

# Computational Identification of Inhibitors of Protein-Protein Interactions

Shijun Zhong, Alba T. Macias and Alexander D. MacKerell Jr.\*

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, 20 Penn Street, Baltimore, MD 21201, USA

**Abstract:** The ability to control protein-protein interactions (PPIs) for therapeutic purposes is attractive since many processes in cells involve such interactions. Recent successes in the discovery of small molecules that target protein-protein interactions for drug development have shown that targeting these interactions is indeed feasible. In the present review the use of computer-aided drug design (CADD) via database screening or docking algorithms for identifying inhibitors of protein-protein interactions is introduced. The principles of database screening and a practical protocol for targeting PPIs are described. The recent applications of these approaches to different systems involving protein-protein interactions, including BCL-2, S100B, ERK and p56lck, are presented and provide valuable examples of inhibitor discovery and design.

**Keywords:** Protein-protein inhibition, virtual database screening, docking, inhibitor identification, computer-aided drug design, BCL-2, S100B, ERK, p56lck.

## INTRODUCTION

Protein-protein interactions (PPIs) regulate numerous cellular processes including signaling pathways, morphogenic pathways, and complex molecular machines, among others [1-4]. Design of inhibitors for PPIs is important for the control and modification of these processes and can be valuable research tools as well as potential therapeutic agents [5]. Indeed the use of chemical agents to block PPIs comprise a large part of the recently introduced field of chemical biology [6, 7].

Low-molecular weight chemical compounds bound to a protein can prevent the formation of protein dimers which are often the active form [8, 9]. Some proteins function in trimeric or higher multimeric states which can be prevented by inhibitors which may bind to the dimer to block the association of the higher order structure [10]. Some inhibitors can also favor dimer dissociation [11]. More significant with respect to chemical biology are protein heterodimers and higher order structures. Blocking the formation of such multimeric systems may be used to alter a variety of vital processes including cellular signaling, gene regulation, immunity and metabolism among others. Blocking PPIs is also useful in the field of chemical genetics [12-14] where small molecules are used to directly change the way proteins work in real time, rather than indirectly by manipulating their structure via genetic manipulation. Accordingly, it is desirable to have approaches to rapidly and systematically obtain inhibitors of PPIs that have both adequate binding affinity and specificity. Towards this goal, computational chemistry methods hold great promise; the remainder of this review will give an overview of the application of computational (aka. *in silico* or virtual) methods to PPI inhibitor identification, with emphasis on low-molecular weight organic compounds as the inhibitors.

Computer-aided drug design (CADD) refers to the application of a variety of computational methods to drug discovery and design. CADD may be divided into two categories. One category is target- or structure-based drug design (SBDD) [15] which use 3D structural and relevant biological information of a target protein to identify chemical entities with a high probability of binding to the target protein. The identified compounds, often referred to as hits, are then subjected to experimental tests to identify active or lead compounds. These methods include database screening based docking or [16, 17] molecular modeling by means of either molecular mechanical (MM) and quantum mechanical (QM) approaches [18, 19]. Alternatively, for a given binding site, novel ligands can be built from fragments that complement the geometric and chemical characteristics of the binding site. This approach is normally referred to as *de novo* ligand design [20]. Ligands designed by *de novo* approaches are usually validated by computational methods like docking at a high level of accuracy followed by synthesis and biological assay. However, the requirement for synthesis at the initial stage of the discovery process typically make *de novo* methods less attractive than database screening or docking methods that employ databases of available compounds.

The second category is ligand-based design methods [21] which use information about known active ligands as the basis for the design of novel lead compounds. These methods include combinatorial library techniques [22-24] which generate hits from chemical building blocks often starting from known peptidic inhibitors, quantitative structure activity relationship (QSAR) [24-26] analyses, which use statistical techniques to derive empirical activity formulas expressed in terms of compound functional groups and physiochemical properties, and pharmacophore approaches [27-30], which match activity with the type and spatial relationship of functional groups in known active compounds. Usually, a *de novo* design process also contains a pharmacophore step, i.e. adding functional groups to provide a given biological activity so that it combines both

\*Address correspondence to this author at the Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, 20 Penn Street, Baltimore, MD 21201, USA; E-mail: alex@outerbanks.umaryland.edu

structure-based and ligand-based designs [31, 32]. While ligand-based approaches have and continue to be of great utility, they require a training set of known compounds for which experimental activity is available. However, in the case of most PPIs of interest such data is not available while structural data of the proteins is accessible. Accordingly, SBDD typically has greater utility for the identification of lead compounds with the potential to be inhibitors of PPIs.

