Chemistry & Biology
Review

Established and Emerging Trends in Computational Drug Discovery in the Structural Genomics Era

Olivier Taboureau,[1] Jonathan B. Baell,[2] Juan Fernández-Recio,[3] and Bruno O. Villoutreix[4,*]
[1]Center for Biological Sequences Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Kongens Lyngby, Denmark
[2]Walter & Eliza Hall Institute of Medical Research, Department of Medical Biology, University of Melbourne, Victoria 3010, Australia
[3]Joint BSC-IRB Research Programme in Computational Biology, Barcelona Supercomputing Center, 08034 Barcelona, Spain
[4]University Paris Diderot, Sorbonne Paris Cite, UMRS 973, MTi, F-75205 Paris, France
*Correspondence: bruno.villoutreix@inserm.fr
DOI 10.1016/j.chembiol.2011.12.007

Bioinformatics and chemoinformatics approaches contribute to hit discovery, hit-to-lead optimization, safety profiling, and target identification and enhance our overall understanding of the health and disease states. A vast repertoire of computational methods has been reported and increasingly combined in order to address more and more challenging targets or complex molecular mechanisms in the context of large-scale integration of structure and bioactivity data produced by private and public drug research. This review explores some key computational methods directly linked to drug discovery and chemical biology with a special emphasis on compound collection preparation, virtual screening, protein docking, and systems pharmacology. A list of generally freely available software packages and online resources is provided, and examples of successful applications are briefly commented upon.

Introduction

Significant efforts are being spent in the pharmaceutical industry and in academic groups to identify small organic hit compounds that could become drugs after simultaneous optimization of interrelated properties, such as bioactivity, absorption-distribution-metabolism-excretion-toxicity (ADMET), and the underlying physicochemical properties. In addition, small molecule probes are also needed to understand biological processes in the health and disease states. Phenotypic and target-based medium-to-high-throughput screening (HTS) of chemical libraries are currently the major technologies used to identify novel hit compounds (Campbell, 2010; Macarrón et al., 2011; Stockwell, 2004; Swinney and Anthony, 2011). Yet, other technologies and methodological developments contribute to the design of better and safer compounds, such as the different "omics" initiatives, next-generation sequencing with ultimately identification of new targets, biomarkers, and genetic variations with the resulting increased awareness of pharmacogenetic and epigenetic events (Kramer et al., 2007; Woollard et al., 2011). Computational approaches are being developed along with experimental technologies to analyze and integrate in vitro and in vivo data, to predict health-related events, and possibly to reduce the amount of experimental work that has to be performed to design a magic bullet. In silico methods can be used at the different stages of the drug discovery process and, for instance, can involve querying genomic data, running comparative genomics, investigating protein and peptide folding, defining protein interaction networks, analyzing the impact of point mutations, and assisting in clinical trial design, to name only a few (Mah et al., 2011; Pierri et al., 2010) (Fernald et al., 2011; Thusberg et al., 2011; Tsai et al., 2009; Tuncbag et al., 2011; Villoutreix, 2002; Woollard, 2010; Woollard et al., 2011). In the present study, we will review some computational approaches and databases that assist the discovery of hit compounds, with an emphasis on strategies used to prepare compound collections and on virtual screening and protein docking and their links with the design of protein-protein interaction modulators. We end the article by briefly discussing systems pharmacology and related off-target prediction methods because we envision that such global integrative approaches can lead to new therapeutic interventions while reducing the risk of failure. Applications of these concepts and methods are briefly illustrated with specific examples integrated in each paragraph.

Compound Collections and Some Related ADMET Considerations

It is now generally accepted that an important prerequisite for successful HTS and in silico screening lies in the use of a high-quality compound collection. The problem is that there is no consensus solution at present to design such an ideal compound library because of our still limited understanding of the complexity of living organisms and because it will obviously be depending on the project, the stage of the project and goals, and whether the intended outcome is a drug or a chemical probe (Workman and Collins, 2010). Lessons learned from the past can assist present decision making and definition of guidelines, keeping in mind that, as knowledge evolves, rules and strategies have to be questioned and revisited (Faller et al., 2011; Gedeck et al., 2010; Segall et al., 2009; Singh et al., 2011; Smith et al., 2010; Stepan et al., 2011). Screening collections of physically available compounds usually contain chemotypes that were synthesized for other projects or were purchased from external vendors. Although they may contain several millions of compounds, they still explore only a minute fraction of the chemical

space, because this one is estimated to be in the range of 10^{18}–10^{200} molecules (Dobson, 2004). Because assembling and maintaining a screening collection is expensive and because the chance of obtaining valuable hits is generally linked to the quality of the compounds, tremendous efforts have been spent in pharmaceutical companies and in academic laboratories to develop protocols facilitating the design of ''better and safer'' libraries while reducing the threat of missing possible active series (Baell, 2010; Bauer et al., 2010; Charifson and Walters, 2002; Davies et al., 2006; Drewry and Macarrón, 2010; Gleeson et al., 2011b; Hert et al., 2009; Hu et al., 2011; Macarrón and Luengo, 2011; Orry et al., 2006; Park et al., 2011; Pitt et al., 2009; Renner et al., 2011; Sperandio et al., 2010b; Stocks et al., 2009). Many chemical databases and compound collections are freely available online (Ekins and Williams, 2010; Ertl and Jelfs, 2007; Gozalbes and Pineda-Lucena, 2011; Richard et al., 2006; Villoutreix et al., 2007). They contain, for instance, collections collated from different vendors, virtual compounds, drugs and experimental molecules or databases of toxic molecules, molecules used in HTS experiments, or metabolites (Table S1 available online). There are also numerous in silico tools that can be used to design a compound collection (Table S1) (Villoutreix et al., 2007), and several general strategies have been recently proposed (Bologa et al., 2006; Brenk et al., 2008; Fourches et al., 2010; Hann, 2011; Huggins et al., 2011; Muchmore et al., 2010; Steinmeyer, 2006). These usually involve physicochemical property filtering based on empirical rules because these properties play many different roles (from compound handling and developability to oral bioavailability up to increased risk of toxicity) (Ganesan, 2008; Khanna and Ranganathan, 2009; Meanwell, 2011; O'Shea and Moser, 2008; Waring, 2009; Waring and Johnstone, 2007). The rule of five (Lipinski, 2000), recently revisited using pharmacokinetic data in rat (Ridder et al., 2011), is a well-known example, but many related rules have been proposed since then, such as the recent 3/75 rule that relates physicochemical properties to in vivo toxicity (Hughes et al., 2008). It is also beneficial when preparing a collection to remove or flag molecules carrying unwanted atoms and functional groups (Axerio-Cilies et al., 2009; Benigni and Bossa, 2011; Enoch et al., 2011; Erve, 2006; Kalgutkar et al., 2005; Kazius et al., 2005; Rishton, 2003). Compounds may be deemed less attractive because they bear toxicophores, such as nitro, aniline, hydantoin, and cyanohydrin, which are associated with metabolism-mediated toxicity. Alternatively, groups such as aldehydes and epoxides may be considered inappropriately electrophilic, whereas others such as thiols are redox active. It has also been recommended to pay attention to compounds such as PAINS that appear to nonspecifically interfere across many different assay formats (Baell and Holloway, 2010) and also to pay attention to unexpectedly electrophilic compounds as detected the ALARM NMR protocol (Huth et al., 2005). It is noted that many of these can be flagged with the recently reported FAF-Drugs2 online server (Lagorce et al., 2011). Some discretion may be required here with regard to assembling a library from commercially available compounds and ''structural alerts,'' because not only is the representation of the diversity space so incomplete, but also because the most rigorous filters may remove up to 98% of a given vendor's collection. In such cases, it may be better to include potential metabolic liabilities rather than to discard valuable diversity space. For example, hydrazones are easy to assemble and anilines are easy to substitute in the activated *ortho* and *para* positions. Both classes are associated with toxic metabolites, but it may be better to include such compounds in screening and simply undertake early isostere replacement in any discovered hit (such as an aminopyrazine for an aniline hit). Similarity and diversity of compounds to add to a collection can be evaluated as distances or coefficients, such as the well-known Tanimoto index, which computes the difference in one or several molecular properties (descriptors) between two compared molecules. The diversity of the compound libraries (Akella and DeCaprio, 2010) can be visualized with, for instance, ChemGPS-NP (Rosén et al., 2009), and graphical comparison of the shapes of the molecules can also be valuable for collection design (Akritopoulou-Zanze et al., 2007). Additional ADMET properties may need to be computed, for instance, before ordering new molecules, before synthesis, or during the hit-to-lead optimization phase. ADMET prediction models are extremely challenging to develop but can help bias medicinal chemistry into ''safer'' areas of the chemical space (Gleeson et al., 2011a; Smith, 2011; Stoll et al., 2011; Valerio, 2009). In silico tools include simple look-up tables, where the query compound is compared with a list of molecules with known experimental data using, for instance, fingerprint methods or other similarity search approaches. Alternatively, or in addition, QSAR models can be built to predict ADMET properties in which molecular structural features (e.g., descriptors) are correlated with observed biological activity by using regression or machine learning methods. Furthermore, direct structural information between a molecule and an ADMET target can be estimated by docking or pharmacophore modeling and can be combined with QSAR modeling (Ekins et al., 2009; Moroy et al., 2011), leading to mechanistic profiling. Where 3D ligand searches are involved, low-energy conformers can be produced ''on the fly'' or can be precalculated and stored as multiconformer databases, but there are several issues associated with 3D structure generation and conformational sampling (Foloppe and Chen, 2009) that we will not dwell on further in this review. Related difficulties include proper enumeration of stereoisomers, tautomers, and protomers, and this is significant because these can have an impact on ADMET predictions. It is possible to find several valuable commercial packages to prepare a collection and to perform ADMET predictions, such as Schrodinger's Ligprep, Openeye's Filter, or tools from MOE, Tripos, Accelrys, ICM, and Molecular Discovery, among many others, but, fortunately, free packages with source codes and online services are also available (Table S1).

