CHAPTER 13

Computational Multitarget Drug Discovery

JEREMY A. HORST, ADRIAN LAURENZI, BRADY BERNARD, and RAM SAMUDRALA

13.1 INTRODUCTION

Pharmaceutical substances have been discovered by means ranging from serendipitous observation [1,2] to specific engineering [3]. The purpose is nearly always to combat one particular disease, and the approach is most often trial and error. The efficiency of these pharmaceutical hunts has been improved greatly by high-throughput pharmaceutical platforms, but the requirement of physical experiment makes these screens scale in expense linearly at best. The expense of discovering a new chemical entity is estimated at US$0.5B–US$2B [4,5].

More recent successes in computational modeling of compound to protein docking open the possibility of nonphysical prelaboratory screens. In our experience this has vastly increased the success rate of bench experiments [6,7] (see Table 13.1, later in this chapter). Computational modeling of protein ligand interactions has been applied to find pharmacologic targets in known drug-disease pairs [8,9]. The more obvious use of these docking methods is to guide discovery of a drug for a disease, as modeling enables design [3]. Design does not need to be limited to one protein target. Searching for one compound for multiple targets in the same pathogen increases odds for successful inhibition of at least one target, and facilitates discovery of multitarget lead inhibitors, [Note 1], which vastly decreases the probability of developing resistance (or habituation) and decreases toxicity via lowered effective dose [6,10–12] (Table 13.1).

Thus far the search for multitarget inhibitors has focused on one organism at a time [6,9,11], but modeling multidisease effects has explained clinical patterns of elimination for two diseases by one drug [13]. The advent of computational multidisease screens will enable access to the most accurate aspects of computational screening, bearing the possibility of vastly reducing barriers to drug development.

Polypharmacology in Drug Discovery, First Edition. Edited by Jens-Uwe Peters.

TABLE 13.1 Summary of More Recent Prospective Drug Discovery Screens

					Bench Experiments			
Disease	Protocol	*In Virtuale*	Targets	*In Vitro* →	Hit	Best (μM)	Type	Reference
Diabetes 2	Throughput:		1	400,000 →	85 <100 μM	4.2	IC_{50}	[47]
Malaria	Throughput:		Organism	2,687 →	19 <1 μM	0.003	IC_{50}	[48]
Malaria	Throughput:		Organism	2,160 →	36 <1 μM	0.010	IC_{50}	[49]
β-Lactamase	Throughput:		1	70,563 →	0 <30 μM	none[a]	IC_{50}	[50]
Malaria	Throughput:		Organism	1,700,000 →	5973 <1.25 μM	NR[b]	IC_{50}	[51]
					648 <100 nM	NR	IC_{50}	
Malaria	Throughput:		Organism	1,986,056 →	13533 <2 μM	NR	IC_{80}	[52]
Diabetes 2	Docking:	150,000	1	7 →	5 <100 μM	21	K_i	[53]
Cancer	Docking:	100,000	1	2 →	2 <100 pM	0.00003	K_d	[54]
Diabetes 2	Docking:	235,000	1	365 →	127 <100 μM	1.7	IC_{50}	[47]
Malaria	QSAR:	12	1[c]	12 →	3 <100 μM	5.7	IC_{50}	[55]
Malaria	QSAR:	9	4	9 →	2 < 1 μM	0.3	IC_{50}	[56]
Malaria	Docking:	241,000	1	100 →	5 < 10 μM	1.0	IC_{50}	[57]
Human GPCRs	Docking:	150,000	5	309 →	50 < 5 μM	0.021	EC_{50}	[32]
Malaria	Docking:	355,000	2	84 →	7 < 10 μM	9.5	IC_{50}	[58]
Depression	Dock,QSAR:	40,000	1	78 →	9 <1 μM(K_i)	0.3	IC_{50}	[41]
β-Lactamase	Docking:	70,563	1	18 →	4 < 200 μM	70[a]	IC_{50}	[59]
Malaria	Docking:	2,344	14	16 →	7 < 1 μM	0.13	IC_{50}	[60]
Tropical	QSAR:	3,665	11,714	2 →	1 NMR	NR	NMR	[61]
Human many	QSAR:[d]	3,665	1,133	30 →	20 < 1 μM	0.001	K_i	[50]
Dengue	Peptides:	14	1	14 →	2 <10 μM	7.0	IC_{50}	[62]

[a] Extensive reversibility studies.
[b] NR—not reported.
[c] The human homolog was also considered as an antitarget.
[d] Search for target; *in vivo* activity already known.

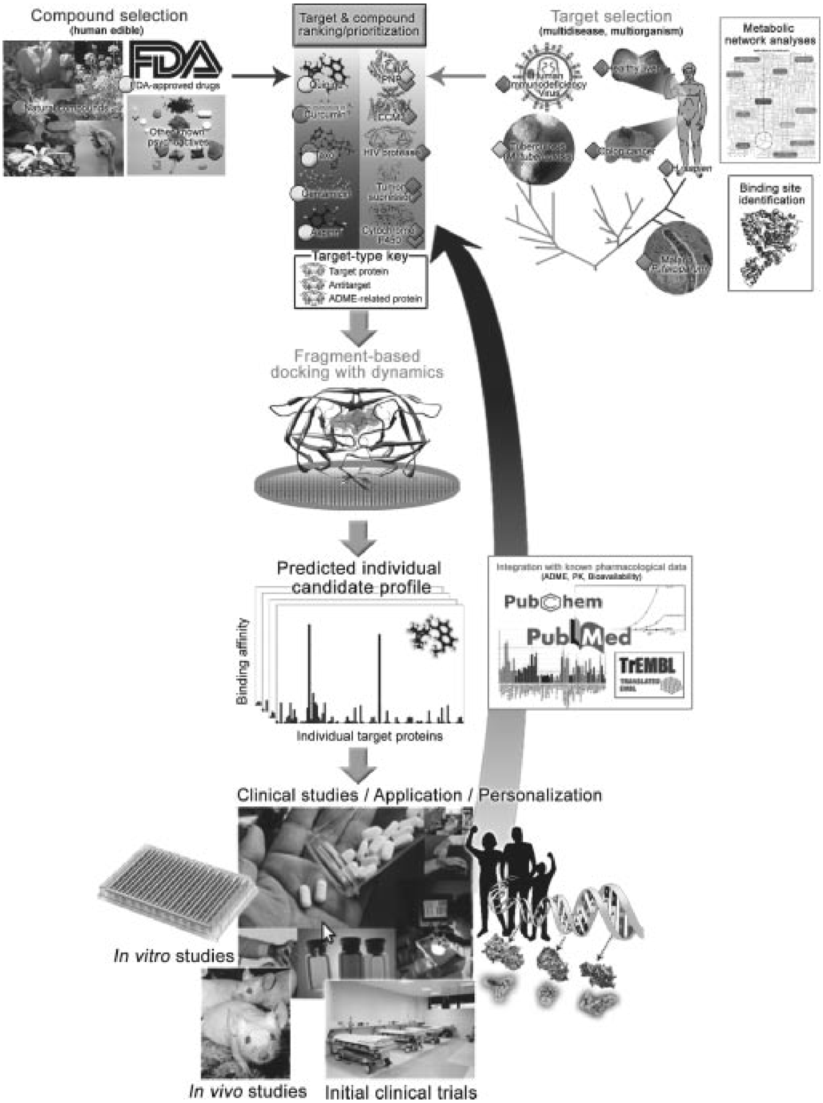

FIGURE 13.1 Computational multidisease multitarget screening pipeline. (See insert for color representation.)

In this chapter we elaborate the conceptual framework underlying rational drug discovery, describe contemporary computational approaches, discuss emerging concepts, and introduce a pipeline (see Figure 13.1) to integrate the array of promising techniques and ideas that are already transforming drug discovery.

The schematic view of our computational multidisease multitarget screening pipeline relates emerging concepts and techniques described in this chapter, which are already transforming drug discovery. The contemporary weaknesses of computational modeling can be overcome to find pharmacologically active substances by careful selection of the protein and compound sets to be used in computational screening (shown on the sides at top). To maximize the chance of bioactivity and safety in humans, compounds to be considered for screening (top left) should be selected from existing drugs (Section 13.6.3) or natural compounds (Section 13.6.1). The selection of protein targets (upper right) that can be exploited to stop a disease is a nontrivial problem requiring extensive analysis (Sections 13.2.1 and 13.3.1). The probability of finding pharmacologically active compounds is heightened by targeting multiple proteins relating to a disease (Section 13.6.10), which can be in the same signaling network (network targeting, Section 13.5.4), or in different disease-associated pathogens (Section 13.6.11); screening against antitarget host proteins can also be performed to control off-target effects (Section 13.6.4). The protein structure (Section 13.5.4) and binding sites (Section 13.5.1) can be predicted using knowledge-based methods (Section 13.4.1). Next, target proteins are prioritized according to the susceptibility of the binding site (Sections 13.3.2 and 13.3.3), the accessibility of the subcellular location, and the similarity of physiologic substrates to the compound set (Section 13.4.6). The set of potential pharmacologic compounds are then prioritized (top center) on the basis of features of the target protein and disease site (Sections 13.3.1 and 13.3.2) and similarity to target substrates (Sections 13.2.4 and 13.4.6). Finally, the compounds are computationally docked to the active sites of the target proteins (upper middle; Sections 13.4.2 and 13.4.5) with small bursts of molecular dynamics (Section 13.5.2), scored (Section 13.4.3), and ranked with respect to each other (Section 13.4.4). Initially a large compound set is evaluated, with subsequent cycling between directed fragment-based optimization, and cycling back to evaluate many similar compounds, which mimics the bench process for discovery of a new chemical entity (Section 13.2.5). The profiles of predicted binding affinities for each compound are compared to titrate selectivity and minimize untoward side effects (lower middle; Section 13.6.10). The use of compounds of known human safety profiles comes to fruition when approaching validation (Section 13.6.1): for diseases with no sufficient model system and no existing cure, existing pharmacologic agents may progress directly to initial clinical trials (center bottom; Section 13.6.3). In addition, the multitarget approach of using compounds that are predicted to be active against multiple pathogen proteins increases the odds of success; if a compound is predicted to inhibit six proteins, there is a good chance that it will actually inhibit at least one (Section 13.6.10). As an extension, computational screens of targets for multiple diseases increases the odds of finding a target for the inhibitor; allowing the discovery process to drive disease selection enables access to the most accurate computational predictions (Section 13.6.11). There are initial indications that computational simulations can be more accurate than high-throughput screening, possibly because they model bioactivity in an explicitly physiological manner whereas the implicit physical interaction model of bench screens is susceptible to nonspecific aggregation, covalent bonding, and promiscuous binding (Section 13.6.6). Meanwhile, sophisticated bench analysis techniques offer the pinnacle of accuracy, particularly the dissection of enthalpic and entropic contributions to the free energy of binding by isothermal titration calorimetry (Section 13.6.5). Protein, whole pathogen, whole animal, and clinical analysis (center bottom) feeds back to improve the accuracy of simulations (large arrow) by integration with existing pharmacologic data (middle right). Modeling the impact of genetic variance on protein structure allows design of generalized inhibitors for rapidly mutating pathogens and cancers, and specification to individual human differences to control side effects (bottom right; Section 13.6.8). Our group and others have demonstrated the early maturity of computational modeling of protein–ligand interactions by predicting compounds for desired pharmacologic activity and testing them in prospective experiments. A philosophy of freely available open-source software has been embraced by many publicly funded groups (Section 13.6.9). These methods not only save time and resources but also are beginning to be more accurate than *in vitro* screening methods (Section 13.6.6). The combination of computational multi-target drug discovery and stringent bench experimentation will lead a new era of effective selective drugs.

13.2 THE PHARMACOLOGICAL HUNT OF YESTERYEAR

13.2.1 Ethnopharmacy

Since antiquity, before written history, humans have sought available substances (mineral, animal, plant) to cure specific ailments. The hundreds of medicinal substances catalogued in the materia medica of various cultures before and during the time of Socrates [14–16] demonstrates that the hunt for pharmacological activity may predate the technology of the scientific method itself. Whether disproving hypotheses or embarking on fishing expeditions, experiences with curative and toxic substances may have conceptually secured the intuitive approach of trial-and-error investigation.

For thousands of years humans have applied trial-and-error experiments, separating out extracts of active agents to increase potency and remove unwanted properties. The earliest records describing pharmacologic safety include descriptions of animal models and progressive increases in dosage to test safety and efficacy [17]. Nonetheless, technological improvements were limited to purifications and altering the design of the trial itself.

For 200 years we have isolated specific pharmacologically active molecules [18]. For a century we have knowingly modified the chemical structure of natural compounds to tune desirable and undesirable effects. These attempts of drug discovery and design have led to one specific molecule at a time to combat microbial infection [19,20] and noninfectious diseases [21].

13.2.2 Protein Targets

With the advent of molecular biology we found the key to rational drug discovery: inhibiting specific protein targets essential to the progression of the disease causing agent. Targets are carefully identified by the consensus of extensive experimentation verified by multiple independent research groups. Thus the major goal of pharmacologic development has emerged as discovering or designing compounds that demonstrate favorable therapeutic activity towards a specific protein target.