SBDD computational methods aim at i) identifying low-molecular weight compounds that bind to the protein (i.e. lead compound identification [33]) and ii) optimization of the lead compounds to improve their affinity, specificity, and ADME (Absorption, Distribution, Metabolism, and Excretion) properties as required to produce viable drug candidates. This review will focus on lead compound identification as performed using virtual database screening, also known virtual or *in silico* high throughput screening. In virtual database screening a million or more molecules, which are often commercially available, are computational "posed" into a putative binding site on the protein-protein interface and assigned a score (see details below), with the top scoring compounds ultimately selected for experimental testing. The power of virtual screening is emphasized in a study identifying inhibitors of tyrosine phosphatase 1B, where the hit rate was improved by 1,700-fold over the use of experimental high throughput screening (HTS) alone [34]. A number of recent reviews address the various docking methodologies and their applications [17, 21, 35-62].

Concerning PPIs several reviews covering inhibitors for particular systems have been presented. For example, Grb2 and Grk family adaptors which are mainly composed of Src homology 2 (SH2) and Src homology 3 (SH3) domains serve as docking sites for signaling proteins including various receptors, cytoplasmic kinases and GTPase regulators [63]. Cyclin-dependent protein kinases (CDKs) are another class which function as essential regulators of cell growth and differentiation related to the treatments for cancer, neurological disorders, and infectious diseases [64]. Other examples include the interaction of Abeta with acetylcholinesterase (AChE), which is one of the several proteins associated with amyloid plaque deposits [65] and the possible effects of HIV-1 protease inhibitors against the severe acute respiratory syndrome (SARS) coronavirus (SARS-CoV) protease [66], Hsp90 and p23 [67].

In addition, more recent work on PPIs using docking methods include the identification of low-molecular weight inhibitors of the tumor necrosis factor-alpha (TNF $\alpha$ ) [10], the calcium-dependent S100B-p53 tumor suppressor interaction [68, 69] and the SH2 domain of p56lck [16, 25]. Alternative screening methods based on similarity analysis (see below) have been used for the validation of compounds inhibiting the SH2 domain of p56lck from peptide binding as true leads for optimization [16, 25]. These and the applications listed above indicate that database screening methods are feasible for identification of the inhibitors of PPIs.

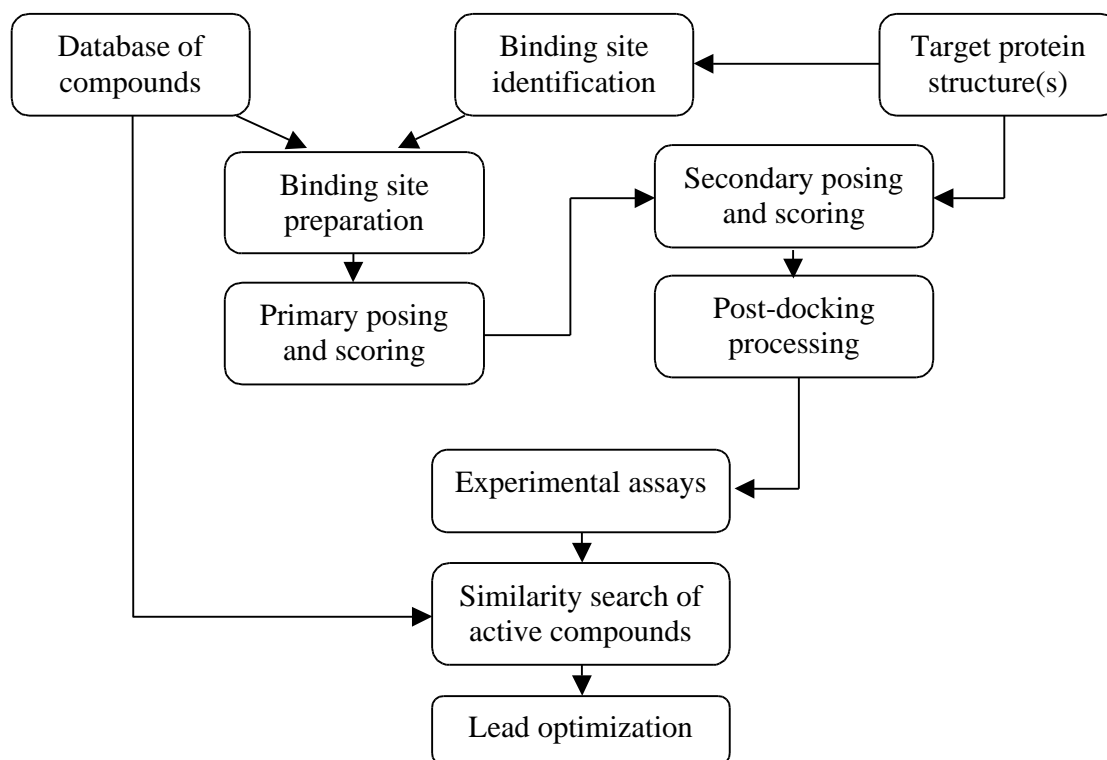
Identification of proteins that are appropriate for inhibition (e.g. "druggable") is one of the key steps in ensuring success for PPI inhibitor development. PPIs often involve large surface areas on the interacting proteins. These surfaces are often relatively flat and, in many cases, the

