Rational approach to selective inhibitor design using multitarget constraints

Thesis Proposal Exam

Steven Albanese
MSKCC // Chodera Lab
Gerstner Sloan Kettering Graduate School of Biomedical Sciences
April 13, 2016

I. Specific Aims

Small molecule kinase inhibitors have become a major focus of drug development for treating cancer, which accounted for 550,000 deaths and 1.5 million diagnoses in the United States in 2015 alone. Currently, there are 31 FDA approved kinase inhibitors. The dominant paradigm for designing such inhibitors has been to optimize maximally selective ligands for a single target. Unfortunately, many such inhibitors fail in clinical trials due to a lack of efficacy and clinical safety. Tumors can evade inhibitors through multiple routes of resistance, including upregulation of a second kinase, mutations in the target kinase or amplification of the target kinase. On the other hand, toxicity arises from on-target inhibition of the wild type kinase or off-target effects of promiscuous small molecules or their metabolites. ATP-competitive kinase inhibitors have great potential for promiscuity, as there are over 500 kinases in the human kinome that each binds a common substrate, ATP. These issues could be addressed through a different paradigm of drug development where small molecules are optimized for selectivity profiles that inhibit desired targets while avoiding binding to unwanted antitargets.

A challenge facing this paradigm is the difficulty of rationally designing a single compound with desired affinities to multiple targets, which is currently a task of extraordinary difficulty for both medicinal and computational chemists. Recent advances in GPU-accelerated alchemical free energy calculations present an opportunity to use computation to both assess the potential to achieve this kind of chemical selectivity with kinase inhibitors and to build a tool to realize it through automated chemical design. We propose to extend a novel computational approach to search chemical space for small molecules that maximize binding affinities to specified positive targets and minimizes affinities for specified antitargets, to identify small molecules with desired selectivity profiles. We consider both a well-studied, experimentally tractable model system—Abl and Src—as well as clinically interesting cases such as EGFR and the HER family of kinases.

Aim 1. Investigate the details necessary to quantitatively predict inhibitor selectivities. To assess the accuracy of GPU-accelerated alchemical free energy calculations for quantitative prediction of binding affinities and instruct which additional chemical details must be added, we will recapitulate experimentally measured free energies of binding to Src and Abl for a large subset of FDA-approved inhibitors. The Chodera lab has developed an expanded ensemble simulation utilizing nonequilibirum candidate Monte Carlo (NCMC), reversible jump Markov Chain Monte Carlo and stochastic approximation-based sampling to predict free energies of binding. If these calculations are not sufficiently accurate to be useful in designing for selectivity, additional details, such as protonation state, phosphorylation state, conformational reorganization or tautomeric state, can be added in a sequential manner and tested systematically within this framework.

Aim 2. Analyze how multitargeted design constraints influence chemical space available to inhibitors. We will investigate how chemical space is narrowed depending on multiple design constraints, rationalizing the chemical features that might be driving the desired selectivity profiles. The above simulation technique will be extended to search through chemical space to find small molecules optimized to satisfy multiple positive and negative design constraints. As chemical space is very large, this process can be biased to only search synthetically accessible space, commercially available kinase-targeted libraries, and drug-like compounds. This algorithm will allow for the proposal of putative small molecules that could satisfy a specific selectivity profile, such as binding to EGFR but not HER2. We will test these predicted binders in the laboratory using biophysical binding assays and examine whether QSAR models are sufficient to rationalize the small molecule features that are important for a given selectivity.

Aim 3. Evaluate the potential for mutant selective small molecule inhibitors On-target toxicity limits the dosage and effectiveness of drugs, while resistance mutations develop quickly in many patients, limiting the length of response to first line treatments. These issues could be addressed if inhibitors selective for oncogenic mutants could be developed. We will investigate whether small molecules can be sufficiently selective for the oncogenic mutant form of kinases, and how chemical space is narrowed by this constraint. We will use the above algorithm to target kinases with oncogenic mutations and antitarget wild type kinases, proposing libraries of molecules that target therapeutically relevant mutants. We will investigate these proposed binders using cheminformatics approaches to learn the features contributing to this selectivity.

II. Background and Significance

Small Molecule Kinase inhibitors are an important area of drug development

In 2015 alone, cancer accounted for 550,000 deaths and 1.5 million new diagnoses in the United States[1]. Since the FDA approval of imatinib in 2001, therapeutics targeting kinases now account for over 50% of current cancer drug discovery and close to 30% of **total** drug discovery efforts[2], with 31 FDA approved small molecule kinase inhibitors (SMKIs) on the market[3]. However, there has been a decrease in productivity using current design strategies, with many drugs failing in late stage clinical trials. By the time a drug fails in Phase III, a typical pharmaceutical company has spent 12 years and almost $1 billion on development[4]. SMKIs can fail late in the development pipeline for two main reasons: safety issues or lack of efficacy. Tumors have multiple routes to resistance, including target amplification, effectively increasing the amount of drug required to get the same level of inhibition. Inhibitor resistance occurs through the presence or upregulation of a redundant pathway, mutation of the target kinase[5], activation of downstream kinases[6], or relief of feedback inhibition[7]. On-target toxicity, from inhibition of wild type kinase, can cause efficacy issues by limiting the maximally tolerated dose (MTD). Safety issues arise from adverse events due off-target toxicity, such as geftinib inhibiting CYP2D6[8] and causing hepatotoxicity in lung cancer patients, or from the on-target toxicity of inhibiting the wild type kinase[9,10].

Kinase inhibitor selectivity can vary dramatically

Each kinase inhibitor has a certain **selectivity profile, or group of biological targets a molecule binds to and inhibits strongly enough to produce a phenotype.** Kinase inhibitors have potential for a great diversity of **selectivity---the number of targets a molecule binds to below a certain K_d threshold.** There are 518[11] members of the human kinome, each with a highly similar, druggable ATP-binding site (**Figure 1A**) [12-16], giving inhibitors targeted to them huge potential for promiscuity, like staurosporine, which inhibits a large percentage of the kinome often with very high K_d. (**Figure 1B**). Even FDA approved drugs have a wide range of selectivities **(Figure 1B)**[17]. In a 2011 paper, Davis *et al.,* characterized the interaction of 72 known kinase inhibitors against a panel of 442 kinases[17] using a competitive binding assay. Of the 72 compounds screened, 70% had a $K_d < 3$ μM for more than 10% of the 442 distinct kinases screened. While this study confirmed that type II inhibitors, SMKIs that bind an active site adjacent pocket exposed in the 'DFG-out' conformation, are more likely to be selective than Type I inhibitors, those that can bind to either the 'DFG-out' or 'DFG-in' conformation, it also found that there are several type II inhibitors that have low selectivity. Conversely, several Type I inhibitors exhibited a high level of selectivity. This suggests that either binding mode is a viable option when seeking to design a selective inhibitor. Additionally, 17 of the 72 compounds bound to fewer than 5 off-target kinases with affinity comparable to their primary target and also had a $K_d < 3$ μM for less than 10% of the assayed kinases. This suggests that it is possible to design compounds that are selective for multiple targets, a strategy that has been suggested as a possible design paradigm termed **targeted polypharmacology**[6,18-20].

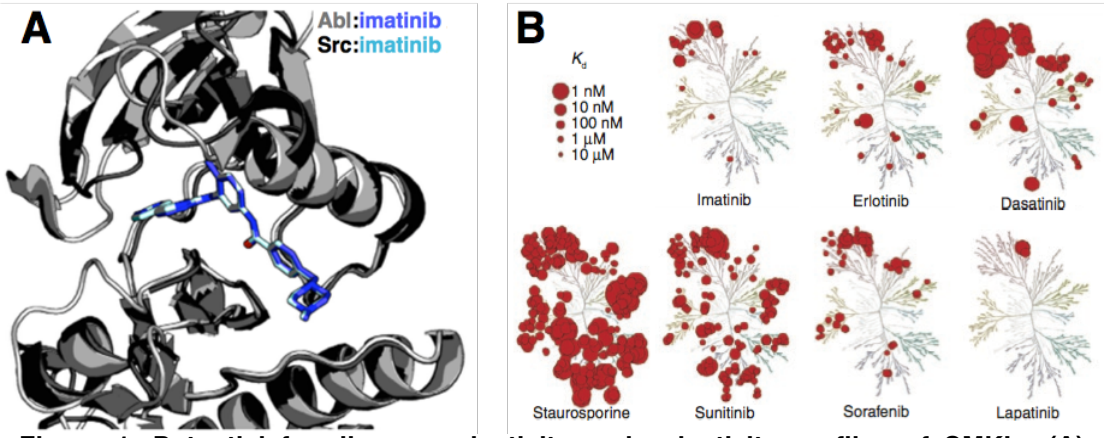

Figure 1. Potential for diverse selectivity and selectivity profiles of SMKIs. (A) Crystal structure of Src (black) [12] and Abl (gray) [13] bound to imatinib, which have almost identical binding sites and poses, highlighting the challenge in achieving selectivity when targeting the ATP binding pocket. **(B)** Selectivity maps [12] highlight the wide range of selectivities SMKIs can have. Even drugs used in the clinic bind multiple kinases.