Under the current paradigm, an attractive target is a protein essential to the infection, onset, or replication of the disease-causing agent, or a protein able to control one of these processes. The protein target should be sufficiently different from homeostatic host proteins such that a drug that inhibits its action would not kill the host. A target should be essential to the metabolism, growth, or reproduction of a pathogen or the progression of a neoplasm, and maximally different from all other antitarget human proteins.

13.2.3 Hitting the Target

Tests of pharmacological efficacy have been refined from observing the signs and symptoms of a disease, to growth of the disease-causing agent (e.g., proliferation of pathogens or cancer cells), to functional assays of specific target proteins. Meanwhile the search for target protein inhibitors has always been governed by the same two approaches, described below.

13.2.3.1 Random Screens Whether one at a time or run in parallel by brute force, many available substances are tested for efficacy. Often a wide net is cast by screening an enormous and diverse compound library (as many as 1.7 million compounds [22]). There is a tendency to test only representatives from a given group of substances; an intelligent step to increase the efficiency of the pharmacologic hunt wherein the "hit" group is explored in further screening. However, reduced screens increase the odds of missing subtle differences that might allow target binding by nonsampled members of the group. Thus, where resources permit, large screens are conducted. From the 1960s to the 1980s *high-throughput screens*, enabled by extraneous technology such as assembly lines and robotics, permitted the pharmaceutical industry to blossom almost strictly according to the paradigm of vast screens [21]. This is still the most common approach used by the pharmaceutical industry today. Without deep understanding of the target chemistry, sampling nature's pharmacopeia may well be the most efficient approach to finding a starting place: a hit compound [6,23,24] (Table 13.1).

13.2.3.2 Directed Exploration Intuitively, the response to finding an agent that has any noticeable desired effect is to seek better effects by similar agents. Intelligent searches for pharmacologically active substances generally follow explorative sampling around successful compounds already discovered in random screens [25]. Similar existing compounds can be tested for more desirable activity, or chemical modifications can be made by substituting, converting, and adding moieties [26].

Those of us who develop computational techniques for drug discovery tend to consider targets from infectious and noninfectious diseases as the same, but in reality they are not. Generally, the goal is to inhibit targets of infectious diseases (to increase the therapeutic index), but a human disorder that is not directly caused by a pathogen may be caused by the malfunction of a protein, so inhibition is not always the goal. For malfunctioning proteins, the goal may be to discover a drug that promotes the active conformation or overcomes the loss of effective signal activity. While computational drug discovery techniques are quite robust, molecular etiology must be considered to select the target and to specify the desired pharmacologic effect.

13.2.4 Similar Active Substances for Rational Selection

Sophistication in understanding the similarity of pharmacological agents was first developed in the ancient processes of chemical extraction. Similar separation in organic solvents indicates similar polarity and hydrophilicity, and often foretells identical chemical moieties. Comparison of compounds with similar chemical properties to compounds with similar pharmacologic effects resulted in the concepts of pharmacophores [27] and quantitative structure–activity relationships (QSARs [28]). These concepts enable intelligent exploration of the chemical and structural space around the natural substrate.

In a case for which the activity profile of a vast drug bank is known for a particular pathogen, analysis of similarly active compounds can facilitate understanding of the basis of molecular recognition between a small molecule and its protein target [22].

13.2.5 Cycling between Random and Directed Searches

Directed exploration requires either identification of the physiologic substrate, a hit compound, or deep knowledge of the target (discussed later in this chapter). Sampling around successful compounds with similar active substances represents an additional round of screening, which can be iterated to attempt improvement. The process of following up an initial hit with rational design is termed *lead optimization*, and is discussed further in Section 13.3.3.

By modifying functional and structural groups to enhance targeting by initial hits, the pharmaceutical industry and the field of organic synthesis generally have massively exploded the available pharmacopeia [29]. Thus directed exploration can optimize a hit compound for a desired effect, and the process also feeds back more bioavailable compounds for random screens generally. Chaotically cycling between the two approaches for the gamut of medical purposes during the past century of drug discovery has clearly resulted in enormous productivity [3], and an evolution of the available pharmacopoeia.

13.2.6 Screening in Current Pharma

Very generally, the approach of major pharmaceutical companies is to run a large chemical compound library against target proteins of interest using a simple protein-based *in vitro* reporter system, or simply high-throughput screening (HTS). The initial hits are then assessed in progressively complex and representative *in vitro* and *in vivo* model systems, whereupon active compounds are considered “leads” to a drug. Finally, the long and arduous process of three phases of clinical trials is undertaken to obtain approval from a governing agency (FDA in the United States).

The traditional cycling between random and directed searches is inefficient since the blinded screens result in a vast number of hits and leads that fail to be effective or safe in humans. The pharmaceutical industry (often colloquially referred to as “big Pharma” or simply “Pharma”) sets prices to derive profit beyond the tremendous overhead [5], and as a result therapeutics are often out of reach to those who need it most. For many infectious diseases there is little or no profit to be made, as the sole prevalence is in impoverished peoples. As a result, many potential drug targets for these diseases are ignored by Pharma [30].

Although much of Pharma follows traditional methods, the economic opportunities within increasingly complex diseases have driven it to make some of the most significant advancements [22,31–33].

13.3 ESTABLISHED TECHNOLOGICAL ADVANCEMENTS

13.3.1 The Exploitable Niche

Many proteins have an enzymatic cleft relatively specific to its substrate(s) by patterns of charge, flexibility, and space [34,35]. Metabolites enter the cleft and emerge with some chemical alteration. Reaction products have lower affinity for the active site, so they dissipate. The physiologic substrate will not bind to the enzyme irreversibly, as the purpose of the interaction is generally to modify ligand, target, or both, and thereafter distribute this change as a signal to the cell or environment. This requirement of physiologic ligand expulsion creates the quintessential exploitable niche for drug discovery.

The protein target is evolved to stabilize a thermodynamically unstable substrate ligand transition state. The protein might bind the ground state, but it stabilizes the reaction intermediate, which decreases the activation energy for the reaction and thereby promulgates the ligand product state. Yet the protein is also evolved to favor egress of the product after the reaction. The protein is most fit to bind the intermediate (rather than ground or product states), but as this state is inherently transient, it should be possible to find substances that are similar to the reaction intermediate but stable in this form. As the transition state is the thermodynamically least favored state, applying a ligand that is thermodynamically stable in a similar form will kinetically overwhelm the protein and thereby inhibit target protein activity [9,26].

13.3.2 Target Dissection for Inhibitor Design

Proteins fold into complex structures. Some parts are evolved to stabilize the topologic fold, while others carry out physiologic interactions, and others yet do both [36]. The chemical structure of the active cleft dictates the function and the range of adoptable structural conformations. Modeling the pattern of tolerated and optimal moieties across the active cleft enables design and virtual selection of pharmacological inhibitors [37]. The presentation of hydrophobicity, polarity, and charge across the surface dictates where complimentary functional groups should be placed.

Affinity can be understood as change in free energy on binding, which represents the sum entropy and enthalpy changes for protein, ligand, and solvent. Significant conformational constraints can decrease entropy of the ligand and protein during binding. The protein attracts binders by the potential energy stored in the hydrated hydrophobic pocket. Matching any nitrogen, oxygen, or fluorine moieties with a hydrogen bond adds further enthalpic drive to the reaction, resulting in a more strongly binding and therefore a more effective inhibitor (see Section 13.6.5 for further understanding of enthalpy and entropy in computational drug discovery). Thus knowledge of the three-dimensional chemical structure of the target active site enables design of strong binders that might be used pharmacologically as inhibitors.

13.3.3 Rational Design and Optimization

As discussed above, the affinity of a hit compound can be improved by strengthening contacts identified by analysis of the active cleft of the protein structure (enthalpic improvement). Successful inhibitors bind a range of active-site conformations, or induce a particularly stable conformation. The natural substrate of the target protein can be studied to understand the contacts that stabilize the physiologic interaction, but the chemical scaffold of the metabolite can rarely be used to design a stable inhibitor. In part for this very reason, a good inhibitor generally avoids covalent modification by the target protein, but the inhibitor may be modified by other proteins to increase affinity (e.g., partial breakdown during first-pass metabolism, or phosphorylation by other enzymes in the targeted pathway).

The goal of optimization is to improve the therapeutic index: to increase activity (efficacy) and decrease toxicity (specificity). Optimization steps can increase affinity or specificity, but seldom improve both simultaneously. Goals for efficacy include outcompeting the physiologic ligand (metabolite), while the more complex goals for toxicity include minimizing other reactions (specificity) and producing a favorable absorption, distribution, metabolism, and excretion (ADME) profile. To balance pharmacokinetic properties during lead optimization, the ADME profile is considered in the context of the clinical indication [38].

Possible modifications to optimize organic inhibitors are nearly infinite. They include adding any chemical group from a single carbon (methyl group) to a heterocyclic, tethering components to force a particular conformation, or swapping atoms to alter ionic or hydrogen bonding, or patterns of hydrophobicity. Changes made to bioactive peptides alone include multimerization and additions of lipid, polyethylene glycol, or peptidomimetic features [39].

Of course, much of the understanding of protein ligand interactions comes from analysis using computational graphics programs. In accord, exploration of affinity optimization can be carried out by hand at the computer terminal, applying experience and intuition to fit specific chemical moieties to concavity forms and electrochemical contacts [26,40]. The optimization process can also be applied by cyclically testing alterations of virtual hits from computational docking [41]. Alternatively, computational methods can be used to produce a group of virtual hits, of which a sufficient number of compounds are tested at the bench to secure multiple submicromolar hit compounds for follow-up animal experiments [6,32,42,43] (Table 13.1). Improvements to the latter approach are the subject of Sections 13.4–13.6 of this chapter.

While many examples of structure-based drug optimization exist, a quintessential example of computationally guided optimization is found in the work of Becker and colleagues, in the production of PRX-00023 as a lead compound for major depressive disorder and generalized anxiety disorder [41] (Table 13.1). The 1-nM K_i hit arylpiperazinylsulfonamide (PRX-93009) was found using purely computational methods by modeling the $5HT_{1A}$ GPCR (serotonin receptor 1A), docking a library of 40,000 compounds, and running 78 virtual hits in an in vitro reporter system [32] (Table 13.1). While the magnitude of target activity demonstrated great success, the compound presented suboptimal selectivity and pharmokinetics. The same group ran the compound arylpiperazinylsulfonamide against 50 other GPCRs *in vitro*, modeled the experimentally derived interactions (α1- and α2-adrenergic receptors and hERG), and optimized selectivity for $5HT_{1A}$ by removing or substituting moieties that strengthened off-target contacts, and by adding many compensatory on-target contacts [41]. The resulting compound, PRX-00023, was sufficiently selective to $5HT_{1A}$, and presented a pharmacological availability profile similar to those of existing drugs for the same indication. The entire process from the computational screen through entry into phase III clinical trials took only 2 years [41]. Unfortunately, although it was tolerated, the efficacy was not sufficient [44,45]. Nonetheless, this adventure demonstrates that computational methods can facilitate lead compound discovery and catalyze the process of getting to the question of real clinical efficacy.

13.3.4 Multitarget Dosing

In many cases, no single drug is sufficiently effective in the therapeutic range to cure the disease, or even to reduce symptoms or recurrence effectively. Thus, multiple drugs can be combined to heighten the effect. Simultaneous effects on multiple targets can decrease therapeutic doses, so that less efficacious and slightly more toxic compounds can be used safely. In addition, pathogens often develop resistance to single-drug therapy, but simultaneous occurrence of multiple resistant mutations are exponentially less prevalent. The *multitarget* concept of targeting more than one protein in a single dose emerged to address these issues.

Perhaps the most successful application of intentional multitarget drug administration is presented in dosing with inhibitors of HIV reverse transcriptase, protease, and integrase in the fight against HIV/AIDS [46]. Multidosing is titrated in a trial-and-error manner, using patient suffering as the error. Because of this undesirable situation, novel approaches have emerged to model synergistic effects of polypharmacology. For example, combinatorial effects have been tested *in vitro* using an automated robotics–informatics pipeline. Pairs of substances that display synergistic inhibition of *Candida albicans* growth, cytokine production, and tumor growth exhibit complex efficacy patterns; highly nonlinear effects are observed in plots of the concentration of one compound versus the other. The complexity is evident of either single-protein targeting by different inhibitors, or more likely, inhibition of multiple proteins involved in the same physiologic process [31]. Further examples, design, and benefits of polypharmacology are discussed throughout this book.

13.4 COMPUTATIONAL DRUG DISCOVERY

The structure-guided computational approach to evaluating protein–ligand interactions generally consists of three steps: (1) conformational sampling of the rotation, translation, and torsion angle degrees of freedom between the protein and ligand; (2) scoring the resulting interactions with a discriminatory function to identify native and near-native complexes from a set of incorrect conformations; and (3) ranking possible ligands to distinguish between strong, weak, and nonbinders. Despite previous successes, limitations persist in structure-guided drug screening and design implementations to date. The principal discrepancy between what computational drug discovery is intended to be and the reality of what it provides is that computational predictions *enrich* for, rather than *design*, compound–protein activity. In the best reported cases there are still many false positives and false negatives (Table 13.1); structure-guided discovery is a rational starting point, but does not yet provide a comprehensive view of biologic interactions.

