; NPC, nasopharyngeal carcinoma; NSCLC, non-small-cell lung cancer; RCC, renal-cell cancer; VEGFR vascular endothelial growth factor receptor

CHAPTER 2. Targeting EGFR Dimerization with Small-Molecule Inhibitors

Abstract

EGFR dimerization is a prerequisite for activation, and the basis to explore the feasibility of inhibiting EGFR activation by blocking dimer formation with small molecules. Initially predicted by virtual screening, subsequent experiments showed that two compounds dose-dependently inhibited EGFR kinase activation. Both compounds were shown to block EGF-stimulated dimer formation in live cells using a real-time luciferase fragment complementation imaging assay. One compound was further shown to inhibit the growth of HeLa cells. These first-generation lead compounds represent the first small-molecule inhibitors of EGF receptors that are not directed against the active sites of the kinase

Introduction

We rationalized that the critical role of extracellular dimerization could be exploited to inhibit receptor activation, and set out to test the feasibility of targeting the dimerization process with small molecules. In this chapter, we report the identification of two small-molecule lead compounds capable of inhibiting the EGFR activation by blocking dimer formation. These inhibitors were initially identified by applying a consensus virtual high-throughput screening (vHTS) protocol to screen the National Cancer Institute Diversity (NCI-Diversity) library[75] for compounds with binding potential to the same pocket that Tyr-246 and Tyr-251 recognize. Subsequent biochemical assays confirmed that these two compounds selectively impaired EGFR dimerization, representing the first members of a new class of small-molecule inhibitors of EGFR activation

Results

The virtual high-throughput screening protocol (vHTS).

The vHTS employed in these experiments used AutoDock 4.0[76, 77] to dock approximately 2000 compounds present in the NCI Diversity database within a 25 Å^3 docking box centered on the Tyr-246/Tyr-251 recognition pocket of the dimerization arm of the EGF receptor. A total of 8 scoring functions were used to independently rank all predicted docking poses, and the consensus of these functions were chosen as the final ranking.

The enrichment of a vHTS protocol is typically measured by its ability to recover true positives as early as possible in a ranked compound library. Protocol evaluation thus depends on the availability of existing reference active compounds. Because there were no existing inhibitors that targeted the Tyr-246/Tyr-251 site, it was not possible to evaluate the enrichment power of our vHTS protocol for the EGF receptor system *a priori*. As a result, robustness, measured as the average enrichment across different protein targets, became a critical criterion for evaluating the performance. Our protocol was applied to four different protein targets: plasmepsin II (PMII), human cyclin-dependent kinase 2 (Cdk2), estrogen receptor (ER), and yeast heat shock protein (Hsp90). Structurally diverse compounds (positives) bound to these protein targets were extracted from co-crystal structures in the Protein Data Bank, and were mixed with 1926 decoy compounds (negatives) to construct a testing library. The ability of the vHTS protocol to recover positives was evaluated using enrichment curve analysis[78].

The protocol recovered at least one true positive within the top 1% of the ranked library for PMII, Cdk2, and ER, and within the top 10% of the library for Hsp90 (Table

2.1, Figure 2.1). On average, this protocol is expected to recover at least one true ligand

within the top 3.5%, and nearly 2/3 of all ligands within the top 15% of the representative

libraries screened.

Table 2.1: Efficacy and robustness of the vHTS protocol.

Targets	Coverage$_{1\%}$[1]	Coverage$_{15\%}$	Coverage$_{30\%}$	Coverage$_{50\%}$	Best [2]
Cdk2	3%	49%	67%	79%	0.05%
PMII	60%	100%	100%	100%	0.65%
ER	69%	81%	94%	100%	0.05%
HSP90	0%	20%	60%	100%	13.21%
Avg	33%	63%	80%	100%	3.5%

[1]Coverage$_{fraction}$ = Number of known actives recovered within the given fraction of the database /
Total number of actives present in the database x 100%
[2]Best = ranking of the best predicted active / database size x 100
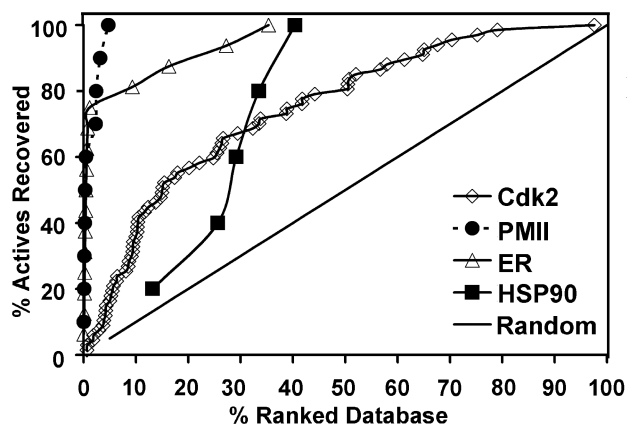
Figure 2.1. Evaluation of the vHTS protocol against four testing cases shown in an enrichment curve analysis. In each case, multiple known ligands were mixed in with ~2000 random compounds to form the screening library. The black diagonal line represents the random distribution of active molecules.

16

Inhibition of EGFR activation as the primary screen.

We applied the vHTS protocol to the EGF receptor and obtained samples of the 80 top-ranked compounds (top 4%) along with 40 randomly chosen compounds from NCI for testing. Of the 80 compounds, 4 were not soluble in water or DMSO, and, therefore, not pursued further. The remaining 76 compounds were tested for their ability to inhibit EGF-stimulated receptor autophosphorylation in cells at a concentration of 100 µM.

Of the 76 compounds tested, 20 produced a significant (>60%) decrease in activation as measured by the level of phosphorylation at Tyr-1173, a major autophosphorylation site on EGF receptors. By contrast, none of the 40 compounds randomly chosen from the same library inhibited receptor phosphorylation when assayed under the same conditions. This highlights the enrichment and the utility of our vHTS protocol in the present system.

Figure 2.2 presents the results for the characterization of what turned out to be the two lead inhibitors. In both cases, the inhibition was dose-dependent. NSC56452 exhibited an IC_{50} value of 400 nM while NSC11241 had an IC_{50} of 12 µM. Despite having the lowest IC_{50} value, NSC56452 was not able to achieve full inhibition at saturating doses.
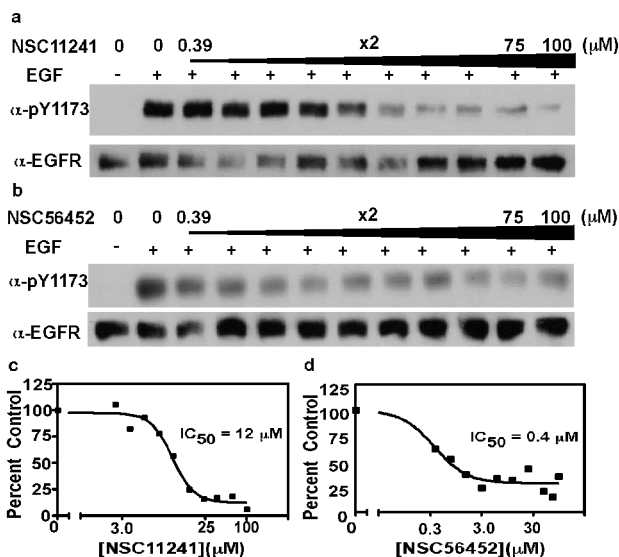
Figure 2.2. Inhibition of EGF receptor autophosphorylation by NSC11241 and NSC56452. Cells were pre-incubated with 1% DMSO for the control, NSC11241 or NSC56452. Inhibition of EGF receptor autophosphorylation of the controls (lane 1 and 2) and increasing doses of **a)** NSC11241 and **b)** NSC56452. Estimated IC_{50} value quantified from blots for **c)** NSC11241 and d) NSC56452.

Specific inhibition of the EGF receptor activation by lead compounds.

To assess the specificity of the 20 candidates, they were tested for their ability to inhibit two related receptor tyrosine kinases, the insulin receptor and the PDGF receptor. For the insulin receptor, insulin-stimulated tyrosine phosphorylation of IRS-1 in differentiated 3T3-L1 cells was assessed[79]. For the PDGF receptor, PDGF-stimulated receptor autophosphorylation in NIH3T3 cells was measured[80]. Neither 3T3-L1 cells nor NIH3T3 cells express the EGF receptor obviating potential problems associated with receptor crosstalk.

Of the 20 compounds that inhibited EGF receptor autophosphorylation, 2 inhibited insulin-stimulated IRS-1 phosphorylation and 4 others inhibited PDGF receptor autophosphorylation (Figure 2.3). An additional 3 compounds markedly enhanced PDGF receptor autophosphorylation and were not pursued further.

a. IRS-1 Phosphorylation

b. PDGFR Phosphorylation
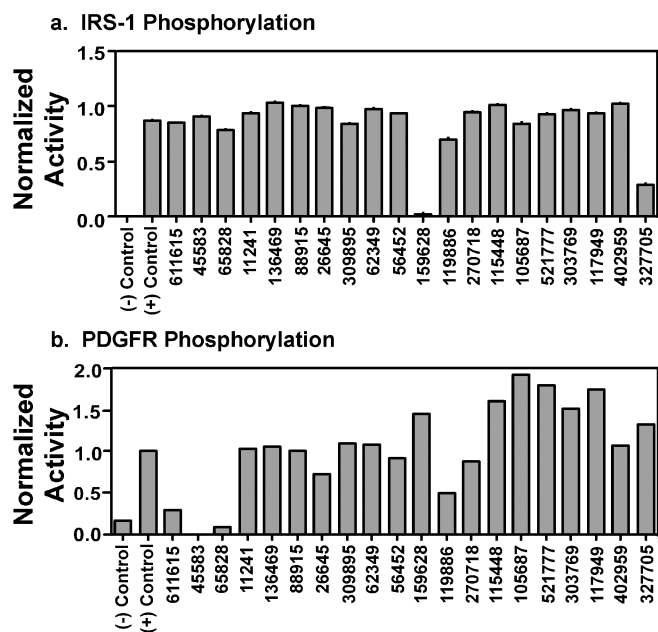
Figure 2.3. Specificity of inhibitors. Cells expressing either the insulin receptor or the PDGF receptor were pre-incubated with 1% DMSO (controls) or 100 μM of each of the 20 lead compounds. **a)** Insulin receptor kinase activity was assessed by measuring the phosphorylation of IRS-1 in response to 3 nM insulin for 1 minute. Representative data of three experiments. **b)** PDGF receptor kinase activity was assessed by measuring autophosphorylation of the PDGF receptor in response to 2 nM PDGF for 3 minutes. Representative data based on two experiments.

Inhibition of EGF receptor dimerization by lead compounds.

Since the lead inhibitors were initially chosen based on their potential to interfere with EGF receptor dimerization, we next determined whether the remaining 11 candidates inhibited EGF receptor autophosphorylation by directly blocking receptor dimerization, as measured by chemical cross-linking. Cells were preincubated with the inhibitors for 15 min at a final concentration of 100 μM. EGF at 25 nM was then added followed by 3 mM BS³, a membrane impermeable chemical cross-linker. Figure 2.4 shows the effect of a subset of these inhibitors on the cross-linking of EGF receptor dimers.

| Compound | - | - | 11241 | 309895 | 56452 | 303769 |
| EGF | - | + | + | + | + | + |

Dimer

Monomer

Figure 2.4. Inhibition of EGF receptor dimerization determined by chemical cross-linking assay. Cells were pre-incubated with 1% DMSO (lane 1 and 2) or 100 μM of lead inhibitors (lane 3-6) prior to stimulation with 25 nM EGF (lane 2-6) for 5 minutes. All cells were then treated with 3 mM of the cross-linking reagent BS³. NSC11241 (lane 3) and NSC56452 (lane 5) significantly inhibit dimer formation. Lane 4 and 6 show compounds that did not inhibit dimer formation.
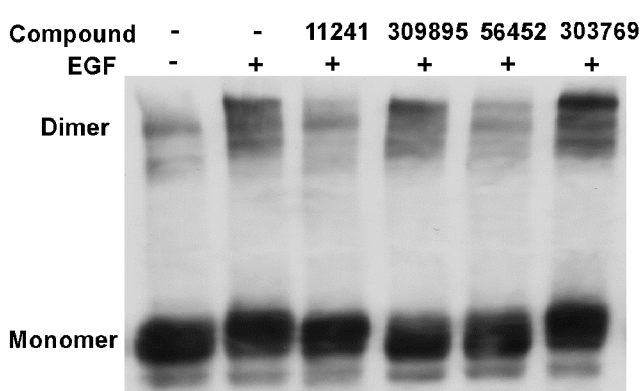
The two lead compounds, NSC11241 and NSC56452, significantly reduced the formation of high molecular weight dimers while NSC309895 and NSC303769 represented two examples of compounds that failed to block dimer formation. None of the other compounds had any effect on the cross-linking of EGF receptor dimers. Because the cross-linker was used at a concentration 30-fold higher than that of the inhibitors (3 mM *vs.* 100 μM), it is unlikely that the observed inhibition was due to quenching of the cross-linking reaction by reaction with the compounds. Consistent with this conclusion, increasing the concentration of cross-linker BS³ to 5 mM yielded the same results. It is possible, however, that false negatives could be obtained if reaction of the compound with cross-linker prevented that compound from binding to the EGF receptor.

The two hits from the cross-linking assay, NSC11241 and NSC56452 (Figure 2.5), were selected for further testing of the hypothesis that they inhibited dimer formation. A real-time live cell imaging assay based on luciferase fragment complementation was utilized to characterize the effect of these compounds on EGF receptor dimerization[81]. In this assay, an EGF receptor lacking the entire intracellular

domain (referred to as ΔC-EGFR) was fused to either an N-terminal (NLuc) or C-terminal (CLuc) fragment of firefly luciferase. Ligand-induced dimerization of the ΔC-EGFR brings the luciferase fragments into close proximity resulting in enzyme complementation and reconstitution of luciferase activity. The rate and extent of receptor dimerization can, therefore, be measured by following photon flux in the presence of luciferin. The absence of the intracellular domain of the EGF receptor from these constructs ensures that compounds affecting luciferase activity without binding to the cytoplasmic portion of the receptor that contains the tyrosine kinase domain.

As expected, EGF stimulated a rapid increase in light production in DMSO-treated control cells consistent with ligand-induced dimer formation. Cetuximab, an FDA-approved antibody-based drug that binds to the extracellular domain of the EGF receptor[59, 82], dramatically decreased EGF-induced luciferase activity, serving as a positive control for inhibition of dimer formation. At a concentration of 25 μM, NSC11241 and NSC56452 each induced a significant decrease in luciferase complementation compared to controls (Figure 2.5).
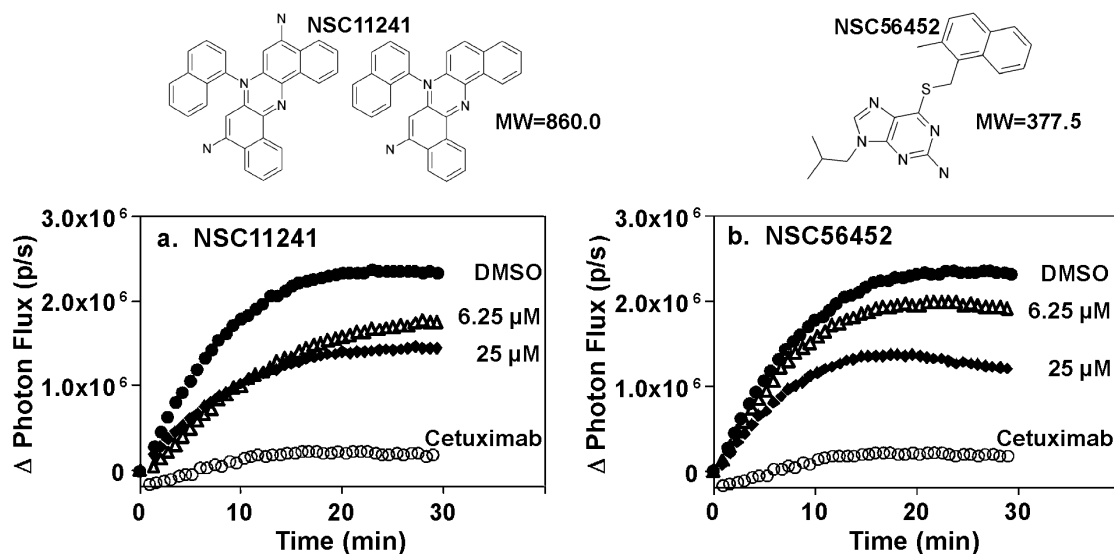
Figure 2.5. Inhibition of EGF receptor dimerization assayed by luciferase fragment complementation. Cells stably expressing ΔC-EGFR-NLuc and ΔC-EGFR-CLuc were pre-treated with DMSO, the indicated concentrations of compounds or 1 μg/ml Cetuximab for 20 min in the presence of 0.6 mg/ml D-luciferin prior to the addition of 3 nM EGF. Data represent the change in photon flux between quadruplicates of EGF-treated and untreated control cells. **a)**, NSC11241; **b)**, NSC56452,

To exclude the possibility that the compounds simply inhibited complementation of the luciferase fragments themselves, they were tested in a similar luciferase fragment system[83] where complementation was induced by the addition of rapamycin to cells expressing FRB-NLuc and its binding partner CLuc-FKBP. Neither compound induced a decrease in luciferase complementation at the tested dose (25 μM) in this control system. This indicates that the observed decrease in complementation in our EGF receptor complementation assay was not due to any direct effect on the NLuc and CLuc fragments themselves

NSC56452 and NSC11241 do not inhibit by interfering with the binding of EGF ligands.

It was still possible that NSC56452 or NSC11241 might inhibit EGF receptor activation by interfering with the binding of EGF to the extracellular domain of the

receptor. To address this possibility, the effect of these compounds on the binding of ^{125}I-EGF to EGF receptor-expressing CHO cells was measured. NSC56452 had little effect on EGF binding suggesting that its effect on phosphorylation and dimerization was not due to blocking the binding of EGF to its receptor (Figure 2.6). NSC11241 had a modest effect on EGF binding, but the magnitude (~30%) was much smaller than the ~90% inhibition of receptor autophosphorylation. Thus, an effect on EGF binding is unlikely to be the predominant mechanism through which NSC11241 exerts its inhibitory effect on EGF receptor phosphorylation, and certainly not an explanation of NSC56452 inhibition.
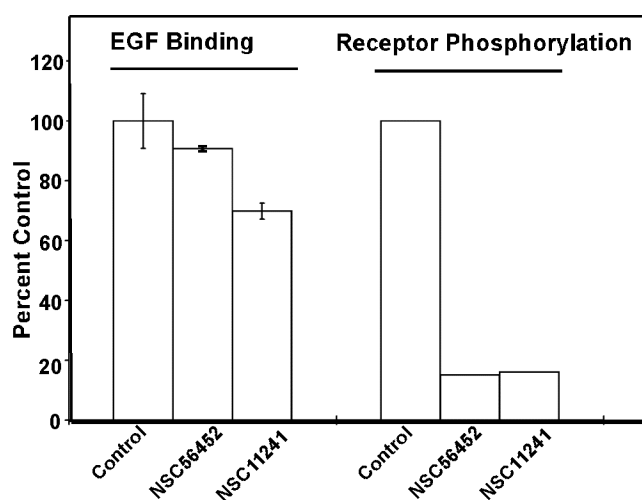
Figure 2.6. Effects of NSC56452 and NSC11241 on ^{125}I-EGF binding and EGF receptor autophosphorylation. ^{125}I-EGF binding and EGF receptor autophosphorylation were assessed as described in Materials and Methods. NSC56452 was tested at 100 μM while NSC11241 was used at 25 μM due to its limited solubility at 4° C, the temperature at which the binding assay was performed.

Growth inhibition of HeLa cells by NSC56452.

To assess the effect of the lead compounds on cancer-cell growth, NSC56452 was tested for its ability to inhibit the proliferation of HeLa cells that express endogenous EGF receptors. NSC11241 was not tested here as its color interfered with the MTS cell-proliferation assay due to overlap in absorption spectra. For comparison, cells were also treated with erlotinib, an EGF receptor kinase inhibitor. By itself, NSC56452 induced significant inhibition of cell growth at 50 μM, but had little to no effect at a lower dose 12.5 μM. Remarkably, the growth inhibition at 12.5 μM NSC56452 was much enhanced

in combination with a sub-effective dose of erlotinib (Figure 2.7). This apparent synergistic observation was consistent with the hypothesis that NSC56452 inhibited EGF receptor activity through a mechanism different from classical kinase inhibitors.
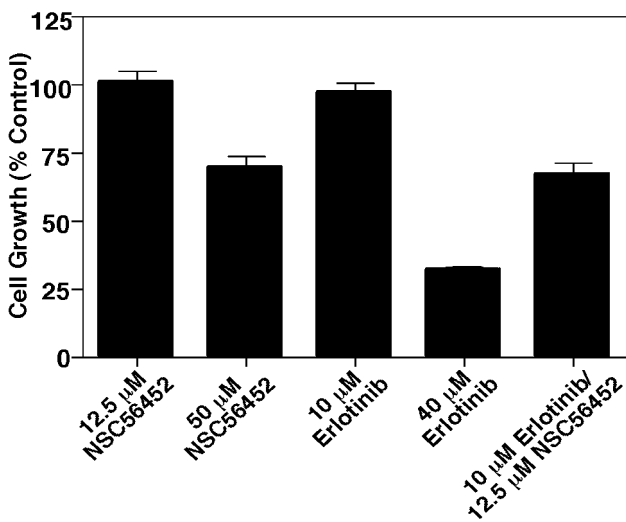

Figure 2.7. Inhibition of HeLa cell growth. Cells were grown in the absence or presence of erlotinib, NSC56452, or a combination of the two inhibitors at the indicated doses. Cell growth was measured by the cellTiter 96 Aqueous One Solution Cell Proliferation Assay after 48 hr incubation with the inhibitors. All experiments were performed in triplicates. All cultures contained 1% DMSO.

Discussion

Systematic screening of compound libraries remains the main strategy to discover novel inhibitors, but lack of appropriate high-throughput bioassays often renders the approach not feasible in practice. In this work, we were able to bypass this challenge by utilizing virtual high-throughput screening in combination with low-throughput biochemical assays to identify lead compounds capable of inhibiting the dimerization and activation of EGF receptor in intact cells.

In identifying these inhibitors, we adopted a structure-based approach and took advantage of the recent structural and mutational data that highlight the critical role of the "dimerization arm" in mediating EGF receptor dimerization[42, 44-46, 84]. Mutations to Tyr-246/Tyr-251 on the arm[42, 46] are sufficient to abolish dimerization, suggesting a stringent structural requirement for these residues to interact with a pair of adjacent hydrophobic pockets on the other monomer in a precise orientation. We hypothesized that this

sensitivity to minor perturbations could be exploited to discover compounds capable of interfering with EGF receptor dimerization.

By testing only 4% of the NCI-diversity library, NSC11241 and NSC56452 were identified as compounds that selectively inhibited EGF-stimulated tyrosine kinase activity while having no effect against the related PDGF and insulin receptor tyrosine kinases. Several lines of evidence suggest that NSC11241 and NSC56452 work by inhibiting EGF receptor dimerization. In chemical cross-linking assays, both compounds significantly reduced the formation of EGF-induced high molecular weight oligomers. This dimer-inhibition mechanism was further confirmed by the luciferase fragment complementation assay[81] where modulation of receptor dimerization by the lead compounds was quantitatively monitored in real time and in live cells. Interestingly, AG1478, an EGF receptor kinase inhibitor, has been shown to increase dimer-formation as measured by both assays[81, 85]. This difference further supports the conclusion that our inhibitors inhibit by a mechanism that is distinct from that used by tyrosine kinase inhibitors. Consistent with this mechanistic difference, NSC56452 and another EGF receptor-specific tyrosine kinase inhibitor, erlotinib, exhibited synergy with respect to their ability to inhibit cancer cell growth. In the recent release of the NCI Cancer Screening Data against 60 cancer cell lines (http://dtp.nci.nih.gov/compare-web-public_compare/login.do, accessed April, 2009), NSC56452 was reported to inhibit the growth of 3 cancer cell lines with statistically significant GI50, TGI, and LC50 values: SK-MEL-5[86] and UO-31[87] cells both overexpress EGF receptors while COLO205 cells overexpress ErbB2[88], a related ErbB family of receptors. The apparent correlation

between sensitivity to NSC56452 and a common molecular target among the cell lines is consistent with our hypothesized mechanism of action.

This class of dimerization inhibitors provides a new set of chemical tools for modulating the activation of the EGF receptor in laboratory settings. For example, one can utilize these compounds to modulate the activation of mutant receptors resistant to kinase inhibitors, or to gain insights into the formation of EGF receptor dimers and high-order oligomers. In addition to having utility as chemical research tools, these two lead compounds serve as possible leads for further development as anticancer agents to complement existing therapeutics.

Materials and Methods

Virtual Screening. Autodock 4.0[76, 77] was used to screen the NCI-diversity database (1990 compounds). The database was initially downloaded from the Autodock website and processed by in-house scripts to fix incorrectly formatted structures, and to exclude structures that contained metals: iron, zinc, mercury and copper (final size = 1926 compounds). A docking box of dimension 25 Å^3 was centered at the Tyr-246/Tyr-Y251 recognition site on monomer A of the extracellular dimer crystal structure (PDB: 1MOX). The Larmackian genetic algorithm with Solis and Wets local search was used to generate 100 docking poses per compound. All poses were subsequently scored using: HP, HM, HS (implemented in X-score 1.2.1[89]), D-score, PMF, G-score, Chem-score (implemented in Sybyl 7.3 CSCORE module), and Dfire[90]. A consensus score for each pose was calculated by summing the rankings given by each of the 8 scoring functions. Three compounds that ranked high using the consensus scores were excluded because

they displayed high rankings against other protein targets suggesting limited specificity for the dimerization site.