binding sites are exposed only after conformational changes associated with multimer formation [70-73]. Developments in computational and experimental methods have allowed the identification of key residues involved in PPIs, with these residues often grouped together in small areas called hot spots [74, 75]. The presence of well-defined grooves or cavities in or near the hot spots makes these sites good putative binding sites for drug development. Careful evaluation of these hot spots, often based on sequence homology, may also allow for control over the specificity of the inhibitor-protein interactions [76]. Since many proteins involved in PPIs have different binding sites in the protein-protein interface which are used to bind to different substrates, adequate specificity of identified inhibitors is obviously essential for their use as both research tools to elucidate, for example, signal transduction pathways, and as therapeutic agents.

The remainder of this review is organized as follows. A general overview of virtual database screening relevant to PPIs is given followed by details of selected docking algorithms. This is followed by a discussion of methodological details associated with docking including binding site selection and preparation and considerations related to conformational sampling. A section is then dedicated to an overview of scoring functions used in virtual screening, following with information on *in silico* databases available for screening, along with approaches required for the preparation of such databases and the utility of chemical fingerprint based approaches for database searching. Finally, several examples of systems involving PPIs that have been subjected to virtual screening, including Bcl-2, p56lck, S100B and ERK, are presented.

## VIRTUAL DATABASE SCREENING

An overview of the virtual database screening process is shown in Fig. (1). To initiate the screening project the 3D structure of the target protein from either X-ray crystallography or NMR spectroscopy is required. If there is no experimental structure available, homologous proteins can be found based on sequence alignment following which a structure may be determined via homology modeling [77-79]. Once a 3D structure is available it is necessary to identify an appropriate binding site. This may be performed in a purely computational fashion, though it is preferable to use as much experimental information as possible to identify the putative binding pocket (see below). Once the structure and binding site are known, a docking procedure is applied to identify compounds with a high probability of binding based on the applied posing approach and the scoring function. In our laboratory, screening of large *in silico* databases (i.e. hundred of thousands of compounds or more) is divided into two steps followed by additional analysis (i.e. post-dock processing). Preliminary or primary docking, from which 20,000 to 50,000 top scoring compounds are selected, is designed to run quickly at the expense of accuracy. Selected compounds are then subjected to secondary docking which uses a higher level of accuracy. In both the primary and secondary docking the ligand is treated as flexible, with the additional level of accuracy in secondary docking obtained via additional structural relaxation. Typically the protein structure is kept rigid due to computational



**Fig. (1).** A flowchart of the virtual database screening protocol.

considerations, though the use of multiple conformations of the protein in secondary docking can partially account for flexibility. For the secondary docking the final compounds are selected for experimental assay, often following some additional selection criteria. The experimental step will identify those selected compounds that have the desired biological activity; these represent the lead compounds obtained from the screening process. The true utility of lead compounds for lead optimization studies may then be determined by obtaining structurally similar compounds via similarity screening where, in the case of a good lead compound suitable for additional optimization, the majority of structurally similar compounds will be active [25].