Multitarget drug design could address the limitations of current design strategies

Current design efforts focus on achieving maximal selectivity for a single target by improving a weak inhibitor through analogue synthesis[21], which is not always rational, or through structure-informed design[21,22], which is difficult because kinases exist as nodes in complex signaling networks[23,24], with feedback inhibition and pathway cross-talk complicating the relationship between binding and signaling. This complicates the notion of inhibiting a single kinase to shutdown a pathway, as alleviating negative feedback can lead to re-activation of the target pathway[7,23], or lead to activation of a secondary pathway previously regulated by inhibitory cross-talk[23,25]. Further, tumors can easily evade inhibition of a single target[6] by mutating the target to ablate inhibitor binding; mutating downstream effectors to bypass the inhibited node in the pathway; up-regulating a redundant signaling pathway or branch; or up-regulating the target kinase to increase the amount of drug needed for efficacious inhibition to occur. Instead, it may be possible to address some of the challenges of SMKI development by using multiple targets as design constraints. This design paradigm could improve selectivity by antitargeting kinases closely related to the desired therapeutic target, such as positively designing an inhibitor for EGFR while antitargeting HER2. Multitarget design could also reduce on-target toxicity, thereby improving the therapeutic window, by targeting the oncogenic mutated kinase and antitargeting the wild type kinase. This could improve upon the success of certain EGFR inhibitors such as gefitinib and erlotinib[26-28] or aid the development of second generation inhibitors for use in treating patients with clinically-acquired resistance mutations, such as the ALK inhibitor alectinib[29]. Additionally, designing inhibitors with desired selectivity profiles could be leveraged to target multiple kinases[30,31]. For example, recent work[24] suggests that combined EGFR and MEK inhibition prevents the emergence of drug resistance, potentially prolonging treatment for EGFR-mutant non-small cell lung cancer patients. This presents an opportunity to harness the promiscuity of SMKIs to develop a molecule that could inhibit both EGFR and MEK. A single multitarget drug might be preferable on practical terms, by avoiding the expense of developing multiple novel molecular entities. However, it is unknown if multitarget design strategies limit chemical space so severely that it becomes impossible to identify small molecules that satisfy the constraints.

Alchemical free energy calculations can be used to predict binding affinities

Virtual screening, a commonly used computational tool to aid rational drug design[32], focuses on enrichment[33] of potential binders after screening large libraries of compounds. While in widespread use, this approach is not true rational design, as docking scoring functions do not correlate with ligand binding affinity[34]. To achieve computational efficiency, virtual screening makes a number of approximations, such as averaging over ligand binding modes or receptor rigidity that lead to inaccuracies in predicting binding affinities. On the other hand, free energy calculations[35] offer a promising route to efficiently calculate the free energy of

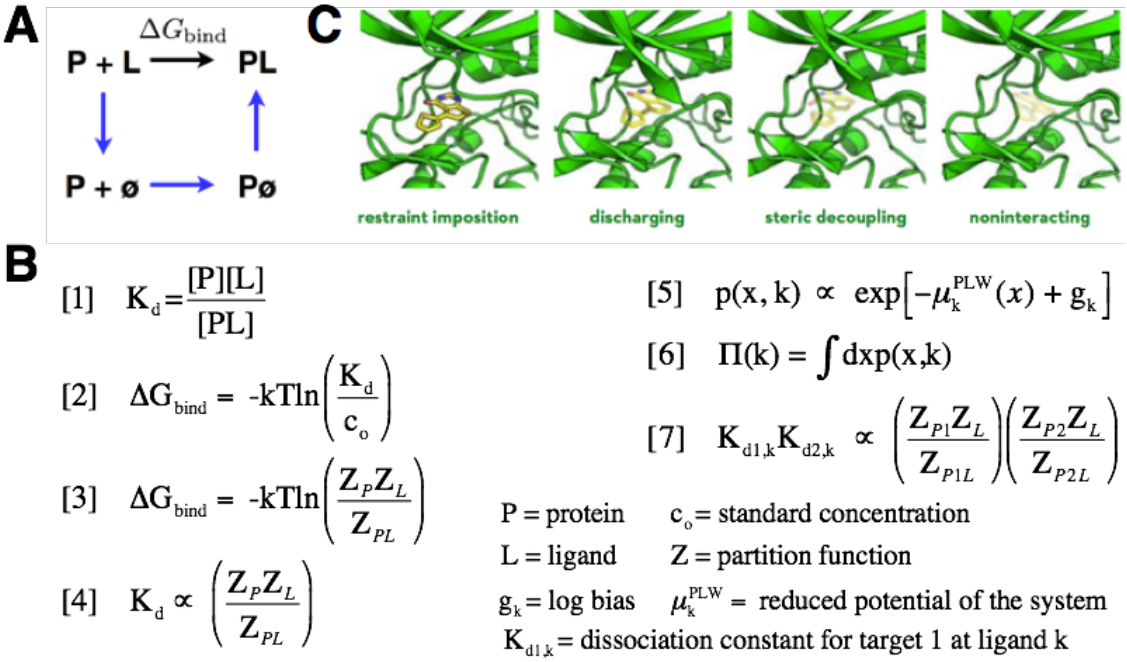

Figure 2. Free energy calculation scheme and equations. (A) Thermodynamic cycle of binding. (B) Equations and theory behind PERSES algorithm (C) Illustration of how alchemical intermediates turn off electrostatic and van der waals interactions to decouple the ligand.

binding, which is related to the dissociation constant **(Figure 2B, equation [1])** as shown in **Figure 2B, equation [2]**. In theory, the free energy of binding could be estimated by simulating multiple binding and unbinding events for a given drug:receptor pairing. However, unbinding is very slow, on the order of hours for typical drugs, which would take ~10^6 years to simulate. Since free energy is a state function and independent of the pathway taken, we can use an alternative pathway **(Figure 2A, blue arrows)** of alchemical intermediates **(Figure 2C)** in which the ligand interactions are gradually turned off or 'decoupled', allowing for a more efficient calculation of the free energy of binding. From statistical mechanics, we can relate the free energy of binding to partition functions **(Figure 2B, equation [3])**, which are quantities that relate the atomistic details of our simulations to the thermodynamic properties we are interested in. Using simulation, we are able to estimate ratios of partition functions, which will be proportional to the affinities we are interested in **(Figure 2B, equation [4])**.

Using this framework, we have developed a simulation scheme **(Figure 3)** where we use the expanded ensemble method to sample from a joint state space **(Figure 2B, equation [5])** containing a flexible receptor and solvent system **x**; and a compound, indexed at **k**. This system is sampled from using Markov chain Monte Carlo (MCMC) (**Figure 3A)**, which propagates the coordinates of the system, **x**. Over the course of the simulation, changes to the chemical identity of the ligand are proposed using reversible jump Markov chain Monte Carlo (RJMCMC)[36] (**Figure 3B)**, which allows for changing degrees of freedom as the number of atoms in the ligand changes. Nonequilibirum candidate Monte Carlo (NCMC)[37], a technique based on nonequilibirum statistical mechanics that allows the acceptance probabilities of Monte Carlo moves to be exponentially enhanced, is used to enhance acceptance rates during Monte Carlo transitions from the original ligand, indexed at **k**, and the newly proposed ligand, **k+1 (Figure 3C)**. Over the course of the simulation, self-adjusted mixture sampling (SAMS) is used to adjust the logarithmic bias, $\mathbf{g_k}$, of the joint state space so that the marginal distribution for compound **k (Figure 2B, equation [6])** is proportional to the binding affinity of the ligand for the receptor (**Figure 3D)**. To accommodate multitarget drug design, multiple coupled simulations can be run simultaneously, such that instead of a single affinity, the simulation will actually predict the product of affinities for two different proteins **(Figure 2B, equation [7])**, in the case of designing a dual target inhibitor. This simulation scheme is flexible and can be

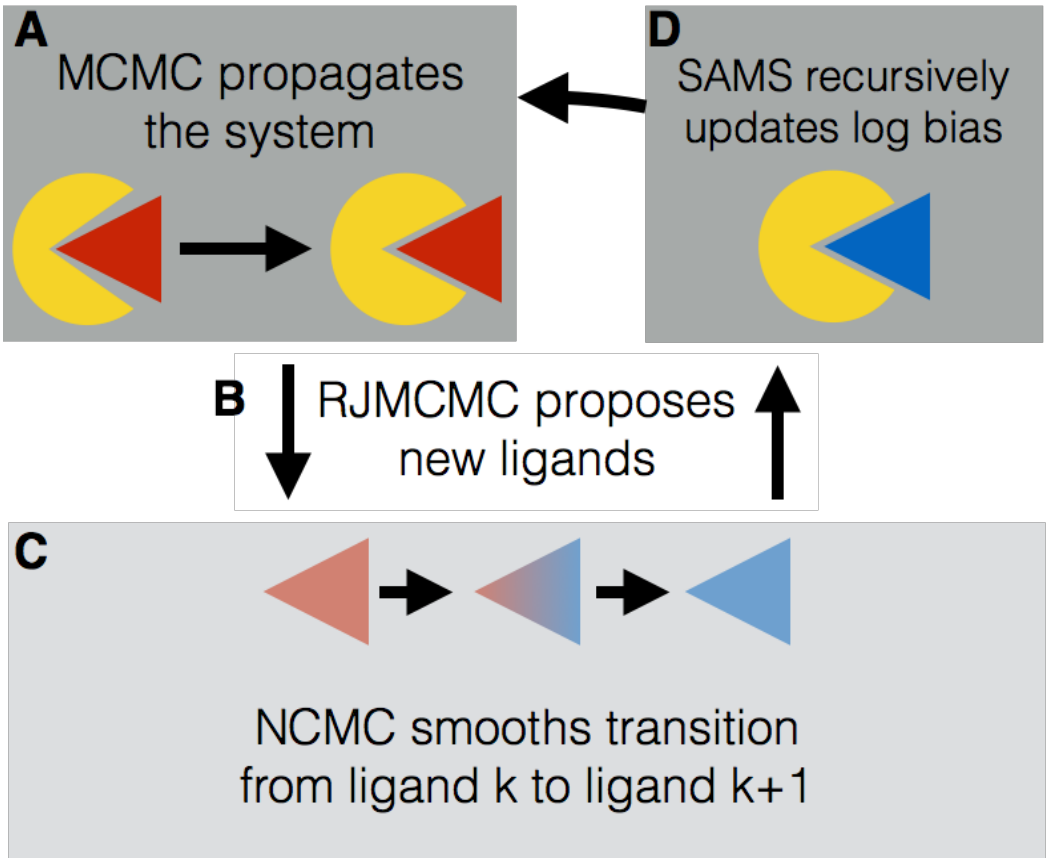

Figure 3. PERSES simulation scheme. (A) MCMC propagates the coordinates of system x **(B)** RJMCMC proposes new chemical species **(C)** NCMC smooths transition between ligands k and k+1 to increase acceptance rates **(D)** SAMS recursively updates the logarithmic bias so that the marginal distribution is proportional to the free energy of binding