EGF receptor autophosphorylation. CHO cells stably expressing wild type EGF receptor were grown to 80% confluency in 35 mm plates in Hams' F-12 containing 10% FBS, penicillin/streptomycin, and 100 μg/ml hygromycin. Prior to use, the cells were incubated for 3 hours in F-12 medium containing 0.1% FBS. For the experiments, cultures were incubated with the test compounds at a final concentration of 100 μM in 1% DMSO for 30 min at 25° C in F-12 containing 1 mg/ml bovine serum albumin and 25mM Hepes, pH 7.2. Control cultures were incubated for the same length of time with 1% DMSO. EGF (Biomedical Technologies, Inc) was then added at a final concentration of 3 nM and the cultures incubated at 25° C for an additional 1 min. Subsequently, the monolayers were washed twice with ice-cold phosphate-buffered saline and scraped into RIPA buffer (10 mM Tris, pH 7.2, 150 mM NaCl, 0.1% sodium dodecyl sulfate, 1% Triton X-100, 17 mM deoxycholate, and 2.7 mM EDTA) containing 1 mM sodium orthovanadate, 20 mM p-nitrophenylphosphate, and protease inhibitors. Equal amounts of protein (BCA assay, Pierce) were separated by electrophoresis on a 9% SDS polyacrylamide gel, and transferred to PVDF or nitrocellulose (Millipore). Western blotting was performed using anti-pY1173 (Cell Signaling), or anti-EGF receptor antibodies (Cell Signaling and Santa Cruz). Time-course and dose-response experiments were done using the same procedure except that the dose or preincubation time with inhibitors was varied. A similar protocol was used for assessing insulin-stimulated phosphorylation of IRS-1 or PDGF-stimulated receptor autophosphorylation except that differentiated 3T3-L1 cells or NIH3T3 cells were used, respectively. In all cases,

27

phosphorylation was quantified using ImageJ and normalized to that observed in control samples.

Chemical cross-linking of the EGF receptor. CHO cells stably expressing EGF receptor were preincubated with the test compounds for 15 min at a final concentration of 100 μM. EGF (25 nM) was then added for 3 min followed by the addition of BS^3 (Pierce) at a final concentration of 3 mM for 30 min. The reaction mixture was buffered at pH 8. The cross-linking reactions were quenched by the addition of glycine to a final concentration of 1 M (pH 7.5). Cells were lysed as above, and equal amounts of protein were loaded onto a 4%-7.5% gradient SDS-polyacrylamide gel. After electrophoresis and transfer to PVDF, EGF receptor dimerization was measured by Western blotting using anti-EGF receptor antibodies.

Luciferase fragment complementation imaging. CHO-K1 Tet-On cells stably expressing ΔC-EGFR-NLuc and ΔC-EGFR-CLuc[81] were plated 48 hrs prior to imaging in DMEM containing 1 μg/ml doxycycline. On the day of imaging, cells were serum-starved for 4 hrs followed by treatment with vehicle, the indicated concentration of each compound, or 1 μg/ml cetuximab for 20 min in the presence of 0.6 mg/ml D-luciferin. 3 nM EGF was then added and the photon flux immediately measured using an IVIS imaging system. Data represent the change in photon flux between EGF-treated cells and control cells. For the control experiments using the FRB-NLuc and CLuc-FKBP system[83, 91], CHO-K1 Tet-On cells were plated 48 hrs prior to use and transiently transfected with the cDNA encoding FRB-NLuc and CLuc-FKBP 24 later. On the day of assay, cells were pre-treated with vehicle or 80 nM rapamycin for 4 hrs. Media was removed and replaced with DMEM lacking phenol red containing 0.6 mg/ml D-luciferin and DMSO, 25 μM

compound NSC11241, or 25 μM compound NSC56452. Photon flux was measured as above.

$^{125-}$*I-EGF binding.* ^{125}I-EGF binding was carried out by incubating the cells with 50 pM ^{125}I-EGF for 24 hr at 4°C, following a previously described protocol[92].

Cell Growth Assay. HeLa cells were grown in Dulbecco's Modified Eagles' Medium with 10% FBS. Cells were plated in triplicate in 96-well plates at 5000 cells per well and allowed to grow for 24 hours before the addition of DMSO (control), erlotinib (Genetech) or NSC56452. All cultures contained 1% DMSO in the final media. Cells were then incubated for 48 hours. The cell growth rate was then measured using the cellTiter 96 Aqueous One Solution Cell Proliferation Assay kit according to the manufacturer's instructions (Promega). Readings were taken at 490 nm after 1 hour incubation with the MTS and PMS solution.

CHAPTER 3. The OpenScreening Project: An Open-Access Virtual Screening Web Server Powered by Distributed Computing Resources and the Xgrid Technology

Abstract

Virtual screening of compound libraries against therapeutic protein targets represents an integral technique in early-stage pharmaceutical drug-discovery. It remains under-utilized within the academic community due to barriers of insufficient technical expertise and computing resources. We aim to address this need by making the technique easily accessible to a broader audience especially those without computational backgrounds. Towards this goal, we implemented the OpenScreening Project, an open-access virtual screening web server powered by a grid-computing infrastructure that integrates an open-source grid manager (GridStuffer) with Xgrid-technology (Apple Inc). OpenScreening is currently powered by the world's largest Xgrid cluster of volunteer CPUs and represents the first open-access virtual screening web application. Functional since September of 2008, over 80 projects have been completed using CPUs from close to 500 client machines. OpenScreening can be freely accessed at http://openmacgrid.org/xvHTS.

Introduction

Rapid progression in proteomics[93-95] has been and will continue to supply a wealth of novel therapeutic targets for drug discovery. Because the process of translating target identification into drug discovery still rests in the hands of pharmaceutical companies, many academic findings of novel targets, particularly those involving rare diseases, suffer a time-delay in making therapeutic impact. One reason for the apparent

bottleneck is that large-scale high-throughput screening (HTS) of existing compound libraries remains the main strategy for identifying lead compounds in the current drug discovery paradigm[96-98]. However, given the typical 1 out of 100,000 hit rate from random screening[98], and the pre-requisite investment in developing target-specific high-throughput assays, wide application of industry-comparable HTS at the academic level remains impractical to most.

Towards alleviating the dependency on HTS, virtual high-throughput screening (vHTS) has emerged as an integral component in drug discovery for prioritizing compounds[99, 100]. The application of vHTS can enrich the hit rate by several orders of magnitude[101] and, therefore, offers an attractive alternative when high-throughput assays or screening facilities are inaccessible[102-104].

A pre-requisite requirement for running vHTS is its intrinsic demand of computing resources, a "bottleneck" that explains why, despite its popularity in industry, the technique remains under-utilized within the non-computational academic community. One emerging solution for accumulating computing power is through distributed computing resources[105-107], and its utility has been demonstrated by the remarkable successes of several recent vHTS applications. These applications were tailored for high-profile projects but were not designed to support community-wide access—utilization of these computing resources is restricted. To address the "bottleneck" by providing unrestricted distributed computing resources in vHTS applications, we have developed OpenScreening, an open-access vHTS web server that operates on idle CPUs harvested from the community. The aims of the project were to (1) provide a validated user-friendly vHTS tool that lowers the barrier to entry, and (2) present a proof-of-concept computing

grid   prototype   easily   transferable   for   embarrassingly   parallel   applications. OpenScreening can be freely accessed at http://openmacgrid.org/xvHTS

Results

Overall workflow. The grid infrastructure is constructed to connect remote client machines (volunteers) over internet to a central Mac OSX server. Server-client communications are established by the Xgrid hardware (Apple Inc) while the trafficking of the communication is maintained by an open-source software, GridStufferr (http://cmgm.stanford.edu/~cparnot/xgrid-stanford/html/goodies/GridStuffer-info.html). Upon submission of a new screening project via the web interface, input files are pre-processed to ensure format consistency and placed in queue for execution on the grid. OpenScreening   currently   operates   on   the   OpenMacGrid (www.macresearch.org/openmacgrid), the world's largest Xgrid cluster. During execution of a screening project, gstuffer dispatches onto available (idle) client machines a set of concurrent Autodock $4.0^{76, 77}$ computations each corresponding to the docking of a compound in the library to the target protein. The process of submission, data retrieval, and re-submission is iterated until the entire library has been screened. Upon completion, the compound ranking and the predicted 3D docked conformations are returned as output (Figure 3.1).
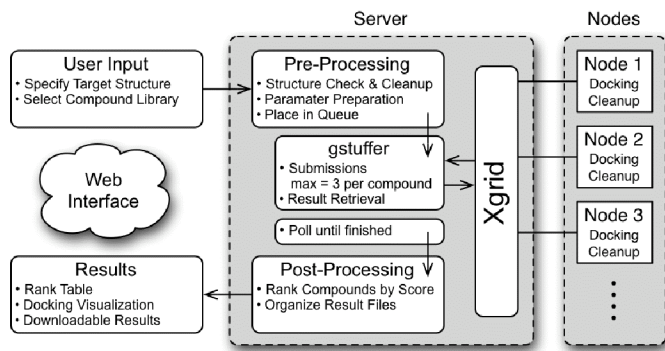
Figure 3.1: Overall workflow of the OpenScreening web server. Inputs and outputs of the screening projects are submitted and retrieved via the web interface, respectively. Pre-processing, post-processing, and analysis are performed by the server scripts. Parallel docking runs are executed on the OMG grid. Grid trafficking and control are handled by GridStuffer.

Project submission. In the simplest case, a 4-letter PDB code is sufficient for submitting a new project. In many cases, however, additional insights about specific binding sites of the target protein are often available for guiding efficient site-directed virtual screening. To accommodate this, the submission interface incorporates an embedded Jmol[108] applet to support interactive visualization of the protein and docking pocket definition (Figure 3.2). Screening of the NCI-diversity (size 1,990) and Maybridge-diversity (size 14,400) compound libraries are currently supported, though uploading of customized databases is possible upon request. OpenScreening currently uses Autodock 4.0 as the backend docking software, which has been shown to be effective in many cases[107, 109, 110]. The choice of run-time parameters is instrumental to success, thus represents a major source of uncertainty among novice users. To facilitate in this aspect, we have empirically derived a set of parameters that displayed a good balance between enrichment power and speed (see below). These parameters are designated as the default parameters for OpenScreening. Users are provided with the option of screening with these default parameters or customized screening parameters. Each project is assigned a password at submission, which is required for subsequent project cancellation and/or data retrieval.
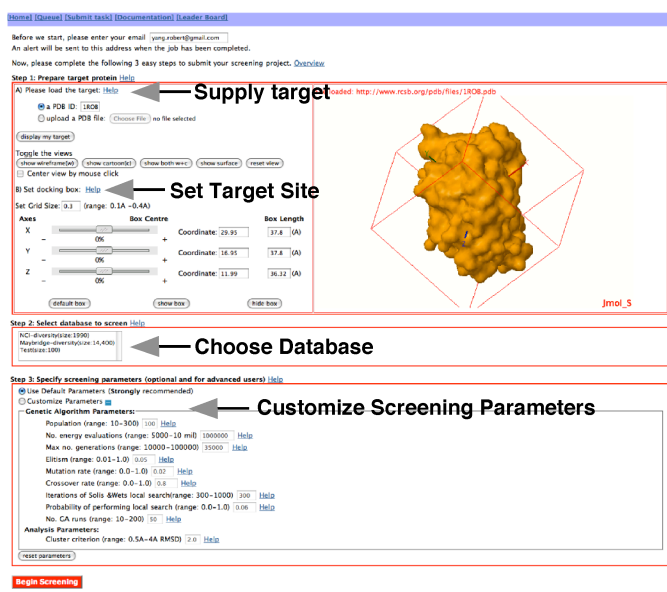
Figure 3.2: Submission interface. 3 easy steps to submit a screening project: 1) visualization and specification of the docking region of the protein target; 2) compound library selection, currently support the NCI-diversity and Maybridge HitFinder libraries. Screening custom libraries available upon request; 3) customization of docking parameters for advanced users (default parameters are strongly recommended for most users).

Result retrieval. Internally, each project is split into dockings of individual compounds to the target protein, which will be independently executed on client machines. These processes can only be executed under the Xgrid partition pre-designated by Macintosh under the lowest privilege to ensure absolute integrity of the client machines. Upon completion, only the result files are transferred back to the server while input and intermediate files are properly deleted from the client machines. To the end-users, the rankings of the top 100 compounds and the visualization of their best-predicted conformation are presented on the result page (Figure 3.3). The complete results, which include the rankings of all compounds and the complete sets of docked conformation in multi-mol2 format, can be downloaded on the same page for further analysis offline.
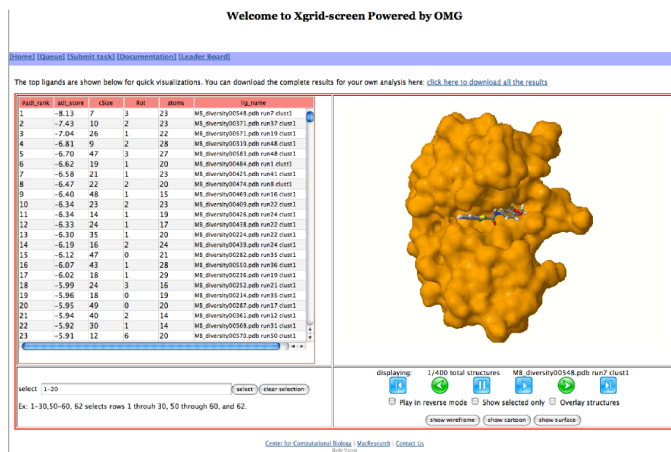
Figure 3.3: Result retrieval interface. The ranking and the corresponding docking pose of the compounds are presented in tabular and graphical forms.

Input preprocessing. Input structure files submitted to the central server through the web interface are first validated to ensure security integrity and format compatibility. They are used to generate a set of protein descriptions in Autogrid maps, the required format that feeds into subsequent docking procedures. In theory, docking maps of the target protein only need to be generated once on the server, transferred to and shared among all client machines during client-side execution. However in practice, it is more efficient to recalculate these maps on-the-fly on client machines rather than through transmission to alleviate the burdens on internet bandwidth. For example, screening against a target box of $25x25x25(\text{Å}^3)$ with grid spacing of 0.2 Å requires 150 megabytes of map files (the map for each atom type is typically 12 megabytes), or 50 megabytes in the zipped format. Repeated transmission of 50 Mb data for every docking process is not practical due to the accumulated amount of input/output communication involving the central server— minimizing server-client transmission is critical for grid stability and growth. We also adhere to the policy of strict cleanup on the client machines and choose not to store these map files remotely. In our implementation, maps are generated once on the server at the pre-processing stage (before putting in the queue), and are repeatedly generated on the clients for every docking calculation during execution. The purpose of

the one-time server-side map generation, besides testing for correct format, provides a benchmark time that will be used to gauge the relative speed of the individual client machines. During a client-side docking calculation, the first step is to generate the maps. The server-side benchmark time is passed along to the clients— machines requiring 5 times longer than the "benchmark" in constructing maps are deemed inefficient and will not proceed to execute the much more computationally intensive docking calculations. The decision to not utilize slow machines on the grid currently is based on the observation that they typically are the last to finish and end up significantly delaying the completion of the entire project. Our decisions to not exclude slow machines on the grid in the first place hinges on the understanding that they are useful for less intensive tasks—for example some machines are not "slow" at calculating tasks that do not require high cache memory. We also reasoned that dynamically testing the speed on-the-fly, rather than keeping a local record about the client machines, is advantageous because the former better accommodates grid flexibility and does not prematurely excludes clients that underwent upgrades.

Parameter sensitivity analysis. A sensitivity analysis was carried out to evaluate and determine the performance of a range of Autodock parameters based on accuracy and time performance. The PDBbind "core set"[111, 112] consisting of 69 non-redundant high-resolution protein-ligand complexes was used for re-docking-- each ligand was extracted from the original complex, their torsion angles randomly scrambled, and re-docked back to its binding protein using different docking parameter combinations (Table 3.1). The docking boxes were centered at the ligand in the crystal structures, and several box dimensions were examined in the analysis. The accuracy was defined by 1) sampling: the

resemblance between the re-docked poses and the original crystal poses as measured by the root-mean-square distance (RMSD), and 2) scoring: the relative ranking of these "correct" ligand poses among all generated poses. The RMSD of 2.0 Angstrom[113] is typically the accepted threshold for defining a successful recovery by the docking protocol though RMSD values in the range of 2.5Å-3.0Å[114, 115] have also been commonly used. For completeness, RMSD thresholds of 1.0, 1.5, 2.0, 2.5, and 3.0 Å were included when determining the success rates of recovering the crystal poses over 69 complexes (Figure 3.4, Table 3.2).

Given the number of parameters (6), systematic scanning of each parameter was impractical—for example, scanning with an interval of just 3 values per parameter would have yielded 3^6 x 69 > 50,000 docking computations over 69 complexes. As an alternative, we chose to test a set of parameter combinations that we have had success with in the past against various targets. For comparison, the default values given by the AutoDockTools[116] were also included as a reference (denoted as Set 1). With the exception of Set 1, the differences in sampling accuracy across parameter sets were subtle under stringent requirement of RMSD 1.0 and 1.5Å (< 10%, Figure 3.4). These results not only outline the boundary of parameter-dependent improvement in sampling, but also reiterate the difficulty in docking—it is hard to recover the exact crystal poses. Thus, for the remaining of our analysis, we focused on the performance using thresholds in the range of RMSD 2.0-3.0 Å. Under these less stringent but widely-used threshold values, the best parameter combination (Set 10) achieved a sampling success rate of close to 80% using RMSD 3.0 Å, and around 60% using RMSD 2.0 Å. However, Set 10 required an average of 21500 seconds (~6 h) per complex (docking 1 ligand to 1 target), an execution

time that we considered to be too long and non-ideal for virtual screening using distributed computing resources—the requirement that a remote client machine needs to stay idle for 6 hours will drastically decrease the amount of useable nodes on the grid. In contrast, Set 6 required an average of 1 hour while still displaying good success rates using thresholds of 2.0-3.0 Å. In addition, Set 2, 3, and 7 also showed good balances between time and performance attractive for virtual screening applications. These parameter sets were further analyzed subsequently. All detailed results are summarized in Figure 3.4 and Table 3.2.

In these initial tests, we kept the dimensions of the docking boxes constant (25 x 25 x 25 Å^3, 126 grid points in all 3 dimension with grid spacing of 0.2 Å). We next evaluated the sensitivity of these parameters to the box dimensions and grid spacing (Figure 3.5, Table 3.3, Table 3.4). It was not surprising that all parameter sets were sensitive to the overall box dimensions since smaller docking boxes translated into improved sampling by eliminating search spaces while larger docking boxes had the opposite effects. The apparent insensitivity to the grid spacing (Grid 1 vs. Grid 3 Figure 3.5) was unexpected since smaller grid spacing was expected to provide a higher resolution representation of the protein target and hence better sampling performance. However, considering that Autodock uses the united-atom representation for the protein targets[76], we reasoned that the difference of 0.1 Å in resolution is not enough to have significant impacts in sampling.

Table 3.1. Parameter Sets.

Set #	pop_size	num_evals	elitism_rate	mutation_rate	crossover_rate	run
1*	150	2.5×10^5	0.01	0.02	0.8	10
2	100	2.5×10^5	0.05	0.02	0.8	50
3	100	2.5×10^5	0.05	0.02	0.8	100
4	100	5.0×10^5	0.05	0.02	0.8	50
5	100	1.0×10^6	0.05	0.02	0.8	50
6	100	5.0×10^5	0.05	0.02	0.8	100
7	100	1×10^6	0.05	0.02	0.8	100
8	100	1×10^6	0.05	0.02	0.1	100
9	100	5×10^6	0.05	0.02	0.8	30
10	100	5×10^6	0.01	0.02	0.8	100
11	100	5×10^6	0.05	0.02	0.8	100
12	30	5×10^6	0.01	0.02	0.8	100

* AutoDockTools default

Table 3.2. Performance of using different parameter sets on 69 protein-ligand complexes.

Set #	Avg. Time (Seconds)	RMSD < 1.0Å	RMSD < 1.5Å	RMSD < 2.0Å	RMSD < 2.5Å	RMSD < 3.0Å
1	120.4	27.5	30.4	37.7	52.2	55.1
2	537.8	26.1	44.9	55.1	65.2	69.6
3	1109.0	31.9	43.5	58.0	62.3	72.5
4	1118.1	26.1	37.7	52.2	62.3	65.2
5	2122.8	26.1	40.6	52.2	60.9	65.2
6	2170.7	30.4	40.6	59.4	68.1	71.0
7	4558.6	31.9	43.5	55.1	63.8	72.5
8	4560.2	31.9	44.9	55.1	63.8	68.1
9	6426.8	26.1	36.2	49.3	58.0	60.9
10	21505.8	30.4	43.5	60.9	68.1	75.4
11	21957.5	30.4	43.5	60.9	66.7	72.5
12	23254.8	27.5	42.0	58.0	66.7	68.1

Table 3.3. Grid Definitions.

Grid Definition #	Grid Spacing (Å)	Grid Points	Box Dimensions ($Å^3$)
1	0.2	126 x 126 x 126	25 x 25 x 25
2	0.3	126 x 126 x 126	37.5 x 37.5 x 37.5
3	0.3	84 x 84 x 84	24.9 x 24.9 x 24.9
4	0.3	66 x 66 x 66	19.5 x 19.5 x 19.5

Table 3.4. Percentage of complexes correctly re-docked using several parameter sets under different grid definitions.

Parameter Set	Grid 1	Grid 2	Grid 3	Grid4
		RMSD < 1.0 Å		
2	26.1	20.3	31.9	29.0
3	31.9	24.6	30.4	34.8
6	30.4	29.0	31.9	33.3
7	31.9	26.1	33.3	31.9
		RMSD < 1.5 Å		
2	44.9	31.9	36.2	42.0
3	43.5	33.3	42.0	49.3
6	40.6	34.8	44.9	46.4
7	43.5	40.6	42.0	43.5
		RMSD < 2.0 Å		
2	55.1	43.5	50.7	60.9
3	58.0	43.5	59.4	60.9
6	59.4	46.4	59.5	69.6
7	55.1	50.7	53.6	59.4
		RMSD < 2.5 Å		
2	65.2	50.7	59.4	71.0
3	62.3	55.1	66.7	73.9
6	68.1	58.0	66.7	75.4
7	63.8	56.5	63.8	68.1
		RMSD < 3.0 Å		
2	69.6	53.6	62.3	79.7
3	72.5	62.3	72.5	79.7
6	71.0	60.9	75.4	79.7
7	72.5	66.7	69.6	76.8
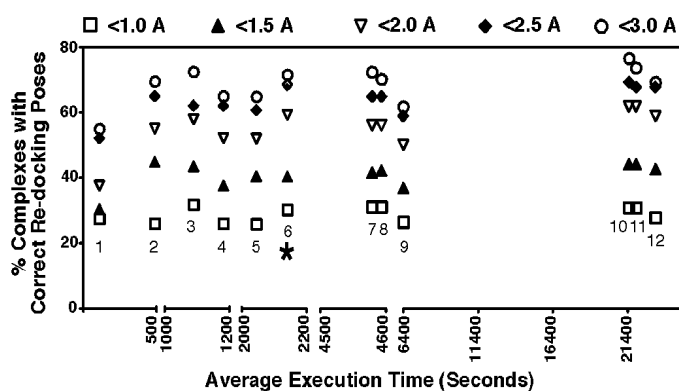
Figure 3.4. Re-docking of 69 protein-ligand complexes to compare sampling accuracy among 12 parameter sets. Sampling accuracy was defined by using a range of RMSD values (1.0-3.0 Å) between the docked ligand poses and the crystal poses. The percentage of complexes with docked poses within the RMSD thresholds of crystal poses are plotted against the average execution time for the docking process. Set 6 displayed a good balance between accuracy and execution time.
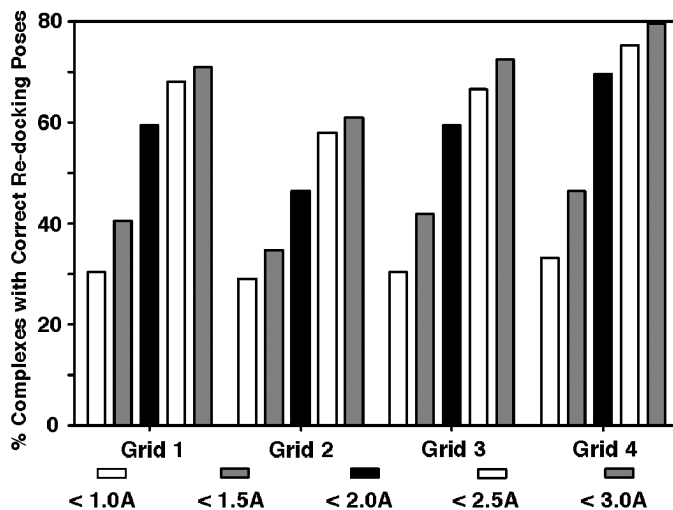
Figure 3.5. Sensitivity of Parameter Set 6 to grid definition. Grid definitions specify the grid spacing and the dimensions of the docking box (see table 3.3). For each definition, the percentage of 69 complexes with at least 1 docking pose within the specified RMSD values of crystal structures are shown in the bar graph.

41

The next level of analysis was focused on the sensitivity to the scoring aspects of Autodock— how well the "correct" docking poses rank relative to the decoy counterparts. All parameter combinations showed good and comparable discriminatory power to favorably rank the "correct" docking poses. On average, these "correct" docking poses were ranked within the top 5-10% of all generated poses on average over 69 complexes (Table 3.5). Given that all but 1 parameters (the "number-of-energy-evaluations" parameter) under consideration deal with the stochastic sampling process (genetic algorithm), these results suggest that the "number-of-energy-evaluations" parameter, despite having little sensitivity in improving scoring, still can impact the docking outcome because of the coupling of sampling and scoring in the genetic algorithm.