Virtual Screening

The term ''virtual screening'' was first reported in the scientific literature in 1997 (Horvath, 1997); it can be defined as a set of computer methods that analyzes large databases or collections of compounds in order to identify likely hit candidates (Sotriffer, 2011b; Walters et al., 1998). This search can be performed on libraries that contain physically existing compounds or on virtual libraries, and thus on compounds that are not yet synthesized. These in silico experiments can be performed to complement HTS (and are indeed often integrated in screening campaigns), prior to experimental screening, or after HTS to rescue some

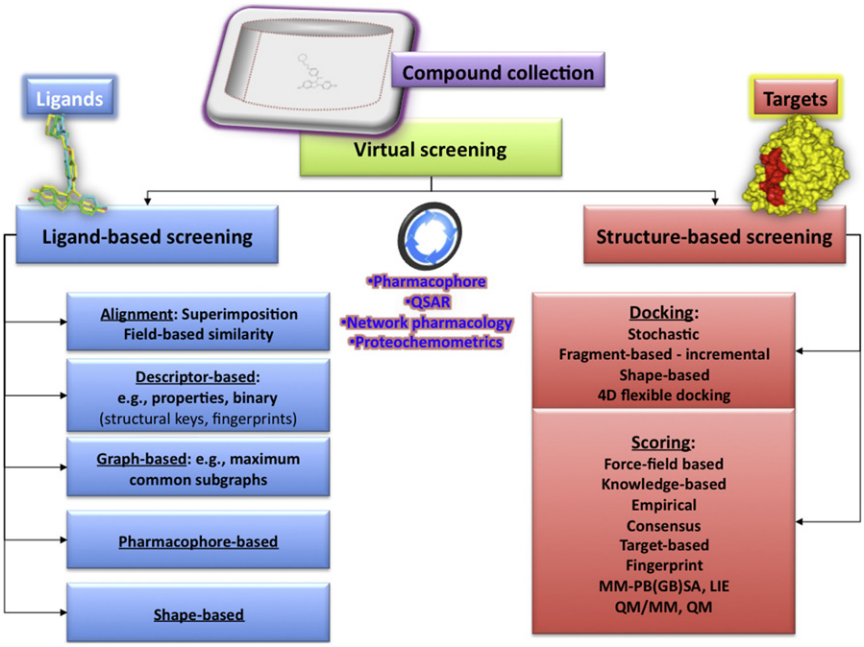

Figure 1. Examples of Virtual Screening Methods
Some approaches can be considered to be at the interface between the two main screening concepts, ligand-based and structure-based, such as pharmacophore modeling using information from cocrystallized target-ligand complexes or in the case of proteochemometric modeling.

compounds potentially missed by the in vitro readouts. The complementarity between HTS and virtual screening has been recently illustrated by screening both in silico and experimentally the same ~198,000-compound collection in an attempt to find new inhibitors for cruzain, a cysteine protease target for Chagas disease (Ferreira et al., 2010). Ultimately, 146 reversible, competitive inhibitors of cruzain belonging to five different classes were identified. Two classes were discovered through the HTS work (they were false negatives in the docking calculations). Another two classes were prioritized because of their favorable docking scores, and the final class was investigated on the basis of high docking scores and HTS results.

Virtual screening approaches have been traditionally subdivided into two main approaches: first, ligand-based screening, in which 2D or 3D chemical structures or molecular descriptors of known actives (and sometimes inactive molecules) are used to retrieve other compounds of interest from a database using some kind of similarity measure or by seeking a common substructure or pharmacophore between the query molecule and the compounds in the database; and second, structure-based (or receptor-based) screening in which compounds from the database are docked into a binding site (or over the entire surface) and are ranked using one or several scoring functions. The process can then be continued if deemed appropriate using different types of postprocessing approaches. The ligand- and structure-based methods can be combined if the necessary information is available. Virtual screening methods are relatively well established, and numerous success stories—in terms of hit identification, contribution to the development of drug candidates, or marketed products—have been recently reviewed (Clark, 2008; Ekins et al., 2007a, b; Köppen, 2009; Rester, 2008; Ripphausen et al., 2011; Ripphausen et al., 2010; Villoutreix et al., 2009; Zhong et al., 2007). This does not mean that the methods have no flaws but that they can help identifying interesting molecules.

Virtual Screening Based on Knowledge About the Bioactive Ligands

Ligand-based methods are, in general, based on the assumption (which is not always true) that similar structures have a greater-than-random chance of having similar biological activities (and the extended idea that states that similar targets usually interact with similar ligands). A multitude of techniques is available covering distinct levels of complexity and computational cost, from topological fingerprints to methods based on geometric or energy field representations of the molecular structures (Figure 1). Most commercial software companies provide ligand-based virtual screening application, such as FlexS, Catalyst, MOE, ROCS, LigandScout, Chemaxon, Daylight, FTtrees, ICM-chemist, and FieldScreen (see for instance Liao et al., 2011), and some (usually) freely available tools are listed in Table S2. Several recent reviews provide extensive description of the methods, performance, and possibility to escape the input chemotype (scaffold hopping) as well as success stories (Cross and Cruciani, 2010; Eckert and Bajorath, 2007; Geppert et al., 2010; Guido et al., 2008; Kirchmair et al., 2008; Reddy et al., 2007; Tropsha and Golbraikh, 2007; Wolber et al., 2008).

Among the different methods, users can find diverse alignment methods in which the detection of similarity is performed by superposing each of the test molecules of the database with a reference molecule (the query compound or compounds if several molecules are used), and the compounds are ranked according to their extent of similarity. Molecular electrostatic potential or other energy fields can also be aligned. Along the same line, shape similarity search algorithms can also align compounds on the basis of shape or of shape colored by atomic properties. There are many descriptor-based screening methods, such as structural keys, fingerprints, feature trees, and pharmacophore, to name only a few. Machine learning can be considered as a form of ligand-based virtual screening approach that builds predictive quantitative structure–activity relationship (QSAR) models based on available experimental data. There, compounds are described with various molecular descriptors to provide numerical representations of the compounds' properties. Popular machine learning techniques are then applied and include self-organizing maps, decision trees, multiple linear regression, artificial neural network, support vector machine approaches, k-nearest neighbors approach, and Bayesian methods (Gedeck et al., 2010; Geppert et al., 2010). Yet, with all these predictive models, the question of their

predictive reach has to be carefully investigated. Therefore, it is important to quantitatively assess the applicability domain of the model. It should also be mentioned that QSAR can be considered as a hybrid approach if, for example, the docking binding energy is integrated into the model (Garg et al., 2010). It should be noted that ligand-based screening may use either generated 3D ligand alignment, as mentioned above, or ligand-superposition directly inferred from the pocket superposition (Pérot et al., 2010).

Numerous ''success stories'' have been reported through the use of ligand-based screening approaches, such as the optimization of a lead compound toward the generation of Aggrastat, a fibrinogen receptor antagonist with anticoagulant properties (Clark, 2008). Many hit compounds have been found with ligand-based approaches, such as the discovery of nonglycoside sodium-dependent glucose cotransporter 2 (diabetes target) using a combination of different methods (pharmacophore with the DISCO/UNITY modules of the Sybyl package, shape-based with ROCS, OpenEye and structure clustering analysis with Discovery Studio, and Accelrys) (Wu et al., 2010) or the identification of orally active acetylcholinesterase inhibitors (Alzheimer disease) (Chaudhaery et al., 2010) using a Catalyst/HypoGen-based pharmacophore (Accelrys) or of new P-glycoprotein inhibitors (multidrug resistance) (Palmeira et al., 2011) by pharmacophore screening with PharmaGist (Schneidman-Duhovny et al., 2008).

Virtual Screening Using Structural Information from the Targets

The significant increase in the number of available structural data of macromolecules (X-ray, NMR, homology models, and the possibility to develop pseudoreceptor models) has accelerated the development of structure-based methods (Cavasotto and Phatak, 2009; Davis et al., 2008; Fan et al., 2009; Tanrikulu and Schneider, 2008; Wlodawer et al., 2008). Although the approaches were essentially used on catalytic sites, they are now more and more applied, with some tuning, to exosites, membrane protein targets, protein-membrane interactions, allosteric mechanisms, modulation of protein-protein interactions, and on nucleic acids (Arkin and Whitty, 2009; Dailey et al., 2009; Kufareva et al., 2011; Laine et al., 2010; Lang et al., 2009; Segers et al., 2007; Sperandio et al., 2010b; Talele et al., 2010; Tautermann, 2011; Tuccinardi, 2011). Just like with ligand-based methods, there are many commercial engines that perform structure-based screening, including ICM, Glide, Surflex, Gold, Molegro Virtual Docker, FlexX, FRED, eHITS, LEADFINDER, and LigandFit (see for instance, Liao et al., 2011 and McInnes, 2007), and there are also several packages that are freely available (the source code is sometimes provided) to academic groups and the private sector (examples in Table S3).

Structure-based screening is divided into two major steps, positioning the ligand into a binding pocket (the docking methods include rigid body docking, shape matching, incremental construction, or stochastic approaches), and the scoring or ranking usually fits into one of these approaches: force-field based, knowledge based, and empirical. The algorithms and protocols have been extensively reviewed elsewhere (Cavasotto and Orry, 2007; Moitessier et al., 2008; Reddy et al., 2007; Sperandio et al., 2006; Yuriev et al., 2011), and we will here briefly comment some recent studies about binding pockets, scoring, postprocessing, and receptor flexibility.

One of the major steps required in structure-based screening is the definition of a binding zone where the ligand will be positioned (unless the process is performed over the entire surface). The delineation of the pocket can have a significant impact on the output, and it is thus critical to analyze this region with care. The predictions are usually performed with geometric or energy-based methods, although some approaches use both. It can also be beneficial to put a druggability score on the identified binding pockets, such as to select the regions that have a greater chance of efficiently binding a small compound. Combining pocket prediction tools and druggability assessment is obviously of interest when screening an orphan target or when looking for an exosite. Furthermore, other computations over these binding regions can be performed like comparing binding pockets independently of the fold, such as to eventually get inspirations from closely related pockets that could already be cocrystallized with a small molecule inside. Many algorithms and methods to predict binding pocket, druggability, and pocket similarities have been reviewed or published recently, and several of them are freely available and are accessible as open source packages or as online servers (Abagyan and Kufareva, 2009; David-Eden et al., 2010; Durrant et al., 2011; Fauman et al., 2011; Huang, 2009; Huang and Jacobson, 2010; Pérot et al., 2010; Ren et al., 2010; Schmidtke and Barril, 2010; Sheridan et al., 2010; Spitzer et al., 2011; Thangudu et al., 2010) (see also the next paragraph and the recently released pocket databases, Kufareva et al., 2012). Most often, if the macromolecule is not too flexible, these predictions are possible and give important insights. In addition, if several experimental structures of the macromolecule are available, it could be important to run the computations on all the conformation, whereas if only one structure is known, simulations can be performed in order to explore further the cavities or even reveal transient pockets (Eyrisch and Helms, 2007). In silico identification of a binding pocket outside the catalytic site and structure-based screening (with Surflex, AutoDock, or Molegro MolDock Virtual Docker) of that region have been performed on the tyrosine kinase SYK (a target for allergic asthma and rhinitis) (Mazuc et al., 2008; Villoutreix et al., 2011). Similarly, a predicted binding pocket on APOBEC3C (inhibitory activity against retroviruses) was shown to interact with RNA, thereby shedding some light onto important molecular mechanisms (Stauch et al., 2009). Other screening protocols that are gaining momentum involve the rational design of multitarget drugs. Most often, the modes of action of these molecules are elucidated retrospectively but designing multitarget inhibitors with predefined biological profiles presents a real challenge. Wei et al. (2008) have developed a computer-assisted strategy to screen for multitarget inhibitors using a combination of molecular docking and common pharmacophore matching and have successfully applied this approach to design dual-target inhibitors against both the human leukotriene A4 hydrolase and the human nonpancreatic secretory phospholipase A2.