A wide range of experimental approaches may be used to evaluate compounds selected from the virtual screening (see Applications section, below). The most common quantities measured in experiments are  $IC_{50}$ ,  $EC_{50}$ , and  $K_i$ .  $IC_{50}$  represents the concentration of an inhibitor that is required for 50% inhibition of an enzyme *in vitro*.  $EC_{50}$  is the effective concentration at which 50% of the maximal biological response is attained.  $K_i$  is the inhibition constant. Any of these values are useful for experimentally ranking ligands, with the value used based on the type of assay being performed.

In general, it is necessary to select an experimental approach that allows for a minimum of 50 compounds from the virtual screen to be assayed. In our experience, approximately 5% of selected compounds show activity in the micromolar range, such that assaying 50 compounds will yield 2-3 hits. This "limited" success rate is due to limita-

tions in the screening process, including limitations in the ability to sample all accessible conformations of the inhibitors and the binding pocket and due to limitations in the scoring function. Accordingly, virtual screening is designed to limit the number of false negatives (i.e. active compounds that are not selected) by selecting a number of false positives (i.e. inactive compounds that are selected) leading to identification of an adequate number of true positives.

## DOCKING

Docking is the general term used for the procedure to generate reasonable bound conformations (i.e. posing) and then assigning a score for the two participants: the target protein and the ligand. Scoring functions, defined according to geometric, chemical, energetic, or any other empirical criteria, are employed to direct the posing process as well as to rank the binding pose of each ligand being tested; these will be discussed below. A docking program is a combination of subroutines for conformational searching and binding pose ranking. In the 1980s, docking became computationally practical although the suggestion of using docking was proposed earlier. The first widely used docking program is DOCK developed by Kuntz in 1982 [80, 81] and is the program used in our laboratory for the majority of screening projects. DOCK initially adopted a geometric scoring criteria and later was extended to energy based scoring functions for ranking the binding poses. As stated above, DOCK as well as most docking programs keeps the protein structure rigid during the docking process. In DOCK ligand flexibility is introduced via an anchor-based approach [81]. In this approach one or more anchors, typically ring

systems, for the ligand being docked are selected. These anchors are then directed into the putative binding site by overlaying them onto a set of spheres that define a negative image of the receptor binding site (i.e. a set of spheres that fill the binding pocket). Once an anchor is placed, the remainder of the molecule is built in a layer by layer fashion, where the new substituents associated with each layer are rotated about the connecting bond to identify the lowest conformation of the substituents in that layer prior to adding the next layer. The process is continued until the entire ligand is built. More rigorous sampling of possible ligand conformations is performed by placing the anchor fragments in multiple orientations prior to the build-up procedure, such that a large number of copies of a given ligand are built, from which the lowest energy ligand is selected. The buildup procedure can be further enhanced by including additional energy minimization of the ligand as the process proceeds. Such additional minimization is performed in the secondary docking approach used in our laboratory; however, this extra minimization increases the computational demand by close to a factor of 10. Once the ligand is built (i.e. posed) a final score is assigned and the next ligand is built into the binding site. It should be noted that the posed ligand can then be rescored using other scoring functions; an approach referred to as consensus scoring uses a combination of results from different scoring functions to perform the final ranking of the ligands [82]. Another relevant point is the inherent parallelizability of database screening. By simply splitting the database across multiple computers each portion of the database may be screened simultaneously, with the final scores for all the ligands collected once all the databases have been screened and the final ranking performed.