Table 3.5. Ranking powers of each parameter sets under different grid definitions.
Average Ranking of the Best-Ranked Correct Docking Pose
Calculated as: (best ranking/number of generated docking poses) x 100%

Param Set	< 1.0 Å*	< 1.5 Å	< 2.0 Å	< 2.5 Å	< 3.0 Å
			Grid 1		
1	8.0	8.0	8.0	11.2	12.6
2	6.6	8.0	8.1	11.3	12.7
3	6.3	10.3	9.8	9.5	12.5
4	4.5	6.4	9.0	10.9	11.9
5	4.6	7.1	8.0	9.7	11.1
6	7.2	7.7	7.9	9.1	9.0
7	8.9	9.3	9.8	11.5	13.4
8	6.6	9.8	8.4	10.4	11.3
9	7.0	7.6	8.4	11.4	11.3
10	7.2	8.8	11.1	12.4	16.3
11	7.5	10.4	10.7	11.1	14.2
12	5.1	8.6	8.3	10.8	10.1
			Grid 2		
2	2.3	4.3	5.3	5.1	5.1
3	2.1	4.4	4.4	6.9	7.9
6	33	4.1	3.9	5.2	6.1
7	4.7	6.8	7.5	7.4	12.5
			Grid 3		
2	6.9	7.0	8.9	10.2	12.1
3	6.9	7.4	9.1	9.2	12.7
6	5.6	5.2	6.3	9.1	9.2
7	5.6	7.4	8.8	9.8	11
			Grid 4		
2	5.6	6.1	9.0	11.7	15.3
3	6.4	8.9	9.1	11.5	13.4
6	7.8	9.9	12.7	14.1	13.7
7	6.9	6.7	8.9	10.6	13.3

* Correct poses are defined by RMSD values between the docked poses and the crystal pose; a range of RMSD thresholds are examined here.
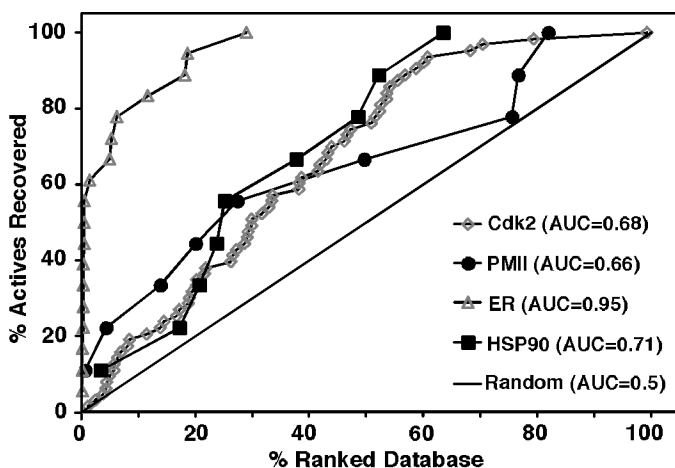
Figure 3.6. Virtual screening of a synthetic compound library (~2000 molecules) against 4 protein targets. Parameters were taken from parameter Set 6. The area-under-curve (AUC) measures the enrichment power in each case.

Further testing of the parameters in the context of vHTS. To evaluate these parameters in a vHTS context, we performed cross-docking experiments using the OpenScreening webserver. A synthetic compound library of 2000 compounds that included known actives was screened against 4 protein targets similar to the protocol described in Chapter 2. The 4 protein targets were plasmepsin II (PMII), cyclin dependent kinase 2 (CDK2), estrogen receptor (ER), heat-shock protein 90 (HSP90). Each protein target had a rich collection of protein-ligand co-crystal complexes in the PDB. These ligands were mixed with 1926 NCI-diversity compounds to form the synthetic compound library. The PDB accession codes used for cross-docking of known ligands to the four protein targets were: plasmepsin II (PMII), 1LEE, 1LF2, 1LF3, 1M43, 1ME6, 1W6H, 1W6I, 1XE5, 2BJU, 2IGX; cyclin dependent kinase 2 (CDK2), 1AQ1, 1CKP, 1DI8, 1DM2, 1E1V, 1E1X, 1E9H, 1FVT, 1G5S, 1GIH, 1GZ8, 1H00, 1H01, 1H07, 1H08, 1H0V, 1H0W, 1H1P, 1H1Q, 1H1R, 1H1S, 1JSV, 1JVP, 1KE5, 1KE6, 1KE7, 1KE8, 1KE9, 1OGU, 1OI9, 1OIQ, 1OIR, 1OIT, 1OIU, 1OIY, 1P2A, 1P5E, 1PF8, 1PKD, 1PXI, 1PXK, 1PXL, 1PXM, 1PXN, 1PXO, 1PXP, 1PYE, 1R78, 1URW, 1V1K,

44

1VYW, 1VYZ, 1W0X, 1WCC, 1Y8Y, 1Y91, 2A4L, 2B52, 2B53, 2B54, 2B55, 2BPM, 2BTR, 2BTS, 2C68, 2C6K, 2C6L, 2C6M, 2CLX, 2FVD, 2UZB, 2UZD, 2UZL; estrogen receptor (ER), 1A52, 1ERR, 1L2I, 1SJ0, 1UOM, 1X7E, 1X7R, 1XP1, 1XP6, 1XP9, 1XPC, 1XQC, 1YIM, 1YIN, 1ZKY, 2AYR, 2B1V, 2FAI, 3ERD, 3ERT; heat-shock protein 90 (HSP90), 1A4H, 1BGQ, 1BRE, 1YC1, 1YC3, 1YC4, 1ZW9, 2BRC, 2CGF. The protein structures with the highest resolution structures in each case were: 1LEE (HSP90), 2B54 (CDK2), 1XPC (ER), and 2BRC (HSP90). All bound small molecules were manually separated from the complex and prepared using OpenBabel v2.0.2[117].

Using the respective highest resolution protein structures, we virtual screened this testing library against all 4 targets with parameter Set 6. The efficiency to rank active compounds higher than the decoy compounds (referred to as the enrichment power) was evaluated by measuring the fraction of active compounds that are present within the top fraction of the entire ranked library. The corresponding enrichment curve (Figure 3.6) plotted the progression of enrichment over a continuous fraction range and was quantified by calculating the area-under-curve (AUC). As shown in Figure 3.6, these parameters displayed a significant trend for favorably ranking the active compounds in all 4 cases—at least one active compound was identified within the top 1-5% of the library. In the best case (ER), half of the actives were recovered within the top 10% of the entire library.

The speed of OpenScreening was assessed by running the same virtual screening experiment against the ER target using a local cluster of 50 nodes. OpenScreening exhibited an overall 10-fold increase in the number of compounds screened at its peak (Fig. 3.7, Table 3.6). Screening ~2000 compounds was completed in less than 24 hours, whereas the same processes took nearly 5 days on the local cluster.
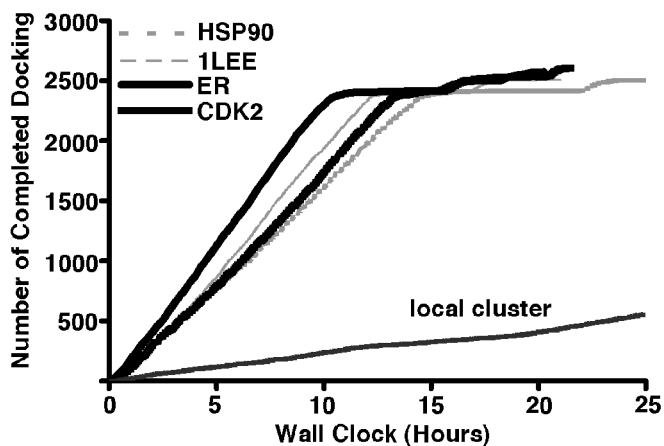
Figure 3.7: Wall-clock time performance of OpenScreening comparing to that of a local computer cluster (50 nodes). The OpenScreening protocol consistently displays 10-fold increase in wall-clock time over the local cluster at its peak. The plateau reflects the bottleneck effects of slow client machines on the grid.

Table 3.6. Bench mark of CPU and wall-clock time.

Screening Target[a]	Total CPU hours	Total Docking Runs[b]	Average CPU hours per Docking Run	Total Wall-clock hours
CDK2	1278.3	2607	0.49	19.1
ER	1350.5	2573	0.52	22.8
PMII	1442.5	2508	0.58	21.7
HSP90	1479.8	2504	0.59	22.9
ER_local cluster	985.2	2415	0.41	116.3

[a] CDK2, cyclin-dependent kinase 2; ER, estrogen receptor; PMII, plasmepsin II; HSP90, heat shock protein 90; ER_local cluster, the same screening project against the estrogen receptor performed on a local computer cluster (the average number of running nodes during the execution was 25).
[b] Total docking runs included resubmissions.

Usability. Over 80 projects spanning several research groups, most having no computational backgrounds, have been performed by OpenScreening since being functional in September of 2008. On average, screening of the NCI-diversity library takes overnight, and screening of the Maybridge library takes 5 days. We have been rigorously testing the stability of OpenScreening in anticipation of a rapid increase in user base upon publication.

Prototype for Xgrid-based applications. VHTS, like all "embarrassingly parallel" applications, has a repetitive nature that demands robust traffic-coordination algorithms that exceed the capability of the primitive Xgrid scheduling system alone. In addition,

intrinsic dependence on remote client-side executions renders these grid applications susceptible to high-error rate that often requires extensive human interventions[118]. In our design, three independent layers of controls by Xgrid, GridStuffer, and the vHTS wrapper script were implemented to address these challenges. Through the Xgrid server-client socket, GridStuffer transfers the Autodock executable, the protein and compound structure files, along with the wrapper script to the Xgrid-designated partition on the client machines. Upon completion, only the result files are transferred back to the server while the rest of the files are properly deleted. The decision to decouple these layers allowed us to independently address, at different abstraction levels, problems that are inherent to distributed computing and specific to vHTS. Xgrid handles errors associated with server-client transmission and notifies GridStuffer of the communication status of individual clients. Based on these client statuses, GridStuffer schedules submission, result- retrieval, and resubmission of failed tasks which are flagged by the return value of the vHTS wrapper scripts. Given that Xgrid and GridStuffer were both designed to be application-agnostic, this protocol can be readily adapted for other embarrassingly parallel applications by changing at the application-specific wrapper script level.

Discussions

The capacity of OpenScreening already exceeds the computing resources of many private corporate counterparts[119, 120], demonstrating the advantage of being supported by the community. However, it clearly cannot compare to the well-established grid-based projects such as FightAIDS@Home[107] (~450,000 nodes) or the Screensaver Project[106] (~3.5 million), nor was it ever our intent. Our goal is to provide a broad research

47

audience with a tool that addresses the need of unrestricted accessibility to vHTS and encourages the integration of computational tools in the non-computing community.

The integration of Xgrid and GridStuffer presents an affordable and easy-to-implement utility to build grids among research labs, departments, institutions, and units of essentially any size. Because Xgrid is an inherent feature of the OS X platform, grid expansion does not require explicit software installations on each machine. Volunteers can join the OpenScreening grid by simply entering the address of the OpenMacGrid grid server (step-by-step instructions on the website). However, this also restricts the client machines to be only Mac machines running OS10.4 or later. The development of middleware to include PC machines is anticipated to further expand the grid size, though efficient recruitment of PC machines would most likely come in the form of pre-existing clusters rather than individual volunteers. The requirement to explicitly install client-side communication portal compatible with the PC middleware demands a level of complication that is better suited for experienced system administrators. Some of the viable open-source middleware frameworks include ACE and TAO from the Center of Distributed Object Computing (http://www.cs.wustl.edu/~schmidt/doc-center.html), and Opal from National Biomedical Computation Resources (http://www.cs.wustl.edu/~schmidt/doc-center.html). Alternatively, collaborations with the drugdiscovery@home project (http://drugdiscoveryathome.com/) that uses the Berkely Open Infrastructure for Networked Computing (BOINC) may be possible if the concept of unrestricted open-access can be preserved.

Parameter Set 6 was set as the default parameters for the OpenScreening webserver for its robust sampling and ranking powers as well as its fast execution time.

Most of the other parameters tested in this chapter also did not result in significant deviations in performance. In particular, Set 2 and Set 3 displayed ¼ and ½ average execution time comparing to that of Set 6, respectively. These parameters therefore are expected as good candidates for screening larger libraries without significantly compensating for accuracy.

Shortcomings in the current implementation will be improved in future updates. One example is the rate-limiting effect of slow client machines on the overall performance. Stagnant performance towards the end stage of a screen project (Figure 3.7, number of docked compounds plateaus towards the end) can dramatically prolong the overall wall-clock time. Despite a 10-fold increase at its peak over the local cluster (Figure 3.7), the increase in the overall wall-clock time to complete a screening project was only 5-fold (Table 3.6, ER vs ER_local). One way to address this issue is to increase the map generation "benchmark time" thresholds from its current 5x to 10x in order to eliminate the use of slow machines. A more elegant solution under testing, however, is to trigger redundant scheduling of multiple docking instances of the same ligand onto several client machines concurrently towards the end of a screening project—the first successful results from these duplications will abort the other instances, thus completing the project.

In addition to using Autodock, we are also exploring the possibility of including additional docking softwares and scoring functions to the OpenScreening webserver. Docking softwares using systematic search algorithm and/or incorporating receptor flexibility may improve the current sampling efficiencies of 60-80%. The benefit of using multiple scoring functions is apparent by comparing the enrichment power of using the

Autodock scoring function alone (Figure 3.6) and that of using the consensus scoring

scheme described in Chapter 2 (Figure 2.1). We expect that by providing the options of a

wide range of docking and scoring softwares, OpenScreening will improve its accuracy

as well as extend its utility to a larger user-base. As we aim to implement additional

components and functions onto OpenScreening, our long-term goal is to stimulate the

integration of computational tools in generating reasonable hypothesis to help with

designing and performing subsequent experiments.

Appendix 3.1: Instructions to set up Xgrid and GridStuffer.
Equipment.

- Mac OS X Server 10.4 or later.

- GridStuffer freely available for download at

  http://cmgm.stanford.edu/~cparnot/xgrid-stanford/html/goodies/GridStuffer-

  info.html.

Equipment setup.

- Configure server according to manual.

- Install GridStuffer and set it to run as a launchd application:

- > launchctl load /Library/LaunchDaemons/org.openmacgrid.xgridstatus.plist

- > launchctl load /Library/LaunchDaemons/org.openmacgrid.gstuffer.plist

Applendix 3.2: User manual of the OpenScreening webserver.

Submission to OpenScreening (Figure 3.2)

1 From the OpenScreening homepage, navigate to the "submit task" tab or go directly to the website (http://omg.phy.umassd.edu/xvHTS/submit/). Accept the applet certificate for Jmol when prompted, this allows the visualization using the Jmol applet.

2 Specify an email address for notification of the job status.

3 Supply a valid target structure in PDB format either by specifying its PDB code or uploading a file. Structure must be displayed to continue.

CRITICAL STEP Validate the format of the file before submitting. It is highly advised that the file only contains coordinates for common amino acids. Ions and other molecules should be deleted from the file. Sometimes water molecules can cause problems as well. Only PDB format is supported at this time.

4 Specify docking site by changing the location and size of the docking box. Small molecules will only be docked onto areas enclosed by the docking box.

5 Select small molecule libraries to screen. Press Shift while clicking to select multiple libraries (depending on individual keyboard setup).

6 Customize screening parameters only if you are an advanced user. Default parameters are highly recommended.

7 Submit and wait for a confirmation number. Successful submissions will be assigned a confirmation number and displayed in the queue.

Cancellation of a project in the queue.

8 From the OpenScreening homepage, navigate to the "queue" tab or go directly to the link (http://omg.phy.umassd.edu/cgi-bin/queue.cgi). Locate the project to be cancelled in

the queue list. Click on the "cancel" link and input the assigned confirmation number when prompted.

Retrieval of screening results (Figure 3.3).

9 From the OpenScreening homepage, navigate to the "queue" tab or go directly to the link (http://omg.phy.umassd.edu/cgi-bin/queue.cgi). Locate the project that has been completed in the queue list. Click on the "view job" link and input the assigned confirmation number when prompted.

TROUBLESHOOTING

Join the grid (voluntary).

10 Follow the link http://www.macresearch.org/contribute_to_openmacgrid for a step-by-step instruction.

Troubleshooting

Table 3.7. Troubleshooting table.

Step	Problem	Solution
3	Structure is not displayed	Make sure it has a valid PDB format. Try another browser with javascript enabled.
7	Submission failed	Delete from the structure file: non-standard amino acids, all HETATM records, ANSIOU records and all alternative coordinates if they exist.
8	Cannot locate the cancel link	Contact us: http://omg.phy.umassd.edu/xvHTS/contact/
	Do not know confirmation number	Contact us: http://omg.phy.umassd.edu/xvHTS/contact/
9	Cannot locate the results	Contact us: http://omg.phy.umassd.edu/xvHTS/contact/
	Do not know confirmation number	Contact us: http://omg.phy.umassd.edu/xvHTS/contact/

CHAPTER 4. Computational Modeling and Experimental Testing of the ErbB2-EGFR Heterodimer Interface

Abstract

Like EGFR homodimers, the formation of ErbB2-EGFR heterodimers is thought to be mediated by inter-molecular interactions between the extracellular domains. Given that the ErbB2-EGFR heterodimers have been associated with elevated transforming potentials, this protein-protein interaction serves as an attractive target to extend the development of dimer inhibitors. However, the molecular interface is unknown in the absence of crystal structures. By using crystal structures of the EGFR homodimer and the ErbB2 monomer as templates, a model of the EGFR-ErbB2 extracellular heterodimer constructed by homology modeling and further refined using molecular dynamics simulation. Comparison between the ErbB2-EGFR heterodimer model and EGFR homodimer revealed subtle difference at the respective dimer interfaces. Based on the heterodimer model, two negatively-charged residues, both unique to ErbB2, were computationally predicted and experimentally confirmed to be critical for the stability of the ErbB2-EGFR interface. The identification of these critical residues and the heterodimer interface model serve as potential target sites for the development of heterodimer inhibitors.

Introduction

ErbB2, a close homolog of the EGFR (64% sequence similarity)[2] in the ErbB receptor family, is the preferred heterodimerization partner of all other ErbB family members[48, 121], but rarely homodimerizes among themselves[122-124]. Like EGFR, ErbB2 has a structural domain arrangement that consists of an extracellular domain, a

transmembrane helix, and an intracellular catalytic kinase domain[122]. In contrast to EGFR and the rest of the ErbB family members, ErbB2 has no known ligand and constitutively adopts the open conformation where the homologous extracellular subdomain II, termed as the "dimerization arm" in EGFR, is readily exposed for interaction with a partner ErbB receptor[123, 124]. As described in Chapter 2, inter-receptor interactions mediated by the dimerization arm drives the formation of EGFR homodimer[44, 45]. Mutations to critical residues on the arm result in complete disruption of homodimers thus underlining the necessity of preserving this interface[42, 46]. This dimer interface centered on the dimerization arm and its respective docking pocket is thought to be universal for the EGFR-ErbB2 heterodimer and among all other ErbB dimers.

Given that EGFR-ErbB2 heterodimers exhibit enhanced signaling and are often associated with aggressive tumors[68, 125], understanding the molecular details of the heterodimer interface carries important implications for developing cancer therapeutics. Crystal structure of the extracellular EGFR-ErbB2 heterodimer has not been solved most likely due to the presence of heterogeneous populations of homo- and hetero-dimers. Crystal structures of the extracellular EGFR in the homodimeric form and ErbB2 as monomers have been solved recently – the monomeric state of these ErbB2 structures reflects the weak propensity of ErbB2 to homodimerize. In an attempt to probe the EGFR-ErbB2 interface, a heterodimer model was built by homology modeling using the crystal structures of the extracellular EGFR homodimer and ErbB2 monomer as templates. The homology model was subsequently refined by an all-atom molecular dynamics (MD) simulation. Computational mutagenesis based on the refined model coupled with experimental site-directed mutagenesis led to the identification of two "hot-

spot" residues located on the ErbB2 dimerization arm that are essential for heterodimerization. The heterodimer model along with the knowledge of the hot-spot residues not only offer structural interpretations of different combinations of dimers, but also serve as roadmaps for the development of heterodimer-disrupting peptidomimetics.

Results

Model refinement

The EGFR-ErbB2 heterodimer homology model was subjected to full-atom molecular dynamic (MD) simulations in explicit water for the purpose of further relieving unfavorable interactions. For reference, the EGFR homodimer crystal structure (PDB 1IVO) was also relaxed using the same MD setup in parallel. In both cases, simulations of these large systems in the short time-frames generated only conformations in the vicinity of a low-energy structure— experimental crystal structure for homodimer, and an energetically stable heterodimer homology model.

Stability and similarity of the global dimer conformations during simulations

Molecular dynamic simulations often require time to relieve unfavorable interactions and equilibrate the starting structures into energetically stable ensembles. Root-mean-square-distance (RMSD) with respect to simulation time is a standard method for separating the initial relaxation phase (rapid increase) from a low-energy local minimum production stage (plateau). Based on the RMSD profiles in Figure 4.1, we chose the production run to start at time 3500 ps for the heterodimer simulation, and 2000 ps for the homodimer simulation.
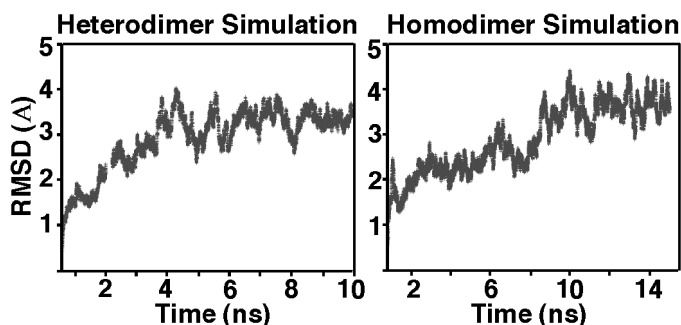
Figure 4.1. Root-mean-square-deviation of homo- and hetero-dimer MD simulations.

The high degree to which secondary structures were preserved from the starting structures during refinement (Figure 4.2) suggested the presence of a low-energy local minimum configuration similar to the starting structures. The number of residues found in well-defined secondary structures was comparable in both homo- and hetero- dimer simulations highlighting an overall similarity in global domain arrangement. These results are consistent with the basis for homology-modeling— the justification to use ErbB2 and EGFR as homology templates because they shared similar global spatial fold. MD simulations, interpreted as relaxations rather than dynamics given the scope of sampling, did not cause significant alterations to the overall global structure during refinement of the starting heterodimer model into a more energetically favorable configuration. Preservation of this global structural similarity between homo- and hetero-dimers can be visualized by a representative snapshot (Figure 4.3).
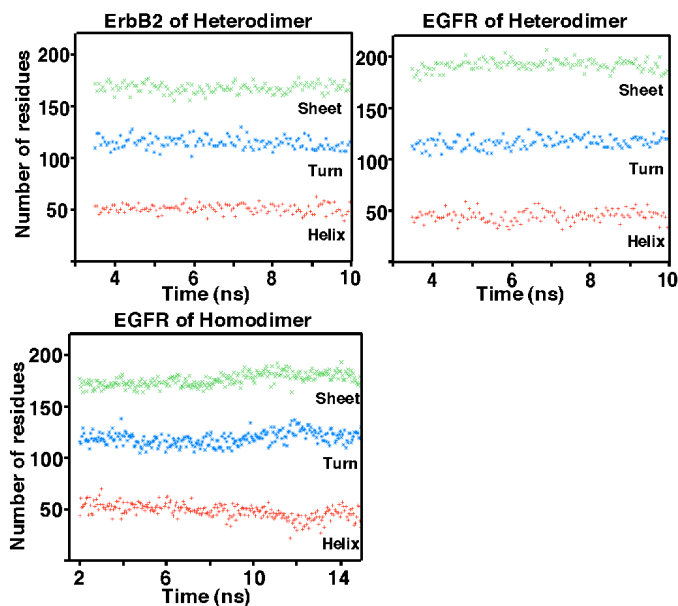
Figure 4.2. Number of residues observed in each secondary structure as a function of time. Values were extracted every 50 picoseconds for clarity.
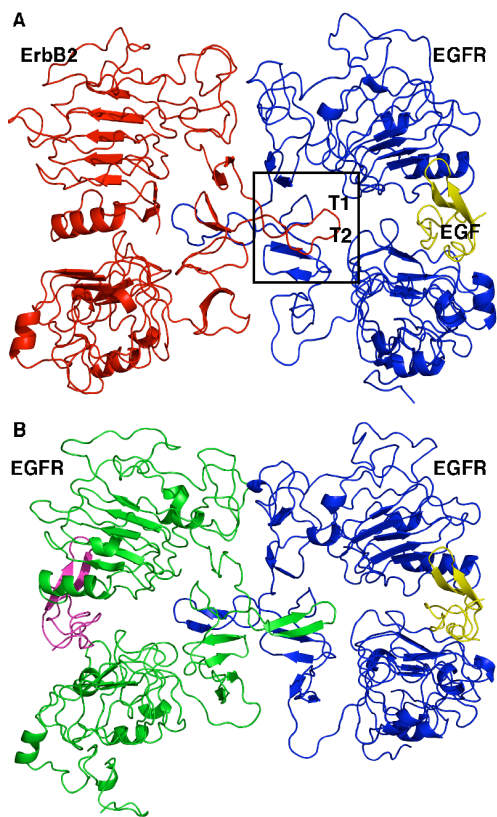
Figure 4.3. Global view of the heterodimer (A) and homodimer (B) extracted from a representative snapshot from MD. The box region indicates the arm interface. The snapshots represent the frames with the lowest energy calculated by the MM/PBSA method. T1 and T2 denote to the location of reverse-Turn 1 and Turn 2.

Differences at the "arm-armpit" interface

The tip of the arm of both ErbB2 and EGFR consisted of five consecutive residues that formed two overlapping reverse-turns at the dimer interface (Figure 4.3). Reverse-turns are among the best-characterized structures for their importance and therapeutic potential as recognition motifs in protein-protein interactions[126]. The molecular conformations of these two reverse-turns, denoted as Turn 1 and Turn 2, were classified and compared at different interfaces: 1) EGFR-arm/ErbB2-armpit, 2) ErbB2-arm/EGFR-armpit, and 3) EGFR-arm/EGFR-armpit. Analyses were carried out using the ensemble of snapshots in the production stage of the simulations, as well as using the static crystal structures (Table 4.1). Both the canonical[127] and Tran[128] reverse-turn definitions were used for classification— the canonical classification is based on backbone dihedral angles while the Tran classification measures the relative side-chain orientations by calculating the $C\alpha$-$C\beta$ vectors.