A next critical step is the docking-scoring phase, possibly in the context of a flexible receptor. Although prediction of the ligand-binding pose is usually possible with the available methods, scoring is still very challenging and it is thus difficult to identify the correct binding pose or to rank compounds. Some of the difficulties with scoring functions come from the fact that several terms and events are difficult to parameterize;

some molecular interactions are not considered, are not known, or are impossible to calibrate accurately; or the noncovalent energy terms are assumed to be additive. Several strategies have been proposed, all with strengths and weaknesses, to try to improve the process (Zhong et al., 2010). Rescoring can be performed with more rigorous scoring functions (de Ruiter and Oostenbrink, 2011; Mitchell and Matsumoto, 2011), using target-specific scoring functions (Seifert, 2009), consensus scoring (Charifson et al., 1999; Feher, 2006), consensus docking and consensus scoring (Lyne, 2002; Miteva et al., 2005), and protein-ligand interaction fingerprints (Brewerton, 2008). Ligand efficiency, kinetic efficiency, or drug efficiency index (depending on the amount of experimental data) evaluation can also assist the process (Holdgate and Gill, 2011; Kawasaki and Freire, 2011; Montanari et al., 2011; Tanaka et al., 2010), and so does, in some circumstances, the addition of water molecules (Huang and Shoichet, 2008; Huggins and Tidor, 2011). In many situations, however, consideration of receptor flexibility should, in theory, improve the results but, in practice, the computations usually become extremely demanding with a risk of diluting the selection of the good compounds because of flaws in the scoring functions. Yet, several recipes have been suggested to deal implicitly or explicitly with receptor flexibility during virtual screening experiments or prior to the runs and include soft potentials, side chain flexibility, local energy minimization, receptor ensemble-based methods (experimental structures or simulated ones), pocket fumigation, removal of some amino acid side chains, and ligand-guided pocket generation among others (Amaro et al., 2008; Bottegoni et al., 2011; B-Rao et al., 2009; Fukunishi, 2010; Ivetac and McCammon, 2011; Rueda et al., 2010; Sotriffer, 2011a; Sperandio et al., 2010a). A key question that is still under debate here is how to select the right subset of receptor conformations that will enhance the performance in a real life scenario. Strategies are still under investigation, yet some guidelines have started to emerge (Bottegoni et al., 2011; Rueda et al., 2010; Sperandio et al., 2010a).

Combining (here essentially understood as multistep protocols but not real integration of the methods) ligand-based and structure-based virtual screening strategies can have utility, as recently reported (Koutsoukas et al., 2011; Tan et al., 2010; Wilson and Lill, 2011). For instance, pharmacophore models derived from the receptor-binding site or from ligand-receptor complexes are being used. Traditionally, a combination of methods implies a sequential use of different methods with, for instance, structure-based screening, experimental validation of the hit list, and ligand-based screening for one round of optimization using the best hits as queries. Alternatively, ligand-based and structure-based scores can be fully integrated, for example, as in the discovery of novel cruzain inhibitors (Wiggers et al., 2011). Fragment-based screening approaches have been gaining momentum over these last 10 years, and both in silico ligand-based and structure-based engines can be used to assist these endeavors (Barelier and Krimm, 2011; Desjarlais, 2011; Zoete et al., 2009). Indeed, it was found that fragments could be docked accurately but that the difficulty was obviously the scoring step (Verdonk et al., 2011). In that same study, the authors also highlighted the role of ligand efficiency in docking performance.

Some Pitfalls of Virtual Screening

Virtual screening is in general more prevalent in academic and small company settings, as opposed to large industries, because it is cheaper than HTS, and, in several cases, is as efficient or more efficient. Often in academia, the most likely scenario is that virtual screening is performed and in vitro screening follows, but the selection of hits and the first round of optimization lack the input of trained medicinal chemists. Thus, if the compound discovered in silico is to be used for chemical biology, it might be of sufficient quality but if the intention is drug discovery, the wrong molecules can be meaninglessly optimized for years. Indeed, many publications mentioning in silico hits contain compounds that industry researchers recognize as being nonspecifically active. These reports taint the credibility of the approach overall and detract from those describing genuinely successful work (such observations can also apply to experimental work; Prinz et al., 2011). It is often not appreciated that virtual screening can discover false-positive results, just as HTS does. A random selection of around 100 vendor-supplied compounds has a good chance of producing an apparent active at the concentrations that are typically tested (low to mid micromolar). If this compound selection was instead defined from an in silico screen, the apparent link between cause and effect can be so coercive that due diligence is not applied, the results are written up, and the manuscript is published. This then becomes cited as an example of a virtual screening success and becomes a self-fulfilling cycle. In fact, most in silico screening groups have been confronted by this situation in the past, yet, things are now changing, in part benefiting from the recent infiltration of industry-savvy medicinal chemists into academic drug discovery and because multidisciplinary projects are becoming the main frame, involving the right expertise and several departments with complementary skills.

Computational Approaches to Predict Protein-Protein Interactions and Their Impact on the Design of Low-Molecular-Weight Modulators

One important challenge in drug discovery is to target macromolecular interaction, such as protein-protein interactions (PPIs). Proteins interact with other proteins and macromolecules to perform their living functions, so any pathological condition is likely to involve a specific interaction between two or more proteins. The use of peptides or therapeutic antibodies has been documented, suggesting that it is possible to target certain PPIs. However, one important challenge is to define strategies to develop orally available small druglike molecules that could modulate specific PPIs in a cost-effective manner (Sperandio et al., 2010b). There are many reasons that make the design difficult, such as the lack of structural information on the location of the interface, protein-protein interfaces are usually large and flat, the hit rate with experimental HTS or in silico screening is usually low, and the paucity of successful compounds to learn from. Clearly, there are several possible avenues to attempt to improve the design of PPI inhibitors, such as exploring a novel area of the chemical space, rationalization of allosteric mechanism, and so forth, but definitively, the prediction of likely binding zones in the interface areas and the modeling of large complexes should play an important role.

One important difficulty in designing PPI modulators is indeed the lack of experimental structural data of protein complexes. To

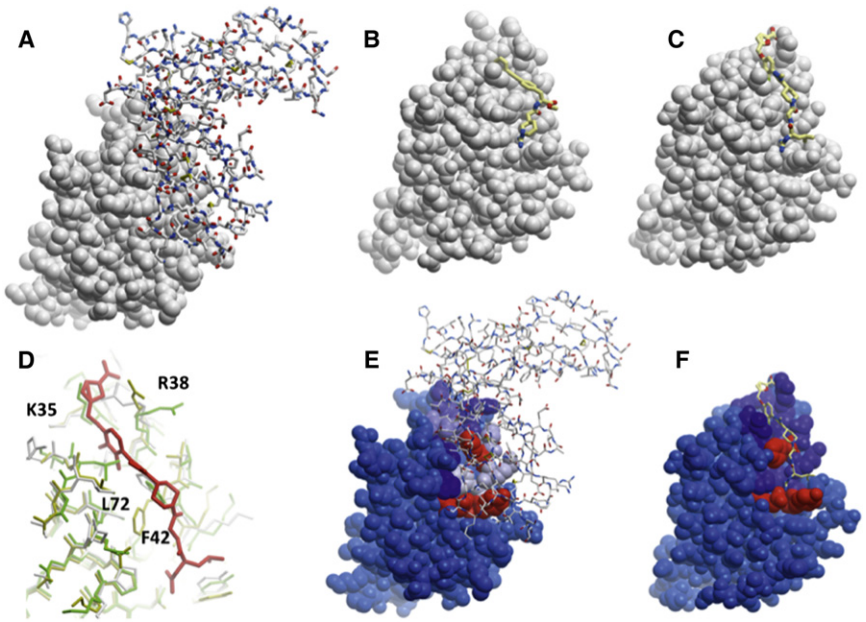

Figure 2. Example of Small Compounds Inhibiting Protein-Protein Association

(A) Structure of complex between IL-2 (white CPK) and its receptor IL-2Ra (ball & stick).

(B) IL-2 bound to its initially developed protein-protein inhibitor Ro26-4550.

(C) IL-2 bound to its much more potent protein-protein inhibitor SP4206.

(D) Detail of IL-2 side chains when unbound (white), bound to IL-2Ra (green), and bound to SP4206 (yellow). SP4206 is shown for comparison (red). Residues R38 and F42 would be clashing with SP4206 binding conformation in the unbound IL-2 structure. Residues K35, F42, and L72 would be clashing with SP4206 binding conformation in the IL-2/IL-2Ra complex structure. None of such structures could have been used to find SP4206 inhibitor without making the IL-2 surface side chains flexible.

(E) Hot-spot analysis of IL-2/IL-2Ra interaction by alanine-scanning mutagenesis, colored according to effect in interaction (hot spots in red: F42, Y45, and E62).

(F) Hot-spot analysis of IL-2/SP4206 interaction by alanine-scanning mutagenesis, colored according to effect in interaction (hot spots in red: F42, Y45, and E62, interestingly, the same key residues as in IL-2/IL-2Ra interaction).

overcome this, different computational approaches have been described to help to identify the interface residues for a given interaction (Fernández-Recio, 2011). Table S4 shows some examples of interface prediction servers. These computational predictions can be used to guide mutational experiments, and they also provide grounds to characterize macromolecular complexes further as a first step in a drug discovery process or chemical biology project. The next goal is to characterize the binding mode between two given proteins that form a complex of interest. Computational docking is a powerful approach for this purpose, and many algorithms have been recently reported (Ritchie, 2008; Vajda and Kozakov, 2009). Table S4 shows a list of public online servers for protein-protein docking. Protein docking still faces many challenges (Lensink and Wodak, 2010; Pons et al., 2010) but it has been applied to specific cases of biomedical interest (e.g., for coagulation proteins; Autin et al., 2006) and even at a genomewide scale (Mosca et al., 2009).

The next difficulty is to design a small molecule that can effectively compete with a large PPI surface. Fortunately, it was shown that not all interface residues contribute equally to the free energy of binding. Only a subset of the interface residues, so called ''hot spots,'' typically contributes above 1–2 kcal/mol to the binding affinity (Clackson and Wells, 1995). It seems reasonable that a small molecule designed to show any effect on a given PPI should target such hot-spot residues, although allosteric pockets could be of interest. Experimental characterization of hot-spot residues by alanine-scanning mutagenesis combined with biophysical experiments is expensive and difficult to extend at a proteomic scale. For this reason, several computational methods to predict hot spots have been reported (Fernández-Recio, 2011), some of which are listed in Table S5. Most of the methods need the atomic 3D structure of the protein-protein complex, but in a realistic scenario, the structure of the complex is unknown. Fortunately, some approaches, such as pyDockNIP (Grosdidier and Fernández-Recio, 2008), which is based on analysis of docking results, can work on the unbound proteins. After characterizing the PPI and identifying putative hot-spot residues, the next steps can be further investigations of the druggability (ligandability) (Edfeldt et al., 2011) of the region (see Table S5 for online servers and databases to predict the druggability of PPIs) and the use of structure-based virtual screening experiments.