13.4.1 Principles and Data Sources

As successful approaches to protein structure prediction do not model any part of the folding process, modeling the physiologic conformation of a bound ligand has little, if anything, to do with the actual physical process of binding. While the hypothetical situation of modeling the wavefunction for each atom in the system could produce a descriptive simulation of ligand binding, this approach is computationally intractable. Again, analogous to the example of protein structure prediction [63], the methods most successful for modeling the stable end state conformations are those that directly consider many measurements of other end-state conformations [64,65]. In essence, physical properties such as interatomic distances, repulsion, or attraction are taken to build models to estimate stability of the protein–ligand system. The strength of computational methods is in automating these analyses across enormous amounts of ligand to protein pairs.

13.4.2 Docking

The term “docking” describes placement of a ligand onto the molecular surface of a protein, in a manner that mimics the real physical interaction as closely as possible. The interaction of any two particles above absolute zero temperature are dynamic, so the protein–ligand physical interaction includes a distribution of conformations that may be clustered extremely tightly (<0.1 Å root mean-squared deviation) or include significantly dynamic protein and ligand movements, as can be found in the range of holo PDB structures [66].

Docking can be performed in such a manner to offer alternative molecules to an initial hit or known physiologic substrate, in which a base molecule provides a starting conformation. Our group recently demonstrated the utility of this approach to peptide inhibitor design, wherein we took as the starting conformation a strand from the physiologic substrate protein (in the PDB structure), and substituted alternate residue sidechains, following a greedy search protocol [7] (Table 13.1).

The complexity of the docking problem expands with the degrees of freedom of each ligand. Unfortunately, while proteins are often treated as rigid surfaces on which to dock a ligand, they are dynamic as well, including movements in response to ligand binding, termed “induced fit” [67].

13.4.2.1 Translation In the most simple case of docking, a roughly spherical ligand (such as a metal ion) is translated about the protein. The translation space sampled can be a grid, limited to a region of the protein or a defined space surrounding the protein, or can be continuous, in which case movements from a starting point must be guided by a scoring function. In either case it is tractable to sample within 0.1 Å of the binding site in a suitable model of the protein structure, and so selection of the real binding site is left to the scoring function (discussed in the next section).

13.4.2.2 Orientation For the anisotropic case of all multiple atom ligands, orientation must be considered. The rigid ligand is rotated about the grid or starting point. To achieve the same 0.1 Å resolution as described above for the isotropic translation search, the requisite search space would be increased 51-fold for a hydrogen molecule (the number of nonredundant 0.1 Å square gridpoints on a 0.76-Å-diameter hemisphere) and exponentially more for ligands of greater size. However, this search is still tractable, and has been applied in various attempts to break down more complex molecules into rigid fragments.

13.4.2.3 Bond Rotation Nonrigid ligands contain rotatable single [sigma (σ)] bonds that dramatically increase the sample space. Simplifications can be made to some rotatable bonds to decrease the impact on sample space, for example, removing bond angles that produce eclipsing of large repulsive chemical groups. However, the existence of multiple rotatable bonds in a ligand generally breaks the tractability of the docking search, and heuristic strategies must be employed. The earliest versions of docking methods simplified flexible ligands as rigid [68], yet even now rotatable bonds not only increase the search space but also decrease the accuracy of all docking methods [64,69]!

Most docking methods combine the three types of movement: translation, orientation, and rotation. The combined movement is generally guided by a scoring function, but the means by which they are applied can be very different [64,70]. For example, the movements from one sampled conformation to the next might be decided by comparing scores for the first and a stochastic progression (Metropolis–Monte Carlo approach), or the trajectory resulting from an estimate of forces in the system (molecular dynamics approach). Therefore, at the heart of the docking protocol is the scoring function.

13.4.3 Scoring and Discriminatory Functions

Functions for evaluating protein–ligand interactions are generally referred to as *scoring functions*. Scoring functions applied to the problem of selecting the most realistic ligand conformation among a set of docked poses is a *discriminatory function*. Protein–ligand scoring functions are categorized into molecular dynamics forcefields, empirical functions, and knowledge-based functions. Forcefields are commonly built to explicitly model physical forces (acceleration) of idealized gas-phase enthalpy, including electrostatics and van der Waals forces (shape complementarity [71]). Often omitted are the contributions of entropy (e.g., torsional) and solvation, while heuristic considerations such as number of hydrogen bonds are most often included [64].

Assignments of the terms *knowledge-based* and *empirical* are historical; both use experimental data to build scores and coefficients. Both perform statistical comparisons of the query case to many bench laboratory-derived binding affinities and/or structural conformations. Generally, empirical functions combine physical terms by regression analysis of experimental binding data, whereas knowledge-based functions derive scores for ranges of spatial parameters (distance, torsion angles, or voxels) from experimentally derived structures without any attempt to divide the underlying physical forces [64,65].

The molecular dynamics forcefield-assisted model building with energy refinement program (AMBER) represents the flagship molecular dynamics function. AMBER models the potential energy of each conformation with a set of terms for covalent bonds, bond angles, torsion angles, electrostatics, and van der Waals energies [72]. AMBER has gone through continual updating by many contributors, to progressively incorporate physics-based models of diverse systems and optimize the coefficients of the formula for specific types of interaction [73,74]. Although molecular dynamics forcefield functions hypothetically have the capacity to direct ligand docking into the lowest energy conformation, using these functions to model an entire protein–ligand system has the tendency to result in models continually expanding out from the physiologically compact state; artificial constraints can be used to hold the model together, but these constraints represent a deviation from the goal of physics-based modeling, are not generalizable, and the results are usually not predictive. Nonetheless, judicious use of a limited progression of molecular dynamics steps guided by these functions can be highly useful for modeling protein–ligand systems [75].

Increased success in developing discriminatory functions have often arisen from specifying the type of protein target, with the presumption that different forces dominate ligand binding by proteins such as transmembrane receptors and transcription factors. However, in 2009 our group developed a generalized knowledge based discriminatory function score to select optimal poses for any type of ligand, within a margin of error that can be sampled by a course lattice method. This knowledge-based function outperforms more than 20 other published ones in several docking decoy tests, by analyzing interatomic distance distributions from the repeating units of high-resolution small-molecule crystallography structures [65]. In part, the strength of this method is the quality of the intermolecular contacts; the crystals of small molecules are much more regular than those of proteins, and so more accurate structures are modeled from the electron density maps. However, exhaustive consideration of the statistical derivation makes this function even better. We considered radial versus normalized frequency distributions, mean versus cumulative reference state, reduced versus complete composition, and the maximum interatomic distance to be considered (cutoff). Across a diverse set of protein interactions with small molecules, other proteins, and DNA, the radial mean reduced derivation performed with the most accuracy [65]. The result is a highly accurate discriminatory function which may provide resolution sufficiently fine to build a continuous function that could act as a forcefield.

Future work to improve scoring functions includes efforts to bolster the accuracy of knowledge-based or empirical functions to address the goals of molecular dynamics approaches. If forces are to be divided into physical contributions, proper handling of entropic and solvation contributions are needed (Section 13.6.5). Further improvements include representing three dimensionality to model the physical intricacies of electron sharing through hydribidized orbitals (e.g. sp^3), and multibody potentials that can account for resonance patterns [76]; there are enough high-resolution structures in the Cambridge structure database [77] to approach these goals [65].

13.4.4 Relative Affinity Ranking

Ultimately there are two roles for the ligand pose selected by a scoring function: to be the representative for ranking among the best scoring poses of other ligands, and to identify the pattern of contacts that might be retained or improved during optimization. Ideally, protein–ligand scoring functions should be able to identify the native or near-native ligand pose from a set of incorrect conformations (i.e., discrimination), and to distinguish between small molecules that do and do not bind a target protein (i.e., relative affinity ranking). This is unfortunately not the case with current methods, as discriminatory functions perform poorly at correlating scores with experimental binding energies. An ideal ranking function would accurately calculate the free energy of binding. Relation to the affinity estimation for another ligand (another drug or physiologic substrate) would be sufficient to estimate biological activity; this thermodynamic understanding would indicate which ligand would outcompete the others by binding strength. The kinetic considerations (e.g., target tissue concentration) could be designed around this understanding. Clearly ranking functions could be extremely useful in computational drug discovery, but currently no function has been shown to consistently reach these goals. An accurate relative affinity ranking function is needed in the field of structure-guided drug screening and design, if these predictive methods are to serve as a useful an complementary tool to prospective experimental investigation.

Knowledge-based functions perform quite well at discrimination [65] but inaccurately provide scores proportional to the size of the ligand, due to their simple additive nature, and therefore may be of limited utility for relative affinity ranking. Empirical scoring functions fitted to experimental binding energies perform rather poorly, especially for classes of molecules not included in the training set, and significantly lack in discriminatory ability. Often experimental complexes are used to correlate scores with experimental binding affinities; in practice, this is not useful, as the objective is to find new compounds that bind to a protein target. The most relevant experiment is to test known inhibitors against alternative protein structures that are not bound by the small molecule of interest, and then evaluate the correlation coefficient (which is invariably lower [78]).

The ability to accurately discriminate the correct binding mode of individual ligands and then to rank the relative binding affinity between different ligands can be treated as distinct computational modeling problems. All protein–ligand scoring functions can be applied as ranking functions, but dissecting apart docking and ranking allows for considerations more important to each problem. For example, counting hydrogen bonds and calculating loss of torsional entropy is essential to ranking ligands, but many conformations of the same ligand can be equivalent for these factors [64]. Therefore, the methods used for discrimination and relative affinity ranking should be separated into distinct functions and developed independently, which has not previously been the case.

13.4.5 Comparison of Docking Methods

Many methods have been created to dock ligands to proteins [64]. But bias and overtraining have impeded attempts of evaluation in the field of computational biology, as demonstrated for protein structure prediction with the solution of the CASP experiments [63]. Blinded or independent examinations are the proper means for unbiasing assessments of predictive methods. Minimizing bias optimizes the estimation of the accuracy in prospective experimentation, which is the purpose of these methods. A more recent experiment performed such an independent test between seven docking programs (Surflex, LigandFit, Glide, GOLD, FlexX, eHiTS, and AutoDock) on 1300 holo structures from the PDBbind database. Ligand conformations were converted through SMILE strings using two different tools (Corina, Omega2). Two commercial products (GOLD, eHITS) outperformed the other methods, with mean accuracy <3.0 Å root mean-square deviation (RMSD) and $>55\%$ of cases <2 Å RMSD. The use of holo rather than apo structures is a caveat to the relevance of these findings to prospective drug discovery. Furthermore, it is likely that the examined methods were trained on some of the same structures as those used to test them, which gives unfair advantage. While prospective experimentation is the only true test of a computational method, this study describes the most independent comparison of methods for drug discovery known to us [69].

13.4.6 Ligand Comparison

Small-molecule structure–activity relationships (SARs) are applied to find active substances similar to initial hits found through bench or computational techniques [28]. The underlying concept follows that the activity of the substrate transition state can be analogized by chemical similarity to any other compound [79]. It follows that the activity of a hit ligand can be analogized by chemical similarity to any other compound. The ligand comparison is calculated by comparing the geometric distribution of electronegativity and hyrophobicity for the hit ligand against a database of existing small molecules. Improvement accurate predictions of the similar active substance are found by limiting the database to known bioactive molecules. This approach is powerful in part because of the small requirement for computational resources compared to docking.

While the structure activity relationship of small-molecule organics has been applied to ligand optimization traditionally, the concept of similar chemical structures having similar bioactivity has recently been applied to discover initial hits [9,11,30,80] (Table 13.1). The rationale here is to use known substrates or predicted ligands in place of the initial hit. It is logical that the physiologic substrate would be a productive starting place for detection of similar active substances. This brand of SAR applications is expected to greatly improve the efficiency of drug discovery and expand our understanding of the coevolution of proteins by their similar physiologic substrates.

Other computational methods compare structural and chemical properties among protein–ligand binding sites directly without considering ligands [81–84]. For example, Das and colleagues dissect a binding site into a profile of probabilities such that a surface patch with a particular physicochemical property will present at a specific distance to another on the binding site surface [82]. When the binding site and tertiary structure is known or predicted, this analysis enables rapid detection of target identification and understanding of multitarget effects, and suggests which compounds to screen for pharmacological inhibition. Binding sites can be predicted by sequence analysis [85,86] or mapping by structural similarity [30].

13.5 MORE RECENT TECHNICAL IMPROVEMENTS

13.5.1 Automated Binding Site Identification

A variety of sequence- and structure-based approaches are used to predict protein–ligand binding sites. For many globular soluble enzymes the binding pocket is easily identified by its characteristic narrowness and depth, which allows harboring of small molecules. This analysis can be automated by geometric measurements; for example, surface concavities can be found by comparing the accessibility of different-sized spheres to the solvent-exposed surface [87]. Meanwhile, many protein active sites are not as obvious from the protein structure; these more difficult problems demand sophisticated bioinformatic tools [88].

Often a ligand can be mapped to the query structure from a holo template protein identified by sequence or structural similarity [30,89,90]. Where ligand mapping is not available, and when results are not consistent, conservation analysis is useful. In particular, proteins from poorly characterized families cannot always be understood by direct similarity analysis. Sequence analysis can evaluate multiple aspects of evolutionary conservation and residue identity to predict binding sites with comparable accuracy to structure-based methods [91–93]. Structural analysis or structure prediction can be combined with conservation calculations to improve interpretation [36,86,90,94]. Our group has found, across many protein active sites, that hidden Markov model estimates of relative conservation entropy is the most accurate single predictor of residue functional importance [36,85].