Table 4.1 Conformation summary of the reverse-turns at different interfaces

	EGFR dimer crystal (PDB: 1IVO)	ErbB2 monomer crystal (PDB: 2A91)	Homodimer simulation	Heterodimer simulation
Canonical Definition				
Turn1 EGFR arm	**IV**	n/a	**IV**(93%), **I**(7%)	**IV**(66%), **I**(34%)
Turn1 ErbB2 arm	n/a	**IV**	n/a	**IV**(99%), **I**(1%)
Tran Definition				
Turn1 EGFR arm	**3**	n/a	**3**(81%), **5**(10%), **1**(9%)	**3**(79%), **1**(10%), **5**(11%)
Turn1 ErbB2 arm	n/a	**3**	n/a	**3**(98%), **6**(2%)
Canonical Definition				
Turn2 EGFR arm	**I**	n/a	**IV**(100%)	**I**(93%), **IV**(7%)
Turn2 ErbB2 arm	n/a	**IV**	n/a	**IV**(97%), **VIII**(3%)
Tran Definition				
Turn2 EGFR arm	**8**	n/a	**8**(66%), **6**(16%), **2**(10%), **5**(7%), **7**(1%)	**8**(81%), **2**(9%), **5**(7%), **6**(3%)
Turn2 ErbB2 arm	n/a	**2**	n/a	**6**(45%), **8**(43%), **2**(1%), **7**(1%)

Turn1 of EGFR arm defined as: 247N 248P 249T 250T
Turn2 of EGFR arm defined as: 248P 249T 250T 251Y
Turn1 of ErbB2 arm defined as: 254N 255T 256D 257T
Turn2 of ErbB2 arm defined as: 255T 256D 257T 258F
Numbers in brackets denote to the percentage of ensembles adopting the corresponding classifications

Because of limited sampling, we restricted our interpretations at the qualitative level. In the crystal structures, Turn 1 of EGFR and ErbB2 both adopted the poorly-defined Type IV backbone conformation by canonical definition—Type IV is defined as "all conformations that do not fit the other Types". The Type IV configuration of Turn 1 was predominantly preserved during homo- and hetero- dimer simulations, although a higher fraction of heterodimer snapshots exhibited variations in adopting dihedral angles characteristic of the well-defined Type I configuration. This observed shift towards ordered backbone configurations may be a product of a less favorable initial heterodimer model and a reflection of the adjustment at this arm-armpit interface. In contrast to the

predominantly "unclassifiable" backbone configurations, the side-chain orientation of Turn 1 could be precisely characterized as Tran Cluster 3 in all cases. This suggested that surface interactions, in both intra- and inter- molecular forms, were well organized and maintained.

Turn 2 of EGFR adopted the Type I configuration in the homodimer crystal structure, but shifted entirely to Type IV during the homodimer simulation. On the other hand, Type I was the predominant configuration for EGFR Turn 2 in the heterodimer simulation, highlighting the difference likely induced by the differences at homo- and hetero- dimer interfaces. Turn 2 of ErbB2 adopted the "unclassifiable" Type IV backbone conformation in crystal structure and stayed "unclassifiable" during heterodimer relaxation. The side-chain orientation of EGFR Turn 2 was preserved as predominantly Tran Cluster 8 in all but one case. The side-chains of ErbB2 Turn 2 had two even Tran clusters during the course of heterodimer simulations suggesting the possibility of relaxed degrees of freedom at the ErbB2-arm/EGFR-armpit interface.
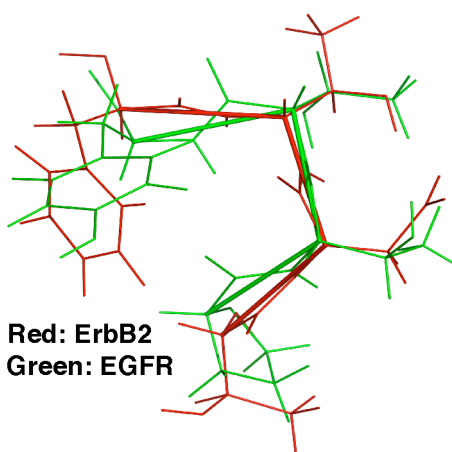
Figure 4.4. Reverse-turn 2 at the tip of the arm extracted from the representative snapshot of the heterodimer simulation. Structures ErbB2 (red) and EGFR (green) were superimposed to outline the difference in both the backbones and side-chains.

To summarize, the reverse-turns of the EGFR and ErbB2 arms adopted different structural configurations at different interfaces. Figure 4.4 shows an example snapshot to compare Turn 2 of ErbB2 with that of EGFR in the heterodimer simulation. Despite

global structural similarity, the reverse-turns at the tip of the dimerization arm exhibited subtle configuration differences. We rationalized that these differences in arm configuration were likely dictated by the nature of the interaction environment provided by the armpit counterparts.

To test this rationalization, inter-receptor contact maps centered on the EGFR and ErbB2 arms were examined in homo- and hetero-dimer simulations (Figure 4.5). In the heterodimer complex, the EGFR arm was shown to contact a larger set of ErbB2 armpit residues than the ErbB2-arm/EGFR-armpit counterparts (Figure 4.5A). This is an indication that despite global symmetry (Figure 4.3), the dimer interface is asymmetric in nature where the EGFR arm extends into a tighter pocket that result in an increased level of heterodimer receptor contacts. This is in contrast with the relatively symmetric interface observed in the homodimer complex (Figure 4.5B). Among the four examined cases, the ErbB2 arm showed the fewest number of contacts with its partner.
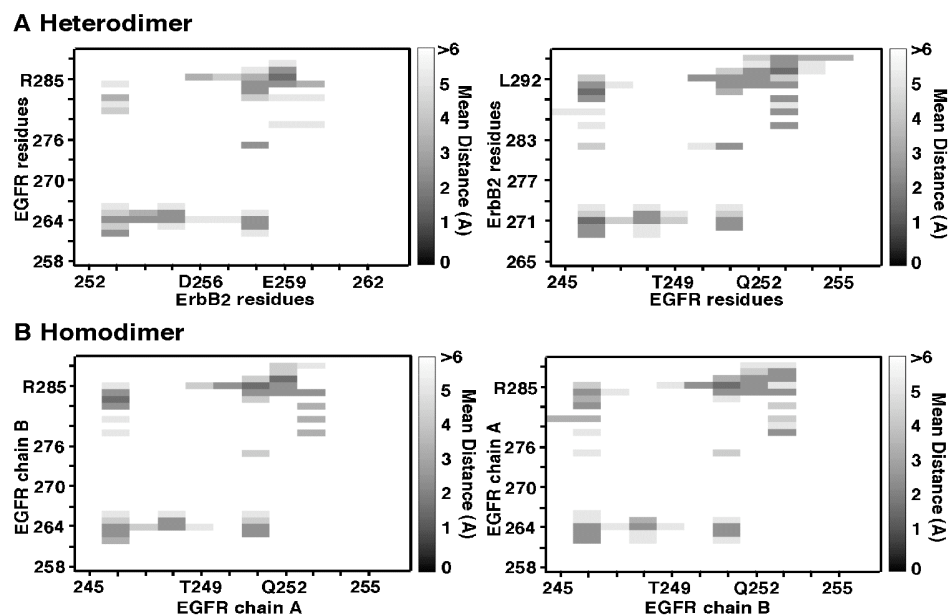
Figure 4.5. Contact maps of A) heterodimer and B) homodimer simulations. In each panel, the X-axis shows residue numbers from the corresponding arm region. The Y-axis shows residue numbers from the respective dimer partner.

Buried solvent-accessible surface area (SASA) was calculated at the interface to further characterize the nature of these inter-receptor contacts (Figure 4.6). Consistent with the contact maps, the EGFR-arm/ErbB2-armpit interface qualitatively had more buried SASA than that of the ErbB2-arm/EGFR-armpit interface. Most of the interactions were hydrophobic in nature. Likewise, the homodimer interface was also dominated by hydrophobic interactions, but had much less buried SASA (Figure 4.6B).
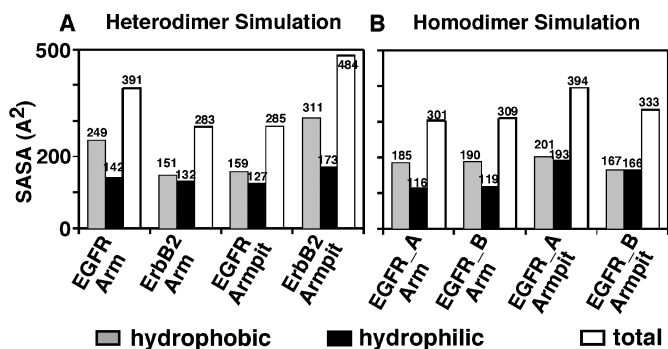
Figure 4.6. Solvent-accessible surface areas of the arm and armpit regions from A) heterodimer simulation and B) homodimer simulation. EGFR_A and EGFR_B denotes to chain A and B in the homodimer.

Based on the reduction in buried hydrophobic SASA at the ErbB2-arm/EGFR-armpit interface, we visualized a few snapshots at this interface and noted the presence of two negatively-charged residues, D256/E259, located at the tip of the ErbB2 dimerization arm. Sequence alignment of the dimerization arm region among all ErbB family members showed that both D256/E259 residues are unique only to ErbB2 (Figure 4.7). We therefore wondered if the two charged residues on the ErbB2 arm may compensate the reduction in intermolecular contacts by participating in favorable salt-bridge and hydrogen-bonding networks.

The presence of hydrogen-bonding network around the ErbB2 arm (residue 252-262) was analyzed. A positively charged residue, R285 in the EGFR armpit was identified as a hydrogen-bond partner with both D256 and E259 in majority of the snapshots. This suggested that these interactions are favorable within the scope of our sampling and consistent with the logic that these interactions can compensate for the loss in hydrophobic interactions. As shown visually in Figure 4.8, the side-chain carboxyl of D256 on the ErbB2 arm forms a salt-bridge interaction with the side-chain guanidino group of R285 in the EGFR armpit, while the side-chain carboxyl of E259 interacts with the backbone amide of R285 (Figure 4.8A). By comparison, the EGFR arm extends deeper into the ErbB2 armpit, thus creating additional surface contacts that, without charged residues, are largely hydrophobic interactions (Figure 4.8B). In the homodimer complex, the presence of R285, but not the corresponding D256 and E259, restricts the size of hydrophobic contact surface and weaker hydrogen bonds are formed to replace the salt-bridge (Figure 4.8C). Interestingly, R285 is present in all ErbB receptors except for ErbB2 where it is substituted with a leucine, a hydrophobic residue (Figure 4.7). It is

possible that this may contribute to explain why ErbB2 receptors do not homodimerize.

```
Arm region                  ↓   ↓
EGFR  246 - YNPTTYQMDVNPEGK - 260
ErbB2 253 - YNTDTFESMPNPEGR - 267
ErbB3       YNKLTFQLEPNPHTK
ErbB4       YNPTTFQLEHNFNAK

Armpit region        ↓
EGFR  282 - SCVRACG - 288
ErbB2 289 - SCTLVCP - 295
ErbB3       SCVRACP
ErbB4       SCVRACP
```

Figure 4.7. Sequence alignment of all 4 ErbB family members around the arm and armpit regions. Arrows indicate substitutions unique to ErbB2. Using ErbB2 residue numbers, they are D256/E259 and L292.
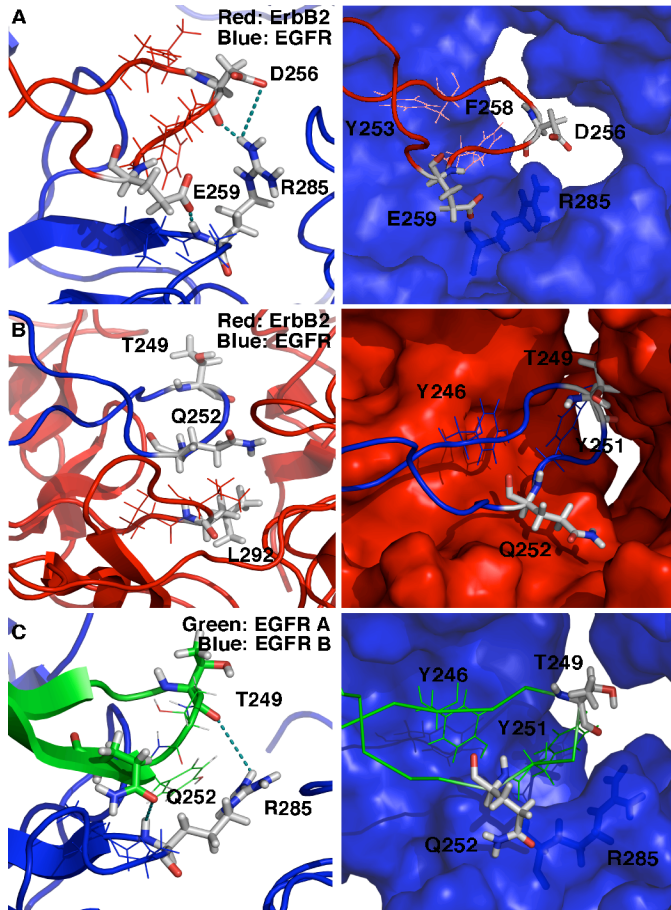
Figure 4.8. Representative snapshot of the arm-armpit interface between A) ErbB2 arm-EGFR armpit, B) EGFR-arm/ErbB2-armpit; and C) EGFR arm- EGFR armpit. A and B are from the heterodimer simulation while C comes from the homodimer simulation.

Computational mutagenesis

To further elucidate the molecular basis of preference for formation of the ErbB2-EGFR heterodimer, we extended the analysis of the dimer models by introducing mutational perturbations to map out residue contributions critical for dimer stability. This analysis is based on the concept that protein interfaces are often dominated by the energetic contributions from only a handful of "hot-spot" residues[129, 130]— as already seen in the mutation studies of Tyr-246/Tyr-251 in the context of EGFR homodimer. Two computational methods, independent in theory and different in computational costs, were utilized to calculate a consensus prediction of mutational effects. The MM/PBSA[131] method calculates the relative binding free energy by decomposing the overall free energy into individual components; in contrast, the Rosetta[132] protocol predicts the relative stability of the dimer complex using fast empirical scoring functions. In both cases, the results were averaged over an ensemble of snapshots extracted randomly from each simulation. It is important to emphasize that these calculations were based on limited sampling with the intention of assessing the stability rather than dynamics of the complexes. As explained below in the control assessments, the interpretations will be strictly at the qualitative level.

Computational mutagenesis by MM/PBSA

The predictability of the protocol was evaluated by conducting calculations on a set of experimentally characterized mutants of the EGFR homodimer[42, 46]: Q194A had little effect on dimer formation; D279A/H280A significantly but not completely disrupted dimer formation; the Y251 and Y246 mutants completely abolished dimer formation. Qualitatively, MM/PBSA predicted the rank order of the mutants in excellent

agreement with the experimental ranking (R^2=0.94) (Figure 4.9A). Although MM/PBSA

has been previously reported to quantitatively reproduce experimental binding energy, we

did not expect the same accuracy due to limited sampling and, therefore, restrict all of our

subsequent interpretation at the qualitative level.
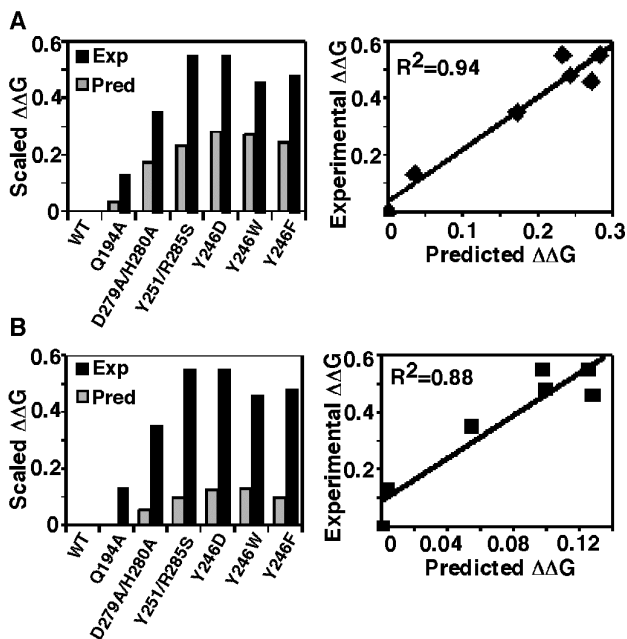
Figure 4.9. Comparison of the predictive powers between A) MM/PBSA and B) Rosetta protocols using mutants with known experimental values. Experimental values were obtained from literatures described in text.

Based on the contact map and hydrogen-bonding analysis, 20 mutants (Table 4.2

and 4.3) around the arm-armpit interface were computationally constructed and

calculated to compare with the wildtype homo- and hetero- dimers. All heterodimer

mutants were predicted to have detrimental effect on dimerization (Table 4.4, Figure

4.10) consistent with the initial design rationale that weakening the electrostatic

interactions at the ErbB2-arm/EGFR-armpit interface and the hydrophobic interactions at

the EGFR-arm/ErbB2-armpit interface would disrupt dimer formation. Qualitatively, the

D256R/E259K mutation on the ErbB2 arm (mut4) and the T249D/Q252E mutation on

the EGFR arm (mut7) were predicted to have the largest detrimental effects. This

supports the rationale that the former shifts a previously favorable electrostatic

interaction with residue R285 of the EGFR armpit to an unfavorable repulsive interaction, while the latter introduces hydrophilic residues into an environment dominated by hydrophobic interactions. Likewise, unfavorable mutations at the corresponding armpit regions (mut10: R285L in EGFR and L292R in ErbB2) were predicted to significantly weaken dimerization.

In contrast, the equivalent electrostatic or hydrophobic interactions are absent at the EGFR homodimer interface (Figure 4.8C). Several putative enhancing mutants aimed at restoring electrostatic or hydrophobic interactions were included to investigate the possibility of interface optimization (Table 4.3). Despite our design, all mutants were predicted to be detrimental in nature by MM/PBSA (Table 4.5 and Figure 4.10A). Qualitatively, mutants 2 and 4 were predicted to have lesser detrimental effect consistent with the initial design purpose to retain hydrophilic-hydrophobic balance. Mutant 6 and 8 were designed to be enhancing by mimicking the ErbB2 armpit, but were instead predicted to have a small detrimental effects. The other putative enhancing mutants, 3 and 7, designed to mimic the charged ErbB2, were predicted unfavorable by MM/PBSA. A detailed look at the energy contributions showed that it was actually the weakening of the electrostatic interactions that contributed the most to the unfavorable binding energy in both cases (Table 4.4).

One possible explanation for the inconsistency between MM/PBSA predictions and our hypothesis could be due to insufficient minimization of the local environment during mutant calculations. As hinted by our earlier analysis on reverse-turns, dynamics of these interface residues likely involved a combination of backbone and side-chain flexibilities. Since our MM/PBSA protocols kept the backbones constant, substitution of

68

a neutral, small wildtype residue with a long, charged residue (T246D) was unlikely to have enough space to relieve steric clashes and establish hydrogen-bonding networks as hypothesized. Towards testing this theory and improving accuracy, we repeated the same sets of mutant calculations using the BACKRUB module[132] of Rosetta, which implemented the backrub motion[133] to account for side-chain-mediated backbone flexibility.
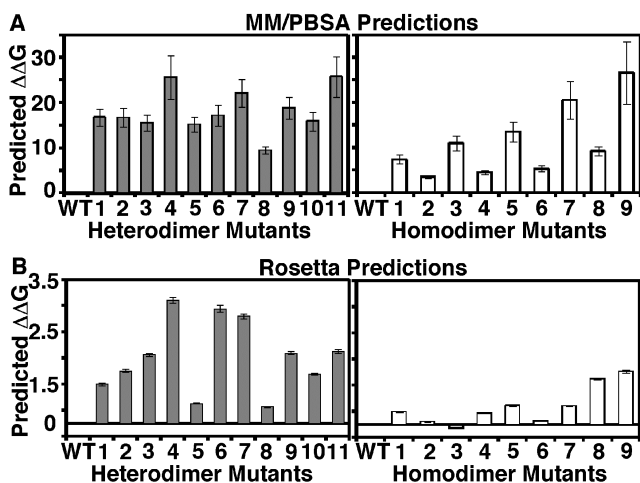
Figure 4.10. Computational mutagenesis using homo- and heterodimer simulations. A) MM/PBSA predictions and B) Rosetta predictions.

Table 4.2. Description of heterodimer mutants

Mutation ID	Mutation details	Region	Rationale	Hypothesis
mut1	B2:D256A/E259A	ErbB2 arm	Alanine-scan	Weaker ErbB2 arm –EGFR armpit interface
mut2	B2:D256N/E259Q	ErbB2 arm	Mutate to non-charged equivalence	Weaker ErbB2 arm –EGFR armpit interface
mut3	B2:D256T/E259Q	ErbB2 arm	Mutate to EGFR equivalence	Weaker ErbB2 arm –EGFR armpit interface
mut4	B2:D256R/E259K	ErbB2 arm	Charge reversal	Much Weaker ErbB2 arm –EGFR armpit interface
mut5	B2:L292R	ErbB2 armpit	Mutate to EGFR equivalence	Weaker EGFR arm –ErbB2 armpit interface
mut6	B2:T291V/L292R/V293A	ErbB2 armpit	Mutate to EGFR equivalence	Weaker EGFR arm –ErbB2 armpit interface
mut7	B1:T249D/Q252E	EGFR arm	Mutate to ErbB2 equivalence	Much weaker EGFR arm –ErbB2 armpit interface
mut8	B1:R285L	EGFR armpit	Mutate to ErbB2 equivalence	Weaker ErbB2 arm –EGFR armpit interface
mut9	B1:V284T/R285L/A286V	EGFR armpit	Mutate to ErbB2 equivalence	Much weaker ErbB2 arm –EGFR armpit interface
mut10	B2:L292R, B1:R285L	EGFR and ErbB2 armpit	Mut5 + Mut8	Weaker at both interfaces
mut11	B2:D256T/E259Q, B1:T249D/Q252E	EGFR and ErbB2 arm	Mut3 + Mut7	Weaker at both interfaces

Table 4.3. Description of homodimer mutants

Mutation ID	Mutation details	Region	Rationale	Hypothesis
mut1	T249A/Q252A	EGFR arm (1 monomer)	Alanine-scan	Weaker interface
mut2	T249N	EGFR arm (1 monomer)	Non-charged mutation	Weaker or little effect on interface
mut3	T249D/Q252E	EGFR arm (1 monomer)	Mutate to ErbB2 equivalence	Stronger interface
mut4	R285L	EGFR armpit (1 monomer)	Mutate to ErbB2 equivalence	Weaker or little effect on interface
mut5	T249D/Q252E/R285L	EGFR arm and armpit	Mut3 + Mut4	Weaker interface
mut6	V284T/R285L/A286V	EGFR armpit (1 monomer)	Mutate to ErbB2 equivalence	Stronger interface
mut7	Double T249D/Q252E	EGFR arm (both)	Mut3 on both monomers	Stronger interface
mut8	Double V284T/R285L/A286V	EGFR armpit (both)	Mut6 on both monomers	Stronger interface
mut9	Double T249D/Q252E/R285L	EGFR arm and armpit	Mut5 on both monomers	Weaker interface

Table 4.4. MM/PBSA predictions on heterodimer mutants

	coul	Vdw	Polar	Apolar	coul+pol	vdw+apol	coul+vdw	pol+apol	Total	Delta
wt	-184.0	-84.8	247.9	-20.1	64.4	-104.9	-265.9	227.4	-39.7	0
	(7.3)	(1.1)	(7.7)	(0.2)	(2.1)	(1.3)	(7.3)	(7.7)	(1.9)	
mut1	-64.6	-86.3	146.3	-19.5	82.8	-105.7	-151.0	126.5	-22.9	16.8
	(6.4)	(1.2)	(6.5)	(0.2)	(3.0)	(1.4)	(6.2)	(6.4)	(2.3)	(1.9)
mut2	-70.2	-89.8	155.4	-20.1	85.8	-109.7	-160.7	135.0	-23.0	16.7
	(8.4)	(1.1)	(7.9)	(0.2)	(2.8)	(1.2)	(7.8)	(7.8)	(2.6)	(2.1)
mut3	-70.0	-89.0	155.0	-20.0	84.0	-109.2	-159.0	134.6	-24.2	15.6
	(7.5)	(1.3)	(8.2)	(0.3)	(3.4)	(1.4)	(7.0)	(8.1)	(2.4)	(1.7)
mut4	-60.5	-91.1	158.9	-21.0	98.4	-111.9	-151.7	137.9	-14.1	25.6
	(7.3)	(1.3)	(7.6)	(0.2)	(3.3)	(1.6)	(7.1)	(7.5)	(2.5)	(4.8)
mut5	-170.7	-88.9	257.8	-20.1	83.4	-109.2	-258.7	237.6	-24.5	15.2
	(7.5)	(1.3)	(7.9)	(0.2)	(3.1)	(1.5)	(8.5)	(7.7)	(2.3)	(1.6)
mut6	-142.9	-88.1	229.7	-20.0	85.1	-108.2	-232.1	208.7	-22.5	17.2
	(7.4)	(1.2)	(8.8)	(0.3)	(3.9)	(1.5)	(7.4)	(8.7)	(2.7)	(2.3)
mut7	-80.1	-88.1	170.2	-20.1	90.5	-108.1	-167.7	150.5	-17.5	22.2
	(7.1)	(1.3)	(8.9)	(0.2)	(3.2)	(1.4)	(7.1)	(9.0)	(2.2)	(3.0)
mut8	-51.0	-88.4	129.6	-19.8	77.7	-108.2	-139.4	109.6	-30.2	9.5
	(7.8)	(1.2)	(8.0)	(0.2)	(2.3)	(1.3)	(8.3)	(7.9)	(2.0)	(0.8)
mut9	-21.1	-90.0	110.9	-19.7	88.8	-109.9	-110.1	92.9	-20.9	18.9
	(8.1)	(1.3)	(8.1)	(0.3)	(3.6)	(1.5)	(8.4)	(8.0)	(2.5)	(2.4)
mut10	-71.1	-89.2	160.3	-19.8	86.6	-108.8	-160.8	140.5	-23.8	15.9
	(7.0)	(1.6)	(7.2)	(0.3)	(3.1)	(1.8)	(7.4)	(6.9)	(2.9)	(2.1)
mut11	-3.8	-89.3	99.4	-19.9	95.1	-108.6	-92.6	78.9	-13.9	25.8
	(7.3)	(1.5)	(7.7)	(0.3)	(3.4)	(1.7)	(8.2)	(7.4)	(2.3)	(4.4)

Table 4.5. MM/PBSA predictions on homodimers

	coul	Vdw	Polar	Apolar	coul+pol	vdw+apol	coul+vdw	pol+apol	Total	Delta
Wt	-119.0 (14.1)	-83.2 (1.6)	180.8 (13.3)	-18.5 (0.8)	59.8 (3.0)	-102.0 (2.4)	-202.0 (15.7)	160.9 (13.8)	-41.3 (3.3)	0
mut1	-73.6 (5.4)	-84.0 (1.7)	143.5 (7.4)	-17.9 (1.0)	68.2 (4.3)	-102.4 (2.7)	-158.8 (6.4)	125.2 (9.7)	-34.0 (3.7)	7.3 (1.0)
mut2	-120.8 (9.9)	-85.2 (2.3)	183.9 (13.1)	-18.5 (0.5)	65.93 (3.0)	-104.2 (2.5)	-203.6 (13.3)	165.8 (12.8)	-37.9 (2.9)	3.5 (0.4)
mut3	-104.8 (6.8)	-86.6 (2.6)	183.0 (8.7)	-18.4 (0.5)	74.2 (3.8)	-105.0 (3.2)	-191.9 (8.5)	164.9 (8.0)	-30.4 (3.7)	10.9 (1.6)
mut4	-85.0 (7.3)	-83.0 (1.5)	147.9 (8.7)	-18.2 (0.4)	63.8 (3.0)	-101.0 (1.6)	-169.0 (8.6)	131.1 (7.6)	-36.9 (2.5)	4.4 (0.5)
mut5	-91.5 (6.0)	-87.3 (2.6)	169.8 (6.7)	-18.1 (0.4)	77.1 (2.7)	-105.6 (3.2)	-178.5 (7.3)	151.0 (7.0)	-27.8 (3.9)	13.5 (2.2)
mut6	-76.7 (7.5)	-84.5 (1.4)	144.5 (10.8)	-18.1 (0.8)	65.3 (4.1)	-102.8 (2.6)	-163.8 (8.7)	124.6 (12.0)	-36.1 (3.3)	5.3 (0.6)
mut7	-120.2 (7.5)	-88.3 (2.7)	200.1 (9.3)	-18.3 (0.5)	85.4 (4.2)	-107.0 (3.1)	-205.6 (9.4)	183.9 (8.1)	-20.8 (3.9)	20.5 (4.2)
mut8	-39.9 (5.5)	-85.9 (1.7)	112.6 (7.3)	-18.1 (0.5)	71.3 (3.0)	-103.8 (2.1)	-126.6 (6.6)	95.2 (7.0)	-32.2 (2.2)	9.2 (1.0)
mut9	3.0 (8.0)	-87.5 (2.3)	83.7 (8.6)	-18.0 (0.5)	90.1 (3.8)	-105.5 (2.7)	-83.1 (9.0)	65.8 (8.4)	-14.7 (3.6)	26.6 (7.0)

Computational mutagenesis by Rosetta

The same set of control mutants was utilized to implement, optimize, and evaluate the Rosetta protocol. Comparable to the MM/PBSA protocol, excellent agreement with qualitative experimental rankings were observed by the Rosetta protocol (R^2=0.88) (Figure 4.10B).