The applications of all or some of the above-described ideas, with or without biophysical methods such as NMR, are starting to pay off (Wells and McClendon, 2007). For example, several molecules are known to inhibit the interaction between the cytokine interleukin-2 (IL-2) and its receptor IL-2Rα. The company Hoffmann-La Roche developed the first inhibitor, Ro26-4550, which was later evolved by Sunesis Pharmaceuticals to a much more potent compound called SP4206, using a fragment-based approach and modeling (Braisted et al., 2003; Thanos et al., 2006) (Figure 2). Docking and energy calculations are typically applied in the lead optimization phase, such as for the design of inhibitors of the Dvl PDZ domain, involved in Wnt signaling pathways (Segall et al., 2009). The discovery of inhibitors for the ZipA-FtsZ interaction at Wyeth was another nice example of integration of experimental and computational tools (Rush et al., 2005; Tsao et al., 2006). The combination of virtual fragment analysis and selection by molecular docking with NMR screening identified a molecule displaying strongly favorable binding enthalpy for the pY pocket of the v-Src tyrosine kinase SH2 domain, which was later nicely explained by computational modeling (Taylor et al., 2007). More recently, inhibitors for the eIF4E/eIF4G interaction have been found, with their binding modes modeled by ligand docking using a scoring function based on the predicted hot spots (Cencic et al., 2011; Kozakov et al., 2011). However, modulating PPI is still highly challenging. As an example, after many efforts, some targets do not seem druggable. Only weak (microM) inhibitors have been found so far to inhibit the ZipA/FtsZ interaction (Tsao et al., 2006) whereas in the case of the

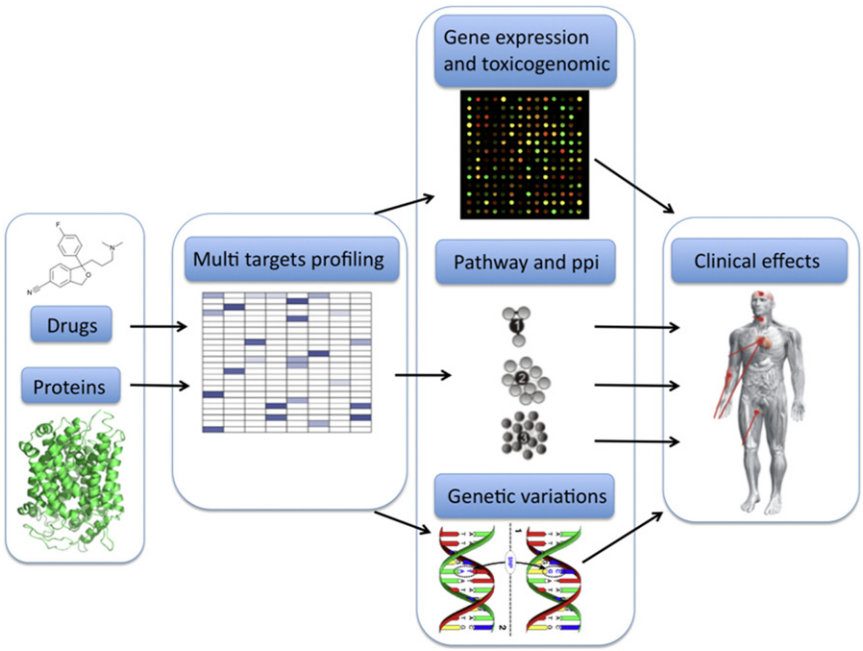

Figure 3. Example of Systems Pharmacology Approaches
In a first step, from a drug or a protein, profiling is performed of multiple targets annotated (from the literature) or predicted (from the tools presented above). Second, integration of diverse "omics" data associated with the ensemble of proteins perturbed by the drug (interactomic, pathway, and genomics) is performed. Finally, analysis of potential clinical effects (therapeutic and adverse effects) associated with the drug is carried out.

interaction between tumor necrosis factor-α (TNFα) and its receptor TNFR1, no synthetic small-molecule inhibitors have been reported. Yet, the very flat VEGF-VEGFR interface has been recently screened in silico and in vitro with some success (Gautier et al., 2011), suggesting that more and more PPIs are going to be modulated by small molecules in the next decade.

Attempts to Move Away from the One Compound-One Target Paradigm: Data Integration, Data Mining, Systems Pharmacology and Off-Target Predictions

For 50 years, the one-target one drug paradigm has been the driving force for developments in biomedicine. Although this strategy allowed bringing new drugs to the market, a significant decrease in the rate of new drug candidates has been observed. The reasons for this attrition are essentially the result of a lack of efficacy and clinical safety or toxicology (Hopkins, 2008). In recent years, it has become apparent that many common diseases, such as cancer, cardiovascular disease, and mental disorders, are much more complex than initially anticipated because they are often caused by multiple molecular abnormalities rather than being the result of a single defect. With the recent advances in molecular biology techniques and the massive biological data available, the focus on drug discovery is shifting from a molecular and cellular level to tissue and biological system level (i.e., network pharmacology) (Berger and Iyengar, 2009; Boran and Iyengar, 2010). At the difference of the single-target approach, the network pharmacology identifies a complex (or combination) of proteins whose perturbation (even indirect) results to the clinical outcome observed. An example of this strategy is depicted in Figure 3. One of the pioneers in this area is the study of Yildirim et al. (2007). They generated a bipartite graph of all known FDA-approved drugs and their targets and provided systematic information about drugs in the context of cellular and disease networks through a physical PPI map. Since then, several databases and tools, such as STITCH (Kuhn et al., 2010b), ChemProt (Taboureau et al., 2011), and Matador (Günther et al., 2008) have been developed (Table S6). In addition to systems pharmacology, others strategies are investigated for a better prediction of clinical response to a drug. Taking advantage of large chemical biology repositories presented above, chemogenomics and proteochemometric approaches (i.e., assessing the overall ligand-target interaction space through ligand-based and structure-based methods) can suggest which ensemble of targets are likely to be hit by a ligand (Cases and Mestres, 2009; Koutsoukas et al., 2011; Ma et al., 2010; Meslamani and Rognan, 2011). Among others, we can mention tools such as iPHACE (Garcia-Serna et al., 2010), SEA (Keiser et al., 2009), PharmMapper (Liu et al., 2010), or ReverseScreen3D (Kinnings and Jackson, 2011) that are dedicated to such profiling.

Another concept is the extraction of biological information from microarrays. Using gene-expression signatures, the genetic perturbation by small molecules can be studied systematically and connected to diseases and toxicity (Lamb et al., 2006). The number of microarray studies is growing rapidly and can be accessible from public databases such as the Gene Expression Omnibus (Barrett et al., 2011). Then, on the basis of similar compound genes signature (Toyoshiba et al., 2009) or in association with chemical structure similarity (Low et al., 2011), chemical-phenotype association can be predicted. One interesting feature about microarray technology is the possibility to detect tissue-specific genes. Groups of genes whose function and expression are preferred in one or more cell types can be gathered from specialized databases (Xiao et al., 2010) and associated with pathology and diseases (Lage et al., 2008). Therefore, we can easily imagine that integration of tissue-specific genes to chemogenomics data could improve the understanding of clinical effects and adverse effects. Side effects and adverse drug reactions (ADRs), as mentioned above, are major bottlenecks in the pharmaceutical industry, and network biology has emerged as an alternative for accurate predictions of promiscuous binders that could cause ADRs (Mendrick, 2011). For example, connecting drugs by side effect similarity can provide insights into the molecular basis of the drug's side effects and allow predicting novel off-targets involved in negative clinical outcomes (Campillos et al., 2008; Yang et al., 2011; Brouwers et al., 2011). Using a large, publicly available bioassay database such as PubChem, the ADRs have also been analyzed

at the organ level (Pouliot et al., 2011). It is possible now to get side and toxicological effects of small molecules from several repositories, such as Sider (Kuhn et al., 2010a), Actor (Judson et al., 2008), or Dailymed (http://dailymed.nih.gov/dailymed), and an in silico approach starts to explore drug repurposing and associations of drugs, targets, and clinical outcomes into an integrated network (Achenbach et al., 2011; Ekins et al., 2011; Kinnings et al., 2009; Oprea et al., 2011; Xie and Bourne, 2011). In this direction, a public-private partnership, named Open Phacts, within the framework of the European Innovative Medicines Initiative (IMI) will provide the pharmacological, pharmacokinetics, ADMET and clinical profiles of drugs and small molecules on a unique and freely accessible platform in the next couple of years (http://www.openphacts.org/). Additionally, genetic polymorphisms and variations in human add an extra level of difficulty on the drug-clinical effect prediction, and computational methods have started to evaluate the impact of single nucleotide polymorphisms (SNPs) in human health and drug discovery. More than 20 million human SNPs have been reported (Sayers et al., 2011), and elucidation of the functional effect of a predisposed SNP is a key factor in deciphering the differences in an individual's drug response and in understanding the mechanism underlying the disease. Several tools can help to predict disease susceptibility to genetic variation, as recently reviewed in these articles (Fernald et al., 2011; Mah et al., 2011; Thusberg and Vihinen, 2009). Overall, computational methods will play a major role to process massive genomic data and will contribute to pharmacogenomics and personalized medicine. However, one of the challenges in systems pharmacology will be to move forward from a qualitative level to a quantitative component, including concentration level and kinetic parameters governing the interactions.

Conclusions

A plethora of computational techniques, including fast lead discovery engines, compound profiling, and protein-protein docking, is available but it is usually not possible to know at the beginning of a project which approach or protocol could be more successful. Today's algorithms are usually efficient, although numerous problems still have to be solved (e.g., scoring, flexibility, designing multitarget inhibitors, predictive ADMET, and so forth). We have summarized several structural bioinformatics and chemoinformatics tools, with a special emphasis on software packages that tend to be freely available to academic users or that are implemented online, as well as a few recent success stories (chemical biology or drug discovery). In line with the recent observations made by Leeson and St-Gallay (2011), who stated that during the past 15 years, although it has become clear that physicochemical properties of drug candidates have an important influence on the likelihood of compound-related attrition during development, a substantial part of the pharmaceutical industry has not modified its drug design practices accordingly and is still producing compounds with suboptimal physicochemical profiles, we suggest that in silico technologies are also not sufficiently applied in drug industry and academia alike, because most of the time, computer groups are too small and overloaded of projects and, as such, are seldom able to deal with all the data in an appropriate manner. We thus hope that this review will encourage researchers in life and health sciences to intensify the usage of computational tools to develop new ideas and to assist decision making, while being aware of the strengths and weaknesses of the methods (this of course holds true for experimental approaches).

SUPPLEMENTAL INFORMATION

Supplemental Information includes six tables and may be found with this article online at doi:10.1016/j.chembiol.2011.12.007.

ACKNOWLEDGMENTS

We apologize to the researchers whose work has not been cited because of space limitations. We thank INSERM, the ANR Piribio, and the University of Paris Diderot for financial support.

REFERENCES

Abagyan, R., and Kufareva, I. (2009). The flexible pocketome engine for structural chemogenomics. Methods Mol. Biol. *575*, 249–279.

Achenbach, J., Tiikkainen, P., Franke, L., and Proschak, E. (2011). Computational tools for polypharmacology and repurposing. Future Med. Chem. *3*, 961–968.

Akella, L.B., and DeCaprio, D. (2010). Cheminformatics approaches to analyze diversity in compound screening libraries. Curr. Opin. Chem. Biol. *14*, 325–330.

Akritopoulou-Zanze, I., Metz, J.T., and Djuric, S.W. (2007). Topography-biased compound library design: the shape of things to come? Drug Discov. Today *12*, 948–952.

Amaro, R.E., Schnaufer, A., Interthal, H., Hol, W., Stuart, K.D., and McCammon, J.A. (2008). Discovery of drug-like inhibitors of an essential RNA-editing ligase in *Trypanosoma brucei*. Proc. Natl. Acad. Sci. USA *105*, 17278–17283.