adapted for both positive and negative design. Ultimately, these simulations will output a collection of ligands and their relative free energies of binding, allowing for the prioritization of molecules that are predicted to be better binders than others. Additional extensions can be made to calculate absolute free energies, which can be directly compared with biophysical binding measurements. The chemical space from which ligand proposals are drawn can be biased in a number of ways. For example, chemical space can be limited to a well-curated library of kinase inhibitors or expanded to include all commercially available molecules.

Src and Abl are a well-understood test system for validating the prediction of affinities and selectivities

The initial success of Imatinib[38], an inhibitor designed to target BCR-Abl and the Abl subfamily of kinases, in treating chronic myeloid leukemia (CML) spurred much initial hope in developing SMKIs for other cancers[39]. Imatinib binds to Abl with about 3,000 times the affinity it binds to Src, its closest relative[13,40]. The difference in affinity was originally attributed to the inability of Src to adopt an imatinib-bound state[41,42]. Surprisingly, a crystal structure was solved showing that Src can adopt an Abl-like conformation and bind imatinib with an almost identical pose[13], while Abl was shown to adopt a Src-like inactive conformation[43]. Subsequent work has involved extensive NMR[44,45], molecular dynamics[46-48], and biochemical[49] characterization of the two kinases with the intent of explaining this difference in binding affinity, which would help inform the design of selective inhibitors. This wealth of crystallographic[50] and biochemical data, known targeted inhibitors[3,51-54], and ability to express the proteins in bacteria[55] make Src and Abl ideal to use as a model system.

EGFR is a clinically interesting test case for rational drug design

Epidermal growth factor receptor (EGFR) is a receptor tyrosine kinase, a member of the ErbB family along with HER2, HER3 and HER4, which transduces mitogenic signals across the plasma membrane and are often deregulated in human cancers[56]. As a monomer, EGFR adopts a Src-like autoinhibited inactive state[57]. Upon binding of an extracellular ligand, such as EGF or TGFα, EGFR forms hetero- or homo- dimers with other members of the ErbB family, including pseudokinase HER3[26,58]. These dimers adopt an asymmetric, head-to-tail conformation in which one kinase domain stabilizes the active conformation of the second[59,60] through a highly hydrophobic interface, which can be disrupted by missense mutations at the interface[59,61,62]. EGFR mutations are present in as much as 26% of NSCLC patients[63]. One-third of all cancer deaths worldwide are caused by lung cancer[64,65] and as many as 80% of all lung cancer patients have NSCLC[65-67]. The most common alterations are a small in-frame deletion in exon 19 and missense mutation L858R[65], which spurred the development of successful first generation and second inhibitors[68] such as erlotinib[69,70], lapatinib[71], and gefitinib[70,72,73]. Extensive structural and biochemical studies have been performed on L858R[27], which was found to adopt an active conformation similar to the wild type protein and have a K_{cat} 50 fold higher[27] than wild type EGFR. This mutant is proposed to activate EGFR by increasing dimerization affinity, thereby shifting the population dynamics of EGFR toward the active state[74]. Both gefitinib and erlotinib have been shown to be selective for this mutant over the wild type kinase[70,75] and growing evidence suggests that these drugs can bind both the active and inactive forms of the kinase, with varying affinities[59,76]. As such, it is possible that differences in the measured affinities, which are a cumulative average of the drug's affinity for each individual kinase conformation, are reflective of a shift in the proportion of the kinase populating the active state for L858R. However, most patients develop resistance to these inhibitors within 9 to 14 months[63,77], through resistance mutations (T790M, 60%)[5,28], activation of secondary kinases (MET[78], FGFRs[79]) or alterations downstream of EGFR that reactivate ERK signaling[24,80]. The mechanism of T790M resistance remains unclear, but it has been suggested to act through increasing affinity for ATP in the L858R double-mutants[81]. Other work suggests that T790M introduces steric hindrance that impedes inhibitor binding[82]. To address resistance, a number of successor inhibitors are in development to target the T790M mutation[66]. This system provides a clinically interesting case that can be served by mutant-specific drugs that target either an oncogenic or resistance mutant more specifically than the WT form of the kinase, as well as a small family within which multitargeted inhibitors, such as lapatinib, would be beneficial.

III. Preliminary Data

kinase	phosphatase	yield (μg/mL)
MK14_HUMAN	Lambda	70.7
VRK3_HUMAN	Lambda	67.5
MK08_HUMAN	Lambda	28.5
CDK16_HUMAN	Lambda	26.9
SRC_HUMAN	YopH	22.0
STK16_HUMAN	Lambda	20.7
ABL1_HUMAN	YopH	2.5
DAPK1_HUMAN	Lambda	2.4
DYRK2_HUMAN	Lambda	2.4

Figure 3. Expression Results by kinase. A selection of kinases showing yield and co-expressed phosphatase. Highlighted in red are Src and Abl.

To aid in the automated high-throughput expression of recombinant human kinases for biophysical binding assays, the lab built a structural informatics pipeline[55] to select domain targets from constructs deposited in the PDB that were expressed in E. *coli*. This lead to the identification of 96 kinase domain constructs that were cloned into 2BT10 plasmids that contained N-terminal His10-TEV tags and coexpressed with Lambda phosphatase (for Ser/Thr kinases) or YopH (for Tyr kinases). These constructs were expressed at the QB3 Macrolab at UC Berkeley in Rosetta2 cells in a 96 well plate in 1mL cultures and purified using Ni-NTA resin in a 96-well plate. Protein yields and purity were determined using LabChip GX II, a microfluidic gel electrophoresis system. As in **Figure 3**, both Src and Abl expressed well and are expected to yield more than 2mg/L when scaled up. This high-throughput and automated protocol will enable rapid production of human kinase for use in biophysical binding experiments.

To generate high quality binding data to compare calculations to experiment, the lab has developed a high-throughput 96-well plate version of a previously reported cuvette-based fluorescence binding assay[83]. **Figure 4A** highlights the direct binding version of this assay, using well-characterized inhibitors, such as gefitinib and bosutinib, as fluorescent probes. By titrating the amount of probe, exciting at 280nm and measuring emission at 480nm, Bayesian analysis can be used to accurately determine ligand binding affinities. Non-fluorescent ligands, such as imatinib, can be accessed by using the competitive version of this assay (**Figure 4B)**, which allows for the characterization of most ATP-competitive inhibitors.