In contrast to the MM/PBSA predictions, Rosetta predicted, among the homodimer mutants, that the T246D/Q252E mutant (mut3) to be an equal, or slightly more favorable mutant than the wildtype (Figure 4.10B, right). The other mutants hypothesized to be either neutral (mut2, 4) or enhancing (mut6, 7) were all predicted to be closer to wildtype than the corresponding MM/PBSA predictions. Mutant 8 was designed to have enhancing mutations on both arms (mutant 6 has the same enhancing mutations on only 1 arm), but was predicted to be unfavorable contrary to our expectations.

In the heterodimer mutants (Figure 4.10B, left), predictions on almost all mutants followed a similar trend as the MM/PBSA predictions, which we interpret as convergence of the two protocols. The glaring exceptions were the armpit mutants L292R of ErbB2 (mut5) and R285L of EGFR (mut8) which Rosetta assessed at a much reduced penalty compared to MM/PBSA. This was likely an example of over-compensating unfavorable interactions by over-estimating flexibility.

Comparison between MM/PBSA and Rosetta

Both MM/PBSA and Rosetta were adequate at predicting loss-of-function mutations (Figure 4.10). The two protocols displayed a high correlation in predicting mutants carrying two or less substitutions (Figure 4.11A). However, it is likely that

because of the differences in flexibility-sampling algorithms, the correlation decreased as the number of simultaneous substitutions increased (Figure 4.11B). Furthermore, backbone movement intrinsic to the BACKRUB module of Rosetta may have been more forgiving at predicting potentially enhancing mutations, but possible overestimation of flexibility also may reduce its discriminatory power rendering the protocol potentially susceptible to false positives.

The magnitudes of the free energy differences by MM/PBSA protocol are in general too high to be interpreted as realistic quantitative predictions (> 10 kcal). This is likely due to its inherent incorporation of several independent softwares at the implementation level. For example, the spatial positions of the mutant residues were sampled and minimized using PLOP forcefield, whereas the corresponding MM energies were calculated using OPLSAA as implemented in Gromacs—the same protein structure will likely yield to different energy magnitude dependent on the forcefield parameters. So while these structures can be considered as locally "favorable" in both forcefields, the degree may be different which translates. For future improvement, we speculate that minimizing the PLOP-generated mutant structures in OPLSAA forcefield before calculating the MM energies should lower the absolutely magnitude of energy difference.
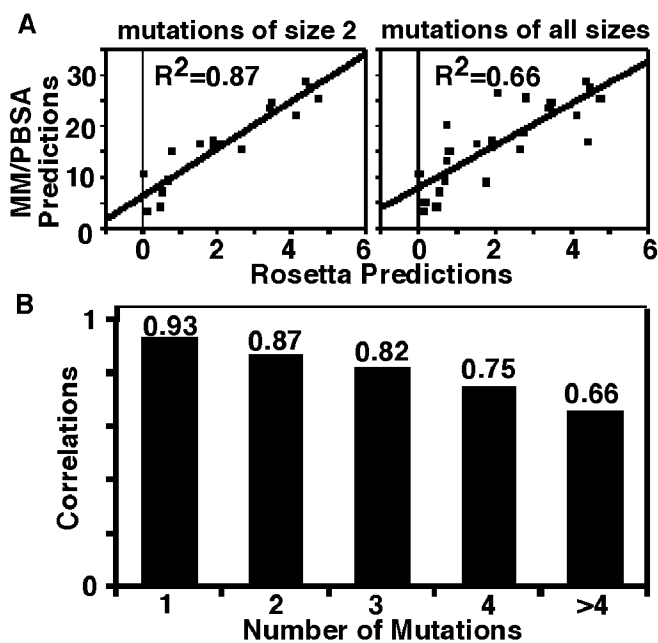
Figure 4.11. Correlation and convergence pattern between the MM/PBSA and Rosetta protocols. A) Correlation graphs of mutants with 2 or less substitutions and all mutations. B) Summary of all mutations categorized by substitution numbers.

Experimental examination of mutant predictions

Four ErbB2 arm mutants (mut1-4, respectively: D256A/E259A, D256N/E259Q, D256T/E259Q, D256R/E259K), predicted to have varying degrees of detrimental effect on heterodimer formation, were experimentally tested for their effect on heterodimer activation and dimerization efficiency in cells. Based on the correlation analysis of the MM/PBSA and Rosetta protocol, we restricted the choice of mutant to two substitutions and chose to focus on the D256/E259 mutants that were unanimously predicted to disrupt dimer formation by both protocols.

Reduced activation of ErbB2 mutant-EGFR heterodimers by Western Blot

To determine the effect of mutation of the ErbB2 arm on activation of the ErbB2-EGFR heterodimer, each ErbB2 mutant was transiently transfected into CHO cells stably expressing the kinase-dead K721A-EGFR[134]. This choice of EGFR mutant ensures that the kinase-dead EGFR can only be activated through heterodimer formation. Activation of the K721A-EGFR in the heterodimer was measured in each of the mutants and was

compared with wild-type ErbB2 by Western blotting for phosphorylation of the K721A-

EGFR (Figure 4.12). We confirmed that phosphorylation could not be detected by

homodimerization of ErbB2 or kinase-dead EGFR alone (Figure 4.12A, last two lanes).

Any detectable EGF-stimulated phosphorylation must therefore be due to the formation

of K721A-EGFR/ErbB2 heterodimers. Compared to cells expressing wildtype ErbB2

receptors, phosphorylation was hindered in all 4 mutants as predicted (Figure 4.12B).

Among the mutants, D256R/E259K almost completely abolished phosphorylation,

supporting both the prediction and hypothesis that electrostatic interactions between the

wildtype ErbB2 arm and the EGFR armpit at this interface was critical for heterodimer

formation.
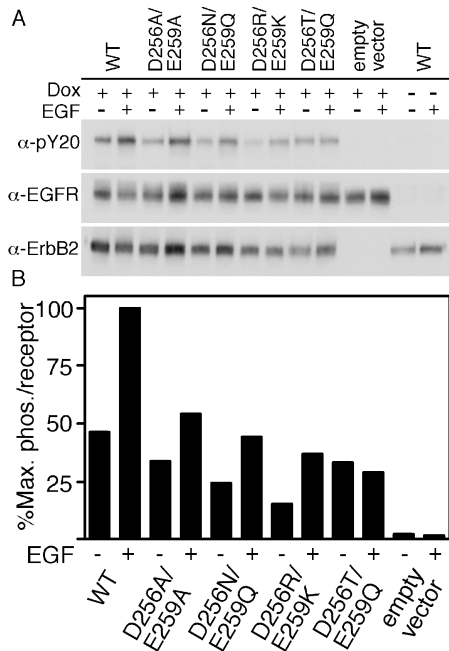
Figure 4.12. Phosphorylation of heterodimers formed by wildtype and mutant ErbB2 receptors. A) Phosphorylation, EGFR expression, and ErbB2 expression in the presence and absence of EGF stimulations. B) Quantification of phosphorylation per receptor level.

Reduced dimerization of the EGFR by luciferase fragment complementation imaging

upon expression of ErbB2

To directly monitor the mutant effect on heterodimer levels, a luciferase fragment

complementation imaging assay[81] was utilized to measure the degree at which the

77

mutants can compete for wildtype EGFR in live cells. The extracellular domain of the EGFR that lacks the entire intracellular domain was fused just beyond the transmembrane domain to an N-terminal luciferase fragment (ΔC-EGFR-NLuc) and the complementing C-terminal luciferase fragment (ΔC-EGFR-CLuc). The ΔC-EGFR-NLuc and ΔC-EGFR-CLuc constructs were stably co-expressed in CHO-K1 Tet-On cells. On the day of imaging, cells were pre-treated with the firefly luciferase substrate, D-Luciferin for 20 min to allow equilibration of the intracellular and extracellular pools of the substrate. A baseline photon flux was then measured, followed by addition of ligand and measurement of photon flux every 30 sec for 30 min.
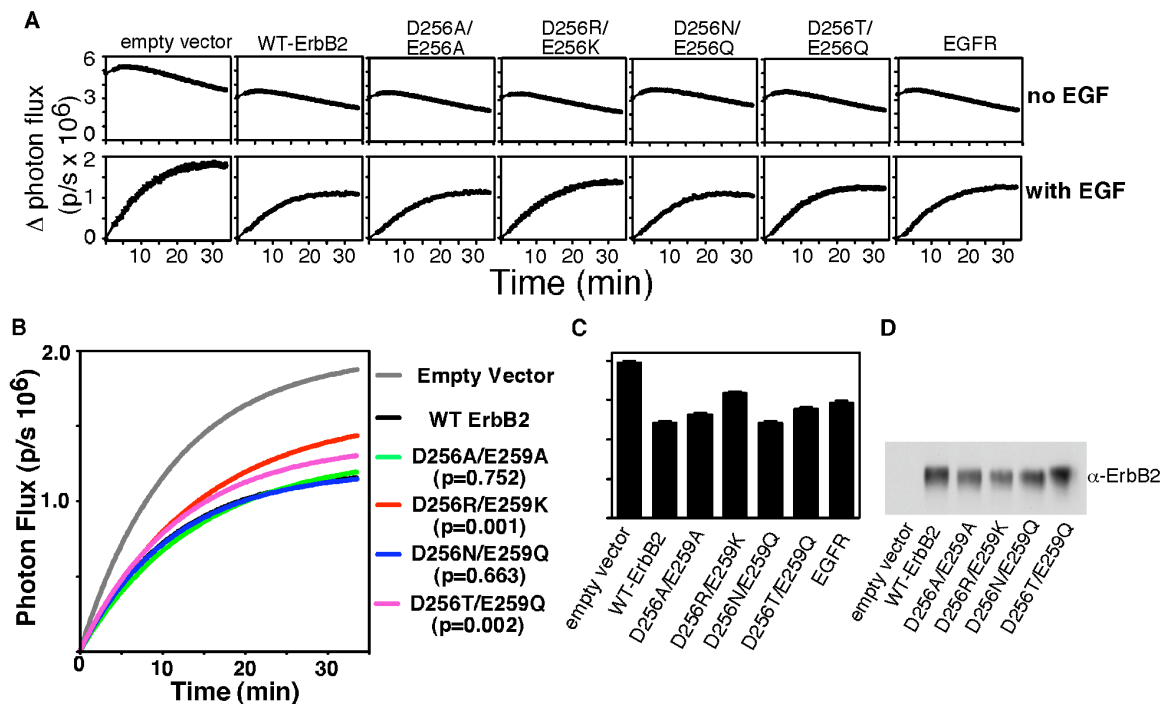
Figure 4.13. Heterodimer competition measured by real-time quantification of EGFR homodimer level in the presence of wildtype and mutant ErbB2 receptors using split-luciferase complementation imaging assay. A) Photon emitted by the complementation of NLuc-EGFR and CLuc-EGFR upon transfection of wildtype and mutant ErbB2 under non-stimulating (top) and stimulating (bottom) conditions. B) Summary of A. C) Quantification of the Y-max from the stimulating conditions. D) ErbB2 receptor level of the transfected cells immediately after imaging.

Upon transient transfection of wildtype or mutant ErbB2 receptors into these cells, formation of either ΔC-EGFR-NLuc/ErbB2 or ΔC-EGFR-CLuc/ErbB2 heterodimers is expected to decrease the population of the signal-emitting ΔC-EGFR-NLuc/ΔC-EGFR-CLuc homodimers. The competing power of the wildtype and mutant ErbB2 receptors were assessed by monitoring their ability to "quench" the photon flux from formation of the ΔC-EGFR-NLuc/ΔC-EGFR-CLuc homodimers. As a reference, the competing power of wildtype EGFR without the luciferase fragment was also assessed. As expected, wildtype ErbB2 was able to quench the photon flux better than the wildtype EGFR, (Figure 4.13) since heterodimer formation is favored over homodimer formation. All ErbB2 mutants, except for one, displayed weaker photon flux quenching than wildtype ErbB2. Consistent with the phosphorylation data, the D256R/E259K mutant had the weakest competing power among all ErbB2 receptors. The expression levels of the wildtype and mutant ErbB2 receptors were measured immediately after imaging using Western blot analysis with an ErbB2 antibody to ensure signal differences were not due to an unequal number of receptors in each case. In the absence of EGF, the basal photon flux was reduced in all cases compared to cells transiently transfected with empty vector. This data is consistent with the notion of pre-formed dimers[135-138] and that these pre-formed dimers are disrupted in a similar competing fashion upon transient transfection with the wildtype or mutant ErbB2 receptors. Interestingly, the magnitude of the decrease in the basal photon flux appears to be identical in all cases, suggesting that the preformed dimers utilize an interface independent of the ErbB2 arm.

Discussion

The series of systematic analysis of the computationally refined EGFR-ErbB2

heterodimer model progressively led to a set of experimentally testable hypothesis centered on the dimerization arm interface. The initial focus on the two ErbB2 charged residues D256 and E259 originated from analyzing the nature of molecular contacts at different dimer interfaces. The contrast at the dimer interfaces led to subsequent sequence alignment where it was observed that D256/E259 residues are unique to ErbB2. Upon mapping these residues onto the tip of the arm, the backbones and side-chains of two overlapping reverse-turns were classified to reveal that Turn 1 of EGFR and ErbB2 adopted similar configurations independent of surrounding residues. By contrast, Turn 2 displayed different range of backbone and side-chain conformations which highlighted a preference of environment-specific flexibility and orientations. This was explained by the contact map analysis (Figure 4.5) where Turn 2 residues were associated with a higher number of inter-receptor contacts, thus requiring specific spatial adjustment to complement corresponding armpit configurations. For example, the ErbB2-arm/EGFR-armpit interface had reduced number of inter-receptor contacts but appeared to be energetically favorable due to the presence of a salt-bridge hydrogen-bond network between D256/E259 of ErbB2 and R285 of EGFR. This observation led us to examine the sequence alignment of the R285 region and subsequently identified the unique substitution of a leucine residue in place of the arginine in ErbB2.

The EGFR-ErbB2 heterodimer model presented here provides, among many biological interests, a structural basis to explain the absence of an ErbB2 homodimer. Using the EGFR homodimer interface as a reference (Figure 4.8C), the EGFR-ErbB2 heterodimer has two energetically more favorable interfaces—electrostatic interactions between D256/E259/R285 at the ErbB2 arm-EGFR armpit interface (Figure 4.8A), and

extensive inter-receptor hydrophobic contacts at the EGFR armpit-ErbB2 arm interface (Figure 4.8B). The hypothetical ErbB2 homodimer on the other hand, is likely to pay a hefty energetic penalty for extending the D256/E259-containing arm of one monomer into the highly hydrophobic armpit of the other monomer where the complementary charged arginine is replaced by leucine. This conclusion that homodimerization of the ErbB2 extracellular domains is energetically unfavorable is in agreement with previous reports that the rare occurrences of full-length ErbB2 homodimers requires additional interactions between the transmembrane[139] and intracellular[37, 140, 141] domains. Although Garret and colleagues had arrived at the same conclusion based on the ErbB2 monomer structure[124], they speculated based on crude estimation of surface electrostatic potential that electrostatic repulsion between ErbB2 arm and armpit was the reason for unfavorable ErbB2 homodimerization. Our current work supports part of this notion but further extends the analysis to pinpoint the unfavorable interactions at the residue level using a relaxed all-atom model.

Among the tested mutants, D256R/E259K displayed the biggest detrimental effect at disrupting heterodimer. This observation is consistent with the analysis that the ErbB2-arm/EGFR-armpit interface is normally stabilized by the presence of strong electrostatic interactions involving residues D256/E259 and the R286 residues. However, replacing the electrostatic attractions with the unforgiving repulsion consequently elevates the energetic barriers of heterodimer formation. D256A/E259A and D256N/E259Q displayed smaller effects in the luciferase competition assay (figure 4.13) most likely because these neutral substitutions did not introduce enough energy penalties to significantly disrupt the interface. Because the substitutions were of smaller (D256A/E259A) or comparable

(D256N/D259Q) size with the wildtype counterparts, they likely did not introduce additional steric penalties in the predominantly hydrophobic EGFR armpit so that the loss of electrostatic interactions alone in these cases was not enough to raise the energetic barrier to the same degree as introducing repulsion penalties. Similar logics should be applicable to the D256T/E259Q mutants. However, in stark contrast to D256A/E259A and D256N/E259Q, we actually observed an apparent lower level of heterodimer "quenching" comparable to that of the D256R/E259K mutant (Figure 4.13). We rationalized that this may be a result of the ability to homodimerize among the D256T/E259Q mutant ErbB2s since this pair of substitutions mimics the EGFR dimerization arm (Figure 4.7) — this is equivalent to the EGFR-arm/ErbB2-armpit interface. Indeed, in the phosphorylation assay we observed an elevated ligand-independent basal phosphorylation for the D256T/E259Q mutants (Figure 4.12) consistent with the concept of spontaneous ErbB2 mutant dimerization.

The experimental confirmation of D256/E259 as "hot-spot" residues serves as good support for the heterodimer model. An immediate application of this work is to design homo- or heterodimer specific therapeutics. Virtually all existing small-molecule inhibitors of ErbB receptors arose from the current paradigm of targeting the kinase catalytic site of individual ErbB receptors[142]. Recent awareness and understanding of the heterodimer-mediated drug-resistance[12, 143-145] has accelerated the development of the dual kinase inhibitor lapatinib[63] and has warranted efforts towards multi-targeting[146]. Understanding the structural differences at the homo- and heterodimer interfaces provides new target sites and roadmaps for designing small-molecule inhibitors to target ErbB family dimers of all combinations. For example, reverse-turn 2 at the tip of the arm can

serve as promising structural templates for designing reverse-turn peptidomimetics directed at the homo- and/or heterodimer interface. Another utility based on these data would be to engineer dimer-enhancing mutants, for example, EGFR mutants with the equivalent ErbB2 D256/E259 substitutions. However, in addition to the fact that none of the hypothesized enhancing mutants were predicted to behave, it is even more challenging to achieve in practice since optimizing an interface by a two-amino acid substitution is fundamentally much less likely to be detectable (even if theoretically correct) than disrupting an interface. While favorable arm-armpit interaction is necessary, there are additional contributions from other domains of the receptor that together, drive the dimerization of full-length ErbB receptors in living cells. We, therefore, limited the utility of this work within the scope of disrupting the interface for therapeutic implications.

Methods

Computational Procedures

Homology modeling

Before homology modeling, flag tag residues DYKDDDDK at the end of the ErbB2 structure (PDB 2A91, 2.5Å) was deleted because they were not part of the native ErbB2 sequence. The resulting PDB file contained residues 1-509. Missing residues 103-106 and all other missing atoms were initially filled in using PLOP v7.7[147] and subsequently minimized using Macromodel v9.1 with the rest of the atoms frozen. Sequence alignment of the residues in 1IVO and 2A91 were constructed using the ClustalW2 webserver[148]. Based on the sequence alignment, ErbB2 was homology modeled onto the EGFR homodimer template (PDB 1IVO) using Modeller v9.5[149] with

loop refinement on the EGFR and ErBb2 arm regions. 10 models were generated and the model with the best DOPE score was selected as the starting structure for MD.

Molecular dynamic simulation

MD simulations were performed using Gromacs v3.3.3[150] using OPLS2001 forcefield[151]. The EGFR-ErbB2 model was solvated in a dodecahedron periodic box of pre-equilibrated TIP5P water molecules[152]. Ions were added to neutralize the system. The protein complex is 14 Å away from the edge of the box in all directions. Following minimization, the system was gradually heated up starting from 50K at intervals of 50K for 30 ps each to reach the final temperature of 300K under NPT (Berendsen temperature coupling every 0.1 ps and Parrinello-Rahman pressure coupling at 1 atm every 1 ps). The initial velocity was randomly assigned from the Maxwell distribution at 50K. All stages of the simulations were performed with the following setup: 2 fs time step; LINCS bond length constraints; PME electrostatic treatment with grid spacing of 0.12; PME order of 6 and damping starting at 9 Å; van der Waal treatment was truncated at 9 Å. The heterodimer simulation was for 9.9 ns and the homodimer simulation was for 15 ns. All simulations were performed on the NSF Teragrid[153] spanning several months.

MM/PBSA

Theory. The Molecular Mechanics/Poisson Boltzmann Solvent Accessible (MM/PBSA) protocol[131, 154], shown to be fast and effective[120, 155-157] in many cases, calculates the relative binding free energies between biomolecular complexes by decomposing it into the sum of solvation and gas-phase terms (AB denotes to complex, A and B denotes to sub-components of the complex):

$$\Delta_{bind}G = \Delta_{gas}G + (\Delta_{solvation}G_{AB} - \Delta_{solvation}G_{A} - \Delta_{solvation}G_{B}) \text{ (equation 1)}$$

Solvation energy is further decomposed into polar and apolar components:

$$\Delta_{solvation}G = \Delta_{PB}G + \gamma\Delta_{SASA}A \text{ (equation 2)}$$

Polar or electrostatic contribution is calculated by the Poisson-Boltzmann method (PB) using a continuum implicit-solvent model. The apolar component is estimated by weighting the change in solvent-accessible surface area (SASA) by a heuristic surface tension term, A.

The gas-phase term is calculated as:

$$\Delta_{gas}G = \Delta_{gas}U + T\Delta S \text{ (equation 3)}$$

The gas-phase potential energy (U) is calculated based on semi-empirical force field energy. Treatment of entropy is assumed to be relatively equivalent across mutants and global structural is preserved.

The final binding energy is the average of the ensemble, and the relative binding energy is the difference between the wildtype ensemble and the mutant ensemble. The relative binding energy is used to predict the relative stability of the mutant complex with respect to the wildtype complex.

Implementations. Representative snapshots of the MD trajectory were extracted every 50 ps starting from 3500 ps for a total of 130 frames for heterodimer simulation, and every 100 ps starting from 2000 ps for a total of 129 frames for homodimer simulation. PDB files were converted to PQR files by PDB2PQR[158] v1.3.0 using OPLS2001 charge and PARSE radii[159]. Mutations were done by using PLOP[147] v15.19 with intra-protein dielectric constant of 2 and solvent dielectric constant of 80. In all cases, side chains within 15 Å of the mutated residue were repacked and minimized. Solvent-accessible surface area was calculated by using the acc module of APBS[160]

v1.1.0. Molecular mechanic energy components were calculated using Gromacs v3.3.3. Because the size of the dimer system, regular PB calculation using APBS required impractical amount of memory and CPU time using the desired grid spacing. Instead, calculations were performed using the asynchronous mode of APBS by dividing the system that resulted in 64 calculations per snapshot, 8320 calculations per mutant, and ~100,000 calculations in total. Each column listed in the tables was calculated as the average of all values calculated from each frame with outliers removed.

The experimental binding energies for the control mutants were calculated based on the K_D[42] and the monomer:dimer ratio[46] using equations: $\Delta G = -RT\ln(K_D)$ and $\Delta G = -RT\ln([monomer]^2/[dimer])$. Normalization with respect to the control case in each experiment was performed to compare the two experiments on the same scale.

Calculation using the backrub module of Rosetta

The same set of snapshots extracted from the MD trajectories were first relaxed by Rosetta[132, 161] v3.0 scoring functions with the backbone atoms fixed. Mutations were constructed and optimized using the BACKRUB module. For each mutation, at least 3 residues were selected for backrub motions and only those residues were subjected for backbone and side-chain repacking. All optimizations were performed on the complex only; monomers were extracted from the optimized complex and not subjected to optimization as this yielded the best results in our experience and as previously suggested[162]. The final stability score of the complex was calculated by subtracting the scores of the monomers from the complex

Experimental Procedures

Reagents

Murine EGF was purchased from Biomedical Technologies, Inc. and was dissolved in sterile water. Doxycycline was purchased from Clontech and was dissolved in sterile water. D-Luciferin (Biosynth) was dissolved in PBS and coelenterazine (Sigma) was dissolved in ethanol. The phosphotyrosine antibody (PY20) was from BD Biosciences. The EGF receptor and ErbB2 antibodies were from Upstate.