Differences in residue identity within otherwise similar binding sites control metabolite specificity and variation in enzymatic reactions [95,96]. Thus the residues that specify ligands are seldom conserved. More advanced analysis is indicated to find these residues; function prediction methods may be useful for selection of atomic contacts to targets during computational drug discovery. Our group has demonstrated that machine learning can be used to transfer dissections of structure and function from many proteins to predict the active sites of highly different query proteins [36,86,93]. Our methods predict protein–ligand binding sites *de novo* using an algorithm that generates metafunctional signatures (MFSs) by combining multiple sources of information reflecting functional importance. MFS can be applied to a protein sequence or structure and has been shown to be more effective in identifying functional sites than have other popular methods [36,86,93].

13.5.2 Docking with Protein Target Dynamics

Biologically active proteins are in continuous motion, yet the majority of protein structure information is limited to the most stable form of a protein when crystallized in artificial conditions. Induced fit is a widely recognized challenge in computational drug screening, wherein the protein undergoes significant conformational changes on ligand binding [67]. As a consequence, traditional rigid protein–ligand docking is insufficient for structure-guided drug screening, and is often misleading. The active cleft surface is treated as rigid, although a conformational shift occurs on binding a physiologic substrate, inhibitor, or interacting protein. This conformational shift brings together the mediator functional groups of the catalytic reaction. The energetic force to bind the reagent metabolite is generally enthalpic, so the bound holo conformational state of the protein is closest to the optimal pharmacologic target. Dynamics simulations increase the possibility of surveying a physiologically relevant conformation beyond using the static crystal structure alone. For example, our group showed that, for a group of HIV-1 protease inhibitors, using molecular dynamics to model changes in the target protein improves the correlation coefficient of predicted score versus measured affinity from 0.35 to 0.88 [75].

Modeling target and ligand flexibility facilitates the multitarget approach. Multiple stable conformations or highly flexible portions of a ligand increase the range of possible target clefts in which the ligand might fit. The benefit of ligand flexibility for action on multiple proteins is exemplified in the difference between first- and second-generation HIV protease inhibitors [55].

Along with our work demonstrating the importance of target protein dynamics in computational docking, many other groups have incorporated target flexibility into their software. For the rotatable bonds of ligands discussed in Section 13.4.2, however, each additional bond considered for rotation dramatically increases the sample space, and so slows down the search. Therefore our approach of using short spans of molecular dynamics (200 steps) appears to be most computationally reasonable, and is used widely [6].

13.5.3 Structure Modeling for Target Docking

The concept of template-based modeling can be understood and applied in different ways. One approach that has been shown to work is to model the query protein using a template, and then dock to this model. However, this is not the only way to make use of a template. It is not always necessary to build a structural model. If there is a known drug or ligand interaction for a template protein, this information may be transferred directly, based on the similarity between the proteins [30]. If docking is indicated, it may be more relevant to dock to the template itself rather than a model built using the template—the accuracy of the template is known, while the model built with the template is guaranteed to be less accurate [78]. A good template will have highly similar binding sites to the query protein, sufficiently similar that the differences in residue identity can be modeled after docking.

The structures of all human GPCRs have been modeled with I-TASSER [97], the best existing protein structure prediction method (albeit an older version), and are freely available [98]. Various publicly available methods are capable of modeling structure and ligand docking for GPCRs. For example, our group combined I-TASSER with our consensus refinement method [99] to perform among the very best groups in a prospective prediction experiment to predict structure and ligand conformation for the second human GPCR X-ray structure [100]. Meanwhile, the proof of concept for all modeling drug discovery for GPCRs was accomplished in 2004 by Becker and colleagues [32], as discussed in Section 13.3.3. Briefly, the authors modeled five GPCRs based on the bovine rhodopsin structure (PDB id 1f88; the only GPCR structure known at the time), used the anchor and grow approach in DOCK4.0 [70] for $\sim$150,000 compounds selected from $\sim$1,600,000 based on physical properties, and ranked the resulting protein–compound pair conformations using in-house software. The outcome of this study includes 50 substances with EC_{50} $<$5-μM activity, a novel EC_{50} $<$100-nM compound for four of the five target GPCRs, and an agonist lead compound [32] (Table 13.1). However, there was no comparison performed to check for enrichment versus docking to the template rhodopsin structure.

A study in 2009 explored the opportunity of template-based modeling and docking for 38 proteins, 2950 ligands of known bioactivity, and 95,316 decoy ligands [78]. The exploration was relatively thorough for protein structure modeling, using templates across a broad range of sequence identity (20–99%). In this study the consensus result of docking against multiple template based models was better than that for docking to the single best model or even the apo structure of the protein (in most cases), and in many cases the consensus model accuracy approached that of docking against the target holo structure. Meanwhile, this study also compared bioactive ligand selection enrichment for docking to the homology model templates versus the actual models. There was a slight trend for holo templates of sequence identity below 40% to more accurately select the bioactive ligands than models derived from the holo template ($R = 0.22$ across sequence identity range). There was no clear range for which it would be better to use homology models. When using apo templates or models derived from them, the correlation for sequence identity dropped ($R = 0.07$): sequence identity is not predictive of whether it is better to use the apo template itself or a model derived from it [78]. On average, docking to templates produced insignificantly higher enrichment for bioactive ligand selection than docking to models of the target protein (Student's paired one-tailed t-test $p = 0.29$). So, on the basis of this study using the latest versions of MODELLER and DOCK, it appears that for the purposes of docking, there is no great benefit to spending the computational resources to build all atom models of target proteins. Meanwhile, the success of the consensus of models suggests that clustering may be useful for finding the best template on which to dock, and that improvement in structure prediction methods may breach the accuracy of docking to homolog holo structures. Nevertheless, the high resolution of the template is, at least for now, a better data source of analysis, whereupon our ability to detect the evolutionary connection between homologous proteins is the most powerful tool.

13.5.4 Ligand–Target Networks

Metabolic systems bring an environmental substrate through a series of reactions that add or remove chemical moieties. The majority of the substrate is often maintained through the process, such that each protein controlling the metabolic network will recognize similar features of the substrate. Therefore, if a drug is selected or designed to inhibit a particular protein target, it is highly likely that the drug will inhibit multiple proteins of the metabolic network [12,101]. Thus many drugs achieve higher efficacy by unintentional pathway multitargeting [102], with benefits described throughout this text.

Network targeting involves activity of a compound across multiple pathways. Multiple routes of attack may be necessary to effectively stop neoplasms or pathogens that have multiple compensatory pathways to allow survival and proliferation. Increasingly, we are learning that simple linear or cyclic pathways are the exception rather than the rule, so even to inhibit a single pathway, it seems that multiple indirectly connected proteins must be inhibited [101]. If one adopts a multitarget philosophy, the principal difference is a need to monitor the interconnectivity of the targets, maximizing relevance to the clinical question.

13.6 EMERGING CONCEPTS

13.6.1 Starting with Nature

The current drug discovery process itself both mimics and expedites the natural evolution of bioactive products. Living organisms have influenced the creation and relative abundance of chemicals on Earth. An example, would be the production of oxygen by conifers, which enabled aerobic metabolism; the cyclic feedback between life and that which is traditionally considered nonlife (small molecule organic compounds) describes a co-evolutionary pattern that can be exploited in drug discovery.

The current diversity of natural chemicals emerged within the same evolutionary soup. This shared evolutionary chemical context sets the stage for various organisms to use the same compounds to control different processes, making one molecule relevant to diverse physiological activity. The observation that structural folds are largely conserved, even when sequence and function are not, provides logical evidence that one compound can be an excellent initial candidate for many different protein targets. The topological forms of proteins (folds) present much more consistency than do those of small molecules. For example, the proteins of various metabolic pathways appear to have evolved from the same template protein, with mutations conferring the ability to perform different chemical alterations. Meanwhile the binding site within a particular pathway is relatively conserved, and a ligand that impedes a reaction in one protein will be promiscuous to the pathway. The result of these patterns of evolutionary divergence is that natural chemicals are highly multitargeting [6,103].

The network of targets for existing drugs reveals physiologic relationships between the proteins within or between proteomes [9]. In particular, not all human disease targets are predicted to be bound by natural small molecules, and it may be that the respective interaction networks are distinct [103]. The relatively unique human drug–target network may be explained as bearing those more unique protein functions for which there are minimal compensatory self-righting mechanisms. The uniqueness of the target proteins seems to coincide with constrictions in the protein interaction network, rather than network hubs that tend to be targetable by natural compounds. Presumably, these network constriction human targets are not canonical enzymes, receptors, or channels—in other words, prototypic natural compounds are not their substrates. Thus, for these targets, natural products and perhaps their derivatives may be insufficient.

Nonetheless, it is clear that there is some piece missing from the immediately preceding argument and referenced data, as 614 of the 974 new chemical entities discovered from 1981 to 2006 were natural products or derivatives thereof, many of which do target host proteins [104]. Leaders in bench drug discovery look to exotic organisms for drug leads continually (e.g., scorpion venom). Natural compounds can be very difficult to prepare outside the source organism, and few exotic organisms are cultivatable on a large scale. These compounds are the products of intricate protein-mediated metabolic pathways seldom understood well enough to be genetically engineered into *E. coli* or yeast. Computational aid to retrosynthetic analysis enabled mass production of natural active products via total synthesis [105].

Thus natural products may not be able to inhibit or activate all host targets, but for any protein that acts on a natural substrate, they likely will be useful. Thousands of years ago we recognized the pharmacological capacity of many natural materials, and over the past few decades nature has still been the greatest source for new drugs. Natural compounds may not comprise ideal decoys for complex substrates such as DNA or other proteins, but we can keep looking to them as one principal source for bioactive compounds. The evolutionary pressure of competition clearly selected for organisms ready to fight other organisms—the resulting arsenal of molecular weapons—is a robust starting point for rational drug discovery.

13.6.2 Peptides and Their Derivatives

Peptides represent a natural modular scaffold that can be easily designed to mimic natural substrates and binding partners for drug discovery. Knowledge-based protein structure prediction methods can be applied by reverse engineering the amino acid sequence of a natural binding partner to optimize binding. For example, our group created peptide inhibitors by redesigning the sequence of the dengue viral entry protein substrate, which prevents infectivity of dengue virus at the micromolar level [7] (Table 13.1).

Peptides present some benefits for computational drug discovery relative to standard organic small molecules. One benefit is the modularity, which enables design, massive replication, and low production cost. Another aspect is that the chemical nature of sidechain and mainchain moieties are evolved to stabilize proteins, and therefore in some cases bind active sites more tightly than organic small molecules. The rapid degradation by endopeptidases is generally seen as a disadvantage because of inactivation and clearance, but protease recognition is designable to some extent, peptide degradation minimizes immunogenicity, and some clinical indications call for rapid clearance.

Disadvantages of peptides also include susceptibility to nonspecific endoproteases (which exist nearly everywhere in the body) and low oral bioavailability. Even with these disadvantages, peptide inhibitor design can be useful as part of an *in vitro* model for finding or verifying targets, and for identifying specific binding site contacts to be targeted by small molecules. However, modifications to overcome disadvantages are chemically straightforward: multimerization [e.g., poly(ethylene glycol)], lipidization, and adding peptidomimetic moieties (e.g., alternate atoms to substitute the amide bonds). Expressible peptides can be modified chemically to produce vast functional diversity suitable for many pharmacologic applications [39].

13.6.3 Off-Label Drug Use

All FDA-approved drugs present advantages similar to those of natural compounds because of their known bioactivity. Added benefits of screening existing drugs include the known safety and ADME profile, demonstration that the compound will get through first-pass metabolism and get to at least some sites of action, and a hint of certainty that they will have the promiscuity of ligand–protein interactions discussed for natural compounds. Perhaps most importantly, since they are already approved for use in humans, the only barrier to clinical trials is demonstration of efficacy [6].

While it appears that use of existing drugs enriches screens for hit compounds, no one has done the proper side-by-side background control of testing a random sample of compounds. Current Pharma compound databases are designed to optimize bioactivity and ADME profiles in the case of presenting a hit inhibitor, such as following Lipinski's rule of 5 [106]. However, four bench screens searching for inhibitors of *Plasmodium falciparum* demonstrate a trend toward enrichment for existing drugs (Table 13.1). Massive screens of $\sim$2 million compounds from the chemical libraries of Novartis [22] and GlaxoSmithKline [107] resulted in 0.35% and 0.68% micromolar hit rates, respectively. In earlier studies, 1000-fold smaller bench screens of $\sim$2000 existing drugs for *Plasmodium falciparum* resulted in 0.71% [23] and 1.7% [24], suggesting slight enrichment. Meanwhile, our computational screen of the same drug database selected 16 compounds, of which 44% are micromolar inhibitors [6] (see Table 13.1 for further details of these studies). Although these giant Pharma companies have put decades of data and analysis into the design of their chemical libraries, similar, if not better, success rates can be achieved on a 1000-fold smaller scale if these screens are simply run with existing drugs. Moreover, our group has shown that publicly available computational methods can vastly enrich this search, and thus recommend existing drugs to be the starting set for any computational drug discovery project.