Arkin, M.R., and Whitty, A. (2009). The road less traveled: modulating signal transduction enzymes by inhibiting their protein-protein interactions. Curr. Opin. Chem. Biol. *13*, 284–290.

Autin, L., Steen, M., Dahlbäck, B., and Villoutreix, B.O. (2006). Proposed structural models of the prothrombinase (FXa-FVa) complex. Proteins *63*, 440–450.

Axerio-Cilies, P., Castañeda, I.P., Mirza, A., and Reynisson, J. (2009). Investigation of the incidence of "undesirable" molecular moieties for high-throughput screening compound libraries in marketed drug compounds. Eur. J. Med. Chem. *44*, 1128–1134.

Baell, J.B. (2010). Observations on screening-based research and some concerning trends in the literature. Future Med. Chem. *2*, 1529–1546.

Baell, J.B., and Holloway, G.A. (2010). New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J. Med. Chem. *53*, 2719–2740.

Barelier, S., and Krimm, I. (2011). Ligand specificity, privileged substructures and protein druggability from fragment-based screening. Curr. Opin. Chem. Biol. *15*, 469–474.

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., et al. (2011). NCBI GEO: archive for functional genomics data sets—10 years on. Nucleic Acids Res. *39* (Database issue), D1005–D1010.

Bauer, R.A., Wurst, J.M., and Tan, D.S. (2010). Expanding the range of 'druggable' targets with natural product-based libraries: an academic perspective. Curr. Opin. Chem. Biol. *14*, 308–314.

Benigni, R., and Bossa, C. (2011). Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology. Chem. Rev. *111*, 2507–2536.

Berger, S.I., and Iyengar, R. (2009). Network analyses in systems pharmacology. Bioinformatics *25*, 2466–2472.

Bologa, C.G., Olah, M.M., and Oprea, T.I. (2006). Chemical database preparation for compound acquisition or virtual screening. Methods Mol. Biol. *316*, 375–388.

Boran, A.D., and Iyengar, R. (2010). Systems pharmacology. Mt. Sinai J. Med. *77*, 333–344.

Bottegoni, G., Rocchia, W., Rueda, M., Abagyan, R., and Cavalli, A. (2011). Systematic exploitation of multiple receptor conformations for virtual ligand screening. PLoS ONE *6*, e18845.

Braisted, A.C., Oslob, J.D., Delano, W.L., Hyde, J., McDowell, R.S., Waal, N., Yu, C., Arkin, M.R., and Raimundo, B.C. (2003). Discovery of a potent small molecule IL-2 inhibitor through fragment assembly. J. Am. Chem. Soc. *125*, 3714–3715.

Brenk, R., Schipani, A., James, D., Krasowski, A., Gilbert, I.H., Frearson, J., and Wyatt, P.G. (2008). Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. ChemMedChem *3*, 435–444.

Brewerton, S.C. (2008). The use of protein-ligand interaction fingerprints in docking. Curr. Opin. Drug Discov. Devel. *11*, 356–364.

Brouwers, L., Iskar, M., Zeller, G., van Noort, V., and Bork, P. (2011). Network neighbors of drug targets contribute to drug side-effect similarity. PLoS ONE *6*, e22187.

B-Rao, C., Subramanian, J., and Sharma, S.D. (2009). Managing protein flexibility in docking and its applications. Drug Discov. Today *14*, 394–400.

Campbell, J.B. (2010). Improving lead generation success through integrated methods: transcending 'drug discovery by numbers'. IDrugs *13*, 874–879.

Campillos, M., Kuhn, M., Gavin, A.C., Jensen, L.J., and Bork, P. (2008). Drug target identification using side-effect similarity. Science *321*, 263–266.

Cases, M., and Mestres, J. (2009). A chemogenomic approach to drug discovery: focus on cardiovascular diseases. Drug Discov. Today *14*, 479–485.

Cavasotto, C.N., and Orry, A.J. (2007). Ligand docking and structure-based virtual screening in drug discovery. Curr. Top. Med. Chem. *7*, 1006–1014.

Cavasotto, C.N., and Phatak, S.S. (2009). Homology modeling in drug discovery: current trends and applications. Drug Discov. Today *14*, 676–683.

Cencic, R., Hall, D.R., Robert, F., Du, Y., Min, J., Li, L., Qui, M., Lewis, I., Kurtkaya, S., Dingledine, R., et al. (2011). Reversing chemoresistance by small molecule inhibition of the translation initiation complex eIF4F. Proc. Natl. Acad. Sci. USA *108*, 1046–1051.

Charifson, P.S., and Walters, W.P. (2002). Filtering databases and chemical libraries. J. Comput. Aided Mol. Des. *16*, 311–323.

Charifson, P.S., Corkery, J.J., Murcko, M.A., and Walters, W.P. (1999). Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. J. Med. Chem. *42*, 5100–5109.

Chaudhaery, S.S., Roy, K.K., Shakya, N., Saxena, G., Sammi, S.R., Nazir, A., Nath, C., and Saxena, A.K. (2010). Novel carbamates as orally active acetylcholinesterase inhibitors found to improve scopolamine-induced cognition impairment: pharmacophore-based virtual screening, synthesis, and pharmacology. J. Med. Chem. *53*, 6490–6505.

Clackson, T., and Wells, J.A. (1995). A hot spot of binding energy in a hormone-receptor interface. Science *267*, 383–386.

Clark, D.E. (2008). What has virtual screening ever done for drug discovery. Expert Opin. Drug Discov. *3*, 841–851. 10.1517/17460441.3.8.841.

Cross, S., and Cruciani, G. (2010). Molecular fields in drug discovery: getting old or reaching maturity? Drug Discov. Today *15*, 23–32.

Dailey, M.M., Hait, C., Holt, P.A., Maguire, J.M., Meier, J.B., Miller, M.C., Petraccone, L., and Trent, J.O. (2009). Structure-based drug design: from nucleic acid to membrane protein targets. Exp. Mol. Pathol. *86*, 141–150.

David-Eden, H., Mankin, A.S., and Mandel-Gutfreund, Y. (2010). Structural signatures of antibiotic binding sites on the ribosome. Nucleic Acids Res. *38*, 5982–5994.

Davies, J.W., Glick, M., and Jenkins, J.L. (2006). Streamlining lead discovery by aligning in silico and high-throughput screening. Curr. Opin. Chem. Biol. *10*, 343–351.

Davis, A.M., St-Gallay, S.A., and Kleywegt, G.J. (2008). Limitations and lessons in the use of X-ray structural information in drug design. Drug Discov. Today *13*, 831–841.

de Ruiter, A., and Oostenbrink, C. (2011). Free energy calculations of protein-ligand interactions. Curr. Opin. Chem. Biol. *15*, 547–552.

Desjarlais, R.L. (2011). Using computational techniques in fragment-based drug discovery. Methods Enzymol. *493*, 137–155.

Dobson, C.M. (2004). Chemical space and biology. Nature *432*, 824–828.

Drewry, D.H., and Macarrón, R. (2010). Enhancements of screening collections to address areas of unmet medical need: an industry perspective. Curr. Opin. Chem. Biol. *14*, 289–298.

Durrant, J.D., de Oliveira, C.A., and McCammon, J.A. (2011). POVME: an algorithm for measuring binding-pocket volumes. J. Mol. Graph. Model. *29*, 773–776.

Eckert, H., and Bajorath, J. (2007). Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. Drug Discov. Today *12*, 225–233.

Edfeldt, F.N., Folmer, R.H., and Breeze, A.L. (2011). Fragment screening to predict druggability (ligandability) and lead discovery success. Drug Discov. Today *16*, 284–287.

Ekins, S., and Williams, A.J. (2010). Precompetitive preclinical ADME/Tox data: set it free on the web to facilitate computational model building and assist drug development. Lab Chip *10*, 13–22.

Ekins, S., Mestres, J., and Testa, B. (2007a). In silico pharmacology for drug discovery: applications to targets and beyond. Br. J. Pharmacol. *152*, 21–37.

Ekins, S., Mestres, J., and Testa, B. (2007b). In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. Br. J. Pharmacol. *152*, 9–20.

Ekins, S., Kortagere, S., Iyer, M., Reschly, E.J., Lill, M.A., Redinbo, M.R., and Krasowski, M.D. (2009). Challenges predicting ligand-receptor interactions of promiscuous proteins: the nuclear receptor PXR. PLoS Comput. Biol. *5*, e1000594.

Ekins, S., Williams, A.J., Krasowski, M.D., and Freundlich, J.S. (2011). In silico repositioning of approved drugs for rare and neglected diseases. Drug Discov. Today *16*, 298–310.

Enoch, S.J., Ellison, C.M., Schultz, T.W., and Cronin, M.T. (2011). A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. Crit. Rev. Toxicol. *41*, 783–802.

Ertl, P., and Jelfs, S. (2007). Designing drugs on the internet? Free web tools and services supporting medicinal chemistry. Curr. Top. Med. Chem. *7*, 1491–1501.

Erve, J.C. (2006). Chemical toxicology: reactive intermediates and their role in pharmacology and toxicology. Expert Opin. Drug Metab. Toxicol. *2*, 923–946.

Eyrisch, S., and Helms, V. (2007). Transient pockets on protein surfaces involved in protein-protein interaction. J. Med. Chem. *50*, 3457–3464.

Faller, B., Ottaviani, G., Ertl, P., Berellini, G., and Collis, A. (2011). Evolution of the physicochemical properties of marketed drugs: can history foretell the future? Drug Discov. Today *16*, 976–984.

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

Fauman, E.B., Rai, B.K., and Huang, E.S. (2011). Structure-based druggability assessment—identifying suitable targets for small molecule therapeutics. Curr. Opin. Chem. Biol. *15*, 463–468.

Feher, M. (2006). Consensus scoring for protein-ligand interactions. Drug Discov. Today *11*, 421–428.

Fernald, G.H., Capriotti, E., Daneshjou, R., Karczewski, K.J., and Altman, R.B. (2011). Bioinformatics challenges for personalized medicine. Bioinformatics *27*, 1741–1748.

Fernández-Recio, J. (2011). Prediction of protein binding sites and hot spots. Comput. Mol. Sci. *1*, 680–698. 10.1002/wcms.45.

Ferreira, R.S., Simeonov, A., Jadhav, A., Eidam, O., Mott, B.T., Keiser, M.J., McKerrow, J.H., Maloney, D.J., Irwin, J.J., and Shoichet, B.K. (2010). Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. J. Med. Chem. *53*, 4891–4905.

Foloppe, N., and Chen, I.J. (2009). Conformational sampling and energetics of drug-like molecules. Curr. Med. Chem. *16*, 3381–3413.

Fourches, D., Muratov, E., and Tropsha, A. (2010). Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. J. Chem. Inf. Model. *50*, 1189–1204.

Fukunishi, Y. (2010). Structural ensemble in computational drug screening. Expert Opin. Drug Metab. Toxicol. *6*, 835–849.

Ganesan, A. (2008). The impact of natural products upon modern drug discovery. Curr. Opin. Chem. Biol. *12*, 306–317.

Garcia-Serna, R., Ursu, O., Oprea, T.I., and Mestres, J. (2010). iPHACE: integrative navigation in pharmacological space. Bioinformatics *26*, 985–986.

Garg, A., Tewari, R., and Raghava, G.P. (2010). KiDoQ: using docking based energy scores to develop ligand based model for predicting antibacterials. BMC Bioinformatics *11*, 125.