DNA Constructs

The D256A/E259A, D256N/E259Q, D256R/E256K, D256T/E259Q-ErbB2 constructs were made using QuikChange site-directed mutagenesis (Stratagene) in the ErbB2 pcDNA3.1(+) construct (kind gift from Dr. Graham Carpenter, Vanderbilt University). The ErbB2 mutants were ligated into pcDNA5.FRT (Invitrogen) using the NheI and XhoI sites. All mutations were verified by sequencing.

Cell Lines

CHO-K1 Tet-On cells (Clontech) were stably transfected with pTK-Hyg (Clontech) and K721A-EGFR (pBI-Tet MCSI) using Lipofectamine 2000 (Invitrogen) as previously described[92]. CHO-K1 Tet-On cells were cotransfected with pTK-Hyg and ΔC-EGFR-NLuc (pBI-Tet MCSI) using Lipofectamine 2000 as previously described[81]. A double-stable cell line was established by transfecting ΔC-EGFR-CLuc (pcDNA4/myc-HisB, Invitrogen) into ΔC-EGFR-NLuc cells using Lipofectamine 2000 and selecting in 400 μg/ml zeocin (Invitrogen). Double-stable lines were maintained in DMEM containing 10% FetalPlex, 1000 μg/ml penicillin/streptomycin, 100 μg/ml G418, 50 μg/ml hygromycin, and 100 μg/ml Zeocin.

Wild-type ErbB2 (pcDNA3.1+), the ErbB2 mutants (pcDNA5.FRT), wild-type EGFR (pcDNA5.FRT), or the empty pcDNA5.FRT vector were transiently transfected

into the K721A-EGFR or the ΔC-EGFR-NLuc/CLuc cell line 24 hr prior to assay using Lipofectamine 2000. For the transfection into the ΔC-EGFR-NLuc/CLuc cell line, cells were co-transfected with renilla luciferase (pRLuc-N1, Packard Bioscience) to assess the transfection efficiency using bioluminescence imaging.

Kinase activation and Western Blotting

K721A-EGFR cells were grown to confluence in 35 mm dishes. Cells were serum-starved in DMEM containing 1 mg/ml BSA for 2 hr. Culture medium was removed and cells were washed twice in ice-cold PBS and then scraped into RIPA buffer (150 mM NaCl, 10 mM Tris pH 7.2, 0.1% sodium dodecysulfate, 1% Triton X-100, 17 mM deoxycholate, and 2.7 mM EDTA) containing 20 mM p-nitrophenyl phosphate, 1 mM sodium orthovanadate, and protease inhibitors. Equal amounts of protein (BCA assay, Pierce) were loaded onto a 9% SDS-polyacrylamide gel and then transferred to PVDF (Millipore). Western blots were blocked for 1 hour in TBST/10% nonfat milk. The blots were incubated in primary antibody for 1 hr, washed in TBST/0.1% BSA, incubated in secondary antibody for 45 min and washed three times in TBST/0.1% BSA. Western blots were detected using the ECL reagent from GE Healthcare.

Luciferase complementation imaging

Luciferase fragment complementation imaging was done as previously described[81]. Briefly, cells were plated 48 hr prior to use at 5×10^3 cells per well in DMEM containing 1 μg/ml doxycycline in a black-walled 96-well plate. On the day of the assay, cells were serum-starved for 2 hr and then incubated for 20 min in 175 μl DMEM without phenol red, containing 1 mg/ml BSA, 25 mM Hepes, and 0.6 mg/ml D-luciferin at 37ºC. To establish a baseline, cell radiance (photons/second/cm^2/sr) was

measured using a cooled CCD camera and imaging system at 37ºC (IVIS 50; Caliper) (30 sec exposure; binning, 8; no filter; f-stop, 1; field of view 12 cm). EGF was added in a volume of 25 μl in the same media (DMEM, 1 mg/ml BSA, 25 mM Hepes, 0.6 mg/ml D-Luciferin). Radiance was measured sequentially as described above. To assess the transient transfection efficiency, renilla luciferase expression was monitored. Media was replaced on cells with DMEM (no phenol red) containing 1 mg/ml BSA, 25 mM Hepes, and 400 nM coelenterazine. Radiance was immediately measured as described above except the filter was set to <510.

Data Analysis

Data was collected in quadruplicate for each condition. A flat-field correction was done to correct for differences in the baseline photon flux. Light production expressed as photon flux (photons/sec) was determined from regions-of-interest defined over wells using LIVINGIMAGE (Xenogen) and IGOR (Wavemetrics) software. Changes in photon flux were calculated by subtracting values from untreated cells from those of EGF-treated cells. Standard errors were determined using the formula for the calculation of the unpooled standard error

CHAPTER 5. Additional Hits and Discussions

Additional hits from structurally similar compounds

As discussed in Chapter 2, one of the advantages of screening the NCI-diversity database was due to its large collection of structurally similar compounds corresponding to each representative compound in the diversity set. We have followed up on one identified lead compound, NSC56452, by searching the original NCI database for additional compounds that share at least 90% Tanimoto similarity[163] in structure and composition. Because the luciferase fragment complementation assay was the faster and superior screening method comparing to Western blotting for activation, our strategy was to screen directly for dimerization inhibitors in the first pass at a single dose, followed by dose-response characterization of subsequent candidates, and finally candidates could be tested by Western blot to confirm inhibitory effect on receptor activation.

In addition to testing the NCI compounds, we also tested a compound set originally synthesized as generic kinase inhibitors by Professor Stefan Laufer[164, 165]. The motivation and the ensuing exciting direction originated from a structural search of our lead compounds in the literature that lead to Professor Laufer's publications. In these articles, they described the synthesis and thorough characterization of the inhibition effect of a set of compounds tested against a wide spectrum of kinases. This was particularly interesting to us because while we are confident that NSC56452 blocks dimer formation (as seen in kinase-deleted luciferase experiment, chapter 2), it would be intriguing to see whether it has or may be modified to have additional activity against the kinase domain of EGFR. Based on the structural resemblance between NSC56452 and the kinase inhibitors synthesized by Laufer *et al*, the direction was then to explore the potentials of

developing multi-targeting agents to increase potency. Given that substantial testing against a wide range of kinase has already been done, the concept of multi-targeting could be extended from targeting multiple domains of EGFR to include targeting multiple kinase targets towards the development of "dirty drugs"[166, 167].

Figure 5.1, Table 5.1 and Table 5.2 summarizes the data obtained so far from testing these compounds. The original intent was to derive a pharmacophore model that can discriminate against structurally close inactive compounds. Pharmacophore models can then be utilized to provide a fast ligand-based filtering layer in our top-down screening approach:

ligand-based virtual screening using pharmacophore models → receptor-based virtual screening by the OpenScreening Project → testing of top candidates by the luciferase fragment complementation imaging assay → biochemical characterization to assess inhibitory activities.

However, preliminary development of these pharmacophore models has so far yielded sub-optimal predicative power largely due to the potentially narrow range in IC_{50} values among compounds. The definition of "actives" and "inactives" in constructing ligand-based models often requires the difference in potency to be of several orders of magnitudes. Testing at a wider dose range is therefore warranted in order to derive more sensitive models

Future direction

The lead inhibitors of EGFR homodimer serve as a proof-of-concept for targeting the dimerization interface with small-molecule inhibitors. I envision two immediate directions to build on this thesis: 1) continuous optimization of the lead compounds with

the help of descriptive pharmacophore model, and 2) peptidomimetic design based on the dimerization arm difference to target homo- and heterodimer interface. The benefit of having the luciferase complementation assay readily available should provide a great platform for immediate testing of potential candidates. The next milestone should be the characterization of the compounds (ideally, second-generation compounds more potent than the initial lead compounds) against purified extracellular domain of EGFR to test the so-far-consistent hypothesis that they target the dimerization arm interface. Crystal complex would be ideal, however, potentially difficult considering the balance between the concentration of extracellular domain needed to form crystals and the concentration of inhibitors necessary to prevent dimer formation
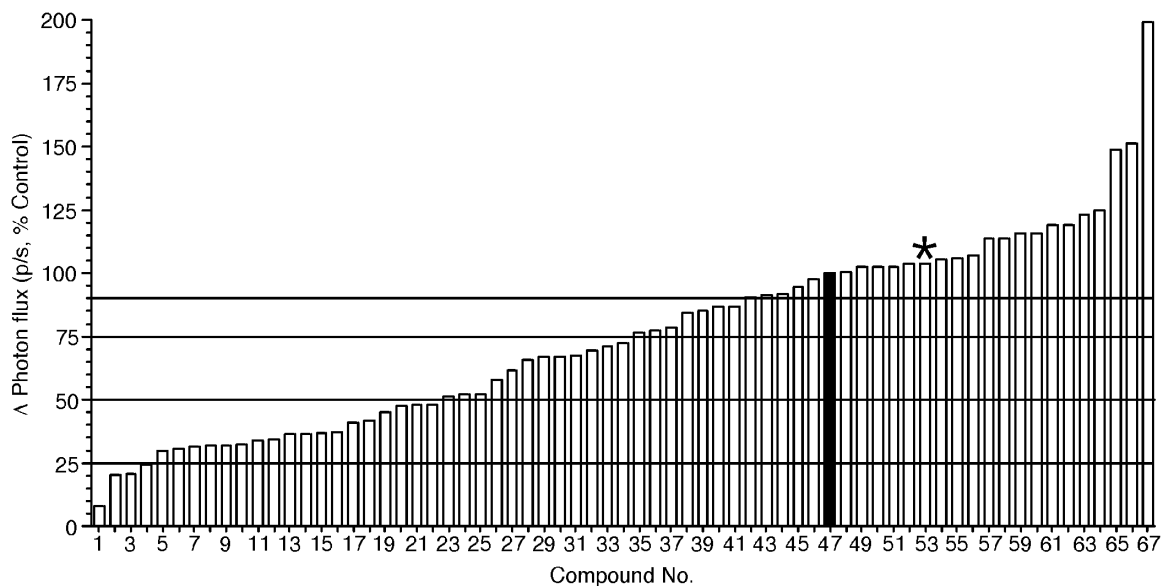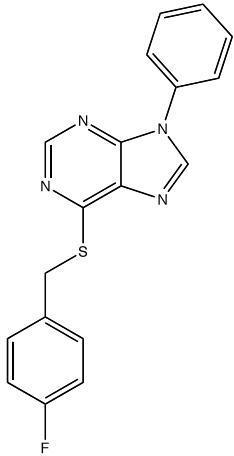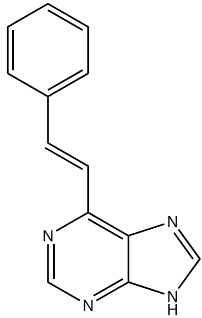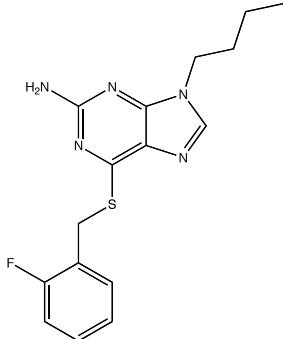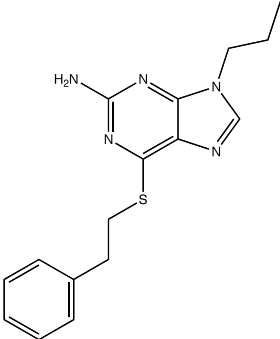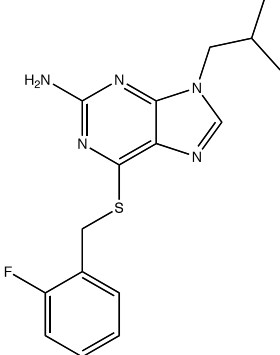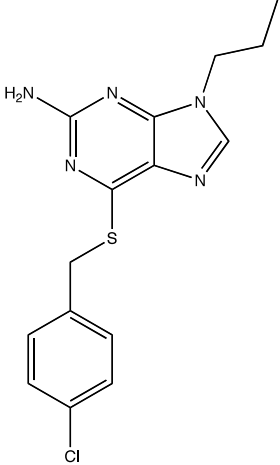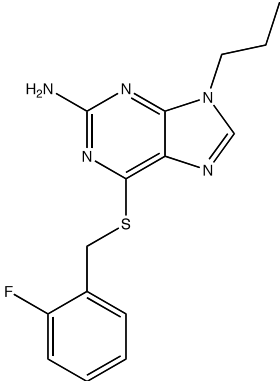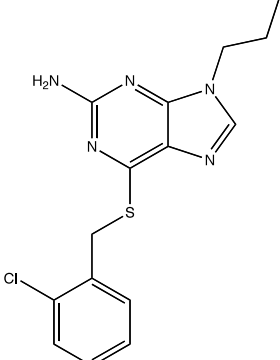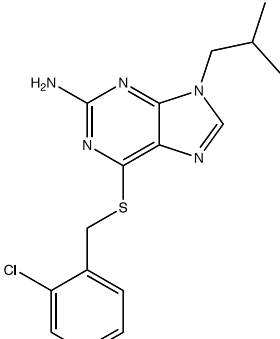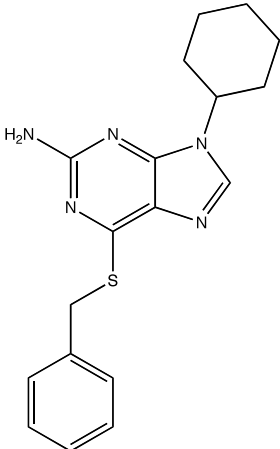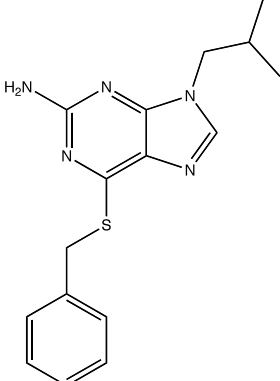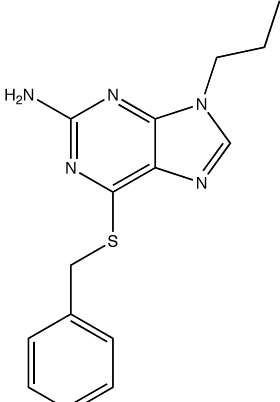
Figure 5.1. Graphical summary of the screening results. Compound 47 is the control with DMSO; compound 53 (*) is the tyrosine kinase inhibitor erlotinib. Compounds showing effect below 25% was denoted as (+++) in Table 5.1, 50% denoted as (++), 75% denoted as (+). Experiment was done in triplicate of 3 wells at concentration of 25 μM compound.
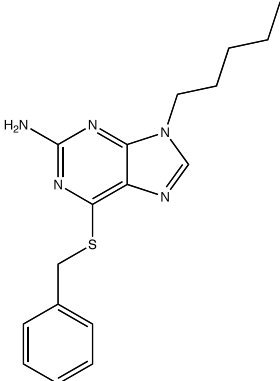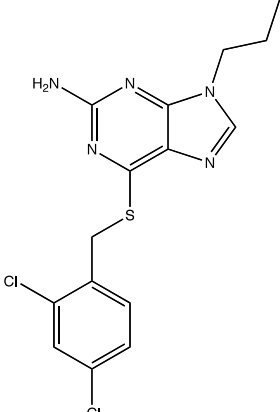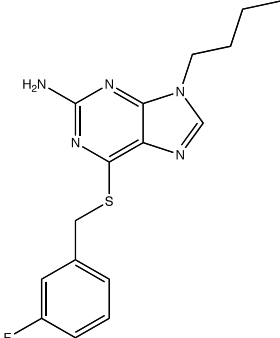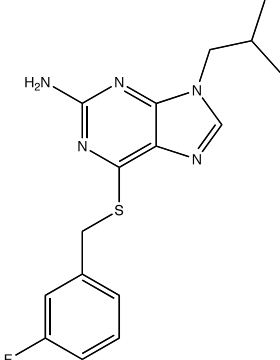
Table 5.1. Screening results of NCI compounds and selected Laufer compounds
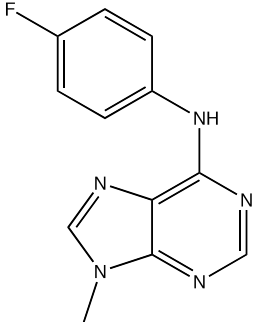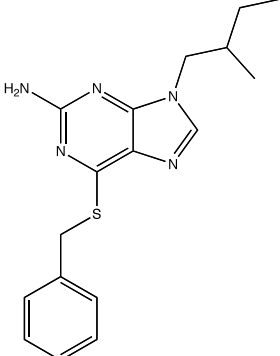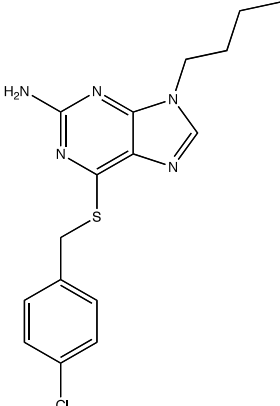
Compound ID	Graph No.	Structure	-/+ or IC50
34489	2		+++
DH193	3		5 µM
49818	4		+++
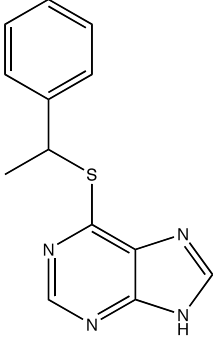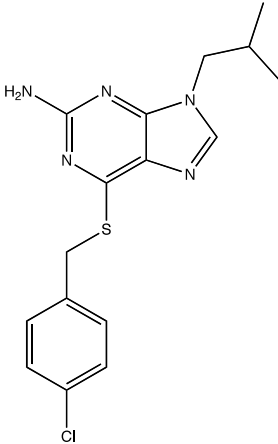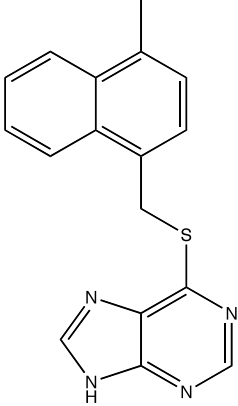
Compound ID	Graph No.	Structure	-/+ or IC50
47777	5		++
48720	6		++
43414	7		++
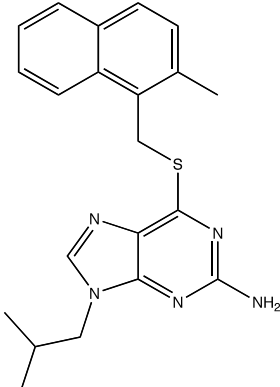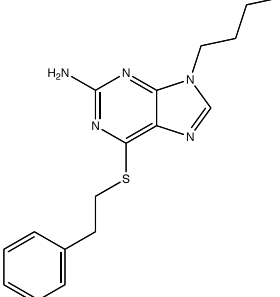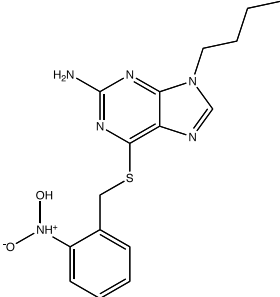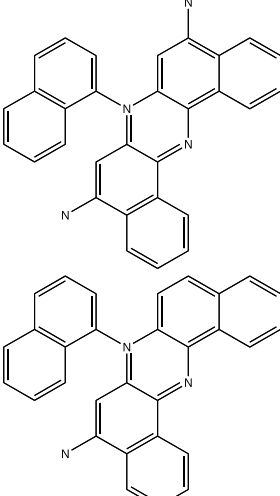
Compound ID	Graph No.	Structure	-/+ or IC50
49817	8		++
44580	9		++
44582	10		++
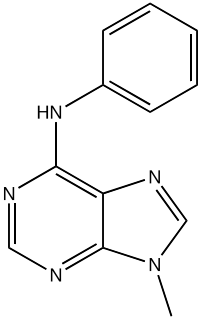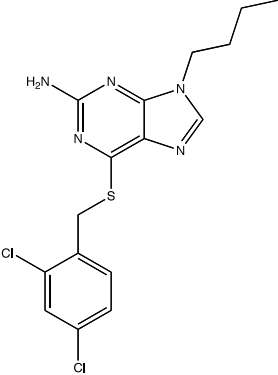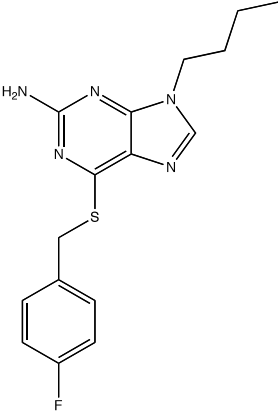
Compound ID	Graph No.	Structure	-/+ or IC50
56453	11		++
42381	12		++
42379	13		++
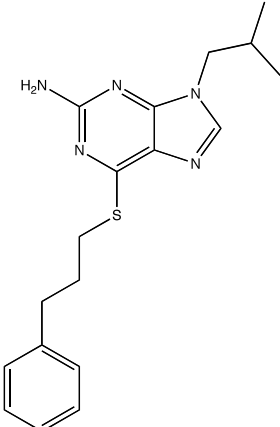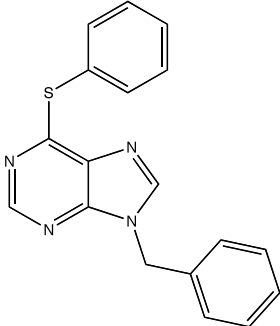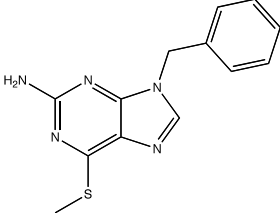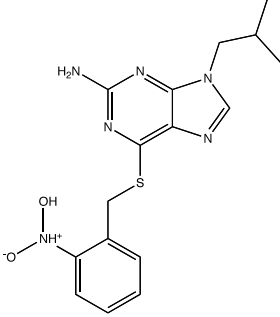
Compound ID	Graph No.	Structure	-/+ or IC50
59485	14		144 nM
44583	15		++
49819	16		++
48724	17		++
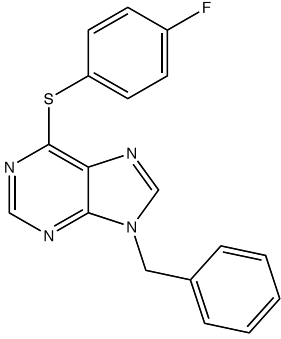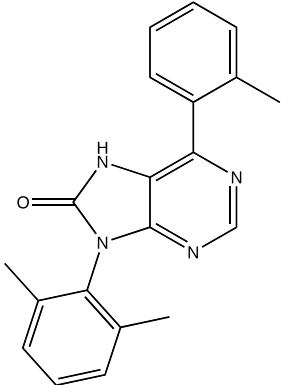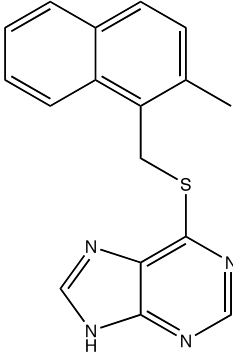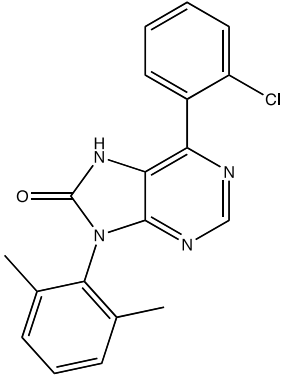
Compound ID	Graph No.	Structure	-/+ or IC50
DH29	18		++
58908	19		++
45156	20		++
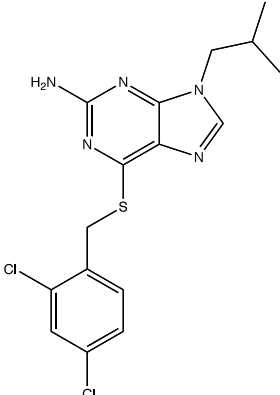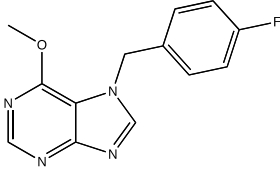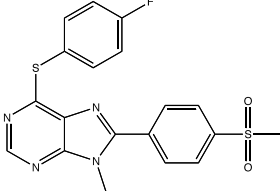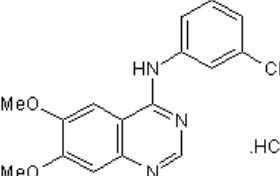
Compound ID	Graph No.	Structure	-/+ or IC50
36828	21		++
43415	22		++
48715	23		++
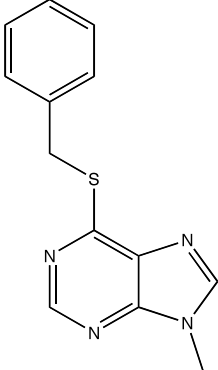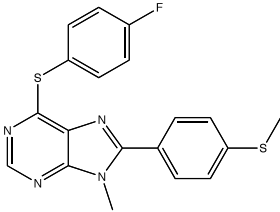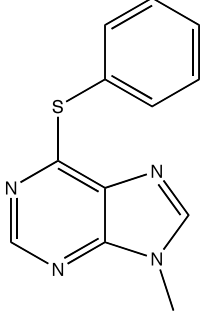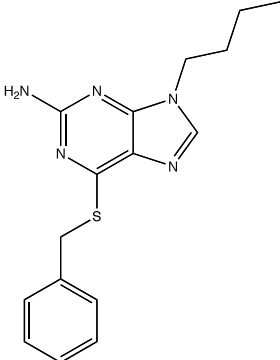
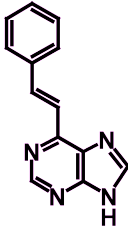| Compound ID	Graph No.	Structure	-/+ or IC50
56452	24		0.6 µM
47779	25		++
45155	26		+
11421	27		+
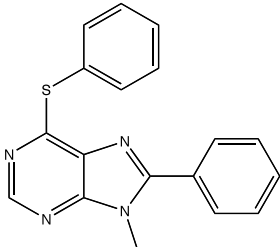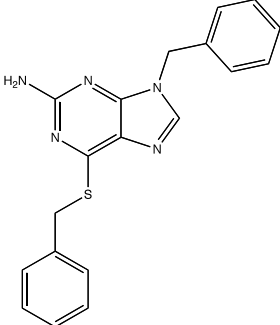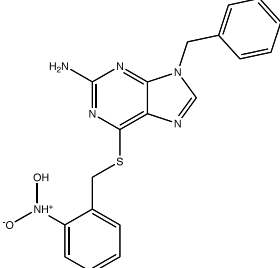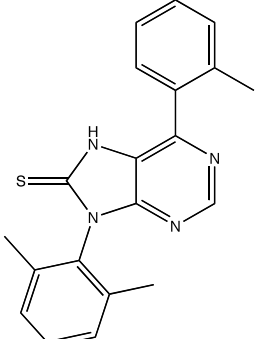
Compound ID	Graph No.	Structure	-/+ or IC50
DH14	28		3 µM
48718	29		+
49820	30		+
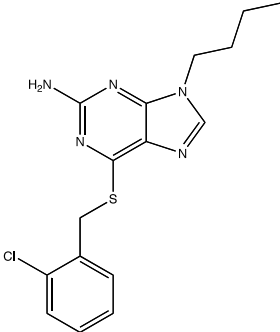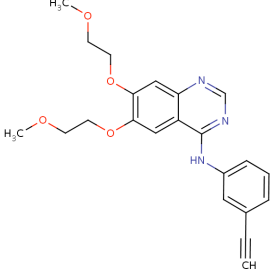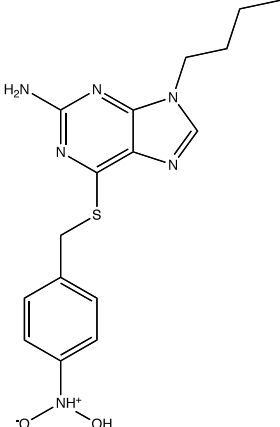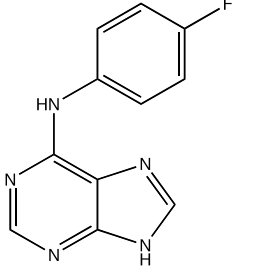
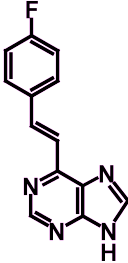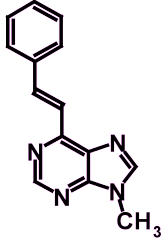| Compound ID	Graph No.	Structure	-/+ or IC50
48722	31		+
DH199	32		4 µM
42380	33		76 µM
47782	34		+
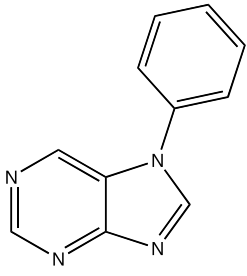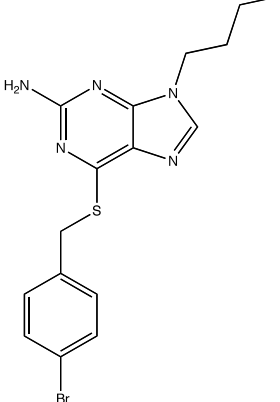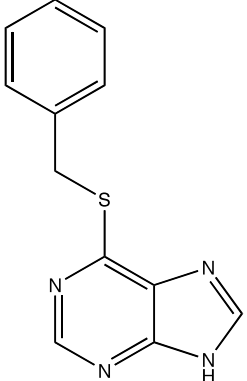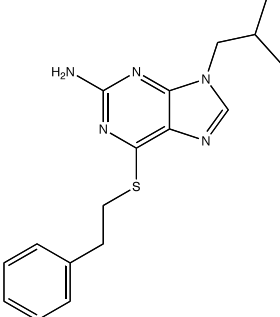
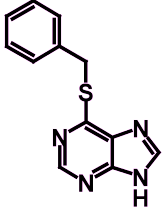| Compound ID	Graph No.	Structure	-/+ or IC50
DH202	35		+
DH250	36		+
48714	37		+
DH262	38		+/-
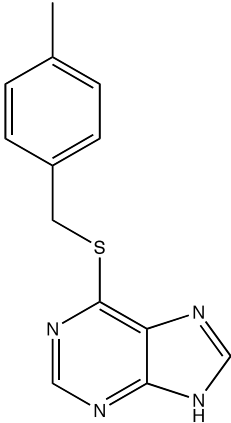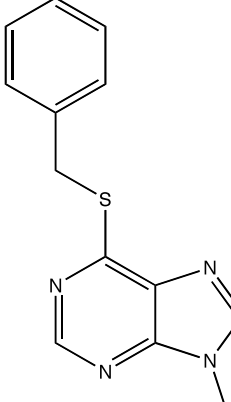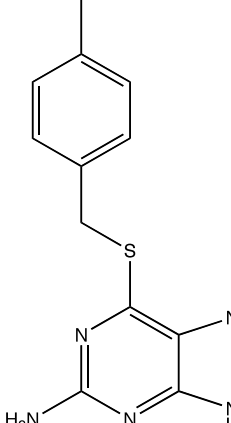
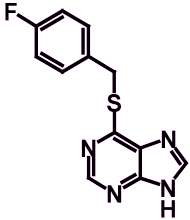| Compound ID	Graph No.	Structure	-/+ or IC50
47786	39		+/-
DH73B	40		+/-
DH172	41		+/-
AG1478	42		+/-
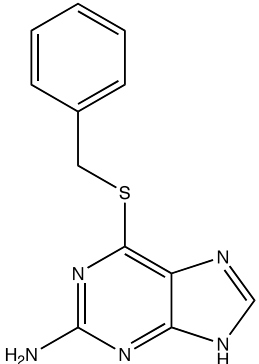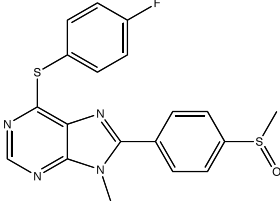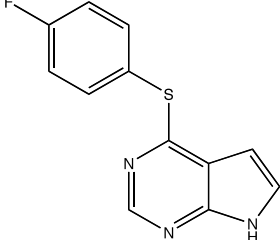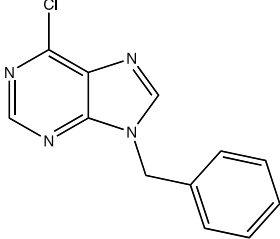
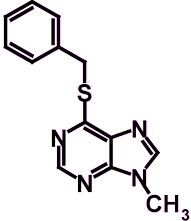| Compound ID	Graph No.	Structure	-/+ or IC50
38303	43		+/-
DH156	44		+/-
DH36	45		-
42382	46		-
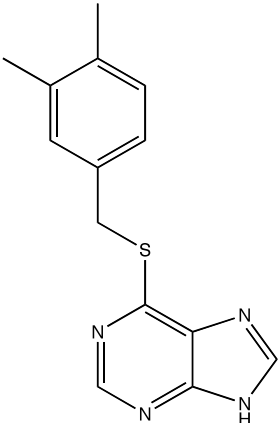
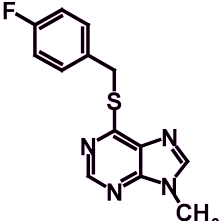| Compound ID	Graph No.	Structure	-/+ or IC50
DH50	48		-
42384	49		-
43412	50		-
DH251	51		-