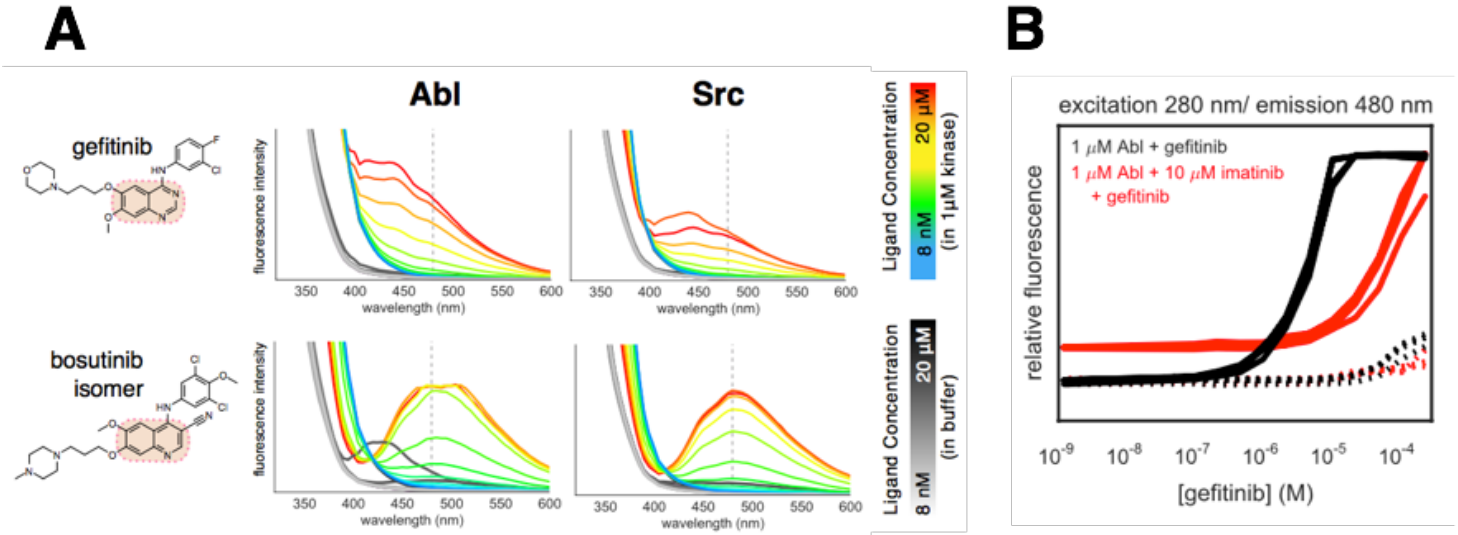

Figure 4. Automated fluorescence assay for label-free measuring of kinase inhibitor binding affinities. (A) *Left:* Gefitinib (quinazoline scaffold) and bosutinib (quinolone scaffold) are two of many fluorescent probes that can be used. *Right:* Fluorescence emission spectra (excitation 280nm) for a wide range of probe concentrations show increase in fluorescence upon binding. Dashed line at 480nm indicates emission wavelength. **(B)** Competition assay using gefitinib as a fluorescent probe to measure affinities for non-fluorescent inhibitors like imatinib. Solid lines indicate the presence of Abl while dashed lines are the molecules alone.

IV. Research Design

Aim 1. Investigate the details necessary to quantitatively predict inhibitor selectivities

Rationale and significance

In order to rationally design small molecules using free energy calculations, we must first assess the current predictive accuracy. What level of detail is required to be accurate enough to impact drug design? An accuracy of less than 2 kcal/mol error is estimated to speed up lead optimization three-fold over traditional medicinal chemistry when only potency optimization is considered[84]. Current methods operate at the molecular mechanics level of detail, with entropy, enthalpy, and conformational flexibility accounted for. However, a number of phenomena are not explicitly considered by the calculation, such as conformational reorganization in kinases; protonation state of both the kinase and the inhibitor; and phosphorylation of the kinase. Should our calculations not reach a useful accuracy threshold as is, we can systematically add details to test if they are required. To do so, we will first use a GPU-accelerated alchemical free energy calculation to estimate the relative free energy of binding of a small set of known kinase inhibitors for Src and Abl. We will then compare the results to biophysical experiments, to gauge the accuracy of our predictions. Using medicinal chemistry data from Abbott[84], we can build a statistical model of medicinal chemists making changes to improve selectivity and affinity, allowing us to determine the utility of our tool based on its accuracy. This will allow us to assess the current calculations and provides a framework to systematically test which details must be added to achieve a useful level of accuracy.

Experimental approaches

Free energy calculations will predict known inhibitor binding affinities

We have developed an algorithm that utilizes MCMC, RJMCMC, NCMC, and SAMS to estimate the relative free energies of binding for ligands using GPU-accelerated simulations. These calculations, using a molecular mechanics forcefield (e.g. AMBER99SB-ILDN[85]), will use a GPU computing cluster at MSKCC. Such alchemical free energy calculations are able to rigorously compute binding affinities to individual conformations[86-90]. This particular approach works by using NCMC to smooth the transition between a series of alchemical intermediates in which the ligand is morphed into a new chemical species. To calculate a relative free energy that includes all relevant statistical mechanical effects, parts of the ligand that are modified are gradually 'decoupled' and morphed into a new chemical species before being 'recoupled'. To test the accuracy of our calculations, Src and Abl will be used as a model system. The free energy calculations will be run using non-covalent FDA-approved SMKIs[51] as the chemical library through which the algorithm searches. This small group will serve as a useful benchmark to test the free energy calculations on, with a dynamic range of selectivities spanning from nM to mM.

Comparing calculation to experiment

In order to determine the accuracy of these calculations, it is important to compare *in silico* work to experimental data. While there is a large amount of publically available affinity data[17,91], much of the data is poorly curated, improperly treated statistically, or from indirect binding assays with varying conditions that impact measured affinities[92-94]. As such, we will collect our own high-quality set of affinity measurements for non-covalent FDA-approved SMKIs to use as a validation set utilizing a high-throughput direct fluorescence readout of binding to recombinantly expressed kinases. First, we will express soluble Src and Abl kinase domains in a high-throughput 96-well, automated fashion. The kinase domains will be cloned into 2BT10 plasmids with an N-terminal His10-TEV cleavable tag and coexpressed with YopH164, a phosphatase that increases expression levels and ensures the kinase remains unphosphorylated[95]. Construct boundaries were identified from the PDB and previous work[55]. These constructs will be expressed in E. *coli,* which is superior to insect cells in terms of cost and convenience compared. Purification will be done using Ni-NTA beads and after tag-cleavage with TEV protease, purity can be assessed using microfluidic gel electrophoresis using a LabChip GX II (Caliper LifeSciences). Phosphorylation state will be confirmed by mass spectrometry.

With the recombinantly expressed kinases, we will use a high-throughput fluorescence assay to measure Src and Abl kinase inhibitor binding affinities. Quinazoline and quinolone scaffold-based ATP-competitive inhibitors greatly increase in fluorescence upon binding kinase[83], providing a label-free and kinase-activity independent method to measure kinase inhibitor affinities. This method can be used to directly measure affinities for fluorescent drugs or for non-fluorescent drugs via competition with fluorescent drug probes. This biophysical experiment takes advantage of the fact that all available ATP-competitive inhibitors will bind to every kinase with varying affinities[96,97]. This assay has several advantages over other methods, requiring only micrograms of protein for a direct binding measurement over a wide range (nM-mM) of affinities. The assay works as a single-wavelength fluorescence measurement with excitation at 280 nm and emission at 480nm, for which Bayesian analysis allows for the accurate determination of inhibitor affinity and its confidence interval, using code developed within the Chodera lab[98]. As a negative control for the direct binding assay, we will use inhibitor in solution without kinase, which will demonstrate that the increase in fluorescence is dependent on binding to the protein and not a function of increasing concentration. The negative controls for the competition assay are the fluorescent probe both alone and with the non-fluorescent ligand of interest in solution without protein. The first control is to show that there is no concentration-dependent increase in fluorescence. The control with both inhibitors demonstrates that the shift in the fluorescence curve is due to the fluorescent probe competing with the non-fluorescent ligand to bind the kinase, and not due to an interaction between the two small molecules.

Possible outcomes and alternative approaches

Using data from Abbott medicinal chemists, we can construct a statistical model that will estimate the utility of our tools for a given level of accuracy. If the current methods are sufficiently accurate for rational design, we can move to a kinase inhibitor targeted library of a few thousand compounds. Importantly, this library would not require the expensive synthesis of large numbers of compounds to test in the lab. If the calculations do not reach the desired accuracy, a number of details can be added and tested in a systematic manner. Discrepancies between the predicted and measured affinities can be attributed to three main sources: (1) missing conformational states of the protein; (2) incomplete treatment of chemical effects such as protonation state changes in the protein[99] or ligand[100], or tautomerization[101]; and (3) deficiencies in the forcefield. Should conformational reorganization or multiple conformations that do not quickly interchange be a required detail, we can run multiple coupled simulations, each one starting with a different conformation of the same kinase. The same schema can be used if multiple phosphorylation states must be accounted for in predicting binding affinity. If ligand protonation state is important, constant-pH methodologies[102] currently under investigation in the Chodera lab can be incorporated into PERSES. With the high-quality data set generated from the fluorescence-based assay, it will be possible to rigorously test which of these potential solutions will bring the error to an acceptable margin. Alternatively, if it is not possible to reach the desired level of accuracy, the free energies can be useful as qualitative scores to rank molecules, in a fashion similar to docking scores. This would still represent an improvement as multiple ligand binding poses and protein conformations are considered. This process would also inform future development efforts by exploring which details are required to approach quantitative accuracy, something that as yet has not been rigorously tested in this framework

The fluorescence assay is limited in the range of affinities it can measure at the low end by the minimal detectable fluorescence and at the high end by the solubility limit of the fluorescent probes (about 50μM). Should this range be insufficient for a given kinase model system, multiple probes can be used to extend this range. If an unbound inhibitor absorbs the emission band of the kinase:probe complex, causing fluorescence interference, we can either monitor emission at multiple wavelengths or chose a different fluorescent probe with a different emission wavelength. As kinase phosphorylation state is important for some inhibitors such as bosutinib, we can allow the kinase to autophosphorylate[103] at high concentration and confirm phosphorylation state with mass spectrometry. Alternatively, we could incubate the kinases with Hck kinase domain[83]. Isothermal titration calorimetry (ITC) [104,105] is a commonly used alternative method. While is possible to use as an alternative to the fluorescence assay, it is not high-throughput and consumes large amounts of protein. This method is also limited by the solubility of the small molecules.

Aim 2. Analyze how multitargeted design constraints influence chemical space available to inhibitors.