Beyond DOCK a wide variety of docking programs are available that use a collection of posing methodologies and scoring functions. For the sake of completeness following is a list of most docking programs and algorithms presented without evaluating their quality or efficiency. These include 3D-Dock Suite (FTDock [83], RPScore [84] and MultiDock [85]), Affinity (Insight II), AutoDock [86], BiGGER [87], CDOCKER [88], ChemScore [89], ClusPro [90], COMBINE [91], CONCOORD/DISCO [92], DOCK [81], DockVision [93], DOT [94], ECEPP/3 [95], FDS [96], FlexX [97], GAPDOCK [98], GasDock [99], GEMDOCK [100], Glide [101, 102], GOLD [103], GRAMM [104], HADDOCK [105], ICM-DISCO [106], ICM-Dock [107], IFREDA [108], LibDock [109], LIGPLOT [110], LigandFit [111], PatchDock and SymmDock [112], PMF\_score [113,114], PSI-DOCK [115], RDOCK [116], SITUS [117], Surflex [118], TreeDock [119], and ZDOCK [120]. Although some comparison studies have been performed on the more commonly used programs [121-124], it should be emphasized that it is difficult to compare different docking programs/algorithms due to the different approaches generally being optimized for a particular purpose. However, if used judiciously, most docking procedures can fulfill the purpose of finding candidate lead compounds from an *in silico* databases within acceptable time limits.

As stated above, one method to account for protein flexibility in docking is via inclusion of multiple conformations of the target protein in the process. Each ligand is docked against each conformation with the best

score for that ligand amongst the protein conformations used for the overall ranking of that ligand. Alternate conformations of a protein may be obtained from different crystal structures, from various structures typically available from an NMR structure determination or from a molecular dynamics simulation performed with CHARMM [125] or another of the available computational chemistry programs [126]. Alternatively, some docking programs do explicitly include protein flexibility, such as FlexX [127], although this will significantly increase the computational requirements for docking each ligand. Another way to account for protein flexibility is soft-core potentials, which allow some overlap of ligand and protein atoms by lowering the Lennard-Jones repulsion between ligand and protein atoms [128]. In the end all of the above methods only partially account for the target-protein flexibility, a limitation whose impact must be considered in all docking studies.

An important step in the development and acceptance of any methodology, such as virtual docking, is a rigorous and independent validation of the approach. Towards this goal was the implementation of a community wide experiment, the Critical Assessment of Predicted Interactions (CAPRI, <http://capri.ebi.ac.uk/>) in 2000 [41, 129-133]. Motivation for CAPRI is based on the Critical Assessment of Techniques for Protein Structure Prediction (CASP) [134, 135]. In CAPRI it was shown that most known small molecule-protein complexes can be reproduced by docking methods although none of docking algorithms is able to predict all of them correctly.

## PROTEIN STRUCTURE CONSIDERATIONS

An essential component of any virtual database screening project is an appropriate protein structure and a well-defined binding pocket. Protein structures, which are experimentally determined by NMR or X-ray diffraction, can be obtained online at the protein data bank (PDB <http://www.rcsb.org>). If the structure of a receptor is not available, homologous protein with known 3D structures may be used to develop a 3D model of the target protein via a variety of modeling approaches [77-79]. These homologous proteins are selected based on sequence alignment by alignment tools including BLAST [136-138], ClustalW [139, 140], MAP [141], and SAM [142].

When using an experimental structure from X-ray crystallography the position of the hydrogens are typically not present or poorly defined. Therefore, the positions of the hydrogen atoms should be optimized through computational means. This may be performed using a variety of computational chemistry packages (e.g. AMBER [143], CHARMM [125], GROMOS [144], Sybyl (Tripos Inc.), or Insight (Accelrys). In the structures from NMR data, often 10 or more structures are given, though in some cases a single average structure is presented. While NMR structures have well defined positions for the majority of hydrogens it is still useful to minimize the position of the hydrogens, especially in the case of averaged structures. Beyond hydrogens, it is often useful to minimize the entire structure of the target protein to remove bad contacts in the experimental structures. Such minimizations are typically done in the presence of harmonic restraints on the non-hydrogen atoms

to insure that the 3D structure does not significantly deviate from the experimental structure.

An essential consideration when preparing a protein for docking is the protonation state of ionizable residues on the protein. While typical  $pK_a$  values for the amino acids are known, the actual protonation state of some amino acid residues can change significantly based on their chemical surroundings. When considering ionization states the most important case is histidine, which can be in one of two neutral states (i.e. the N 1 or the N 2 atom protonated) or in the fully charged +1 state. Other residues that must be considered are the acidic residues Asp and Glu as well as cysteine, tyrosine, lysine and arginine, although these residues will typically be in their standard ionization states due to their  $pK_a$  being relatively far from pH 7. However, when any ionizable amino acid is located in the binding site being targeted, possible alternate ionization states should be considered. Online  $pK_a$  prediction algorithms are available (e.g. <http://biophyscs.cs.vt.edu/H++/>); however, their accuracy is difficult to gauge.