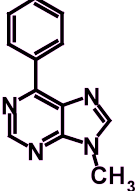| Compound ID	Graph No.	Structure	-/+ or IC50
42383	52		-
Tarceva	53		-
43411	54		-
DH2	55		-

Compound ID	Graph No.	Structure	-/+ or IC50
DH12	56		-
49821	57		-
29421	58		18 mM
47783	59		-
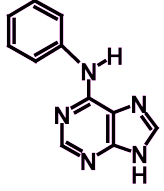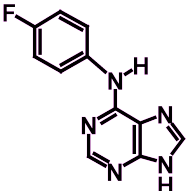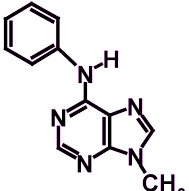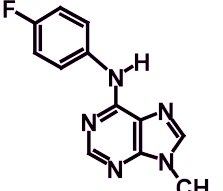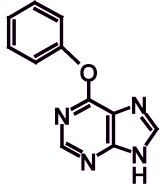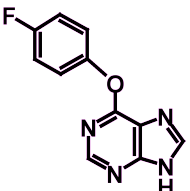
Compound ID	Graph No.	Structure	-/+ or IC50
46383	60		-
DH42	61		-
46384	62		-

Compound ID	Graph No.	Structure	-/+ or IC50
15747	63		-
DH168	64		-
DH144	65		-
DH58A	66		-
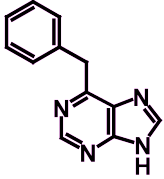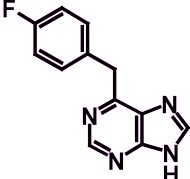
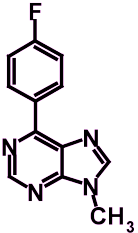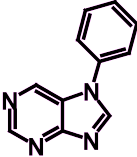| Compound ID	Graph No.	Structure	-/+ or IC50
48713	67		-

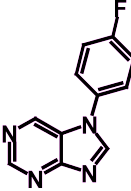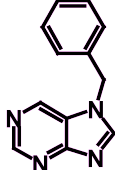Compound 1 is positive control, Cetuximab; Compound 47 is control with DMSO; Compound 53 is negative control, Tarceva.
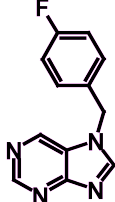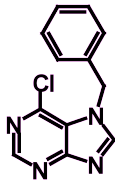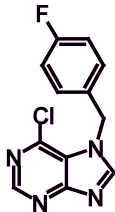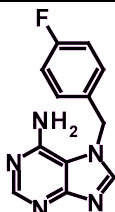
Table 5.2. Screening results of all Laufer compounds at dose of 20 µM.
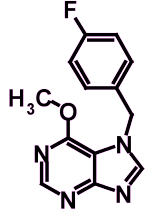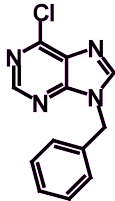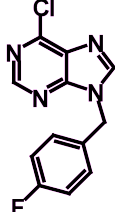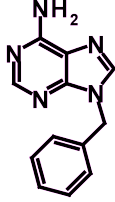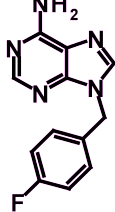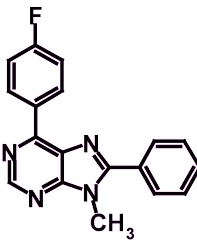
IUPAC	Publication ID	Internal ID	Structures	Activity
Phenyl-(9*H*-purin-6-yl)-amin	2 (1)	DH 1		-
(4-Fluor-phenyl)-(9H-purin-6-yl)-amin	3 (2)	DH 2		-
(9-Methyl-9H-purin-6-yl)-phenyl-amin	4 (5)	DH 14		++
(4-Fluor-phenyl)-(9-methyl-9H-purin-6-yl)-amin	5 (6)	DH 29		++
6-Phenoxy-9*H*-purin	9 (3)	DH 3		-
6-(4-Fluor-phenoxy)-9*H*-purin	10 (4)	DH 4		-

113
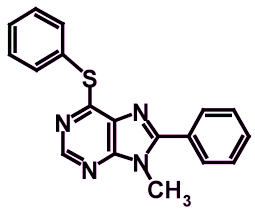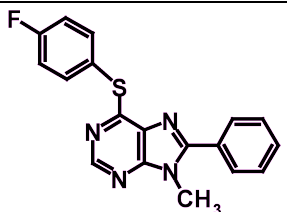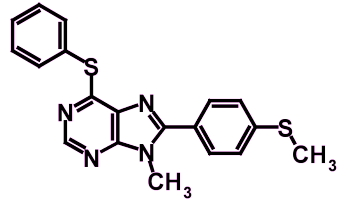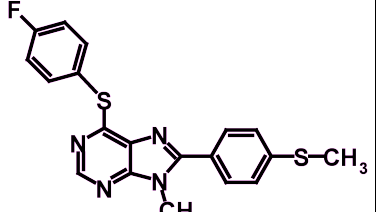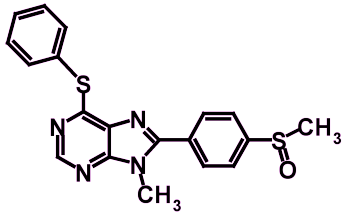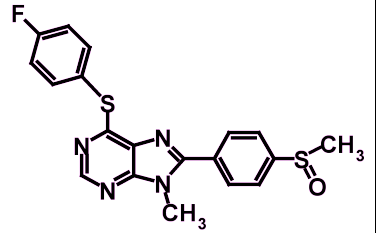
6-Phenylsulfanyl-9*H*-purin	11 (7)	DH 33		-
6-(4-Fluor-phenylsulfanyl)-9*H*-purin	12 (8)	DH 34		-
9-Methyl-6-phenylsulfanyl-9*H*-purin	13 (11)	DH 36		+/-
6-(4-Fluor-phenylsulfanyl)-9-methyl-9*H*-purin	14 (12)	DH 35		-
4-Phenylsulfanyl-7*H*-pyrrolo[2,3-*d*]pyrimidin	20 (9)	DH 143		-
4-(4-Fluor-phenylsulfanyl)-7*H*-pyrrolo[2,3-*d*]pyrimidin	21 (10)	DH 144		-

114
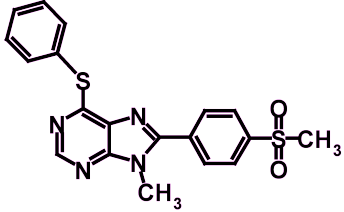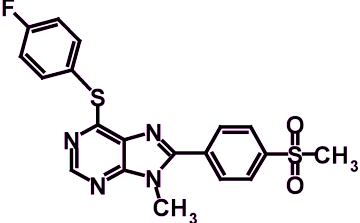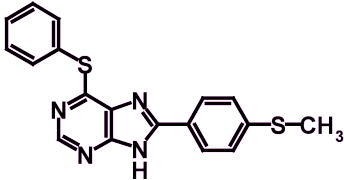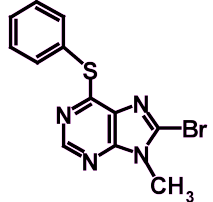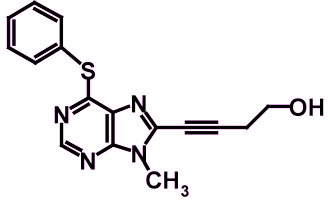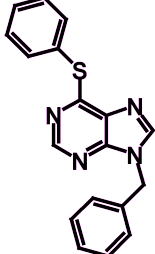
6-Benzyl-9*H*-purin	25 (29)	DH 181		-
6-(4-Fluor-benzyl)-9*H*-purin	26 (30)	DH 196		-
6-Styryl-9H-purin	27 (31)	DH 193		++
6-[2-(4-Fluor-phenyl)-vinyl]-9*H*-purin	28	DH 240		-
9-Methyl-6-styryl-9*H*-purin	29 (32)	DH 128		-
6-Benzylsulfanyl-9*H*-purin	31 (13) nsc29421	DH 17		-
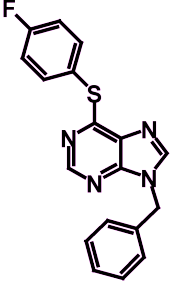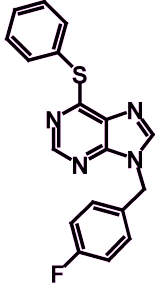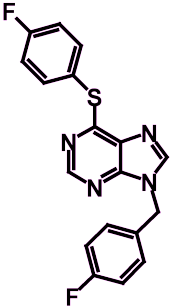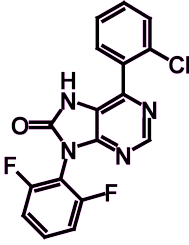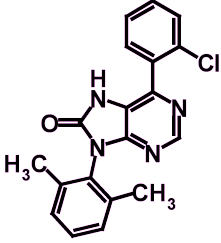
6-(4-Fluor-benzylsulfanyl)-9*H*-purin	32 (14)	DH 18		-
6-Benzylsulfanyl-9-methyl-9*H*-purin	33 (15) nsc38303	DH 42		-
6-(4-Fluor-benzylsulfanyl)-9-methyl-9*H*-purin	34 (16)	DH 41		-
9-Methyl-6-phenyl-9*H*-purin	41 (39)	DH 70		-
6-(4-Fluor-phenyl)-9-methyl-9*H*-purin	42 (40)	DH 69		-
7-Phenyl-7*H*-purin	45 (19)	DH 12		-
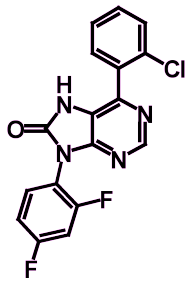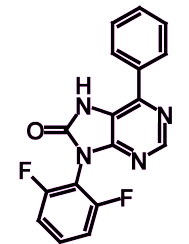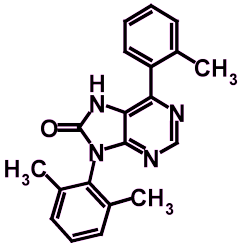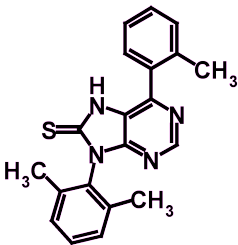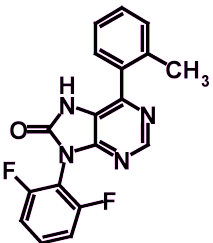
7-(4-Fluor-phenyl)-7*H*-purin	46 (20)	DH 11		-
7-Benzyl-7*H*-purin	49 (17)	DH 53		-
7-(4-Fluor-benzyl)-7*H*-purin	(18)	DH 55		-
7 Benzyl 6 chlor 7*H* purin	51 (22)	DH 58B		-
6 Chlor (4 fluor benzyl) 7*H* purin	52 (24)	DH 59B		-
7-(4-Fluor-benzyl)-7*H*-purin-6-ylamin	53 (27)	DH 73A		-
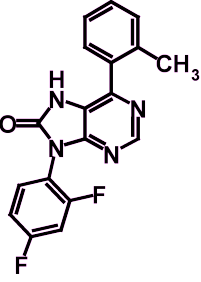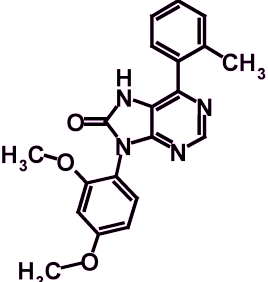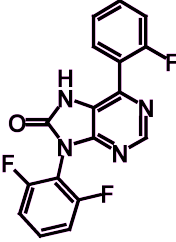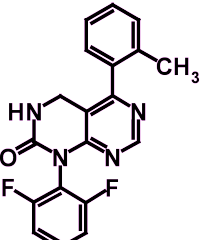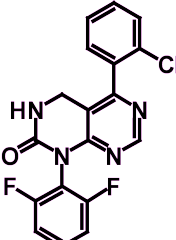
7 (4 Fluor benzyl) 6 methoxy 7*H* purin	54 (28)	DH 73B		+/-
9-Benzyl-6-chlor-9*H*-purinund	55 (21)	DH 58A		-
6-Chlor-9-(4-fluor-benzyl)-9*H*-purin	56 (23)	DH 59A		-
9-Benzyl-9*H*-purin-6-ylamin	57 (25)	DH 74		-
9-(4-Fluor-benzyl)-9*H*-purin-6-ylamin	58 (26)	DH 72		-
6-(4-Fluor-phenyl)-9-methyl-8-phenyl-9*H*-purin	59 (41)	DH 71		-

9-Methyl-8-phenyl-6-phenylsulfanyl-9*H*-purin	66 (44)	DH 50		-
6-(4-Fluor-phenylsulfanyl)-9-methyl-8-phenyl-9*H*-purin	67 (45)	DH 57		-
9-Methyl-8-(4-methylsulfanyl-phenyl)-6-phenylsulfanyl-9*H*-purin	68 (49)	DH 154		-
6-(4-Fluor-phenylsulfanyl)-9-methyl-8-(4-methylsulfanyl-phenyl)-9*H*-purin	69 (50)	DH 156		-
8-(4-Methansulfinyl-phenyl)-9-methyl-6-phenylsulfanyl-9*H*-purin	70 (51)	DH 167		-
6-(4-Fluor-phenylsulfanyl)-8-(4-methan-sulfinyl-phenyl)-9-methyl-9*H*-purin	71 (52)	DH 168		-

8-(4-Methansulfonyl-phenyl)-9-methyl-6-phenylsulfanyl-9*H*-purin	72 (53)	DH 171		-
6-(4-Fluor-phenylsulfanyl)-8-(4-methansulfonyl-phenyl)-9-methyl-9*H*-purin	73 (54)	DH 172		-
8-(4-Methylsulfanyl-phenyl)-6-phenylsulfanyl-9H-purin	76	DH 180		-
8-Brom-9-methyl-6-phenylsulfanyl-9*H*-purin	78 (55)	DH 216		-
4-(9-Methyl-6-phenylsulfanyl-9*H*-purin-8-yl)-but-3-yn-1-ol	79 (56)	DH 217		-
9-Benzyl-6-phenylsulfanyl-9*H*-purin	80 (57)	DH 199		++

9-Benzyl-6-(4-fluor-phenylsulfanyl)-9*H*-purin	81 (58)	DH 202		++
9-(4-Fluor-benzyl)-6-phenylsulfanyl-9*H*-purin	82 (59)	DH 203		+/-
9-(4-Fluor-benzyl)-6-(4-fluor-phenylsulfanyl)-9*H*-purin	83 (60)	DH 204		-
6-(2-Chlor-phenyl)-9-(2,6-difluor-phenyl)-7,9-dihydro-purin-8-on	104 (21, ref2)	DH 227		-
6-(2-Chlor-phenyl)-9-(2,6-dimethyl-phenyl)-7,9-dihydro-purin-8-on	105 (22, ref2)	DH 262		+/-

6-(2-Chlor-phenyl)-9-(2,4-difluor-phenyl)-7,9-dihydro-purin-8-on	106 (23, ref2)	DH 238		-
9-(2,6-Difluor-phenyl)-6-phenyl-7,9-dihydro-purin-8-on	107 (24, ref2)	DH 233		-
9-(2,6-Dimethyl-phenyl)-6-o-tolyl-7,9-dihydro-purin-8-on	108 (25, ref2)	DH 250		-
9-(2,6-Dimethyl-phenyl)-6-o-tolyl-7,9-dihydro-purin-8-thion	109	DH 251		-
9-(2,6-Difluor-phenyl)-6-o-tolyl-7,9-dihydro-purin-8-on	110 (26, ref2)	DH 237		-

9-(2,6-Difluoro-phenyl)-6-o-tolyl-7,9-dihydro-purin-8-on	111 (27, ref2)	DH 244		-
9-(2,4-Dimethoxy-phenyl)-6-o-tolyl-7,9-dihydro-purin-8-on	112 (28, ref2)	DH 249		-
9-(2,6-Difluor-phenyl)-6-(2-fluor-phenyl)-7,9-dihydro-purin-8-on	113 (29, ref2)	DH 275		-
1-(2,6-Difluor-phenyl)-5-o-tolyl-3,4-dihydro-1H-pyrimido[4,5 d]pyrimidin-2-on	121	DH 321		-
5-(2-Chlor-phenyl)-1-(2,6-difluor-phenyl)-3,4-dihydro-1H pyrimido[4,5 d]pyrimidin-2-on	122	DH 332		-

Publication ID numbers are given by Professor Laufer, ID in brackets are the actual IDs assigned in the final print of their publication[164]. Some compounds were published in a separate publication[165], denoted as ref2 in the table. Compounds were initially tested as a pair, and pairs that tested positive were then subjected for individual testing.

References

1. Y. Yarden and M. X. Sliwkowski. Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell. Biol.* **2**, 127-137 (2001).
2. R. N. Jorissen, et al. Epidermal growth factor receptor: mechanisms of activation and signalling. *Exp Cell Res* **284**, 31-53 (2003).
3. E. M. Bublil and Y. Yarden. The EGF receptor family: spearheading a merger of signaling and therapeutics. *Curr Opin Cell Biol* **19**, 124-134 (2007).
4. M. A. Olayioye, R. M. Neve, H. A. Lane and N. E. Hynes. The ErbB signaling network: receptor heterodimerization in development and cancer. *Embo J* **19**, 3159-67 (2000).
5. A. Ullrich, et al. Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature* **309**, 418-25 (1984).
6. M. A. Lemmon. Ligand-induced ErbB receptor dimerization. *Exp Cell Res* **315**, 638-648 (2009).
7. K. Oda, Y. Matsuoka, A. Funahashi and K. Hiroaki. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol Syst Biol* **1**(2005).
8. N. E. Hynes and H. A. Lane. ErbB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer* **5**, 341-354 (2005).
9. A. Gschwind, O. M. Fischer and A. Ullrich. The discovery of receptor tyrosine kinases: targets for cancer therapy. *Nat Rev Cancer* **4**, 361-370 (2004).
10. N. Hynes. Targeting ERBB Receptors in Cancer. *Recent Res Cancer* **172**, 45-57 (2007).
11. A. B. Motoyama, N. E. Hynes and H. A. Lane. The Efficacy of ErbB Receptor-targeted Anticancer Therapeutics Is Influenced by the Availability of Epidermal Growth Factor-related Peptides. *Cancer Res* **62**, 3151-3158 (2002).
12. N. V. Sergina, et al. Escape from HER-family tyrosine kinase inhibitor therapy by the kinase-inactive HER3. *Nature* **445**, 437-441 (2007).
13. R. Levi-Montalcini. EFFECTS OF MOUSE TUMOR TRANSPLANTATION ON THE NERVOUS SYSTEM. *Ann NY Acad Sci* **55**, 330-343 (1952).
14. S. Cohen and R. Levi-Montalcini. Purification and Properties of a Nerve Growth-promoting Factor Isolated from Mouse Sarcoma 180. *Cancer Res* **17**, 15-20 (1957).
15. R. Levi-Montalcini and S. Cohen. EFFECTS OF THE EXTRACT OF THE MOUSE SUBMAXILLARY SALIVARY GLANDS ON THE SYMPATHETIC SYSTEM OF MAMMALS. *Ann NY Acad Sci* **85**, 324-341 (1960).
16. S. Cohen. Isolation of a Mouse Submaxillary Gland Protein Accelerating Incisor Eruption and Eyelid Opening in the New-born Animal. *J. Biol. Chem.* **237**, 1555-1562 (1962).
17. S. Cohen and G. A. Elliot. The Stimulation of Epidermal Keratinization by a Protein Isolated from the Submaxillary Gland of the Mouse1. *J. Invest. Dermatol.* **40**, 1-5 (1963).
18. S. Cohen. The stimulation of epidermal proliferation by a specific protein (EGF). *Dev Biol* **12**, 394-407 (1965).