Gautier, B., Miteva, M.A., Goncalves, V., Huguenot, F., Coric, P., Bouaziz, S., Seijo, B., Gaucher, J.-F., Broutin, I., Garbay, C., et al. (2011). Targeting the proangiogenic VEGF-VEGFR protein-protein interface with drug-like compounds by in silico and in vitro screening. Chem. Biol. *18*, 1631–1639.

Gedeck, P., Kramer, C., and Ertl, P. (2010). Computational analysis of structure-activity relationships. Prog. Med. Chem. *49*, 113–160.

Geppert, H., Vogt, M., and Bajorath, J. (2010). Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. J. Chem. Inf. Model. *50*, 205–216.

Gleeson, M.P., Hersey, A., and Hannongbua, S. (2011a). In-silico ADME models: a general assessment of their utility in drug discovery applications. Curr. Top. Med. Chem. *11*, 358–381.

Gleeson, M.P., Hersey, A., Montanari, D., and Overington, J. (2011b). Probing the links between in vitro potency, ADMET and physicochemical parameters. Nat. Rev. Drug Discov. *10*, 197–208.

Gozalbes, R., and Pineda-Lucena, A. (2011). Small molecule databases and chemical descriptors useful in chemoinformatics: an overview. Comb. Chem. High Throughput Screen. *14*, 548–558.

Grosdidier, S., and Fernández-Recio, J. (2008). Identification of hot-spot residues in protein-protein interactions by computational docking. BMC Bioinformatics *9*, 447.

Guido, R.V., Oliva, G., and Andricopulo, A.D. (2008). Virtual screening and its integration with modern drug design technologies. Curr. Med. Chem. *15*, 37–46.

Günther, S., Kuhn, M., Dunkel, M., Campillos, M., Senger, C., Petsalaki, E., Ahmed, J., Urdiales, E.G., Gewiess, A., Jensen, L.J., et al. (2008). SuperTarget and Matador: resources for exploring drug-target relationships. Nucleic Acids Res. *36* (Database issue), D919–D922.

Hann, M.M. (2011). Molecular obesity, potency and other addictions in drug discovery. Med. Chem. Commun. *2*, 349–355. 10.1039/C1MD0017A.

Hert, J., Irwin, J.J., Laggner, C., Keiser, M.J., and Shoichet, B.K. (2009). Quantifying biogenic bias in screening libraries. Nat. Chem. Biol. *5*, 479–483.

Holdgate, G.A., and Gill, A.L. (2011). Kinetic efficiency: the missing metric for enhancing compound quality? Drug Discov. Today *16*, 910–913.

Hopkins, A.L. (2008). Network pharmacology: the next paradigm in drug discovery. Nat. Chem. Biol. *4*, 682–690.

Horvath, D. (1997). A virtual screening approach applied to the search for trypanothione reductase inhibitors. J. Med. Chem. *40*, 2412–2423.

Hu, Y., Stumpfe, D., and Bajorath, J. (2011). Lessons learned from molecular scaffold analysis. J. Chem. Inf. Model. *51*, 1742–1753.

Huang, B. (2009). MetaPocket: a meta approach to improve protein ligand binding site prediction. OMICS *13*, 325–330.

Huang, N., and Shoichet, B.K. (2008). Exploiting ordered waters in molecular docking. J. Med. Chem. *51*, 4862–4865.

Huang, N., and Jacobson, M.P. (2010). Binding-site assessment by virtual fragment screening. PLoS ONE *5*, e10109.

Huggins, D.J., and Tidor, B. (2011). Systematic placement of structural water molecules for improved scoring of protein-ligand interactions. Protein Eng. Des. Sel. *24*, 777–789.

Huggins, D.J., Venkitaraman, A.R., and Spring, D.R. (2011). Rational methods for the selection of diverse screening compounds. ACS Chem. Biol. *6*, 208–217.

Hughes, J.D., Blagg, J., Price, D.A., Bailey, S., Decrescenzo, G.A., Devraj, R.V., Ellsworth, E., Fobian, Y.M., Gibbs, M.E., Gilles, R.W., et al. (2008). Physiochemical drug properties associated with in vivo toxicological outcomes. Bioorg. Med. Chem. Lett. *18*, 4872–4875.

Huth, J.R., Mendoza, R., Olejniczak, E.T., Johnson, R.W., Cothron, D.A., Liu, Y., Lerner, C.G., Chen, J., and Hajduk, P.J. (2005). ALARM NMR: a rapid and robust experimental method to detect reactive false positives in biochemical screens. J. Am. Chem. Soc. *127*, 217–224.

Ivetac, A., and McCammon, J.A. (2011). Molecular recognition in the case of flexible targets. Curr. Pharm. Des. *17*, 1663–1671.

Judson, R., Richard, A., Dix, D., Houck, K., Elloumi, F., Martin, M., Cathey, T., Transue, T.R., Spencer, R., and Wolf, M. (2008). ACToR—Aggregated Computational Toxicology Resource. Toxicol. Appl. Pharmacol. *233*, 7–13.

Kalgutkar, A.S., Gardner, I., Obach, R.S., Shaffer, C.L., Callegari, E., Henne, K.R., Mutlib, A.E., Dalvie, D.K., Lee, J.S., Nakai, Y., et al. (2005). A comprehensive listing of bioactivation pathways of organic functional groups. Curr. Drug Metab. *6*, 161–225.

Kawasaki, Y., and Freire, E. (2011). Finding a better path to drug selectivity. Drug Discov. Today *16*, 985–990.

Kazius, J., McGuire, R., and Bursi, R. (2005). Derivation and validation of toxicophores for mutagenicity prediction. J. Med. Chem. *48*, 312–320.

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., et al. (2009). Predicting new molecular targets for known drugs. Nature *462*, 175–181.

Khanna, V., and Ranganathan, S. (2009). Physiochemical property space distribution among human metabolites, drugs and toxins. BMC Bioinformatics *10* (*Suppl 15*), S10.

Kinnings, S.L., and Jackson, R.M. (2011). ReverseScreen3D: a structure-based ligand matching method to identify protein targets. J. Chem. Inf. Model. *51*, 624–634.

Kinnings, S.L., Liu, N., Buchmeier, N., Tonge, P.J., Xie, L., and Bourne, P.E. (2009). Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. PLoS Comput. Biol. *5*, e1000423.

Kirchmair, J., Distinto, S., Schuster, D., Spitzer, G., Langer, T., and Wolber, G. (2008). Enhancing drug discovery through in silico screening: strategies to increase true positives retrieval rates. Curr. Med. Chem. *15*, 2040–2053.

Köppen, H. (2009). Virtual screening: what does it give us? Curr. Opin. Drug Discov. Devel. *12*, 397–407.

Koutsoukas, A., Simms, B., Kirchmair, J., Bond, P.J., Whitmore, A.V., Zimmer, S., Young, M.P., Jenkins, J.L., Glick, M., Glen, R.C., et al. (2011). From in silico target prediction to multi-target drug design: current databases, methods and applications. J. Proteomics. *74*, 2554–2574.

Kozakov, D., Hall, D.R., Chuang, G.Y., Cencic, R., Brenke, R., Grove, L.E., Beglov, D., Pelletier, J., Whitty, A., and Vajda, S. (2011). Structural conservation of druggable hot spots in protein-protein interfaces. Proc. Natl. Acad. Sci. USA *108*, 13528–13533.

Kramer, J.A., Sagartz, J.E., and Morris, D.L. (2007). The application of discovery toxicology and pathology towards the design of safer pharmaceutical lead candidates. Nat. Rev. Drug Discov. *6*, 636–649.

Kufareva, I., Rueda, M., Katritch, V., Stevens, R.C., and Abagyan, R.; GPCR Dock 2010 participants. (2011). Status of GPCR modeling and docking as

reflected by community-wide GPCR Dock 2010 assessment. Structure *19*, 1108–1126.

Kufareva, I., Ilatovskiy, A.V., and Abagyan, R. (2012). Pocketome: an encyclopedia of small-molecule binding sites in 4D. Nucleic Acids Res. *40*, D535–D540.

Kuhn, M., Campillos, M., Letunic, I., Jensen, L.J., and Bork, P. (2010a). A side effect resource to capture phenotypic effects of drugs. Mol. Syst. Biol. *6*, 343.

Kuhn, M., Szklarczyk, D., Franceschini, A., Campillos, M., von Mering, C., Jensen, L.J., Beyer, A., and Bork, P. (2010b). STITCH 2: an interaction network database for small molecules and proteins. Nucleic Acids Res. *38* (Database issue), D552–D556.

Lage, K., Hansen, N.T., Karlberg, E.O., Eklund, A.C., Roque, F.S., Donahoe, P.K., Szallasi, Z., Jensen, T.S., and Brunak, S. (2008). A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. Proc. Natl. Acad. Sci. USA *105*, 20870–20875.

Lagorce, D., Maupetit, J., Baell, J., Sperandio, O., Tufféry, P., Miteva, M.A., Galons, H., and Villoutreix, B.O. (2011). The FAF-Drugs2 server: a multistep engine to prepare electronic chemical compound collections. Bioinformatics *27*, 2018–2020.

Laine, E., Goncalves, C., Karst, J.C., Lesnard, A., Rault, S., Tang, W.J., Malliavin, T.E., Ladant, D., and Blondel, A. (2010). Use of allostery to identify inhibitors of calmodulin-induced activation of *Bacillus anthracis* edema factor. Proc. Natl. Acad. Sci. USA *107*, 11277–11282.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., et al. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science *313*, 1929–1935.

Lang, P.T., Brozell, S.R., Mukherjee, S., Pettersen, E.F., Meng, E.C., Thomas, V., Rizzo, R.C., Case, D.A., James, T.L., and Kuntz, I.D. (2009). DOCK 6: combining techniques to model RNA-small molecule complexes. RNA *15*, 1219–1230.

Leeson, P.D., and St-Gallay, S.A. (2011). The influence of the 'organizational factor' on compound quality in drug discovery. Nat. Rev. Drug Discov. *10*, 749–765.

Lensink, M.F., and Wodak, S.J. (2010). Docking and scoring protein interactions: CAPRI 2009. Proteins *78*, 3073–3084.

Liao, C., Sitzmann, M., Pugliese, A., and Nicklaus, M.C. (2011). Software and resources for computational medicinal chemistry. Future Med. Chem. *3*, 1057–1085.

Lipinski, C.A. (2000). Drug-like properties and the causes of poor solubility and poor permeability. J. Pharmacol. Toxicol. Methods *44*, 235–249.

Liu, X., Ouyang, S., Yu, B., Liu, Y., Huang, K., Gong, J., Zheng, S., Li, Z., Li, H., and Jiang, H. (2010). PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. Nucleic Acids Res. *38* (Web Server issue), W609–W614.

Low, Y., Uehara, T., Minowa, Y., Yamada, H., Ohno, Y., Urushidani, T., Sedykh, A., Muratov, E., Kuz'min, V., Fourches, D., et al. (2011). Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches. Chem. Res. Toxicol. *24*, 1251–1262.

Lyne, P.D. (2002). Structure-based virtual screening: an overview. Drug Discov. Today *7*, 1047–1055.