### BINDING SITE CONSIDERATIONS

Computational determination of putative binding pockets typically requires determination of the protein surface followed by identification of concave regions into which ligands may potentially bind. The surface representation can be the van der Waals (vdW) surface [145] which is defined by a collection of points based on vdW radii of the receptor atoms. The program MIDAS [146], which was recently replaced by Chimera [147], creates the vdW surface around each atom followed by removal of areas of overlap to create a Corey-Pauling-Koltun (CPK) type model. The currently used solvent-accessible surface by the DOCK program, which can be generated by the program dms in the MIDAS package, is defined by rolling a theoretical water 'probe' sphere around the vdW surface of the molecule and using the contact and reentrant points to determine the surface. A helpful algorithm for analytically calculating smooth three-dimensional contour was proposed by Connolly [148, 149]. Therefore, the solvent-accessible surface generated by the Connolly algorithm, referred to as a Connolly surface, involves a network of convex, concave, and saddle shape surfaces. In other words, it is comprised of spherical contact faces, spherical reentrant faces and toric reentrant faces. A coarse approximation of the Connolly surface is the Alpha shape, which is made of a set of "alpha-exposed" points. Each pair of "alpha-exposed" points forms a circle of radius alpha, which contains no other points. All "alpha-exposed" points define an enclosed region to represent the molecular shape. Smaller radius alpha increases the resolution of the region. The alpha shape is used for fast surface matching because there are few points to consider. In addition, a number of algorithms have been proposed to generate the molecular surface using a fixed solvent probe radius or a variable-radius solvent probe [150-160].

Based on a molecular surface, putative binding pockets may be identified by a variety of algorithms. The Lenhoff technique computes a complementary surface for the receptor to determine possible positions for locating the atom centers of the ligand [161]. Kuntz and coworkers generate clustered spheres using the program SPHGEN. Each sphere

is related to a pair of points *i* and *j* that lie on the molecular surface and is centered on the normal at point *i*. Therefore, the cavities on the receptor can be represented by clusters of overlapping spheres which are used to place the anchor used in ligand placement described above.

While computational approaches alone can delineate binding sites, ideally, such methods should be used in conjunction with experimental data to select the final binding pocket for virtual screening. In the case of systems where an experimental structure of a ligand-protein complex is available (e.g. inhibitors in enzyme active sites) the process is trivial; however, for PPIs the binding site must be selected from an extensive protein-protein binding interface. Useful experimental data for this purpose often comes from mutational studies. Mutation(s) may cause significant changes in the affinity of PPIs and thus indicate important residues. Examples of the utility of mutation data in defining binding pockets are given in the applications section below.

For inhibitors of PPIs, the binding site should be located on the protein-protein interface [2]. These interfaces have contact areas varying from 550-4900Å<sup>2</sup> and are typically dominated by steric and hydrophobic interaction, although there also exist electrostatic interactions and hydrogen bonds [162]. The latter types of interactions may be considered desirable to include in a binding pocket as they may facilitate the selection of more specific inhibitors. Visual inspection of the protein surface is helpful for selecting a binding pocket, especially following the placement of spheres as used in DOCK. Such inspection along with consideration of residue types, experimental mutation data and homology data should then be used to select several potential sites for docking.

Once several possible sites are selected, test docking may be undertaken to identify a suitable binding pocket for the full docking study. In our laboratory test docking involves docking a set of 1000 diverse compounds. The spatial distribution of the ligands along with the extent of the ligands overlapping with the binding site are then used to rate the site via a term referred to as the binding pose [163]. A larger binding pose is associated with sites in which a significant portion of the initial ligands remains in the binding pocket and the binding orientation of all the ligands are similar. Comparing the binding pose between the different sites being tested is then used to select the final site for docking. Alternatively, one can perform docking calculations simultaneously on several putative binding sites and rank the ligands by scores adjusted by a statistical correction with respect to the possible sites, a method referred to as the multiple active site correction (MASC) [164]. Another approach related to this correction is a cluster analysis of the distance between sites [165]. These MASC treatments can be also used to consider cross docking of multiple receptor structures.