Rationale and significance

The feasibility of multitarget drug design depends on how multiple constraints narrow the chemical space available to a small molecule that satisfy the criteria, making small molecule discovery difficult or limiting us to less drug-like molecules that have undesirable physical properties. We can test this by running calculations with multiple constraints and rationalizing the molecular features that drive the desired selectivity. Using multiple positive targets as constraints to search through a kinase-targeted inhibitor library, we can propose dual-target inhibitors, such as lapatinib[57,106] which targets both EGFR and HER2 and is highly selective compared to the other FDA-approved SMKIs[17,91]. This design strategy may prove useful in designing inhibitors for EGFR and MEK, which would be useful in prolonging response to SMKIs in non-small cell lung cancer[24]. We can also seek to improve selectivity within a family of kinases by positively targeting one kinase and antitargeting the others. This would be useful for creating molecules selective for a single member of the ErbB family, while minimizing its affinity for the other ErbB kinases. This approach can also be adapted for the purposes of designing Type I or Type II inhibitors, which differ on whether they bind to the DFG-in or DFG-out conformation of a kinase[51,107], by using multiple conformations of a single kinase as design constraints. We will use alchemical free energy calculations to propose libraries of molecules that meet a desired selectivity profile and analyze them to identify potential new fragments or scaffolds that drive such selectivity using cheminformatics techniques. This methodology will also allow us to potentially identify new binding modes or poses that could facilitate achieving the desired selectivity.

Experimental approaches

Free energy calculations using multitarget design constraints

To utilize multiple design constraints, the alchemical free energy calculations will be expanded to use multiple coupled simulations, each seeded with a different kinase structure as a design constraint. These simulations will be coupled by the SAMS re-weighting step, which adjusts the logarithmic bias, $\mathbf{g_k}$, **(Figure 2B, eqn. [5])**. To maximize binding affinity to both targets, we will adopt an expanded objective such as **Figure 2B, eqn. [7]**. By searching chemical space to maximize this objective, we will propose and rank molecules based on **the product of ligand binding affinities for both targets.** Maximizing the **ratio of binding affinities for each target** would minimize binding affinity for the antitarget while maximizing the affinity for the positive design target. Using this scheme, we will perform three different calculations: targeting both EGFR and HER2, targeting EGFR and antitargeting HER2, antitargeting EGFR and targeting HER2. These calculations will be run on MSKCC's GPU cluster using the AMBER99SB-ILDN forcefield and the kinase inhibitor library KINA (ChemBridge Research Labs), which contains 3200 compounds available through the RNAi/HTSC core at MSKCC. These simulations will output three libraries of molecules proposed based on each of the three different selectivity profiles.

Validate proposed libraries using fluorescence-based assays

We will validate our predictions by selecting a subset of molecules to measure binding affinities experimentally. To determine whether to use the direct or competition assay, we will perform a full spectrum analysis using Src and Abl binding. The concentration of each molecule will be varied between 8nM and 20uM, both with and without kinase. We will use a known fluorescent molecule as a positive control to ensure that fluorescence is being detected. The negative control for each new compound will be the series of spectra collected without kinase present, which will tell us whether any changes in fluorescence or dependent on ligand binding to protein. These experiments will determine whether we use a direct or competition based assay, as described in Aim 1, for each of the molecules we chose to validate. These binding affinity assays will use recombinantly expressed wild type EGFR and HER2 kinase domains, which were used in the alchemical free energy calculations that proposed these molecules. We will express soluble EGFR and HER2 in a high-throughput 96-well, automated fashion. The kinase domains will be expressed using 2BT10 plasmids with an N-terminal

His10-TEV cleavable tag and coexpressed with YopH164 phosphatase. These kinases will be expressed in E. *coli* and purified using Ni-NTA beads. After tag-cleavage with TEV protease, purity can be assessed using microfluidic gel electrophoresis using a LabChip GX II (Caliper LifeSciences). The small molecules can be obtained by request from the RNAi core at MSKCC and will not require extensive synthesis.

Test whether linear additive QSAR can rationalize selectivity

Cheminformatics[108-111] tools provide a powerful method to analyze the physical determinants driving each selectivity profile. Each of the molecules in the proposed library will be encoded with affinity data from the alchemical free energy calculations. We will then generate radial molecular fingerprints[112], which are representations of the connectivity of a compound in 32-bit form. These fingerprints are topological descriptors and depend only on the 2D structure of the chemical. Using these fingerprints, we will generate predictive linear additive 2D quantitative structure-activity relationship (QSAR) models using the direct kernel-based Partial Least Squares (KPLS) [113,114] supervised learning algorithm implemented in Canvas (Schrodinger) [115,116]. These models will be generated using a training set randomly selected from the proposed libraries, while the rest of the library will form the test set. The model will be assessed on its ability to recapitulate the affinities of the test set molecules. A bootstrapping scheme will be employed to get an estimate of the variability of the model based on how the test set is drawn, by randomly sampling members from the training set, building a model and predicting the activity of a molecule in the test set. Test set molecules with features that correspond to highly variable parameters and change depending on how test set is drawn will be predicted with high uncertainty. Assessing QSAR models in this manner is critical to ensuring that the features important in the model are not artifacts of the training set selection process. Another important feature of this methodology is that the contributions of individual atoms to the model can be mapped backed on to the molecules. This allows for easy interpretation of the model and enables identification of scaffolds or fragments that are important for driving selectivity.

Possible outcomes and alternative approaches

Based on preliminary work with the algorithm, we expect that we will be able to search through a library of 3200 compounds and proposed libraries that satisfy each of the design constraints. KPLS is effective at creating models using a small number of molecules[113], and if the contributions to selectivity are linear, it should be possible to generate statistically meaningful models using this size library. These QSAR models can help identify fragments or scaffolds that are important for driving selectivity. Additionally, we will have a number of proposed good binders based on each selectivity profile and can prioritize interesting hits for future optimization.

A kinase-targeted library covers only a small subset of chemical space and there may be significant overlap between the molecules proposed for each design constraint. If this happens, it will be difficult to create predictive models that can discriminate between features important to the different selectivity profiles. Further, it is possible that the design constraints severely restrict chemical space and only a small number of molecules, if any, are capable of satisfying them. To address this, we can use a larger library of molecules, such as eMolecules, which will allow us to sample a larger subset of chemical space while still allowing us to purchase molecules for experimental work. In addition to increasing the length of time it would take to adequately sample the chemical space of the library, large sets of compounds have the potential to have multiple clusters of chemicals that are difficult to sample between or contain molecules that are not drug-like and have unfavorable pharmacodynamics and pharmacokinetic properties. As an alternative, we could generate a virtual fragment-based library[117], which would provide a large sample of chemical space that is expected to be drug-like. However, molecules proposed from this library would be difficult to test experimentally, as they would likely require synthesis.

While Src and Abl are easily expressed in bacteria, EGFR has low bacterial expression[55] and HER2 has only been expressed in insect cells. To recombinantly express these proteins, we will try a number of solubility promoting tags[118] and search for new construct domains that might promote bacterial expression[119]. Because

the fluorescence assays are not dependent on kinase activity, we can use site-directed mutagenesis to introduce point mutations to inactivate the kinases, which has been shown to increase bacterial yields[120]. Additionally, both EGFR and HER2 kinase domains are commercially available from a number of sources, such as Carna Biosciences, and can be purchased to facilitate these studies.

There are a number of alternative methods for developing QSAR models. While radial fingerprints for 2D models are in common use in the field, studies have shown that dendritic fingerprints are a viable alternative for use with KPLS[121,122] and can easily be tried alongside radial fingerprints. Should 2D QSAR models prove unable to rationalize the determinants of selectivity, failure can be attributed to not taking into account the 3D shape of the ligands, or the presence of large numbers of stereoisomers in the dataset. In topological fingerprints, stereoisomers appear identical, but may have drastically different activities. These so called activity cliffs[123,124], where molecules look similar but have very different activities, pose problems in developing quality QSAR models. Stereoisomers can be addressed by either hand-curating these datasets to remove them or using 3D pharmacophores[125] instead of 2D fingerprints. 2D fingerprints are typically considered superior because they do not require generating 3D structures or ligand alignment, which introduces a significant bottleneck to the QSAR workflow for large numbers of molecules. Activity cliffs resulting from non-stereoisomer pairs are more difficult to address, but techniques such as MODI[124] can be employed to *a priori* calculate whether there are too many activity cliffs in a data set to generate a QSAR model. A drawback of this method is the need to classify activity data, which is quantitative, into a categorical format. There are also a number of more advanced methods that can be used to develop QSAR models, such as 4D QSAR[126,127]. However, these methods generally use noninterpretable descriptors making them difficult to use to rationalize the feature driving selectivity. Further, it is also possible that the contributions to selectivity are inherently not linear additive, in which case QSAR models will be insufficient to predict measured selectivity. In this case, we learned about the nonadditivity of these contributions and have an alchemical free energy tool that can address such nonadditive contributions.

Aim 3. Evaluate the potential for mutant selective small molecule inhibitors

Rationale and significance

EGFR-targeted inhibitors are typically given in non-small cell lung cancer patients to target the activated, mutated form of the kinase. While inhibitors like gefitinib and erlotinib have been shown to be selective for the mutant type over the wild type, they still bind to the wild type kinase with high affinity. Inhibiting the wild type pathway in non-cancer cells can cause on-target toxicity with deleterious side effects in patients, which limits the amount of inhibitor that can be safely given to a patient. In non-small cell lung cancer, EGFR inhibitors have been associated with skin rash[128-131], diarrhea[132], and ocular adverse events[133] that have been attributed to inhibition of wild type EGFR. These adverse events, while not lethal, often lead to interruption of dose-modification for these patients. Further, patients treated with EGFR inhibitors develop drug resistance in 9-14 months[63,77], most commonly with a secondary missense mutation, T790M[5,134]. While the mechanism of this mutation is not entirely clear, it has been shown to increase the ATP affinity of L858R-mutant EGFR but not in the context of wild type kinase. This mutant has also been proposed to introduce a steric clash that prevents inhibitor binding, similar to the Src T315I gatekeeper mutation, which has been suggested to both stabilize the active conformation as well as sterically interfere with inhibitor binding[135,136]. A second generation of inhibitors, while initially effective in treating EGFR SMKI-resistance patients, was unsuccessful due to dose-related, on-target toxicities[66]. As such, new generations of drugs that target both L858R and L858R/T790M mutant EGFR but not the wild type kinase are required. Using the multitarget design scheme developed in Aim 2, we will begin to ask whether chemical space allows for molecules that satisfy these criteria.