19. R. H. Starkey, S. Cohen and D. N. Orth. Epidermal growth factor: identification of a new hormone in human urine. *Science* **189**, 800-802 (1975).

20. H. Gregory. Isolation and structure of urogastrone and its relation to epidermal growth factor. *Nature* **257**, 325-327 (1975).

21. S. Cohen and G. Carpenter. Human epidermal growth factor: isolation and chemical and biological properties. *Proc Natl Acad Sci U S A* **72**, 1317-1321 (1975).

22. M. D. Hollenberg and P. Cuatrecasas. Epidermal Growth Factor: Receptors in Human Fibroblasts and Modulation of Action by Cholera Toxin. *Proc Natl Acad Sci U S A* **70**, 2964-2968 (1973).

23. G. Carpenter, K. J. Lembach, M. M. Morrison and S. Cohen. Characterization of the binding of 125-I-labeled epidermal growth factor to human fibroblasts. *J. Biol. Chem.* **250**, 4297-4304 (1975).

24. G. Carpenter, L. King and S. Cohen. Epidermal growth factor stimulates phosphorylation in membrane preparations in vitro. *Nature* **276**, 409-410 (1978).

25. S. Cohen, G. Carpenter and L. King, Jr. Epidermal growth factor-receptor-protein kinase interactions. Co- purification of receptor and epidermal growth factor-enhanced phosphorylation activity. *J. Biol. Chem.* **255**, 4834-4842 (1980).

26. S. Cohen, H. Ushiro, C. Stoscheck and M. Chinkers. A native 170,000 epidermal growth factor receptor-kinase complex from shed plasma membrane vesicles. *J. Biol. Chem.* **257**, 1523-1531 (1982).

27. S. Cohen, R. Fava and S. Sawyer. Purification and characterization of epidermal growth factor receptor/protein kinase from normal mouse liver. *Proc Natl Acad Sci U S A* **79**, 6237-6241 (1982).

28. A. L. Schechter, et al. The neu oncogene: an erb-B-related gene encoding a 185,000-Mr tumour antigen. *Nature* **312**, 513-516 (1984).

29. L. Coussens, et al. Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene. *Science* **230**, 1132-1139 (1985).

30. M. H. Kraus, W. Issing, T. Miki, N. C. Popescu and S. A. Aaronson. Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor receptor family: evidence for overexpression in a subset of human mammary tumors. *Proc Natl Acad Sci U S A* **86**, 9193-9197 (1989).

31. G. D. Plowman, et al. Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal growth factor receptor family. *Proc Natl Acad Sci U S A* **90**, 1746-1750 (1993).

32. C. Chang and P. W. Sternberg. C. elegans Vulval Development as a Model System to Study the Cancer Biology of EGFR Signaling. *Cancer Metast Rev* **18**, 203-213 (1999).

33. R. Schweitzer and B.-Z. Shilo. A thousand and one roles for the Drosophila EGF receptor. *Trends Genet* **13**, 191-196 (1997).

34. A. Reich and B.-Z. Shilo. A new ligand of the Drosophila epidermal growth factor receptor, undergoes two modes of cleavage. *J EMBO* **21**, 4287-4296 (2002).

35. D. E. Klein, S. E. Stayrook, F. Shi, K. Narayan and M. A. Lemmon. Structural basis for EGFR ligand sequestration by Argos. *Nature* **453**, 1271-1275 (2008).

36.   Y. Yarden and J. Schlessinger. Epidermal growth factor induces rapid, reversible aggregation of the purified epidermal growth factor receptor. *Biochemistry* **26**, 1443-51 (1987).

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

38.   R. Perez-Soler. HER1/EGFR targeting: refining the strategy. *Oncologist* **9**, 58-67 (2004).

39.   S. K. Pal and M. Pegram. Epidermal growth factor receptor and signal transduction: potential targets for anti-cancer therapy. *Anticancer Drugs* **16**, 483-94 (2005).

40.   N. Prenzel, O. M. Fischer, S. Streit, S. Hart and A. Ullrich. The epidermal growth factor receptor family as a central element for cellular signal transduction and diversification. *Endocr Relat Cancer* **8**, 11-31 (2001).

41.   C. C. Bancroft, et al. Effects of pharmacologic antagonists of epidermal growth factor receptor, PI3K and MEK signal kinases on NF-kappaB and AP-1 activation and IL-8 and VEGF expression in human head and neck squamous cell carcinoma lines. *Int J Cancer* **99**, 538-48 (2002).

42.   J. P. Dawson, et al. Epidermal growth factor receptor dimerization and activation require ligand-induced conformational changes in the dimer interface. *Mol. Cell. Biol.* **25**, 7734-7742 (2005).

43.   K. M. Ferguson, et al. EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Mol. Cell.* **11**, 507-517 (2003).

44.   T. P. J. Garrett, et al. Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor [alpha]. *Cell* **110**, 763-773 (2002).

45.   H. Ogiso, et al. Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* **110**, 775-787 (2002).

46.   F. Walker, et al. CR1/CR2 interactions modulate the functions of the cell surface epidermal growth factor receptor. *J. Biol. Chem.* **279**, 22387-22398 (2004).

47.   E. Tzahar, et al. A hierarchical network of interreceptor interactions determines signal transduction by Neu differentiation factor/neuregulin and epidermal growth factor. *Mol Cell Biol* **16**, 5276-87 (1996).

48.   D. Graus-Porta, R. R. Beerli, J. M. Daly and N. E. Hynes. ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. *Embo J* **16**, 1647-55 (1997).

49.   X. Qian, C. M. LeVea, J. K. Freeman, W. C. Dougall and M. I. Greene. Heterodimerization of epidermal growth factor receptor and wild-type or kinase-deficient Neu: a mechanism of interreceptor kinase activation and transphosphorylation. *Proc Natl Acad Sci U S A* **91**, 1500-1504 (1994).

50.   P. M. Guy, J. V. Platko, L. C. Cantley, R. A. Cerione and K. L. Carraway. Insect cell-expressed p180erbB3 possesses an impaired tyrosine kinase activity. *Proc Natl Acad Sci U S A* **91**, 8132-8136 (1994).

51.   D. Slamon, et al. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177-182 (1987).

52. F. Ciardiello. Epidermal growth factor receptor inhibitors in cancer treatment. *Future Oncol* **1**, 221-234 (2005).

53. J. Marshall. Clinical implications of the mechanism of epidermal growth factor receptor inhibitors. *Cancer* **107**, 1207-1218 (2006).

54. J. Reis-Filho, et al. Metaplastic breast carcinomas exhibit EGFR, but not HER2, gene amplification and overexpression: immunohistochemical and chromogenic in situ hybridization analysis. *Breast Cancer Res* **7**, R1028 - R1035 (2005).

55. J. R. Grandis and J. C. Sok. Signaling through the epidermal growth factor receptor during the development of malignancy. *Pharmacol Ther* **102**, 37-46 (2004).

56. D. S. Salomon, R. Brandt, F. Ciardiello and N. Normanno. Epidermal growth factor-related peptides and their receptors in human malignancies. *Crit Rev Oncol Hemat* **19**, 183-232 (1995).

57. J. Dancer, H. Takei, J. Y. Ro and L.-N. Mary. Coexpression of EGFR and HER-2 in pancreatic ductal adenocarcinoma: a comparative study using immunohistochemistry correlated with gene amplification by fluorescencent in situ hybridization. *Oncol Rep* **18**, 151-155 (2007).

58. N. Normanno, et al. Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene* **366**, 2-16 (2006).

59. P. Kirkpatrick, J. Graham and M. Muhsin. Cetuximab. *Nat. Rev. Drug. Discov.* **3**, 549-550 (2004).

60. L. Saltz, C. Easley and P. Kirkpatrick. Panitumumab. *Nat Rev Drug Discov* **5**, 987-988 (2006).

61. M. A. Molina, et al. Trastuzumab (Herceptin), a Humanized Anti-HER2 Receptor Monoclonal Antibody, Inhibits Basal and Activated HER2 Ectodomain Cleavage in Breast Cancer Cells. *Cancer Res* **61**, 4744-4749 (2001).

62. J. Dowell, J. D. Minna and P. Kirkpatrick. Erlotinib hydrochloride. *Nat Rev Drug Discov* **4**, 13-14 (2005).

63. B. Moy, P. Kirkpatrick, S. Kar and P. Goss. Lapatinib. *Nat Rev Drug Discov* **6**, 431-432 (2007).

64. M. Muhsin, J. Graham and P. Kirkpatrick. Gefitinib. *Nat Rev Drug Discov* **2**, 515-516 (2003).

65. W. Pao, et al. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med* **2**, e73 (2005).

66. S. V. Sharma, D. W. Bell, J. Settleman and D. A. Haber. Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer* **7**, 169-81 (2007).

67. D. W. Bell, et al. Inherited susceptibility to lung cancer may be associated with the T790M drug resistance mutation in EGFR. *Nat Genet* **37**, 1315-1316 (2005).

68. D. Abd El-Rehim, et al. Expression and co-expression of the members of the epidermal growth factor receptor (EGFR) family in invasive breast carcinoma. *Br J Cancer* **91**, 1532-1542 (2004).

69. A. J. Wong, et al. Structural alterations of the epidermal growth factor receptor gene in human gliomas. *Proc Natl Acad Sci U S A* **89**, 2965-2969 (1992).

70. L. Zhan, B. Xiang and S. K. Muthuswamy. Controlled Activation of ErbB1/ErbB2 Heterodimers Promote Invasion of Three-Dimensional Organized

Epithelia in an ErbB1-Dependent Manner: Implications for Progression of ErbB2-Overexpressing Tumors. *Cancer Res* **66**, 5201-5208 (2006).

71.   F. F. Vajdos, et al. Comprehensive Functional Maps of the Antigen-binding Site of an Anti-ErbB2 Antibody Obtained with Shotgun Scanning Mutagenesis. *J Mol Biol* **320**, 415-428 (2002).

72.   J. Bostrom, et al. Variants of the Antibody Herceptin That Interact with HER2 and VEGF at the Antigen Binding Site. *Science* **323**, 1610-1614 (2009).

73.   T. Holbro and N. E. Hynes. ERBB RECEPTORS: Directing Key Signaling Networks Throughout Life. *Annual Review of Pharmacology and Toxicology* **44**, 195-217 (2004).

74.   N. E. Hynes and G. MacDonald. ErbB receptors and signaling pathways in cancer. *Curr Opin Cell Biol* **21**, 177-184 (2009).

75.   http://dtp.nci.nih.gov/branches/dscb/diversity_explanation.html.

76.   G. M. Morris, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639-1662 (1998).

77.   R. Huey, G. M. Morris, A. J. Olson and D. S. Goodsell. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **28**, 1145-1152 (2007).

78.   H. Chen, P. D. Lyne, F. Giordanetto, T. Lovell and J. Li. On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors. *J. Chem. Inf. Model.* **46**, 401-415 (2006).

79.   C. Semenkovich, M. Wims, L. Noe, J. Etienne and L. Chan. Insulin regulation of lipoprotein lipase activity in 3T3-L1 adipocytes is mediated at posttranscriptional and posttranslational levels. *J. Biol. Chem.* **264**, 9030-9038 (1989).

80.   S. Nakata, et al. Regulation of platelet-derived growth factor receptor activation by afadin through SHP-2: implications for cellular morphology. *J. Biol. Chem.* **282**, 37815-37825 (2007).

81.   K. S. Yang, M. X. G. Ilagan, D. Piwnica-Worms and L. J. Pike. Luciferase Fragment Complementation Imaging of Conformational Changes in the Epidermal Growth Factor Receptor. *J. Biol. Chem.* **284**, 7474-7482 (2009).

82.   S. Li, et al. Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* **7**, 301-311 (2005).

83.   K. E. Luker, et al. Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proc. Natl. Acad. Sci. USA.* **101**, 12288-93 (2004).

84.   K. M. Ferguson, P. J. Darling, M. J. Mohan, T. L. Macatee and M. A. Lemmon. Extracellular domains drive homo- but not hetero-dimerization of ErbB receptors. *EMBO J* **19**, 4632–4643 (2000).

85.   H. K. Gan, et al. The EGFR tyrosine kinase inhibitor AG1478 increases the formation of inactive untethered EGFR dimers: Implications for combination therapy with mab 806. *J Biol Chem* **282**, 2840-2850 (2007).

86.   W. Liu, et al. Relationship of EGFR Mutations, Expression, Amplification, and Polymorphisms to Epidermal Growth Factor Receptor Inhibitors in the NCI60 Cell Lines. *Clin. Cancer. Res.* **13**, 6788-6795 (2007).

87.   P. C. Bishop, et al. Differential sensitivity of cancer cells to inhibitors of the epidermal growth factor receptor family. *Oncogene* **21**, 119-127 (2002).

88.   D. Vernimmen, M. Gueders, S. Pisvin, P. Delvenne and R. Winkler. Different mechanisms are implicated in ERBB2 gene overexpression in breast and in other cancers. *Br. J. Cancer.* **89**, 899-906 (2003).

89.   R. Wang, L. Lai and S. Wang. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided. Mol. Des.* **16**, 11-26 (2002).

90.   C. Zhang, S. Liu, Q. Zhu and Y. Zhou. A knowledge-based energy function for protein-ligand, protein-protein, and protein-DNA complexes. *J. Med. Chem.* **48**, 2325-2335 (2005).

91.   V. Villalobos, S. Naik and D. Piwnica-Worms. Detection of Protein-Protein Interactions in Live Cells and Animals with Split Firefly Luciferase Protein Fragment Complementation. in *Genomics Protocols* Vol. 439 339-352 (Humana Press, 2008).

92.   J. L. Macdonald and L. J. Pike. Heterogeneity in EGF-binding affinities arises from negative cooperativity in an aggregating system. *Proc. Natl. Acad. Sci. USA* **105**, 112-7 (2008).

93.   International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945 (2004).

94.   H. M. Berman, et al. The Protein Data Bank and the challenge of structural genomics. *Nat Struct Mol Biol* **7**, 957-959 (2000).

95.   S. K. Burley, et al. Structural genomics: beyond the Human Genome Project. *Nat Genet* **23**, 151-157 (1999).

96.   K. H. Bleicher, H.-J. Bohm, K. Muller and A. I. Alanine. Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov* **2**, 369-378 (2003).

97.   Y. Tanrikulu and G. Schneider. Pseudoreceptor models in drug design: bridging ligand- and receptor-based virtual screening. *Nat Rev Drug Discov* **7**, 667-677 (2008).

98.   R. W. Spencer. High-throughput screening of historic collections: Observations on file size, biological targets, and file diversity. *Biotechnol Bioeng* **61**, 61-67 (1998).

99.   W. P. Walters, M. T. Stahl and M. A. Murcko. Virtual screening - an overview. *Drug Discov Today* **3**, 160-178 (1998).

100. N. Brooijmans and I. D. Kuntz. Molecular Recognition and Docking Algorithms. *Annu Rev Biophy Biom* **32**, 335-373 (2003).

101. B. K. Shoichet. Virtual screening of chemical libraries. *Nature* **432**, 862-865 (2004).

102. M. Schapira, et al. Discovery of diverse thyroid hormone receptor antagonists by high-throughput docking. *P Natl Acad Sci USA* **100**, 7354-7359 (2003).

103. S. Betzi, et al. Protein-protein interaction inhibition (2P2I) combining high throughput and virtual screening: Application to the HIV-1 Nef protein. *P Natl Acad Sci USA* **104**, 19256-19261 (2007).

104. T. I. Oprea and H. Matter. Integrating virtual screening in lead discovery. *Curr Opin Chem Biol* **8**, 349-358 (2004).

105. M. J. Levesque, K. Ichikawa, S. Date and J. H. Haga. Design of a grid service-based platform for in silico protein-ligand screenings. *Comput Meth Prog Bio* **93**, 73-82 (2009).

106. W. G. Richards. Virtual screening using grid computing: the screensaver project. *Nat Rev Drug Discov* **1**, 551-555 (2002).

107. M. W. Chang, W. Lindstrom, A. J. Olson and R. K. Belew. Analysis of HIV Wild-Type and Mutant Structures via in Silico Docking against Diverse Ligand Libraries. *J. Chem. Inf. Model.* **47**, 1258-1262 (2007).

108. A. Herráez. Biomolecules in the computer: Jmol to the rescue. *Biochem Mol Biol Edu* **34**, 255-261 (2006).

109. L. Olsen, I. Pettersson, L. Hemmingsen, H.-W. Adolph and F. S. Jogensen. Docking and scoring of metallo-beta-lactamases inhibitors. *J Comput Aid Mol Des* **18**, 287-302 (2004).

110. J. R. Schames, et al. Discovery of a Novel Binding Trench in HIV Integrase. *J. Med. Chem.* **47**, 1879-1881 (2004).

111. R. Wang, X. Fang, Y. Lu and S. Wang. The PDBbind Database: Collection of Binding Affinities for Proteinâˆ'Ligand Complexes with Known Three-Dimensional Structures. *J Med Chem* **47**, 2977-2980 (2004).

112. R. Wang, X. Fang, Y. Lu, C.-Y. Yang and S. Wang. The PDBbind Database: Methodologies and Updates. *J Med Chem* **48**, 4111-4119 (2005).

113. T. Cheng, X. Li, Y. Li, Z. Liu and R. Wang. Comparative Assessment of Scoring Functions on a Diverse Test Set. *J Chem Inf Mod* **49**, 1079-1093 (2009).

114. S. Renner, S. Derksen, S. Radestock and F. Morchen. Maximum Common Binding Modes (MCBM): Consensus Docking Scoring Using Multiple Ligand Information and Interaction Fingerprints. *J Chem Inf Mod* **48**, 319-332 (2008).

115. B. Gorelik and A. Goldblum. High quality binding modes in docking ligands to proteins. *Proteins: Structure, Function, and Bioinformatics* **71**, 1373-1386 (2008).

116. G. M. Morris, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **9999**, NA (2009).

117. R. Guha, et al. The Blue Obelisk Interoperability in Chemical Informatics. *J Chem Inf Mod* **46**, 991-998 (2006).

118. N. Jacq, et al. Grid-enabled Virtual Screening Against Malaria. *J Grid Comput* **6**, 29-43 (2008).

119. D. Bullard, A. Gobbi, M. A. Lardy, C. Perkins and Z. Little. Hydra: A Self Regenerating High Performance Computing Grid for Drug Discovery. *J. Chem. Inf. Model.* **48**, 811-816 (2008).

120. S. P. Brown and S. W. Muchmore. High-Throughput Calculation of Protein-Ligand Binding Affinities: Modification and Adaptation of the MM-PBSA Protocol to Enterprise Grid Computing. *J. Chem. Inf. Model.* **46**, 999-1005 (2006).

121. D. Karunagaran, et al. ErbB-2 is a common auxiliary subunit of NDF and EGF receptors: implications for breast cancer. *EMBO J* **15**, 254-64 (1996).

122. A. W. Burgess, et al. An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol Cell* **12**, 541-52 (2003).

123. H.-S. Cho, et al. Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* **421**, 756-760 (2003).

124. T. P. J. Garrett, et al. The Crystal Structure of a Truncated ErbB2 Ectodomain Reveals an Active Conformation, Poised to Interact with Other ErbB Receptors. *Mol Cell* **11**, 495-505 (2003).

125. S. Paik and E. T. Liu. HER2 as a predictor of therapeutic response in breast cancer. *Breast Disease* **11**, 91-102 (2000).

126. J. D. A. Tyndall, B. Pfeiffer, G. Abbenante and D. P. Fairlie. Over One Hundred Peptide-Activated G Protein-Coupled Receptors Recognize Ligands with Turn Structure. *Chemical Reviews* **105**, 793-826 (2005).

127. E. G. Hutchinson and J. M. Thornton. A revised set of potentials for beta-turn formation in proteins. *Protein Science* **3**, 2207-2216 (1994).

128. T. Tran, et al. Topological side-chain classification of β-turns: Ideal motifs for peptidomimetic development. *J Comput Aid Mol Des* **19**, 551-566 (2005).

129. J. A. Wells. Systematic mutational analyses of protein-protein interfaces. **202**, 390-411 (1991).

130. T. Clackson and J. A. Wells. A hot spot of binding energy in a hormone-receptor interface. **267**, 383-6 (1995).

131. I. Massova and P. A. Kollman. Computational Alanine Scanning To Probe Protein-Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem. Soc.* **121**, 8133-8143 (1999).

132. C. A. Smith and T. Kortemme. Backrub-Like Backbone Simulation Recapitulates Natural Protein Conformational Variability and Improves Mutant Side-Chain Prediction. *J Mol Biol* **380**, 742-756 (2008).

133. I. W. Davis, W. B. Arendall III, D. C. Richardson and J. S. Richardson. The Backrub Motion: How Protein Backbone Shrugs When a Sidechain Dances. *Structure* **14**, 265-274 (2006).

134. A. M. Honegger, A. Schmidt, A. Ullrich and J. Schlessinger. Evidence for epidermal growth factor (EGF)-induced intermolecular autophosphorylation of the EGF receptors in living cells. *Mol Cell Biol* **10**, 4035-44 (1990).

135. J. Schlessinger. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **103**, 211-225 (2000).

136. X. Yu, K. D. Sharma, T. Takahashi, R. Iwamoto and E. Mekada. Ligand-independent dimer formation of epidermal growth factor receptor (EGFR) is a step separable from ligand-induced EGFR signaling. *Mol Biol Cell* **13**, 2547-57 (2002).

137. S. Saffarian, Y. Li, E. L. Elson and L. J. Pike. Oligomerization of the EGF receptor investigated by live cell fluorescence intensity distribution analysis. *Biophys J* **93**, 1021-31 (2007).

138. R.-H. Tao and I. N. Maruyama. All EGF(ErbB) receptors have preformed homo- and heterodimeric structures in living cells. *J Cell Sci* **121**, 3207-3217 (2008).

139. J. M. Mendrola, M. B. Berger, M. C. King and M. A. Lemmon. The Single Transmembrane Domains of ErbB Receptors Self-associate in Cell Membranes. *J. Biol. Chem.* **277**, 4704-4712 (2002).

140. K. W. Thiel and G. Carpenter. Epidermal growth factor receptor juxtamembrane region regulates allosteric tyrosine kinase activation. *Proc Natl Acad Sci U S A* **104**, 19238-43 (2007).

141. E. Penuel, R. W. Akita and M. X. Sliwkowski. Identification of a region within the ErbB2/HER2 intracellular domain that is necessary for ligand-independent association. *J Biol Chem* **277**, 28468-73 (2002).

142. C. L. Arteaga and J. Baselga. Tyrosine kinase inhibitors: Why does the current process of clinical development not apply to them? *Cancer Cell* **5**, 525-531 (2004).

143. K. D. Miller. The Role of ErbB Inhibitors in Trastuzumab Resistance. *Oncologist* **9**, 16-19 (2004).

144. S. E. Wang, et al. HER2 kinase domain mutation results in constitutive phosphorylation and activation of HER2 and EGFR and resistance to EGFR tyrosine kinase inhibitors. *Cancer Cell* **10**, 25-38 (2006).

145. D. L. Wheeler, et al. Mechanisms of acquired resistance to cetuximab: role of HER (ErbB) family members. *Oncogene* **27**, 3944-3956 (2008).

146. J. Baselga and C. L. Arteaga. Critical Update and Emerging Trends in Epidermal Growth Factor Receptor Targeting in Cancer. *J Clin Oncol* **23**, 2445-2459 (2005).

147. M. P. Jacobson, R. A. Friesner, Z. Xiang and B. Honig. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *J Mol Biol* **320**, 597-608 (2002).

148. M. A. Larkin, et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).

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

150. D. V. D. Spoel, et al. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **26**, 1701-1718 (2005).

151. W. L. Jorgensen. OPLS force fields. **3**, 1986–1989 (1998).

152. M. W. Mahoney and W. L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J Chem Phys* **112**, 8910-8922 (2000).

153. J. Mervis. ADVANCED COMPUTING: NSF Launches TeraGrid for Academic Research. *Science* **293**, 1235-1237 (2001).

154. P. A. Kollman, et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **33**, 889-897 (2000).

155. L. T. Chong, Y. Duan, L. Wang, I. Massova and P. A. Kollman. Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proc Natl Acad Sci U S A* **96**, 14330-14335 (1999).

156. T. Hou, L. Zhu, L. Chen and X. Xu. Mapping the Binding Site of a Large Set of Quinazoline Type EGF-R Inhibitors Using Molecular Field Analyses and Molecular Docking Studies. *J. Chem. Inf. Model.* **43**, 273-287 (2003).

157. C. Luo, et al. Computational analysis of molecular basis of 1:1 interactions of NRG-1 wild-type and variants with ErbB3 and ErbB4. *Proteins* **59**, 742-756 (2005).

158. T. J. Dolinsky, et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* **35**, W522-5 (2007).

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

160. N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041 (2001).

161. B. Kuhlman, et al. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **302**, 1364-1368 (2003).

162. D. W. Sammond, et al. Structure-based Protocol for Identifying Mutations that Enhance Protein-Protein Binding Affinities. *J Mol Biol* **371**, 1392-1404 (2007).

163. P. Willett. Similarity and Clustering in Chemical Information Systems. (1987).

164. S. A. Laufer, D. M. Domeyer, T. R. F. Scior, W. Albrecht and D. R. J. Hauser. Synthesis and Biological Testing of Purine Derivatives as Potential ATP-Competitive Kinase Inhibitors. *J Med Chem* **48**, 710-722 (2005).

165. D. R. J. Hauser, T. Scior, D. M. Domeyer, B. Kammerer and S. A. Laufer. Synthesis, Biological Testing, and Binding Mode Prediction of 6,9-Diarylpurin-8-ones as p38 MAP Kinase Inhibitors. *J Med Chem* **50**, 2060-2066 (2007).

166. J. Baselga. Targeting Tyrosine Kinases in Cancer: The Second Wave. *Science* **312**, 1175-1178 (2006).

167. S. Frantz. Drug discovery: Playing dirty. *Nature* **437**, 942-943 (2005).