Ma, X.H., Shi, Z., Tan, C., Jiang, Y., Go, M.L., Low, B.C., and Chen, Y.Z. (2010). In-silico approaches to multi-target drug discovery : computer aided multi-target drug design, multi-target virtual screening. Pharm. Res. *27*, 739–749.

Macarrón, R., and Luengo, J.I. (2011). Yin and yang in medicinal chemistry: what does drug-likeness mean? Future Med. Chem. *3*, 505–507.

Macarrón, R., Banks, M.N., Bojanic, D., Burns, D.J., Cirovic, D.A., Garyantes, T., Green, D.V., Hertzberg, R.P., Janzen, W.P., Paslay, J.W., et al. (2011). Impact of high-throughput screening in biomedical research. Nat. Rev. Drug Discov. *10*, 188–195.

Mah, J.T., Low, E.S., and Lee, E. (2011). In silico SNP analysis and bioinformatics tools: a review of the state of the art to aid drug discovery. Drug Discov. Today *16*, 800–809.

Mazuc, E., Villoutreix, B.O., Malbec, O., Roumier, T., Fleury, S., Leonetti, J.P., Dombrowicz, D., Daeron, M., Martineau, P., and Dariavach, P. (2008). A novel druglike spleen tyrosine kinase binder prevents anaphylactic shock when administered orally. J. Allergy Clin. Immunol. *122*, 188–194.

McInnes, C. (2007). Virtual screening strategies in drug discovery. Curr. Opin. Chem. Biol. *11*, 494–502.

Meanwell, N.A. (2011). Improving drug candidates by design: a focus on physicochemical properties as a means of improving compound disposition and safety. Chem. Res. Toxicol. *24*, 1420–1456.

Mendrick, D.L. (2011). Transcriptional profiling to identify biomarkers of disease and drug response. Pharmacogenomics *12*, 235–249.

Meslamani, J., and Rognan, D. (2011). Enhancing the accuracy of chemogenomic models with a three-dimensional binding site kernel. J. Chem. Inf. Model. *51*, 1593–1603.

Mitchell, W., and Matsumoto, S. (2011). Large-scale integrated supercomputing platform for next generation virtual drug discovery. Curr. Opin. Chem. Biol. *15*, 553–559.

Miteva, M.A., Lee, W.H., Montes, M.O., and Villoutreix, B.O. (2005). Fast structure-based virtual ligand screening combining FRED, DOCK, and Surflex. J. Med. Chem. *48*, 6012–6022.

Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., and Corbeil, C.R. (2008). Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. Br. J. Pharmacol. *153* (*Suppl 1*), S7–S26.

Montanari, D., Chiarparin, E., Gleeson, M.P., Braggio, S., Longhi, R., and Rossi, T. (2011). Application of drug efficiency index in drug discovery. Expert Opin. Drug Discov. *6*, 913–920. 10.1517/17460441.2011.602968.

Moroy, G., Martiny, V.Y., Vayer, P., Villoutreix, B.O., and Miteva, M.A. (2011). Towards in silico structure-based ADMET prediction in drug discovery. Drug Discov. Today, [Epub ahead of print].

Mosca, R., Pons, C., Fernández-Recio, J., and Aloy, P. (2009). Pushing structural information into the yeast interactome by high-throughput protein docking experiments. PLoS Comput. Biol. *5*, e1000490.

Muchmore, S.W., Edmunds, J.J., Stewart, K.D., and Hajduk, P.J. (2010). Cheminformatic tools for medicinal chemists. J. Med. Chem. *53*, 4830–4841.

O'Shea, R., and Moser, H.E. (2008). Physicochemical properties of antibacterial compounds: implications for drug discovery. J. Med. Chem. *51*, 2871–2878.

Oprea, T.I., Nielsen, S.K., Ursu, O., Yang, J.J., Taboureau, O., Mathias, S.L., Kouskoumvekaki, I., Sklar, L.A., and Bologa, C.G. (2011). Associating drugs, targets and clinical outcomes into an integrated network affords a new platform for computer-aided drug repurposing. Mol. Inform. *30*, 100–111. 10.1002/minf.201100023.

Orry, A.J., Abagyan, R.A., and Cavasotto, C.N. (2006). Structure-based development of target-specific compound libraries. Drug Discov. Today *11*, 261–266.

Palmeira, A., Rodrigues, F., Sousa, E., Pinto, M., Vasconcelos, M.H., and Fernandes, M.X. (2011). New uses for old drugs: pharmacophore-based screening for the discovery of P-glycoprotein inhibitors. Chem. Biol. Drug Des. *78*, 57–72.

Park, B.K., Boobis, A., Clarke, S., Goldring, C.E., Jones, D., Kenna, J.G., Lambert, C., Laverty, H.G., Naisbitt, D.J., Nelson, S., et al. (2011). Managing the challenge of chemically reactive metabolites in drug development. Nat. Rev. Drug Discov. *10*, 292–306.

Pérot, S., Sperandio, O., Miteva, M.A., Camproux, A.C., and Villoutreix, B.O. (2010). Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. Drug Discov. Today *15*, 656–667.

Pierri, C.L., Parisi, G., and Porcelli, V. (2010). Computational approaches for protein function prediction: a combined strategy from multiple sequence alignment to molecular docking-based virtual screening. Biochim. Biophys. Acta *1804*, 1695–1712.

Pitt, W.R., Parry, D.M., Perry, B.G., and Groom, C.R. (2009). Heteroaromatic rings of the future. J. Med. Chem. *52*, 2952–2963.

Pons, C., Grosdidier, S., Solernou, A., Pérez-Cano, L., and Fernández-Recio, J. (2010). Present and future challenges and limitations in protein-protein docking. Proteins *78*, 95–108.

Pouliot, Y., Chiang, A.P., and Butte, A.J. (2011). Predicting adverse drug reactions using publicly available PubChem BioAssay data. Clin. Pharmacol. Ther. *90*, 90–99.

Prinz, F., Schlange, T., and Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? Nat. Rev. Drug Discov. *10*, 712.

Reddy, A.S., Pati, S.P., Kumar, P.P., Pradeep, H.N., and Sastry, G.N. (2007). Virtual screening in drug discovery—a computational perspective. Curr. Protein Pept. Sci. *8*, 329–351.

Ren, J., Xie, L., Li, W.W., and Bourne, P.E. (2010). SMAP-WS: a parallel web service for structural proteome-wide ligand-binding site comparison. Nucleic Acids Res. *38* (Web Server issue), W441–W444.

Renner, S., Popov, M., Schuffenhauer, A., Roth, H.J., Breitenstein, W., Marzinzik, A., Lewis, I., Krastel, P., Nigsch, F., Jenkins, J., and Jacoby, E. (2011). Recent trends and observations in the design of high-quality screening collections. Future Med. Chem. *3*, 751–766.

Rester, U. (2008). From virtuality to reality—virtual screening in lead discovery and lead optimization: a medicinal chemistry perspective. Curr. Opin. Drug Discov. Devel. *11*, 559–568.

Richard, A.M., Gold, L.S., and Nicklaus, M.C. (2006). Chemical structure indexing of toxicity data on the internet: moving toward a flat world. Curr. Opin. Drug Discov. Devel. *9*, 314–325.

Ridder, L., Wang, H., de Vlieg, J., and Wagener, M. (2011). Revisiting the rule of five on the basis of pharmacokinetic data from rat. ChemMedChem *6*, 1967–1970.

Ripphausen, P., Nisius, B., and Bajorath, J. (2011). State-of-the-art in ligand-based virtual screening. Drug Discov. Today *16*, 372–376.

Ripphausen, P., Nisius, B., Peltason, L., and Bajorath, J. (2010). Quo vadis, virtual screening? A comprehensive survey of prospective applications. J. Med. Chem. *53*, 8461–8467.

Rishton, G.M. (2003). Nonleadlikeness and leadlikeness in biochemical screening. Drug Discov. Today *8*, 86–96.

Ritchie, D.W. (2008). Recent progress and future directions in protein-protein docking. Curr. Protein Pept. Sci. *9*, 1–15.

Rosén, J., Lövgren, A., Kogej, T., Muresan, S., Gottfries, J., and Backlund, A. (2009). ChemGPS-NP(Web): chemical space navigation online. J. Comput. Aided Mol. Des. *23*, 253–259.

Rueda, M., Bottegoni, G., and Abagyan, R. (2010). Recipes for the selection of experimental protein conformations for virtual screening. J. Chem. Inf. Model. *50*, 186–193.

Rush, T.S., 3rd, Grant, J.A., Mosyak, L., and Nicholls, A. (2005). A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. J. Med. Chem. *48*, 1489–1495.

Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Federhen, S., et al. (2011). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. *39* (Database issue), D38–D51.

Schmidtke, P., and Barril, X. (2010). Understanding and predicting druggability: a high-throughput method for detection of drug binding sites. J. Med. Chem. *53*, 5858–5867.

Schneidman-Duhovny, D., Dror, O., Inbar, Y., Nussinov, R., and Wolfson, H.J. (2008). PharmaGist: a webserver for ligand-based pharmacophore detection. Nucleic Acids Res. *36* (Web Server issue), W223–W228.

Segall, M., Champness, E., Obrezanova, O., and Leeding, C. (2009). Beyond profiling: using ADMET models to guide decisions. Chem. Biodivers. *6*, 2144–2151.

Segers, K., Sperandio, O., Sack, M., Fischer, R., Miteva, M.A., Rosing, J., Nicolaes, G.A., and Villoutreix, B.O. (2007). Design of protein membrane interaction inhibitors by virtual ligand screening, proof of concept with the C2 domain of factor V. Proc. Natl. Acad. Sci. USA *104*, 12697–12702.

Seifert, M.H. (2009). Robust optimization of scoring functions for a target class. J. Comput. Aided Mol. Des. *23*, 633–644.

Sheridan, R.P., Maiorov, V.N., Holloway, M.K., Cornell, W.D., and Gao, Y.D. (2010). Drug-like density: a method of quantifying the "bindability" of a protein target based on a very large set of pockets and drug-like ligands from the Protein Data Bank. J. Chem. Inf. Model. *50*, 2029–2040.

Singh, J., Petter, R.C., Baillie, T.A., and Whitty, A. (2011). The resurgence of covalent drugs. Nat. Rev. Drug Discov. *10*, 307–317.

Smith, D.A. (2011). Discovery and ADMET: where are we now. Curr. Top. Med. Chem. *11*, 467–481.

Smith, D.A., Di, L., and Kerns, E.H. (2010). The effect of plasma protein binding on in vivo efficacy: misconceptions in drug discovery. Nat. Rev. Drug Discov. *9*, 929–939.

Sotriffer, C.A. (2011a). Accounting for induced-fit effects in docking: what is possible and what is not? Curr. Top. Med. Chem. *11*, 179–191.

Sotriffer, C.A. (2011b). Virtual screening: principles, challenges and practical guidelines, *Volume 48* (Weinheim, Germany: Wiley-VCH Verlag GmbH).

Sperandio, O., Miteva, M.A., Delfaud, F., and Villoutreix, B.O. (2006). Receptor-based computational screening of compound databases: the main docking-scoring engines. Curr. Protein Pept. Sci. *7*, 369–393.