Experimental approaches

Free energy calculations to propose mutant selective small molecules

To determine whether mutant selectivity is achievable using multitarget design constraints, we will use the alchemical free energy calculations developed in Aim 2 to propose libraries of small molecules. We will run a series of alchemical free energy calculations using mutant EGFR (L858R or L858R/T790M) as a positive design constraint and wild type EGFR as a negative design constraint. The calculations will be run using the GPU computing cluster at MSKCC, using a forcefield such as AMBER99SB-ILDN. We will use a commercially available compound database, such as eMolecules. These alchemical free energy calculations will search chemical space to maximize the **ratio of ligand affinity for mutant EGFR over ligand affinity for wild type EGFR,** and output libraries of molecules proposed to be more selective for the mutant form over the wild type kinase.

Experimentally test free energy predictions

To ensure that the predictions are accurate, we will assess several of the proposed small molecules using a biophysical fluorescence based binding assay to measure affinities for wild type and mutant EGFR. First, we will determine whether these molecules are fluorescent or not by performing a full spectrum analysis using Src and Abl binding. The concentration of each compound will be varied between 8nM and 20uM, both with and without Src or Abl, using a known fluorescent molecule as a positive control to ensure that fluorescence is being read. The negative control for each new compound will be the series of spectra collected without kinase, showing that any observed changes are dependent on the ligand binding to the protein are not a function of increasing concentration. These experiments will determine whether we use a direct or competition based assay, as described in Aim 1, for each of the molecules we chose to validate. We will recombinantly express wild type EGFR, L858R mutant and L858R/T790M mutant kinase domains for using in the binding assays. Both EGFR mutants will be generated using QuickChange Site-direct mutagenesis (Agilent) and expressed in bacteria with an N-terminal His10-TEV cleavable tag with coexpressed YopH164. Purification will be done using Ni-NTA beads and purity will be assessed after tag-cleavage with TEV protease via microfluidic gel electrophoresis using a LabChip GX II (Caliper LifeSciences).

Analyze libraries for features driving selectivity

To rationalize the features of the small molecules contributing to the desired selectivity profile, we will use cheminformatics tools to build predictive QSAR models. Using radial molecular fingerprints, which are representations of the connectivity of a compound in 32-bit form, we will create 2D QSAR models using the direct kernel-based Partial Least Squares (KPLS) supervised learning algorithm, implemented in Canvas (Schrodinger). A training set will be randomly drawn from the proposed libraries, with the rest of the molecules forming the test set. The validity of the QSAR model will be measured using a bootstrapping method, where multiple models are created by randomly drawing from the training set and used to predict the activity of molecules in the test set. This will identify features in the model that are highly dependent on how the test set is drawn. Once a valid QSAR model is developed, the contributions of each atom will be mapped back onto each chemical in the library. Visualizing these contributions will allow us to identify scaffolds and modifications that favorably contribute to mutant selectivity.

Possible outcomes and alternative approaches

We expect to be able to predict a number of compounds that are mutant-selective, both for the L858R and L858R/T790M mutants, compared to the wild type. We also expect to be able to build a statistically valid QSAR model that will rationalize the physical determinants of mutant selectivity. However, it is possible that the design constraints narrow chemical space extensively and a larger library is needed. To address this, we can use a combinatorial or synthetically accessible library. These libraries, while covering a much larger subset of chemical space, can contain molecules that are difficult to obtain or synthesize. Such molecules would be difficult to experimentally validate. Additionally, many molecules would not be drug-like or feasible for use due to difficulty to synthesize.

As stated in Aim 2, wild type EGFR does have low bacterial expression[55] and the mutant forms have only been expressed in insect cells. To increase bacterial expression of these proteins, we will try a number of solubility

promoting tags[118] or search for new construct domains[119]. Since the fluorescence assays are independent of kinase activity, we can inactivate the kinase using site-directed mutagenesis to increase bacterial yields[120]. Additionally, both EGFR and the mutants of interest are commercially available from a number of sources, such as Carna Biosciences, and can be purchased to facilitate these studies if bacterial expression is intractable.

Should the proposed 2D QSAR models be unable to rationalize the differences in selectivity, we will move to 3D pharmacophores[125], which contain information about the 3D shape of a ligand. While 2D fingerprints are typically considered superior because their use does not require the generation of 3D structures or ligand alignment, 3D pharmacophores have had success in limiting the effect of activity cliffs. There are also a number of more advanced methods that can be used to develop QSAR models, such as 4D QSAR[126,127]. However, these methods often use noninterpretable descriptors, making rationalizing chemical features driving selectivity difficult.

V. Conclusion

Since the approval of imatinib in 2001, there has been a great deal of interest and research into developing small molecule inhibitors of kinases to treat cancer. Despite this, the pharmaceutical industry has seen a downturn in productivity, as many molecules fail in late stage clinical trials due to safety and efficacy issues. Tumors have multiple routes of resistance, including upregulation of a second kinase, mutations in the target kinase or amplification of the target kinase. On the other hand, toxicity arising from on-target inhibition of the wild type kinase can limit maximal tolerated dosage and reduce the effectiveness of the drug. Such issues might be addressed by designing drugs to maximize binding affinity for targets and minimizing binding for antitargets. In this study, we propose using GPU-accelerated alchemical free energy calculations to both assess the potential for achieving this kind of chemical selectivity with kinase inhibitors and to build a tool to realize it through automated chemical design. We first aim to investigate the accuracy of these calculations as quantitative predictors of binding affinity, and assess what chemical phenomena must be considered to achieve utility for rational drug design. We will then extend these alchemical free energy calculations to multitarget design, where we will assess the affect of multiple design constraints on the chemical space available to inhibitors, both for multiple kinases and oncogenic mutated kinases. To rationalize the physical determinants of these selectivity profiles, we will employ cheminformatics tools to develop 2D QSAR models. Ultimately, this study will provide a tool to rationally design drugs to improve selectivity and minimize on-target toxicity.

VI. References

1. American Cancer Society. Cancer Facts & Statistics. 1–56 (2015).
2. Cohen, P. & Tcherpakov, M. Will the Ubiquitin System Furnish as Many Drug Targets as Protein Kinases? *Cell* **143,** 686–693 (2010).
3. Wu, P., Nielsen, T. E. & Clausen, M. H. FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol. Sci.* **36,** 422–439 (2015).
4. Paul, S. M. *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* **9,** 203–214 (2010).
5. Pao, W. *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).
6. Knight, Z. A., Lin, H. & Shokat, K. M. Targeting the cancer kinome through polypharmacology. *Nat. Rev. Cancer* **10,** 130–137 (2010).
7. Chandarlapaty, S. *et al.* AKT Inhibition Relieves Feedback Suppression of Receptor Tyrosine Kinase Expression and Activity. *Cancer Cell* **19,** 58–71 (2011).
8. Kijima, T. *et al.* Safe and successful treatment with erlotinib after gefitinib-induced hepatotoxicity: difference in metabolism as a possible mechanism. *J. Clin. Oncol.* **29,** e588–90 (2011).
9. Rudmann, D. G. On-target and off-target-based toxicologic effects. *Toxicol Pathol* **41,** 310–314 (2013).
10. Liu, S. & Kurzrock, R. Toxicity of targeted therapy: Implications for response and impact of genetic polymorphisms. *Cancer Treatment Reviews* **40,** 883–891 (2014).
11. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **298,** 1912–1934 (2002).
12. Cowan-Jacob, S. W. *et al.* Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. *Acta Crystallogr Sect D Biol Crystallogr* **63,** 80–93 (2007).
13. Seeliger, M. A. *et al.* c-Src Binds to the Cancer Drug Imatinib with an Inactive Abl/c-Kit Conformation and a Distributed Thermodynamic Penalty. *Structure* **15,** 299–311 (2007).
14. Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109,** 275–282 (2002).
15. Harrison, S. C. Variation on an Src-like theme. *Cell* **112,** 737–740 (2003).
16. Nagar, B. c-Abl tyrosine kinase and inhibition by the cancer drug imatinib (Gleevec/STI-571). *J. Nutr.* **137,** 1518S–1523S– discussion 1548S (2007).
17. Davis, M. I. *et al.* Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **29,** 1046–1051 (2011).
18. Apsel, B. *et al.* Targeted polypharmacology: discovery of dual inhibitors of tyrosine and phosphoinositide kinases. *Nat. Chem. Biol.* **4,** 691–699 (2008).
19. Hopkins, A. L., Mason, J. S. & Overington, J. P. Can we rationally design promiscuous drugs? *Curr. Opin. Struct. Biol.* **16,** 127–136 (2006).
20. Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* **4,** 682–690 (2008).
21. Zhang, J., Yang, P. L. & Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **9,** 28–39 (2009).
22. Huggins, D. J., Sherman, W. & Tidor, B. Rational approaches to improving selectivity in drug design. *J. Med. Chem.* **55,** 1424–1444 (2012).
23. Mendoza, M. C., Er, E. E. & Blenis, J. The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem. Sci.* **36,** 320–328 (2011).
24. Tricker, E. M. *et al.* Combined EGFR/MEK Inhibition Prevents the Emergence of Resistance in EGFR-Mutant Lung Cancer. *Cancer Discov* **5,** 960–971 (2015).
25. Bailey, S. T. *et al.* mTOR Inhibition Induces Compensatory, Therapeutically Targetable MEK Activation in Renal Cell Carcinoma. *PLoS ONE* **9,** e104413 (2014).
26. Littlefield, P. *et al.* Structural analysis of the EGFR/HER3 heterodimer reveals the molecular basis for activating HER3 mutations. *Sci Signal* **7,** ra114–ra114 (2014).