Understanding the biologic activity of known drugs, of course, makes it easier to repurpose them for desired physiologic effects. It is important to note here, within this chapter on automated tools for drug discovery, that deep understanding of existing drugs and the disease of interest enable enrichment far beyond that currently available with contemporary computational methods.

Accordingly, off-label uses are continuously being discovered. Carbamazepine, a widely used anticonvulsant and mood stabilizer, seems to combat hepatic fibrosis [108]. A lead for polycystic kidney disease has been discovered by intuiting the target, for which an inhibitor was already developed in effort to treat diabetes [33].

The trend for drugs approved for treatment of one disease to effectively treat another one underscores the importance of epidemiologic studies to track disease patterns in medicated patients. Clinical informatics is an emerging field intended to handle issues such as this. Meanwhile, the reward for repurposing an existing drug is highly similar to discovering its first use. In the United States, intellectual property and patents are defined by the purpose; if you can figure out a new use for a hula hoop, you can patent it. A new use for an existing chemical entity is unique intellectual property. The only successful generalization of profit for a drug has been through manufacture of the physical drug itself. Thus, opportunity awaits in repurposing old drugs to new tricks.

13.6.4 Off-Target Effects

Virtual drug screening methods have been employed to help identify sources of off-target drug effects and investigate their potential to cause adverse or desirable side effects [8,9]. Desirable off-target effects include unintended multitargeting of other proteins in the target pathogen [6], fighting other infectious agents [8,109,110], and balancing untoward effects of other drugs being used in a polypharmacologic regimen. Through proper screening of relevant host and pathogen proteins and metabolites, current methods can enrich the design of off-target pharmacology.

Off-target effects can be predicted by ligand-docking methods [6,8,109–111], ligand structure–activity relationships [9,112,113], and comparison of protein binding sites [81–83,111]. After decades of development [28], SAR methods are emerging as clinically useful [9] (Table 13.1). Meanwhile, methods to compare protein-binding sites and affinity-ranking methods are still in their infancy, yet the latter has already demonstrated clinically significant utility [8].

Although virtual screening methods have been useful to inform drug design, many current methods are not able to account for off-target drug effects because they require structural information that is not available for most of the human proteome [114]. Further, because of the difficulty of crystallizing membrane proteins, structures for these proteins are highly underrepresented, making up less than 1% of the structures in the PDB [115]. Nonetheless, nearly half of available drugs act on G-protein-coupled receptors, a major class of membrane signal receptors [116]. Therefore it is important to consider membrane proteins in the identification of off-target drug interactions. Although structural data are lacking, protein sequence data cover nearly the entire human proteome [117]. Therefore it may be useful to develop computational protein sequence analysis methods to identify the similarity of protein–ligand binding sites through their meta-functional signatures [93], which could model drug toxicity explicitly across human and pathogen proteomes. The most useful off-target screening methods will combine comparative analysis of ligand structure, protein structure, protein sequence, and the types of interactions between protein and ligand.

13.6.5 Affinity, Entropy, Enthalpy, and Optimization

Dissecting the contributions of entropy and enthalpy to changes in the free energy of a system through bench calorimetry has enabled a much more rationalizable approach to computational drug discovery. This work, led by the Freire group, stems from the universal approach of balancing losses in entropy with gains in enthalpy. The novelty is both the focus on enthalpic improvements, and using isothermal titration calorimetry as a tool by which to separately measure the enthalpic and entropic contributions to affinity [118–120].

Affinity is improved with larger losses in free energy, such that either gains in entropy or loss in enthalpy could drive a reaction. Improvements in one (entropy or enthalpy) can overcome deleterious effects on the other. Meanwhile, scientists traditionally measure only affinity (K_d) or inhibition (IC_{50}, EC_{50}, K_i). These are one-dimensional measures of binding strength, which are highly useful, but it can be difficult to interpret the correspondence of ligand structural changes to whole affinity differences. By dissecting the contributions of enthalpy and entropy to the gains or loss in affinity, one can see how the changes are made; ligand changes that effect conformational freedom represent entropic changes, while improved interactions and fit are enthalpic [119]. By separating the measurement of these effects in bench studies, the optimization process obtains direct logical feedback. A change to the ligand designed to improve enthalpic contributions might have much more severe entropic consequences than anticipated.

Attempts to gain affinity driven by entropy might not make a significant change because of constraining the protein for an entropic loss. Without the separation of analyses afforded by calorimetry, the lack of improved affinity might be misinterpreted as enthalpic losses, which would misdirect further attempts at optimization. Thus the relatively simple concept of separating affinity measures into enthalpic and entropic contributions through isothermal titration calorimetry enables feedback for straight forward rational design [119,120].

Decrease of conformational restrictions in the protein or ligand correspond to favorable entropic changes. Entropy estimations are useful to interrelate affinities between different ligands (affinity ranking). However, it has been argued that optimization efforts are better spent on improving the enthalpy of binding [121]. Considerations for design include the fact that every added hydrogen bond has both enthalpy of desolvation and of binding, and that each 1.4 kcal/mol of enthalpy change drives the reaction thermodynamically by an order of magnitude. These considerations are so important that Freire has suggested that binding enthalpy should be measured by isothermal titration calorimetry whenever a new hydrogen bond donor or acceptor is considered [121].

Separate measures of enthalpy and entropy can enable better estimates of both contributions [118–120,122], but what should go into the enthalpy calculation? Many types of enthalpic contributions are understood and well approximated. Details such as the contribution of hydrogen bonds are modeled by comparing the docked donor–acceptor distance to the ideal distance for proton sharing, in the context of the similar interactions available in the solvent. Binding enthalpy was estimated for 25 ligands in 7 proteins within a standard error of 0.4 kcal/mol, by supplementing estimates of conformational enthalpy change, with estimations of changes in solvent accessibility for solvent molecules in shells $\leq$ 5–7 Å away from the ligand, and a correction for protonation [118]. Modeling changes in enthalpy across different ligands may therefore be possible, and useful for estimating affinity rank.

The contribution of space filling to enthalpy had not advanced substantially since the shape complementarity analysis of Lennard-Jones [71]. In 2010 the Freire group presented a study on how filling an empty protein cavity affects enthalpy [122]. For the example of filling clefts in the binding pocket of HIV-1 protease, a pattern of effects emerged across a limited spectrum of moiety size. When the cavity was not completely filled by the ligand moiety, van der Waals forces benefitted enthalpy, but at the cost of entropy. When the moiety was enlarged, the protein accommodated more optimal filling of the cavity space, adjusting around the ligand to reach a more enthalpically favored conformation. Entropy increased, driving the reaction. The interactions enabled by optimal space filling may have allowed the protein to stably undergo pivot motions around this region, such that stabilizing interactions at the ligand interface allow other areas of the protein to be more flexible, and thus the reaction becomes entropically favored. There is an apparent overstretching point at which the ligand pushes the protein into a more strained set of conformations, which penalize by both entropy and enthalpy. Thus proper filling of the space can add entropic and enthalpic driving force to binding [122].

Through the analysis provided by the Freire group during 2000–2010, we have gained the ability to dissect very basic contributions of designed ligand moieties. Bench isothermal titration calorimetry analysis enables specific feedback to improve our estimates of entropy and enthalpy, and inform changes for computational design. This combination of a relatively simple but highly accurate bench technique with computational modeling is an emerging tool that can carry us forward to the next generation of drug discovery.

13.6.6 False Hits

The concept of false hits was demonstrated elegantly by the work of the Shoichet group, in showing that hit compounds can inhibit protein activity by pathological mechanisms [123]. The "false hit" inhibitory mechanisms of β-lactamase inhibitors discovered by high-throughput techniques include many aggregators, covalent bonders, and promiscuous inhibitors. Poignantly, none of the 1274 initial hits were found to be specific reversible inhibitors, which are pharmacologically desirable. Meanwhile, 2 of 16 computationally derived hits were specific reversible micromolar inhibitors (Table 13.1). Thus, the approach of computational screens is bolstered by the fact that they model bioactivity in an explicitly physiological manner, whereas wet-lab systems model the physical interaction and therefore can get sidetracked by irrelevant behavior—a behavior that could be highly dangerous to the host [123]!

13.6.7 Finding Targets of Known Inhibition

Many drugs have no known mechanism. For many more drugs, the mechanistic basis of side effects is not understood. Mechanisms are the deep understanding of an interaction that enable improved design and analogy to less understood cases. They let us understand the exceptions, such as variable response.

Target elucidation allows us to understand clinical paired disease patterns. From observations that the opportunistic pathogen CMV is cleared from AIDS patients undergoing antiretroviral therapy, one might anticipate the nonspecific mechanism of HIV-1 inhibition allowing return of immunity and nonspecific clearing of CMV [124]. However, CD4 T-lymphocyte counts do not correlate with clearance [125]. Our docking study predicts that amprenavir and indinavir target the CMV protease specifically [110]. Our group presented a similar descriptive prediction for HIV-1 inhibition by the common antibiotic minocycline being through HIV-1 integrase [8].

Moreover, we can understand the interrelation between bioactive compounds (metabolites and drugs) and the relevant proteome through the network of overlapping target–ligand interactions. A more recent tour de force was applied to predict the interactions of all drugs to the human proteome. The resulting network is a roadmap for polypharmacologic effects—leads and suggestions for caution [9].

In the case for which the activity profile of a vast drug library is known for a particular pathogen, which is becoming more common, analysis of similarly active compounds can facilitate understanding of the targetable aspects of the pathogen. Targets can be selected according to the profile of activity across the library [22] (Table 13.1). Depending on the clinical indication, a target may be selected for uniqueness of library activity relative to the host and commensal organisms, or perhaps for similarity to targets of other diseases to maximize the chances of discovery of an existing drug multitargeting the disease of interest.

13.6.8 Personalized Pharmacology

As the accuracy for models of protein–ligand interactions improves, along with it comes the ability to personalize these predictions. In models of individual susceptibility versus resistance, or predictions of disease progression, differences in genotype have already been modeled with a high degree of accuracy. The most common difference relevant to this problem is the nonsynonymous single-nucleotide mutation or polymorphism. The change of one or more residues by mutation alters specific contacts to increase or decrease affinity, thereby rendering the mutant organism susceptible or resistant, respectively.

Our group designed a sequence analysis tool to predict the significance for this type of mutation [36], but much work remains. Our group also created a group of tools to take a patient's HIV-1 protease and reverse transcriptase sequence mutations and predict the profile of resistance versus susceptibility to the commonly used antiretroviral medications [109,126–128], and integrated them into a freely available web server that uses the consensus of the structural and logistic regression techniques to select the optimal drug for HIV-1 patients (this web server has handled over 1000 separate queries (available at `http://protinfo.compbio.washington.edu/pirspred`; [37]).

Other personalization includes screening for untoward side effects, such as inhibition of CYP450 proteins or monamine oxidases. Additionally, we differ not only in our human genotype but also that of our symbiotic bacteria. Personalized pharmacology may one day include identifying an *E. coli* strain by genotyping stool samples, so an antibiotic regimen can be selected that will not cause imbalance to one's enteral flora.

Finally, dosage can be prescribed by employing models of enteral uptake using the genes that code for microvilli intercellular junctions and models of metabolism based on the CYP450 genes, and immunogenicity by the antibodies of memory T cells and mast cells. Dosage can also be prescribed by gene copy number variant, and relative susceptibility.

13.6.9 Open-Source Drug Discovery

Through the development of robust, free, and publicly available computational methods for drug discovery, we can increase efficiency and decrease costs for researchers and institutions involved in drug discovery worldwide. Computational methods have demonstrated the ability to greatly reduce the cost of hit and lead compound discovery [6,7,30,41,123] (Table 13.1). Therefore they have the potential to enable the development and distribution of drugs to combat diseases that disproportionately affect impoverished nations (also known as *tropical* or *third-world diseases*), such as malaria and dengue fever. Since tropical diseases mostly affect the poor, the historical perspective has been that there is little to no incentive for pharmaceutical companies to invest in the development of these drugs. Nonetheless, it should be noted that some of the largest Pharma companies have more recently devoted massive resources to join the fight against malaria, including Novartis [22] and GlaxoSmithKline [107] (Table 13.1). In addition, from the standpoint of computational methodology, in head-to-head comparisons the best-performing computational methods for drug discovery are not freely available nor publicly funded software [69,100]. Reasons for a partial shift to open publication and application of resources to minimally profitable diseases are intriguing, but beyond the scope of this text; for now these are the exceptions rather than the rule. The importance of reducing drug development costs through computation is unwavering.

Although many existing tools used in drug discovery are freely available, the skills necessary to use them and interpret the output typically require a large amount of knowledge, which represents an obstacle to widespread use. It is rare even for medical scientists capable of performing animal studies and clinicians capable of performing clinical trials to possess the knowledge necessary to use computational predictive methods. In response to these barriers, a trend to release the identity of predicted compound-target interactions has emerged among publicly funded computational research groups [6–9,30,42,43,109–111] (Table 13.1). Moreover, the trend has been to share the outcome for initial experiments among these leads publicly. For example, since 2000 our group has been committed to making all of our software, ideas, and data freely available to advance the science, and to release our predicted hit compounds in a way that maximizes impact and availability.