Sperandio, O., Mouawad, L., Pinto, E., Villoutreix, B.O., Perahia, D., and Miteva, M.A. (2010a). How to choose relevant multiple receptor conformations for virtual screening: a test case of Cdk2 and normal mode analysis. Eur. Biophys. J. *39*, 1365–1372.

Sperandio, O., Reynès, C.H., Camproux, A.C., and Villoutreix, B.O. (2010b). Rationalizing the chemical space of protein-protein interaction inhibitors. Drug Discov. Today *15*, 220–229.

Spitzer, R., Cleves, A.E., and Jain, A.N. (2011). Surface-based protein binding pocket similarity. Proteins *79*, 2746–2763.

Stauch, B., Hofmann, H., Perkovic, M., Weisel, M., Kopietz, F., Cichutek, K., Münk, C., and Schneider, G. (2009). Model structure of APOBEC3C reveals a binding pocket modulating ribonucleic acid interaction required for encapsidation. Proc. Natl. Acad. Sci. USA *106*, 12079–12084.

Steinmeyer, A. (2006). The hit-to-lead process at Schering AG: strategic aspects. ChemMedChem *1*, 31–36.

Stepan, A.F., Walker, D.P., Bauman, J., Price, D.A., Baillie, T.A., Kalgutkar, A.S., and Aleo, M.D. (2011). Structural alert/reactive metabolite concept as applied in medicinal chemistry to mitigate the risk of idiosyncratic drug toxicity: a perspective based on the critical examination of trends in the top 200 drugs marketed in the United States. Chem. Res. Toxicol. *24*, 1345–1410.

Stocks, M.J., Wilden, G.R., Pairaudeau, G., Perry, M.W., Steele, J., and Stonehouse, J.P. (2009). A practical method for targeted library design balancing lead-like properties with diversity. ChemMedChem *4*, 800–808.

Stockwell, B.R. (2004). Exploring biology with small organic molecules. Nature *432*, 846–854.

Stoll, F., Göller, A.H., and Hillisch, A. (2011). Utility of protein structures in overcoming ADMET-related issues of drug-like compounds. Drug Discov. Today *16*, 530–538.

Swinney, D.C., and Anthony, J. (2011). How were new medicines discovered? Nat. Rev. Drug Discov. *10*, 507–519.

Taboureau, O., Nielsen, S.K., Audouze, K., Weinhold, N., Edsgärd, D., Roque, F.S., Kouskoumvekaki, I., Bora, A., Curpan, R., Jensen, T.S., et al. (2011). ChemProt: a disease chemical biology database. Nucleic Acids Res. *39* (Database issue), D367–D372.

Talele, T.T., Khedkar, S.A., and Rigby, A.C. (2010). Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. Curr. Top. Med. Chem. *10*, 127–141.

Tan, L., Batista, J., and Bajorath, J. (2010). Computational methodologies for compound database searching that utilize experimental protein-ligand interaction information. Chem. Biol. Drug Des. *76*, 191–200.

Tanaka, D., Tsuda, Y., Shiyama, T., Nishimura, T., Chiyo, N., Tominaga, Y., Sawada, N., Mimoto, T., and Kusunose, N. (2010). A practical use of ligand

efficiency indices out of the fragment-based approach: ligand efficiency-guided lead identification of soluble epoxide hydrolase inhibitors. J. Med. Chem. Published online December 30, 2010. 10.1021/jm101273e.

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

Tautermann, C.S. (2011). The use of G-protein coupled receptor models in lead optimization. Future Med. Chem. *3*, 709–721.

Taylor, J.D., Gilbert, P.J., Williams, M.A., Pitt, W.R., and Ladbury, J.E. (2007). Identification of novel fragment compounds targeted against the pY pocket of v-Src SH2 by computational and NMR screening and thermodynamic evaluation. Proteins *67*, 981–990.

Thangudu, R.R., Tyagi, M., Shoemaker, B.A., Bryant, S.H., Panchenko, A.R., and Madej, T. (2010). Knowledge-based annotation of small molecule binding sites in proteins. BMC Bioinformatics *11*, 365.

Thanos, C.D., DeLano, W.L., and Wells, J.A. (2006). Hot-spot mimicry of a cytokine receptor by a small molecule. Proc. Natl. Acad. Sci. USA *103*, 15422–15427.

Thusberg, J., and Vihinen, M. (2009). Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. Hum. Mutat. *30*, 703–714.

Thusberg, J., Olatubosun, A., and Vihinen, M. (2011). Performance of mutation pathogenicity prediction methods on missense variants. Hum. Mutat. *32*, 358–368.

Toyoshiba, H., Sawada, H., Naeshiro, I., and Horinouchi, A. (2009). Similar compounds searching system by using the gene expression microarray database. Toxicol. Lett. *186*, 52–57.

Tropsha, A., and Golbraikh, A. (2007). Predictive QSAR modeling workflow, model applicability domains, and virtual screening. Curr. Pharm. Des. *13*, 3494–3504.

Tsai, C.J., Ma, B., and Nussinov, R. (2009). Protein-protein interaction networks: how can a hub protein bind so many different partners? Trends Biochem. Sci. *34*, 594–600.

Tsao, D.H., Sutherland, A.G., Jennings, L.D., Li, Y., Rush, T.S., 3rd, Alvarez, J.C., Ding, W., Dushin, E.G., Dushin, R.G., Haney, S.A., et al. (2006). Discovery of novel inhibitors of the ZipA/FtsZ complex by NMR fragment screening coupled with structure-based design. Bioorg. Med. Chem. *14*, 7953–7961.

Tuccinardi, T. (2011). Binding-interaction prediction of RNA-binding ligands. Future Med. Chem. *3*, 723–733.

Tuncbag, N., Gursoy, A., and Keskin, O. (2011). Prediction of protein-protein interactions: unifying evolution and structure at protein interfaces. Phys. Biol. *8*, 035006.

Vajda, S., and Kozakov, D. (2009). Convergence and combination of methods in protein-protein docking. Curr. Opin. Struct. Biol. *19*, 164–170.

Valerio, L.G., Jr. (2009). In silico toxicology for the pharmaceutical sciences. Toxicol. Appl. Pharmacol. *241*, 356–370.

Verdonk, M.L., Giangreco, I., Hall, R.J., Korb, O., Mortenson, P.N., and Murray, C.W. (2011). Docking performance of fragments and druglike compounds. J. Med. Chem. *54*, 5422–5431.

Villoutreix, B.O. (2002). Structural bioinformatics: methods, concepts and applications to blood coagulation proteins. Curr. Protein Pept. Sci. *3*, 341–364.

Villoutreix, B.O., Renault, N., Lagorce, D., Sperandio, O., Montes, M., and Miteva, M.A. (2007). Free resources to assist structure-based virtual ligand screening experiments. Curr. Protein Pept. Sci. *8*, 381–411.

Villoutreix, B.O., Eudes, R., and Miteva, M.A. (2009). Structure-based virtual ligand screening: recent success stories. Comb. Chem. High Throughput Screen. *12*, 1000–1016.

Villoutreix, B.O., Laconde, G., Lagorce, D., Martineau, P., Miteva, M.A., and Dariavach, P. (2011). Tyrosine kinase syk non-enzymatic inhibitors and potential anti-allergic drug-like compounds discovered by virtual and in vitro screening. PLoS ONE *6*, e21117.

Walters, W.P., Stahl, M.T., and Murcko, M.A. (1998). Virtual screening: an overview. Drug Discov. Today *3*, 160–178. 10.1016/S1359-6446(97)01163-X.

Waring, M.J. (2009). Defining optimum lipophilicity and molecular weight ranges for drug candidates: molecular weight dependent lower logD limits based on permeability. Bioorg. Med. Chem. Lett. *19*, 2844–2851.

Waring, M.J., and Johnstone, C. (2007). A quantitative assessment of hERG liability as a function of lipophilicity. Bioorg. Med. Chem. Lett. *17*, 1759–1764.

Wei, D., Jiang, X., Zhou, L., Chen, J., Chen, Z., He, C., Yang, K., Liu, Y., Pei, J., and Lai, L. (2008). Discovery of multitarget inhibitors by combining molecular docking with common pharmacophore matching. J. Med. Chem. *51*, 7882–7888.

Wells, J.A., and McClendon, C.L. (2007). Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature *450*, 1001–1009.

Wiggers, H.J., Rocha, J.R., Cheleski, J., and Montanari, C.A. (2011). Integration of ligand- and target-based virtual screening for the discovery of cruzain inhibitors. Mol. Inform. *30*, 565–578. 10.1002/minf.201000146.

Wilson, G.L., and Lill, M.A. (2011). Integrating structure-based and ligand-based approaches for computational drug design. Future Med. Chem. *3*, 735–750.

Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. FEBS J. *275*, 1–21.

Wolber, G., Seidel, T., Bendix, F., and Langer, T. (2008). Molecule-pharmacophore superpositioning and pattern matching in computational drug design. Drug Discov. Today *13*, 23–29.

Woollard, P.M. (2010). Asking complex questions of the genome without programming. Methods Mol. Biol. *628*, 39–52.

Woollard, P.M., Mehta, N.A., Vamathevan, J.J., Van Horn, S., Bonde, B.K., and Dow, D.J. (2011). The application of next-generation sequencing technologies to drug discovery and development. Drug Discov. Today *16*, 512–519.

Workman, P., and Collins, I. (2010). Probing the probes: fitness factors for small molecule tools. Chem. Biol. *17*, 561–577.

Wu, J.S., Peng, Y.H., Wu, J.M., Hsieh, C.J., Wu, S.H., Coumar, M.S., Song, J.S., Lee, J.C., Tsai, C.H., Chen, C.T., et al. (2010). Discovery of non-glycoside sodium-dependent glucose co-transporter 2 (SGLT2) inhibitors by ligand-based virtual screening. J. Med. Chem. *53*, 8770–8774.

Xiao, S.J., Zhang, C., Zou, Q., and Ji, Z.L. (2010). TiSGeD: a database for tissue-specific genes. Bioinformatics *26*, 1273–1275.

Xie, L., and Bourne, P.E. (2011). Structure-based systems biology for analyzing off-target binding. Curr. Opin. Struct. Biol. *21*, 189–199.

Yang, L., Wang, K.J., Wang, L.S., Jegga, A.G., Qin, S.Y., He, G., Chen, J., Xiao, Y., and He, L. (2011). Chemical-protein interactome and its application in off-target identification. Interdiscip. Sci. *3*, 22–30.

Yildirim, M.A., Goh, K.I., Cusick, M.E., Barabási, A.L., and Vidal, M. (2007). Drug-target network. Nat. Biotechnol. *25*, 1119–1126.

Yuriev, E., Agostino, M., and Ramsland, P.A. (2011). Challenges and advances in computational docking: 2009 in review. J. Mol. Recognit. *24*, 149–164.

Zhong, S., Macias, A.T., and MacKerell, A.D., Jr. (2007). Computational identification of inhibitors of protein-protein interactions. Curr. Top. Med. Chem. *7*, 63–82.

Zhong, S., Zhang, Y., and Xiu, Z. (2010). Rescoring ligand docking poses. Curr. Opin. Drug Discov. Devel. *13*, 326–334.

Zoete, V., Grosdidier, A., and Michielin, O. (2009). Docking, virtual high throughput screening and in silico fragment-based drug design. J. Cell. Mol. Med. *13*, 238–248.