27. Yun, C.-H. *et al.* Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* **11,** 217–227 (2007).
28. Gajiwala, K. S. *et al.* Insights into the aberrant activity of mutant EGFR kinase domain and drug recognition. *Structure* **21,** 209–219 (2013).
29. Song, Z., Wang, M. & Zhang, A. Alectinib: a novel second generation anaplastic lymphoma kinase (ALK) inhibitor for overcoming clinically-acquired resistance. *Acta Pharmaceutica Sinica B* **5,** 34–37 (2015).
30. Fan, Q.-W. *et al.* A dual phosphoinositide-3-kinase alpha/mTOR inhibitor cooperates with blockade of epidermal growth factor receptor in PTEN-mutant glioma. *Cancer Res.* **67,** 7960–7965 (2007).
31. Fan, Q.-W. *et al.* A dual PI3 kinase/mTOR inhibitor reveals emergent efficacy in glioma. *Cancer Cell* **9,** 341–349 (2006).
32. Norman, R. A., Toader, D. & Ferguson, A. D. Structural approaches to obtain kinase selectivity. *Trends Pharmacol. Sci.* **33,** 273–278 (2012).
33. Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **432,** 862–865 (2004).
34. Gregory L Warren *et al.* A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **49,** 5912–5931 (2005).
35. Chodera, J. D. *et al.* Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **21,** 150–160 (2011).
36. GREEN, P. J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82,** 711–732 (1995).
37. Nilmeier, J. P., Crooks, G. E. & Minh, D. Nonequilibrium candidate Monte Carlo is an efficient tool for equilibrium simulation. in (2011).
38. Deininger, M., Buchdunger, E. & Druker, B. J. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood* **105,** 2640–2653 (2005).
39. Cohen, P. & Alessi, D. R. Kinase Drug Discovery – What's Next in the Field? *ACS Chem. Biol.* **8,** 96–104 (2012).
40. Agafonov, R. V., Wilson, C., Otten, R., Buosi, V. & Kern, D. Energetic dissection of Gleevec's selectivity toward human tyrosine kinases. *Nature Publishing Group* **21,** 848–853 (2014).
41. Xu, W., Harrison, S. C. & Eck, M. J. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* **385,** 595–602 (1997).
42. Williams, J. C. *et al.* The 2.35 A crystal structure of the inactivated form of chicken Src: a dynamic molecule with multiple regulatory interactions. *Journal of Molecular Biology* **274,** 757–775 (1997).
43. Levinson, N. M. *et al.* A Src-like inactive conformation in the abl tyrosine kinase domain. *PLoS Biol.* **4,** e144 (2006).
44. Vogtherr, M. *et al.* NMR characterization of kinase p38 dynamics in free and ligand-bound forms. *Angew. Chem. Int. Ed. Engl.* **45,** 993–997 (2006).
45. Vajpai, N. *et al.* Solution conformations and dynamics of ABL kinase-inhibitor complexes determined by NMR substantiate the different binding modes of imatinib/nilotinib and dasatinib. *J. Biol. Chem.* **283,** 18292–18302 (2008).
46. Lovera, S. *et al.* The different flexibility of c-Src and c-Abl kinases regulates the accessibility of a druggable inactive conformation. *J. Am. Chem. Soc.* **134,** 2496–2499 (2012).
47. Lin, Y.-L., Meng, Y., Jiang, W. & Roux, B. Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 1664–1669 (2013).
48. Aleksandrov, A. & Simonson, T. Molecular dynamics simulations show that conformational selection governs the binding preferences of imatinib for several tyrosine kinases. *J. Biol. Chem.* **285,** 13807–13815 (2010).
49. Dar, A. C., Lopez, M. S. & Shokat, K. M. Small molecule recognition of c-Src via the Imatinib-binding conformation. *Chem. Biol.* **15,** 1015–1022 (2008).
50. Cowan-Jacob, S. W. *et al.* The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation. *Structure* **13,** 861–871 (2005).
51. Wu, P., Nielsen, T. E. & Clausen, M. H. Small-molecule kinase inhibitors: an analysis of FDA-approved

drugs. *Drug Discov. Today* (2015). doi:10.1016/j.drudis.2015.07.008
52. Weisberg, E. *et al.* Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. *Cancer Cell* **7,** 129–141 (2005).
53. Hochhaus, A. & Kantarjian, H. The development of dasatinib as a treatment for chronic myeloid leukemia (CML): from initial studies to application in newly diagnosed patients. *J. Cancer Res. Clin. Oncol.* **139,** 1971–1984 (2013).
54. Sanford, D. S. *et al.* The role of ponatinib in Philadelphia chromosome-positive acute lymphoblastic leukemia. *Expert Review of Anticancer Therapy* **15,** 365–373 (2015).
55. Parton, D. L., Hanson, S. M. & Rodríguez-Laureano, L. An open library of human kinase domain constructs for automated bacterial expression. *bioRxiv* (2016). doi:10.1101/038711
56. Yarden, Y. The EGFR family and its ligands in human cancer. *European Journal of Cancer* **37,** 3–8 (2001).
57. Wood, E. R. *et al.* A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res.* **64,** 6652–6659 (2004).
58. Hynes, N. E. & Lane, H. A. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat. Rev. Cancer* **5,** 341–354 (2005).
59. Zhang, X., Gureasko, J., Shen, K., Cole, P. A. & Kuriyan, J. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* **125,** 1137–1149 (2006).
60. Stamos, J., Sliwkowski, M. X. & Eigenbrot, C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J. Biol. Chem.* **277,** 46265–46272 (2002).
61. Zhang, X. *et al.* Inhibition of the EGF receptor by binding of MIG6 to an activating kinase domain interface. *Nature* **450,** 741–744 (2007).
62. Jura, N. *et al.* Catalytic Control in the EGF Receptor and Its Connection to General Kinase Regulatory Mechanisms. *Mol. Cell* **42,** 9–22 (2011).
63. Rosell, R. *et al.* Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* **13,** 239–246 (2012).
64. Wao, H., Mhaskar, R., Kumar, A., Miladinovic, B. & Djulbegovic, B. Survival of patients with non-small cell lung cancer without treatment: a systematic review and meta-analysis. *Syst Rev* **2,** 10 (2013).
65. Sharma, S. V., Bell, D. W., Settleman, J. & Haber, D. A. Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer* **7,** 169–181 (2007).
66. Juchum, M., Günther, M. & Laufer, S. A. Fighting cancer drug resistance: Opportunities and challenges for mutation-specific EGFR inhibitors. *Drug Resist. Updat.* **20,** 12–28 (2015).
67. Visbal, A. L. *et al.* Gender differences in non-small-cell lung cancer survival: an analysis of 4,618 patients diagnosed between 1997 and 2002. *Ann. Thorac. Surg.* **78,** 209–15– discussion 215 (2004).
68. Lee, C.-C., Shiao, H.-Y., Wang, W.-C. & Hsieh, H.-P. Small-molecule EGFR tyrosine kinase inhibitors for the treatment of cancer. *Expert Opin Investig Drugs* **23,** 1333–1348 (2014).
69. Tsao, M.-S. *et al.* Erlotinib in lung cancer - molecular and clinical predictors of outcome. **353,** 133–144 (2005).
70. Pao, W. *et al.* EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib. *PNAS* **101,** 13306–13311 (2004).
71. Rusnak, D. W. *et al.* The characterization of novel, dual ErbB-2/EGFR, tyrosine kinase inhibitors: potential therapy for cancer. *Cancer Res.* **61,** 7196–7203 (2001).
72. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350,** 2129–2139 (2004).
73. Paez, J. G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304,** 1497–1500 (2004).
74. Shan, Y. *et al.* Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. *Cell* **149,** 860–870 (2012).