In addition to making all data publicly available, it would be useful to develop an easily accessible public web sever usable by nonscientists and scientists alike to expedite communication of knowledge to advance the discovery of novel drugs. Using a web server could be as simple as uploading the structure or sequence of a single target protein or set of related target proteins. A comprehensive analysis of the target(s) would predict inhibitors and substrates of the target(s). Antitargets with the potential to interact with each of the lead compounds could also be identified and presented to the user. Potential compounds tested for activity against the target(s) would come from a library of existing bioactive small molecules. When available, experimental data such as ADME, bioavailability, or binding affinity could be stored for each compound and presented to the user in a standardized way. A few open-source drug discovery projects have begun to address these goals to promote the discovery and development of novel therapies to neglected diseases [6,30].

13.6.10 Multitarget Design

While we have argued that the search for a compound with a desired activity can be expedited by evaluating multitargeting compounds, we have not yet elaborated the principle of a single compound multitargeting a single disease. This relates to the off-target properties discussed above, according to the concept that natural compounds and known drugs are more likely to be multitargeting. Here we extend the assertion that compounds can be selected to target other proteins in the same disease. This concept was perhaps first formalized by Erlich, who described a magic bullet that would inhibit cancer by multiple mechanisms [129]. One such example is Gleevec (i.e., imatinib, STI-571), which serendipitously targets both BCR-Abl and c-Abl, inhibiting the two principal known causes of cell proliferation in chronic myelogenous leukemia (CML [130]). Gleevec has been the most widely used treatment for CML since 2002.

The most effective drugs in humans (e.g., aspirin, Gleevec) inevitably interact with and bind to multiple proteins, a feature that traditional models based on single-target drugs fail to take into account. Yet there is substantial evidence that these multitarget compounds have a higher incidence of untoward side effects than do single-target compounds [131]. The multitarget approach is necessary because every drug has to be effective at its site of action (e.g., HIV-1 protease inhibitors have to bind and inhibit the protease molecule) and readily metabolized by the body (e.g., the cytochrome P450 enzymes, which are responsible for metabolizing the majority of drugs). Computational screening for multitarget binding and inhibition is effective because it exploits the evolutionary fact that protein structure is conserved much more in nature than is function or sequence.

It is ironic and surprising that reduced affinity sometimes corresponds to higher efficacy. This appears to be due to weak linkage of multiple target proteins within a particular physiologic network [10]. Low-affinity multitarget drugs may perturb networks more efficiently than high-affinity, single-hit drugs [12]. Simultaneous effects on multiple targets can decrease the therapeutic dose, so that untoward side effects can be handled by lower doses; simply, a compound with three targets of similar affinity will be effective at one-third the tissue concentration. The effects of salicylates on multiple proinflammatory signals exemplify the fact that multiple mechanisms causing homeostatic imbalance can be targeted by a single drug; the low effective dose facilitated by multitargeting has made aspirin one of the most popular drugs in the world [132].

Pathogens and cancers develop resistance to single-drug therapy. Inhibitor resistance is largely overcome in the multitargeting approach by the exponentially decreased probability of resistant mutations simultaneously arising in genes encoding proteins corresponding to all targets. The multitarget approach can be extended to incorporate the variability of target proteins across a disease pathogen population [55].

Computational predictions are obviously not perfect. The accuracy of recent docking with dynamics and structure activity relationship predictions contain less than 50% true positive hits at best (Table 13.1). However, if one compound is predicted to hit multiple targets, the odds increase for actually inhibiting at least one target. Thus we have taken the approach to target as many essential proteins as there are crystal structures for a specific pathogen or disease [6]. The complexity of possible multitarget effects indicate that occasionally it may be relevant to test in whole-disease-organism screens or even animal models of disease before evaluating which of the predicted multitarget interactions actually occur physically.

13.6.11 Multidisease Screens and Reversing the Disease–Drug Search

Old Western movies keep alive the iconography of "cure alls" popularized into nostalgia by traveling salesmen of the mid nineteenth century. These tinctures were intended to solve any medical problem, or at least a group of quite unrelated problems. In this chapter we share some examples of single drugs that combat multiple diseases. We also preach the repurposing of existing drugs, exploration of natural compounds, and the use of chemical derivatives of each; that is, we continue with the concept of exploiting existing bioavailable, nontoxic, nonimmunogenic, multitargeting compounds. So it would be logical to test the ability of all these compounds to target any and all disease targets.

Given the limited set of compounds that we propose to be used, the chance of finding a drug for one particular disease might not be great, but with contemporary methods the chance of finding a disease for a particular drug is extremely probable. Multidisease screens can find the opportunities that do exist; the screening process can drive the drug-disease selection, rather than the disease (tradition). This concept represents a reversal of the conceptual framework underlying drug discovery, wherein we play to our strengths. At each point of the modeling process we rely on the best scoring instances from the scoring functions. While somewhat ambiguous instances arise for all methods, scoring functions make it easy to know when the models are of little or great utility. Thus, if we scan for instances for which the accuracy estimates indicate useful models, rather than searching for the best model for one's pet project, we may truly access those diseases, targets, and compounds that are most realistically modeled with existing computational methods.

Obviously computational drug discovery methods work in some cases. Obviously computational drug discovery methods do not work in all cases. One approach to solving this problem is to improve the methods; while that process continues, should we not also work to find the cases for which the methods work? One captivating feature of this paradigm shift is that it minimizes the need for improved ranking functions, which, as discussed above, is the part of drug modeling in which the field has made the least progress.

The Sali group [30, 61] presented a project in which they let the available pharmacopoeia (FDA-approved compounds in DrugBank) be the driving force to choosing the organism and protein to target. Specifically, they started with 10 disease-associated genomes, modeled as many of the proteins as feasible with template-based modeling, predicted protein–ligand matches by ligand mapping from template proteins, analogized protein–ligand matches to protein–drug matches by QSAR analysis, and finally ran four protein–drug pairs that appeared promising and relevant; three of the four demonstrate specific reversible binding [30] (Table 13.1). In abstraction, the project used only the best-scoring predictions of full modeling on a widely cast net. While the analysis already done in this work may hold other therapeutically relevant hits or leads, it already represents evidence that bolstering computational predictions over many possible targets can be expected to be productive if the decisions are made by the scoring functions.

13.7 SUMMARY

Incurable or untreatable diseases comprise a salient group of applications for computational drug discovery. Etiologies for incurable diseases include pathogens (e.g., acquired immunodeficiency syndrome, ebola, polio, human papilloma virus), neoplasms (i.e., cancers), genetic abnormalities (e.g. Down, Creutzfeldt–Jakob, and Proteus syndromes), autoimmunity (e.g., lupus erythematosus, asthma, multiple sclerosis), and inappropriate response to environment (e.g., prions, type 2 diabetes mellitus). Of those for which treatment exists, therapy manages symptoms but does not remove recurrence of disease on cessation of treatment (e.g., treatment of AIDS). Many life-threatening diseases have no treatment whatsoever. The motivation for computational approaches to drug discovery is to spur the bench and clinical studies to find cures for all diseases and alleviate human suffering. Amid these great successes in pharmacological discovery, it is important to consider that cures exist for many chronic and opportunistic diseases in the form of proper preventive behaviors (e.g., diet, exercise, hygiene), for which psychology is perhaps a more relevant solution than pharmacology.

The opportunity addressed by computational techniques is to abstract the knowledge from the many instances of physiologic interactions chronicled over the past century, to the clinical situations that plague humanity. The links that allow these abstractions are the genetic code, which helps us find the most relevant instances, and the structural models which help us predict how the interactions will occur.

Our research group, the groups of Shoichet, Freire, Becker, Avery, Sali, and others, have demonstrated the early maturity of computational modeling of protein–ligand interactions by predicting compounds for desired pharmacological activity and testing them in prospective experiments. These methods not only save time and resources but are also becoming more accurate than *in vitro* screening methods [6,47,123] (Table 13.1).

POSTSCRIPT

It should be noted that some authors, including ourselves, often discuss the goal of drug discovery only in the context of inhibitors. However, pharmacological activators are desired, particularly for nonpathogenic ailments such as depression and pain, so all discussions of pharmacological inhibitors here and elsewhere should be understood to be generalized to all pharmacologically active substances. Meanwhile, depending on the target, it may be more difficult to design an activator (agonist) or inhibitor (antagonist); for example, the types of contacts and similarity to the physiologic substrate may be exploited differently by each.

REFERENCES

1. Fleming, A. (1929), On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae, *Br. J. Exp. Pathol. 10*, 226–236.
2. Fleming, A.; Voureka, A.; Kramer, I. R.; Hughes, W. H. (1950), The morphology and motility of Proteus vulgaris and other organisms cultured in the presence of penicillin, *J. Gen. Microbiol. 4*, 257–269.
3. Schneider, G.; Fechner, U. (2005), Computer-based de novo design of drug-like molecules, *Nature Rev Drug Discov. 4*, 649–663.
4. DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. (2003), The price of innovation: new estimates of drug development costs, *J Health Econ. 22*(2), 151–185.
5. Adams, C.; Brantner, V. (2006), Estimating the cost of new drug development: Is it really 802 million dollars? *Health Aff. (*Millwood) *25*(2), 420–428.
6. Jenwitheesuk, E.; Horst, J. A.; Rivas, K. L.; Van Voorhis, W. C.; Samudrala, R. (2008), Novel paradigms for drug discovery: Computational multitarget screening, *Trends Pharmacol. Sci. 29*(2), 62–71.
7. Costin, J. M.; Jenwitheesuk, E.; Lok, S.-M.; Hunsperger, E.; Conrads, K. A.; Fontaine, K. A.; Rees, C. R.; Rossmann, M. G.; Isern, S.; Samudrala, R.; Michael, S. F. (2010), Structural optimization and de novo design of dengue virus entry inhibitory peptides, *PLoS Negl. Trop. Dis. 4*, e721.
8. Jenwitheesuk, E.; Samudrala, R. (2007), Identification of potential HIV-1 targets of minocycline, *Bioinformatics 23*(20), 2797–2799.
9. Keiser, M. J.; Setola, V.; Irwin, J. J.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijer, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. (2009), Predicting new molecular targets for known drugs, *Nature 462*, 175–181.
10. Rogawski, M. A. (2000), Low affinity channel blocking (uncompetitive) NMDA receptor antagonists as therapeutic agents—towards an understanding of their favorable tolerability, *Amino Acids 19*, 133–149.
11. Nezami, A.; Kimura, T.; Hidaka, K.; Kiso, A.; Liu, J.; Kiso, Y.; Goldberg, D. E.; Freire, E. (2003), High-affinity inhibition of a family of Plasmodium falciparum proteases by a designed adaptive inhibitor, *Biochemistry 42*(28), 8459–8464.
12. Csermely, P.; Agoston, V.; Pongor, S. (2005), The efficiency of multi-target drugs: The network approach might help drug design, *Trends Pharmacol. Sci. 26*, 187–182.

13. Samudrala, R.; Jenwitheesuk, E. (2007), Identification of potential HIV-1 targets of minocycline, *Bioinformatics 23*, 2797–2799.
14. de Vos, P. (2010), European materia medica in historical texts: Longevity of a tradition and implications for future use, *J. Ethnopharmacol. 132*(1), 28–47.
15. Campbell, J.; Campbell, J.; David, R. (2005), An insight into the practice of pharmacy in ancient Egypt, *Pharm. Hist. 35*(4), 62–68.
16. Manniche, L., *An Ancient Egyption Herbal,* British Museum Publications Ltd., London, 1989.
17. Huff, T., *The Rise of Early Modern Science: Islam, China, and the West,* Cambridge University Press, 2003.
18. Hamilton, G. R.; Baskett, T. F. (2000), In the arms of Morpheus the development of morphine for postoperative pain relief, *Can. J. Anaesth. 47*(4), 367–374.
19. Ehrlich, P. (1910), Die behandlung der syphilis mit dem ehrlichschen präparat 606, *Dsch. Med. Wochenschrift* 1893–1896.
20. Lloyd, N. C.; Morgan, H. W.; Nicholson, B. K.; Ronimus, R. S. (2005), The composition of Ehrlich's salvarsan: Resolution of a century-old debate, *Angew. Chem. Intnatl. Ed. Engl. 44*(6), 941–4.
21. Strebhardt, K.; Ullrich, A. (2008), Paul Ehrlich's magic bullet concept: 100 years of progress, *Nature Rev. Cancer 8*(6), 473–480.
22. Plouffe, D.; Brinker, A.; McNamara, C.; Henson, K.; Kato, N.; Kuhen, K.; Nagle, A.; Adrián, F.; Matzen, J. T.; Anderson, P.; Nam, T. G.; Gray, N. S.; Chatterjee, A.; Janes, J.; Yan, S. F.; Trager, R.; Caldwell, J. S.; Schultz, P. G.; Zhou, Y.; Winzeler, E. A. (2008), In silico activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen, *Proc. Natl. Acad. Sci. USA 105*, 9059–9064.
23. Chong, C. R.; Chen, X.; Shi, L.; Liu, J. O.; Sullivan, D. J., Jr., (2006), A clinical drug library screen identifies astemizole as an antimalarial agent, *Nature Chem. Biol. 2*, 415–416.
24. Weisman, J. L.; Liou, A. P.; Shelat, A. A.; Cohen, F. E.; Guy, R. K.; DeRisi, J. L. (2006), Searching for new antimalarial therapeutics amongst known drugs, *Chem. Biol. Drug Design 67*, 409–416.
25. Schreiber, S.L. (2000), Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science, 287*(5460), 1964–1969.
26. Abdi, M. H.; Beswick, P. J.; Billinton, A.; Chambers, L. J.; Charlton, A.; Collins, S. D.; Collis, K. L.; Dean, D. K.; Fonfria, E.; Gleave, R. J.; Lejeune, C. L.; Livermore, D. G.; Medhurst, S. J.; Michel, A. D.; Moses, A. P.; Page, L.; Patel, S.; Roman, S. A.; Senger, S.; Slingsby, B.; Steadman, J. G.; Stevens, A. J.; Walter, D. S. (2010), Discovery and structure-activity relationships of a series of pyroglutamic acid amide antagonists of the P2X7 receptor, *BioMed Central Lett. 20*, 5080–5084.
27. Cammarata, A.; Menon, G. K. (1976), Pattern recognition. Classification of therapeutic agents according to pharmacophores, *J. Med. Chem. 19*(6), 739–748.
28. Hansch, C. (1969), Quantitative approach to biochemical structure-activity relationships, *Acc. Chem. Res. 2*(8), 232–239.
29. Cupido, T.; Tulla-Puche, J.; Spengler, J.; Albericio, F. (2007), The synthesis of naturally occurring peptides and their analogs, *Curr. Opin. Drug. Discov. Devel. 10*(6), 768–83.
30. Ortí, L.; Carbajo, R. J.; Pieper, U.; Eswar, N.; Maurer, S. M.; Rai, A. K.; Taylor, G.; Todd, M. H.; Pineda-Lucena, A.; Sali, A.; Marti-Renom, M. A. (2009), A kernel for open source drug discovery in tropical diseases, *PLoS Negl. Trop. Dis. 3*(4), e418.

31. Borisy, A. A.; Elliott, P. J.; Hurst, N. W.; Lee, M. S.; Lehar, J.; Price, E. R.; Serbedzija, G.; Zimmermann, G. R.; Foley, M. A.; Stockwell, B. R.; Keith, C. T. (2003), Systematic discovery of multicomponent therapeutics, *Proc. Natl. Acad. Sci. USA 100*: 7977–7982.

32. Becker, O. M.; Marantz, Y.; Shacham, S.; Inbal, B.; Heifetz, A.; Kalid, O.; Bar-Haim, S.; Warshaviak, D.; Fichman, M.; Noiman S. (2004), G protein-coupled receptors: In silico drug discovery in 3D, *Proc. Natl. Acad. Sci. USA 101*(31), 11304–11309.

33. Natoli, T. A.; Smith, L. A.; Rogers, K. A.; Wang, B.; Komarnitsky, S.; Budman, Y.; Belenky, A.; Bukanov, N. O.; Dackowski, W. R.; Husson, H.; Russo, R. J.; Shayman, J. A.; Ledbetter, S. R.; Leonard, J. P.; Ibraghimov-Beskrovnaya, O. (2010), Inhibition of glucosylceramide accumulation results in effective blockade of polycystic kidney disease in mouse models, *Nature Med. 16*(7), 788–792.

34. Jensen, R. A. (1976), Enzyme recruitment in evolution of new function, *Annu. Rev. Microbiol. 30*, 409–425.

35. Khersonsky, O.; Roodveldt, C.; Tawfik, D. S. (2006), Enzyme promiscuity: Evolutionary and mechanistic aspects, *Curr. Opin. Chem. Biol. 10*(5), 498–508.

36. Horst, J. A.; Wang, K.; Horst, O. V.; Cunningham, M. L.; Samudrala, R. (2010), Disease risk of missense mutations using structural inference from predicted function, *Curr. Protein Pept. Sci. 11*(7), 500–515.

37. Jenwitheesuk, E.; Wang, K.; Mittler, J. E.; Samudrala, R. (2005), PIRSpred: A web server for reliable HIV-1 protein-inhibitor resistance/susceptibility prediction, *Trends Microbiol. 13*(4), 150–151.

38. Ekins, S. (2005), Systems-ADME/Tox: Resources and network approaches, *J. Pharmacol. Toxicol. Meth. 53*(1), 38–66.

39. Bellmann-Sickert, K.; Beck-Sickinger, A. G. (2010), Peptide drugs to target G protein-coupled receptors, *Trends Pharmacol. Sci. 9*, 434–441.

40. Noble, M. E.; Endicott, J. A.; Johnson, L. N. (2004), Protein kinase inhibitors: insights into drug design from structure, *Science. 303*(5665), 1800–1805.

41. Becker, O. M.; Dhanoa, D. S.; Marantz, Y.; Chen, D.; Shacham, S.; Cheruku, S.; Heifetz, A.; Mohanty, P.; Fichman, M.; Sharadendu, A.; Nudelman, R.; Kauffman, M.; Noiman, S. (2006), An integrated in silico 3D model-driven discovery of a novel, potent, and selective amidosulfonamide 5-HT1A agonist (PRX-00023) for the treatment of anxiety and depression, *J. Med. Chem. 49*(11), 3116–3135.

42. Desai, P. V.; Patny, A.; Sabnis, Y.; Tekwani, B.; Gut, J.; Rosenthal, P.; Srivastava, A.; Avery, M. (2004), Identification of novel parasitic cysteine protease inhibitors using virtual screening. 1. The ChemBridge database, *J. Med. Chem. 47*(26), 6609–6615.

43. Desai, P. V.; Patny, A.; Gut, J.; Rosenthal, P. J.; Tekwani, B.; Srivastava, A.; Avery, M. (2006), Identification of novel parasitic cysteine protease inhibitors by use of virtual screening. 2. The available chemical directory, *J. Med. Chem. 49*(5), 1576–1584.

44. de Paulis, T.; Reinhard, J. F., Jr., ; Oshana, S.; Kauffman, M.; Donahue, S. (2007), Tolerability, pharmacokinetics, and neuroendocrine effects of PRX-00023, a novel 5-HT1A agonist, in healthy subjects *J. Clin. Pharmacol. 47*, 817–824.

45. Rickels, K.; Mathew, S.; Banov, M. D.; Zimbroff, D. L.; Oshana, S.; Parsons, E. C. Jr.; Donahue, S. R.; Kauffman, M.; Iyer, G. R.; Reinhard, J. F. Jr., (2008), Effects of PRX-00023, a novel, selective serotonin 1A receptor agonist on measures of anxiety and

depression in generalized anxiety disorder: Results of a double-blind, placebo-controlled trial, *J. Clin. Psychopharmacol. 28*, 235–239.

46. Hirschel, B.; Francioli, P. (1998), Progress and problems in the fight against AIDS, *New Engl. J. Med. 338*(13), 906–908.
47. Doman, T. N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly, D. T.; Shoichet, B. K. (2002), Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B, *J. Med. Chem. 45*(11), 2213–2221.
48. Chong, C. R.; Chen, X.; Shi, L.; Liu, J. O.; Sullivan, D. J., Jr., (2006), A clinical drug library screen identifies astemizole as an antimalarial agent, *Nat. Chem. Biol. 2*(8), 415–416.
49. Weisman, J. L.; Liou, A. P.; Shelat, A. A.; Cohen, F. E.; Guy, R. K.; DeRisi, J. L. (2006), Searching for new antimalarial therapeutics amongst known drugs, *Chem. Biol. Drug. Des. 67*(6), 409–416.
50. Babaoglu, K.; Simeonov, A.; Irwin, J. J.; Nelson, M. E.; Feng, B.; Thomas, C. J.; Cancian, L.; Costi, M. P.; Maltby, D. A.; Jadhav, A.; Inglese, J.; Austin, C. P.; Shoichet, B. K. (2008), Comprehensive mechanistic analysis of hits from high-throughput and docking screens against beta-lactamase, *J. Med. Chem. 51*(8), 2502–2511.
51. Kato, N.; Sakata, T.; Breton, G.; Le Roch, K. G.; Nagle, A.; Andersen, C.; Bursulaya, B.; Henson, K.; Johnson, J.; Kumar, K. A.; Marr, F.; Mason, D.; McNamara, C.; Plouffe, D.; Ramachandran, V.; Spooner, M.; Tuntland, T.; Zhou, Y.; Peters, E. C.; Chatterjee, A.; Schultz, P. G.; Ward, G. E.;, Gray, N.; Harper, J.; Winzeler, E. A. (2008), Gene expression signatures and small-molecule compounds link a protein kinase to Plasmodium falciparum motility, *Nat. Chem. Biol. 4*(6), 347–356.
52. Gamo, F. J.; Sanz, L. M.; Vidal, J.; de Cozar, C.; Alvarez, E.; Lavandera, J. L.; Vanderwall, D. E.; Green, D. V.; Kumar, V.; Hasan, S.; Brown, J. R.; Peishoff, C. E.; Cardon, L. R.; Garcia-Bustos, J. F. (2010), Thousands of chemical starting points for antimalarial lead identification, *Nature. 465*(7296), 305–310.
53. Sarmiento, M.; Wu, L.; Keng, Y. F.; Song, L.; Luo, Z.; Huang, Z.; Wu, GZ.; Yuan, AK.; Zhang, Z. Y. (2000), Structure-based discovery of small molecule inhibitors targeted to protein tyrosine phosphatase 1B, *J. Med. Chem. 43*(2), 146–155.
54. Grzybowski, B. A.; Ishchenko, A. V.; Kim, C. Y.; Topalov, G.; Chapman, R.; Christianson, D. W.; Whitesides, G. M.; Shakhnovich, E. I. (2002), Combinatorial computational method gives new picomolar ligands for a known enzyme, *Proc. Natl. Acad. Sci. U. S. A. 99*(3), 1270–1273.
55. Nezami, A.; Luque, I.; Kimura, T.; Kiso, Y.; Freire, E. (2002), Identification and characterization of allophenylnorstatine-based inhibitors of plasmepsin II, an antimalarial target, *Biochemistry. 41*(7), 2273–2280.
56. Nezami, A.; Kimura, T.; Hidaka, K.; Kiso, A.; Liu, J.; Kiso, Y.; Goldberg, DE.; Freire, E. (2002), High-affinity inhibition of a family of Plasmodium falciparum proteases by a designed adaptive inhibitor. *Biochemistry*. *42*(28), 8459–8464.
57. Desai, PV.; Patny, A.; Sabnis, Y.; Tekwani, B.; Gut, J.; Rosenthal, P.; Srivastava, A.; Avery, M.; (2008), Identification of novel parasitic cysteine protease inhibitors using virtual screening. 1. The ChemBridge database. *J. Med. Chem. 47*(26), 6609–6615.

58. Desai, P. V.; Patny, A.; Gut, J.; Rosenthal, P. J.; Tekwani, B.; Srivastava, A.; Avery, M.; (2003), Identification of novel parasitic cysteine protease inhibitors by use of virtual screening. 2. The available chemical directory. *J. Med. Chem. 49*(5), 1576–1584.

59. Babaoglu, K,; Simeonov, A.; Irwin, J. J.; Nelson, M. E.; Feng, B.; Thomas, C. J.; Cancian, L.; Costi, M. P.; Maltby, D. A.; Jadhav, A.; (2008), Inglese J, Austin C. P, Shoichet BK. Comprehensive mechanistic analysis of hits from high-throughput and docking screens against beta-lactamase. *J. Med. Chem. 51*(8), 2502–2511.

60. Jenwitheesuk, E.; Horst, J. A.; Rivas, K. L.; Van, Voorhis, W. C.; Samudrala, R., (2008), Novel paradigms for drug discovery: computational multitarget screening. *Trends. Pharmacol. Sci. 29*(2), 62–71.

61. L. Orti, R. J.; Carbajo, U.; Pieper, N.; Eswar, S. M.; Maurer, A. K.; Rai, G.; Taylor, M. H.; Todd, A.; Pineda-Lucena, A.; Sali, M.A.; Marti-Renom, (2009), A kernel for the Tropical Disease Initiative. *Nat. Biotechnol.* 27, 320–321.

62. Costin, J. M.; Jenwitheesuk, E.; Lok, S. M.; Hunsperger, E.; Conrads, K. A.; Fontaine, K. A.; Rees, C. R.; Rossmann, M. G.; Isern, S.; Samudrala, R.; Michael, S. F.; (2010), Structural optimization and de novo design of dengue virus entry inhibitory peptides. *PLoS. Negl. Trop. Dis. 4*(6), e721.

63. Moult, J. (2005), A decade of CASP: Progress, bottlenecks and prognosis in protein structure prediction, *Curr. Opin. Struct. Biol. 15*(3), 285–289.

64. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. (2004), Docking and scoring in virtual screening for drug discovery: Methods and applications, *Nature Rev. Drug Discov. 3*, 935–949.

65. Bernard, B.; Samudrala, R. (2009), A generalized knowledge-based discriminatory function for biomolecular interactions, *Proteins 76*: 115–128.

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

67. Koshland, D. E. (1958), Application of a theory of enzyme specificity to protein synthesis, *Proc. Natl. Acad. Sci. USA 44*(2), 98–104.

68. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. (1982), A geometric approach to macromolecular-ligand interactions, *J. Mol. Biol. 161*, 269–288.

69. Plewczynski, D.; Laźniewski, M.; Augustyniak, R.; Ginalski K. (2010), Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database, *J. Comput. Chem.* 4, 742–755.

70. Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. (2001), DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases *J. Comput.-Aid. Mol. Design 15*, 411–428.

71. Lennard-Jones, J. E. (1924), On the determination of molecular fields, *Proc. Roy. Soc. Lond. A 106*(738), 463–477.

72. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr., ; Weiner, P. J. (1984), A new force field for molecular mechanical simulation of nucleic acids and proteins, *J. Am. Chem. Soc. 106*, 765–784.

73. Ponder, J. W.; Case, D. A. (2003), Force fields for protein simulations, *Adv. Protein Chem. 66*, 27–85.

74. Case, D. A.; Cheatham, T. E. 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. (2005), The Amber biomolecular simulation programs, *J. Comput. Chem. 26*(16), 1668–1688.

75. Jenwitheesuk, E.; Samudrala, R. (2003), Identifying inhibitors of the SARS coronavirus proteinase, *BioMed Central Lett. 13*(22), 3989–3992.

76. Ngan, S.-C.; Inouye, M.; Samudrala, R. (2006), A knowledge-based scoring function based on residue triplets for protein structure prediction, *Protein Eng. Design Select. 19*, 187–193.

77. Allen, F. H. (2002), The Cambridge Structural Database: A quarter of a million crystal structures and rising, *Acta Crystallogr. B. 58*(3), 380–388.

78. Fan, H.; Irwin, J. J.; Webb, B. M.; Klebe, G.; Shoichet, B. K.; Sali, A. (2009), Molecular docking screens using comparative models of proteins, *J. Chem. Inform. Model. 49*(11):2512–2527.

79. Abdel-Rahman, H. M.; Kimura, T.; Hidaka, K.; Kiso, A.; Nezami, A.; Freire, E.; Hayashi, Y.; Kiso, Y. (2004), Design of inhibitors against HIV, HTLV-I, and Plasmodium falciparum aspartic proteases *Biol. Chem. 385*(11), 1035–1039.

80. Nezami, A.; Laque, I.; Kimura, T.; Kiso, Y.; Freire, E. (2002), Identification and characterization of allophenylnorstatine-based inhibitors of plasmepsin II, an antimalarial target, *Biochemistry 41*, 2273–2280.

81. Gold, N. D.; Jackson, R. M. (2006), SitesBase: Fold independent structural comparisons of protein-ligand binding sites for exploring functional relationships, *J. Mol. Biol. 355*, 1112–1124.

82. Das, S.; Kokardekar, A.; Breneman, C. M. (2009), Rapid comparison of protein binding site surfaces with property encoded shape distributions, *J. Chem. Inform. Model. 49*, 2863–2872.

83. Weill, N.; Rognan, D. (2009), Development and validation of a novel protein-ligand fingerprint to mine chemogenomic space: Application to G protein-coupled receptors and their ligands, *J. Chem. Inform. Model. 49*(4), 1049–1062.

84. Chen, B. Y.; Honig, B. (2010), VASP: A volumetric analysis of surface properties yields insights into protein-ligand binding specificity, *PLoS Comput. Biol. 6*, e1000881.

85. Wang, K.; Samudrala, R. (2006), Incorporating background frequency improves entropy-based residue conservation measures, *BioMed Central Bioinform. 7*, 385.

86. Horst, J. A.; Samudrala, R. (2010), A protein sequence meta-functional signature for calcium binding residue prediction, *Pattern Recogn. Lett. 31*(14), 2103–2112.

87. Greaves, R.; Warwicker, J. (2005), Active site identification through geometry-based and sequence profile-based calculations: Burial of catalytic clefts, *J. Mol. Biol. 349*(3), 547–557.

88. Gutteridge, A.; Thornton, J. M. (2005), Understanding nature's catalytic toolkit, *Trends Biochem. Sci. 30*, 622–629. Grzybowski, B. A.; Ishchenko, A. V.; Kim, C. Y.; Topalov, G.; Chapman, R.; Christianson, D. W.; Whitesides, G. M.; Shakhnovich, E. I. (2002), Combinatorial computational method gives new picomolar ligands for a known enzyme, *Proc. Natl. Acad. Sci. USA 99*(3), 1270–1273.

89. Lopez, G.; Ezkurdia, I.; Tress, M. L. (2009), Assessment of ligand binding residue predictions in CASP8, *Proteins 77* (Suppl. 9), 138–146.

90. Roy, A.; Kucukural, A.; Zhang Y. (2010), I-TASSER: A unified platform for automated protein structure and function prediction, *Nature Protoc. 5*, 725–738.

91. Berezin, C.; Glaser, F.; Rosenberg, Y.; Paz, I.; Pupko, T.; Fariselli, P.; Casadio, R.; Ben-Tal, N. (2004), ConSeq: The identification of functionally and structurally important residues in protein sequences, *Bioinformatics 20*, 1322–1324.
92. Fischer, J. D.; Mayer, C. E.; Söding, J. (2008), Prediction of protein functional residues from sequence by probability density estimation, *Bioinformatics 24*(5), 613–620.
93. Wang, K.; Horst, J. A.; Cheng, G.; Nickle, D. C.; Samudrala, R. (2008), Protein meta-functional signatures from combining sequence, structure, evolution, and amino acid property information. *PLoS. Comput. Biol. 4*(9), e1000181.
94. Landau, M.; Mayrose, I.; Rosenberg, Y.; Glaser, F.; Martz, E.; Pupko, T.; Ben-Tal, N. (2005), ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures, *Nucl. Acids Res. 33*, W299–W302.
95. Ashworth, J.; Havranek, J. J.; Duarte, C. M.; Sussman, D.; Monnat, R. J, Jr.; Stoddard, B. L.; Baker, D., (2006), Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature. 441*(7093), 656–659.
96. Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Ršthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas C. F., 3rd, Hilvert, D.; Houk, K. N.; Stoddard, BL.; Baker, D, (2008), De novo computational design of retro-aldol enzymes. *Science. 319*(5868), 1387–1391.
97. Zhang, Y. (2008), I-TASSER server for protein 3D structure prediction, *BioMed Central. Bioinform. 9*, 40.
98. Zhang, Y.; Devries, M. E.; Skolnick, J. (2006), Structure modeling of all identified G protein-coupled receptors in the human genome, *PLoS Comput. Biol. 2*(2), e13.
99. Liu, T.; Horst, J. A.; Samudrala, R. (2009), A novel method for predicting and using distance constraints of high accuracy for refining protein structure prediction, *Proteins 77*, 220–234.
100. Michino, M.; Abola, E; Brooks, C. L., 3rd, Dixon, J. S.; Moult, J.; Stevens, R. C. GPCR Dock 2008 participants; (2009), Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008, *Nature Rev. Drug Discov. 8*(6), 455–463.
101. Hopkins, A. L. (2008), Network pharmacology: The next paradigm in drug discovery, *Nature Chem. Biol. 4*: 682–690.
102. Kohanski, M. A.; Dwyer, D. J.; Collins, J. J.; How antibiotics kill bacteria: from targets to networks. *Nat Rev Microbiol. 8*(6), 423–35.
103. Dancík, V.; Seiler, K. P.; Young, D. W.; Schreiber, S. L.; Clemons, P. A. (2010), Distinct biological network properties between the targets of natural products and disease genes, *J. Am. Chem. Soc. 132*(27), 9259–9261.
104. Newman, D. J.; Cragg, G. M. (2007), Natural products as sources of new drugs over the last 25 years, *J. Nat. Prod. 70*(3), 461–477.
105. Corey, E. J.; Long, A. K.; Rubenstein, S. D.; (1985) Computer-assisted analysis in organic synthesis. *Science. 228*(4698), 408–418.
106. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., (1997), Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 2001 *Mar.* 46(1–3):3–26.
107. Gamo, F. J.; Sanz, L. M.; Vidal, J.; de Cozar, C.; Alvarez, E.; Lavandera, J. L.; Vanderwall, D. E.; Green, D. V.; Kumar, V.; Hasan, S.; Brown, J. R.; Peishoff, C. E.; Cardon, L. R.;

Garcia-Bustos J. F. (2010), Thousands of chemical starting points for antimalarial lead identification, *Nature 465*(7296), 305–310.

108. Hidvegi, T.; Ewing, M.; Hale, P.; Dippold, C.; Beckett, C.; Kemp, C.; Maurice, N.; Mukherjee, A.; Goldbach, C.; Watkins, S.; Michalopoulos, G.; Perlmutter, D. H. (2010), An autophagy-enhancing drug promotes degradation of mutant alpha1-antitrypsin Z and reduces hepatic fibrosis, *Science 329*(5988), 229–232.
109. Jenwitheesuk, E.; Samudrala, R. (2003), Improved prediction of HIV-1 protease-inhibitor binding energies by molecular dynamics simulations, *BioMed Centeral Structural Biology, 3*, 2.
110. Jenwitheesuk, E.; Samudrala, R. (2005), Identification of potential multitarget antimalarial drugs, *JAMA 294*(12), 1490–1491.
111. Xie, L.; Wang, J.; Bourne, P. E. (2007), In silico elucidation of the molecular mechanism defining the adverse effect of selective estrogen receptor modulators, *PLoS Comput. Biol. 3*, e217.
112. Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, BK.; (2007), Relating protein pharmacology by ligand chemistry. *Nat Biotechnol. 25*(2), 197–206.
113. Eckert, H.; Bajorath, J. (2007), Molecular similarity analysis in virtual screening: Foundations, limitations and novel approaches, *Drug Discov. Today 12*, 225–233.
114. Xie, L.; Bourne, P. E. (2005), Functional coverage of the human genome by existing structures, structural genomics targets, and homology models, *PLoS Comput. Biol. 1*, e31.
115. Walian, P.; Cross, T. A.; Jap. B. K.; (2004), Structural genomics of membrane proteins. *Genome Biol. 5*(4), 215.
116. Gudermann, T.; N Ÿrnberg, B.; Schultz, G. (1995), Receptors and G proteins as primary components of transmembrane signal transduction. Part 1. G-protein-coupled receptors: structure and function. *J. Mol Med. 73*(2), 51–63.
117. UniProt, Consortium.; (2007) The universal protein resource (UniProt). *Nucleic Acids Res.* 36(Database issue), D190-5.
118. Luque, I.; Freire, E. (2002), Structural parameterization of the binding enthalpy of small ligands, *Proteins 49*(2), 181–190.
119. Freire, E. (2009), A thermodynamic approach to the affinity optimization of drug candidates, *Chem. Biol. Drug Design 74*(5), 468–72.
120. Ladbury, J. E.; Klebe, G.; Freire, E. (2010), Adding calorimetric data to decision making in lead discovery: A hot tip, *Nature Rev. Drug Discov. 9*(1), 23–7.
121. Freire, E. (2008), Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov. Today 13*(19–20), 869–874.
122. Kawasaki, Y.; Chufan, E. E.; Lafont, V.; Hidaka, K.; Kiso, Y.; Mario Amzel, L.; Freire, E. (2010), How much binding affinity can be gained by filling a cavity? *Chem. Biol. Drug Design 75*(2), 143–151.
123. Babaoglu, K.; Simeonov, A.; Irwin, J. J.; Nelson, M. E.; Feng, B.; Thomas, C. J.; Cancian, L.; Costi, M. P.; Maltby, D. A.; Jadhav, A.; Inglese, J.; Austin, C. P.; Shoichet, B. K. (2008), Comprehensive mechanistic analysis of hits from high-throughput and docking screens against beta-lactamase, *J. Med. Chem. 51*(8), 2502–2511.
124. Deayton, J.; Mocroft, A.; Wilson, P.; Emery, V. C.; Johnson, M. A.; Griffiths, P. D. (1999), Loss of cytomegalovirus (CMV) viraemia following highly active antiretroviral therapy in the absence of specific anti-CMV therapy, *AIDS 13*(10), 1203–1206.

125. Reed, J. B.; Morse, L. S. (1998), CMV retinitis and the controversies associated with highly active antiretroviral therapy and the immune recovery hypothesis, *AIDS Patient Care STDs 12*(3), 181–185.

126. Wang, K.; Jenwitheesuk, E.; Samudrala, R.; Mittler, J. E. (2004), Simple linear model provides highly accurate genotypic predictions of HIV-1 drug resistance, *Antivir. Ther. 9*(3), 343–352.

127. Jenwitheesuk, E.; Wang, K.; Mittler, J. E.; Samudrala, R. (2004), Improved accuracy of HIV-1 genotypic susceptibility interpretation using a consensus approach, *AIDS 18*(13), 1858–1859.

128. Jenwitheesuk, E.; Samudrala, R. (2005), Virtual screening of HIV-1 protease inhibitors against human cytomegalovirus protease using docking and molecular dynamics, *AIDS 19*(5), 529–531.

129. Ehrlich, P. (1911), Aus theorie und praxis der chemotherapie, *Folia Serologica 7*, 697–714.

130. Kaelin, W. G., Jr., (2004), Gleevec: Prototype or outlier? *Sci STKE. 16* (225), pe12.

131. Peters, J. U.; Schnider, P.; Mattei, P.; Kansy, M. (2009), Pharmacological promiscuity: Dependence on compound properties and target specificity in a set of recent Roche compounds, *ChemMedChem 4*(4), 680–686.

132. Huang, S. (2002), Rational drug discovery: What can we learn from regulatory networks? *Drug Discov. Today 7*, S163–S169.