75. Kim, Y. *et al.* Temporal resolution of autophosphorylation for normal and oncogenic forms of EGFR and differential effects of gefitinib. *Biochemistry* **51,** 5212–5222 (2012).
76. Lu, C., Mi, L.-Z., Schürpf, T., Walz, T. & Springer, T. A. Mechanisms for kinase-mediated dimerization of the epidermal growth factor receptor. *J. Biol. Chem.* **287,** 38244–38253 (2012).
77. Mok, T. S. *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. **361,** 947–957 (2009).
78. Engelman, J. A. *et al.* MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* **316,** 1039–1043 (2007).
79. Ware, K. E. *et al.* A mechanism of resistance to gefitinib mediated by cellular reprogramming and the acquisition of an FGF2-FGFR1 autocrine growth loop. *Oncogenesis* **2,** e39 (2013).
80. Ercan, D. *et al.* Reactivation of ERK signaling causes resistance to EGFR kinase inhibitors. *Cancer Discov* **2,** 934–947 (2012).
81. Yun, C.-H. *et al.* The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 2070–2075 (2008).
82. Song, Z. *et al.* Challenges and Perspectives on the Development of Small-Molecule EGFR Inhibitors against T790M-Mediated Resistance in Non-Small-Cell Lung Cancer. *J. Med. Chem.* acs.jmedchem.5b00840 (2016). doi:10.1021/acs.jmedchem.5b00840
83. Levinson, N. M. & Boxer, S. G. Structural and spectroscopic analysis of the kinase inhibitor bosutinib and an isomer of bosutinib binding to the Abl tyrosine kinase domain. *PLoS ONE* **7,** e29828 (2012).
84. Shirts, M. R., Mobley, D. L. & Brown, S. P. Free energy calculations in structure-based drug design. *Drug Design: Structure-and …* (2010).
85. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78,** 1950–1958 (2010).
86. Chodera, J. D. *et al.* Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **21,** 150–160 (2011).
87. Wang, K., Chodera, J. D., Yang, Y. & Shirts, M. R. Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *J. Comput. Aided Mol. Des.* **27,** 989–1007 (2013).
88. Annual Reports in Computational Chemistry. (2007).
89. Boyce, S. E. *et al.* Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *Journal of Molecular Biology* **394,** 747–763 (2009).
90. Michel, J. & Essex, J. W. Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J. Comput. Aided Mol. Des.* **24,** 639–658 (2010).
91. Anastassiadis, T., Deacon, S. W., Devarajan, K., Ma, H. & Peterson, J. R. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat. Biotechnol.* **29,** 1039–1045 (2011).
92. Knight, Z. A. & Shokat, K. M. Features of selective kinase inhibitors. *Chem. Biol.* **12,** 621–637 (2005).
93. Kramer, C., Kalliokoski, T., Gedeck, P. & Vulpetti, A. The Experimental Uncertainty of Heterogeneous Public Ki Data. *J. Med. Chem.* **55,** 5165–5173 (2012).
94. Sutherland, J. J., Gao, C., Cahya, S. & Vieth, M. What general conclusions can we draw from kinase profiling data sets? *Biochim. Biophys. Acta* **1834,** 1425–1433 (2013).
95. Seeliger, M. A. *et al.* High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. *Protein Sci.* **14,** 3135–3139 (2005).
96. Fabian, M. A. *et al.* A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* **23,** 329–336 (2005).
97. Karaman, M. W. *et al.* A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **26,** 127–132 (2008).
98. Hanson, S. M. & Chodera, J. D. AssayTools.
99. Shan, Y. *et al.* A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 139–144 (2009).
100. Czodrowski, P., Sotriffer, C. A. & Klebe, G. Protonation Changes upon Ligand Binding to Trypsin and

Thrombin: Structural Interpretation Based on pKa Calculations and ITC Experiments. *Journal of Molecular Biology* **367,** 1347–1356 (2007).
101. Martin, Y. C. Let's not forget tautomers. *J. Comput. Aided Mol. Des.* **23,** 693–704 (2009).
102. Chen, Y. & Roux, B. Constant-pH Hybrid Nonequilibrium Molecular Dynamics–Monte Carlo Simulation Method. *J. Chem. Theory Comput.* **11,** 3919–3931 (2015).
103. Lochhead, P. A. Protein Kinase Activation Loop Autophosphorylation in Cis: Overcoming a Catch-22 Situation. *Sci Signal* **2,** pe4–pe4 (2009).
104. Holdgate, G. A. & Ward, W. H. J. Measurements of binding thermodynamics in drug discovery. *Drug Discov. Today* **10,** 1543–1550 (2005).
105. O'Neill, M. A. A. & Gaisford, S. Application and use of isothermal calorimetry in pharmaceutical development. *Int J Pharm* **417,** 83–93 (2011).
106. Bose, R. *et al.* Activating HER2 mutations in HER2 gene amplification negative breast cancer. *Cancer Discov* **3,** 224–237 (2013).
107. Wang, Q., Zorn, J. A. & Kuriyan, J. A structural atlas of kinases inhibited by clinically approved drugs. *Meth. Enzymol.* **548,** 23–67 (2014).
108. Kubinyi, H. QSAR and 3D QSAR in drug design Part 1: methodology. *Drug Discov. Today* **2,** 457–467 (1997).
109. Kubinyi, H. QSAR and 3D QSAR in drug design Part 2: applications and problems. *Drug Discov. Today* **2,** 538–546 (1997).
110. Nantasenamat, C. & Isarankura-Na-Ayudhya, C. A practical overview of quantitative structure-activity relationship. *EXCLI J* (2009).
111. Duffy, B. C., Zhu, L., Decornez, H. & Kitchen, D. B. Early phase drug discovery: cheminformatics and computational techniques in identifying lead series. *Bioorg. Med. Chem.* **20,** 5324–5342 (2012).
112. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J Chem Inf Model* **50,** 742–754 (2010).
113. Bennett, K. P. & Embrechts, M. J. An optimization perspective on kernel partial least squares regression. *Nato Science Series sub …* (2003).
114. Rosipal, R. & Trejo, L. J. Kernel partial least squares regression in reproducing kernel hilbert space. *The Journal of Machine Learning Research* **2,** 97–123 (2002).
115. Schrodinger. Canvas. (2016).
116. An, Y., Sherman, W. & Dixon, S. L. Kernel-Based Partial Least Squares: Application to Fingerprint-Based QSAR with Model Visualization. *J Chem Inf Model* **53,** 2312–2321 (2013).
117. Congreve, M., Chessari, G., Tisi, D. & Woodhead, A. J. Recent Developments in Fragment-Based Drug Discovery. *J. Med. Chem.* **51,** 3661–3680 (2008).
118. Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L. & McCafferty, J. Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. *BMC Biotechnol.* **4,** 32 (2004).
119. Elloumi-Mseddi, J., Jellali, K. & Aifa, S. In vitro activation and inhibition of recombinant EGFR tyrosine kinase expressed in Escherichia coli. *ScientificWorldJournal* **2013,** 807284 (2013).
120. Williams, D. M., Wang, D. & Cole, P. A. Chemical rescue of a mutant protein-tyrosine kinase. *J. Biol. Chem.* **275,** 38127–38130 (2000).
121. Duan, J., Dixon, S. L., Lowrie, J. F. & Sherman, W. Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. *J. Mol. Graph. Model.* **29,** 157–170 (2010).
122. Sastry, M., Lowrie, J. F., Dixon, S. L. & Sherman, W. Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments. *J Chem Inf Model* **50,** 771–784 (2010).
123. Maggiora, G., Vogt, M., Stumpfe, D. & Bajorath, J. Molecular Similarity in Medicinal Chemistry. *J. Med. Chem.* **57,** 3186–3204 (2014).
124. Golbraikh, A., Muratov, E., Fourches, D. & Tropsha, A. Data set modelability by QSAR. *J Chem Inf Model* **54,** 1–4 (2014).
125. Cherkasov, A. *et al.* QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.*

57, 4977–5010 (2014).
126. Potemkin, V. & Grishina, M. Principles for 3D/4D QSAR classification of drugs. *Drug Discov. Today* **13,** 952–959 (2008).
127. Myint, K. Z. & Xie, X.-Q. Recent advances in fragment-based QSAR and multi-dimensional QSAR methods. *Int J Mol Sci* **11,** 3846–3866 (2010).
128. Liu, H.-B. *et al.* Skin rash could predict the response to EGFR tyrosine kinase inhibitor and the prognosis for patients with non-small cell lung cancer: a systematic review and meta-analysis. *PLoS ONE* **8,** e55128 (2013).
129. Petrelli, F., Borgonovo, K., Cabiddu, M., Lonati, V. & Barni, S. Relationship between skin rash and outcome in non-small-cell lung cancer patients treated with anti-EGFR tyrosine kinase inhibitors: a literature-based meta-analysis of 24 trials. *Lung Cancer* **78,** 8–15 (2012).
130. Bollag, G. *et al.* Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **467,** 596–599 (2010).
131. Rudin, C. M. *et al.* Pharmacogenomic and pharmacokinetic determinants of erlotinib toxicity. *J. Clin. Oncol.* **26,** 1119–1127 (2008).
132. Liu, G. *et al.* Epidermal growth factor receptor polymorphisms and clinical outcomes in non-small-cell lung cancer patients treated with gefitinib. *Pharmacogenomics J.* **8,** 129–138 (2008).
133. Huillard, O. *et al.* Ocular adverse events of molecularly targeted agents approved in solid tumours: a systematic review. *Eur. J. Cancer* **50,** 638–648 (2014).
134. Kobayashi, S. *et al.* EGFRMutation and Resistance of Non–Small-Cell Lung Cancer to Gefitinib. *N Engl J Med* **352,** 786–792 (2005).
135. Azam, M., Seeliger, M. A., Gray, N. S., Kuriyan, J. & Daley, G. Q. Activation of tyrosine kinases by mutation of the gatekeeper threonine. *Nature Structural & Molecular Biology* **15,** 1109–1118 (2008).
136. Daub, H., Specht, K. & Ullrich, A. Strategies to overcome resistance to targeted protein kinase inhibitors. *Nat Rev Drug Discov* **3,** 1001–1010 (2004).