### CONFORMATIONAL SAMPLING

As discussed above docking is typically performed by explicitly including ligand flexibility, with the protein structure typically treated as rigid. The latter, undesirable condition is generally necessary due to computational speed limitations. Despite this numerous methods do include protein flexibility to varying extents. The flexibility of the

protein structure can be considered in three ways. The first is cross-docking, based on docking to multiple conformations of a protein. These conformations can be those found in different crystals, from a single crystal that contains multiple, non-symmetric copies of the protein, from multiple NMR structures or generated by molecular dynamics simulations [166, 167]. The second approach is docking against an ensemble of the receptor conformations, which again requires the collection of an ensemble of conformations [168, 169]. The third method is direct flexible docking, where the protein conformation is sampled along with that of the ligand during the posing process [106, 108, 170-179]. While this may be considered the most "correct" approach it is typically the most expensive method because of the change of the coordinates of important parts on the receptor. While such methods allow for limited adjustments of the side-chains and backbone conformation in the vicinity of the binding pocket during ligand posing, they generally do not consider larger scale conformational changes that can occur upon ligand binding.

A variety of approaches have been used to include protein flexibility in docking. These include fast Fourier transform surface matching algorithms [180, 181] (MolFit, MULTIDOCK, DOT, GRAMM, ZDOCK), geometric hashing [182, 183] (PPD, PatchDock, BUDDA), genetic algorithms [86, 98, 99, 103, 184-192] (GAsDock [99], GAPDOCK [98]), Monte Carlo sampling [93, 95, 193-196] (ECEPP/3 [95], ICM-DISCO, MCDOCK [194]), and molecular or Brownian-type mechanics procedures (SmoothDock, TSCF). Recently, the impact of the inclusion of receptor flexibility in several docking methods has been presented [121].

Concerning ligand flexibility, several methods beyond the anchor-based method used in DOCK (see above) have been developed. These include surface matching methods, genetic algorithms and Monte Carlo based approaches. Generally, the "better" an approach, as defined by the ability to more rigorously sample conformational space, the computationally more demanding the approach. This compromise requires the selection of a balance between sampling ability and number of compounds being screened. Thus, for an extensive screen of a database of 1 million or more compounds, the computationally efficient anchor search method may be the most appropriate. However, in cases where a smaller number of ligands are to be docked and it is desirable to maximize the potential of identifying more accurate poses alternate methods, including those described below should be considered.

The surface matching methods aim at finding the transformation that will yield the maximum number of matching surface points for the receptor and the ligand. Geometric hashing algorithms [182, 183] solve the surface matching problem by a series of special steps. First, all triplets of atoms on the surface of both the receptor and the ligand are computed and their unique coordination systems are kept in a so-called Hash table. Second, the triplets from the receptor are compared to the ones from the ligand. Then, the set of triplets that yields the maximum number of matches are used to derive the transformation, T, that results in the best least-squares match between corresponding

triplets. Finally, the transformation T is applied to see if the resulting transformation abides by geometric and energetic constraints. If not, another receptor triplet will be used for the search. In Surflex, a flexible molecular docking algorithm, the search engine relies on a surface-based molecular similarity method as a means to rapidly generate suitable putative poses for molecular fragments [118].

Genetic algorithms as general tools for optimization were introduced in the 1970s and developed in the past decade [86, 98, 99, 103, 184-192]. First, alternative ligand positions and conformations are generated, forming an initial population of solutions. The initial population is then evolved, meaning that only the "most fit" or lowest energy ligand positions survive into the next generation. Then, new "individuals" or ligand states are generated through mutation and crossover. With each successive generation, "more evolved" or lower energy ligand positions are produced, thereby finally achieving the "most evolved" or lowest energy binding modes.

In the Monte Carlo (MC) methods [93, 95, 193-196], ligand positions near the active site are randomly generated and subsequently optimized, with the selection of steps (i.e. conformational changes) performed via the Metropolis criteria. Use of this approach allows for increases and decreases in interaction energy with the protein, where the former are important to overcome energy barriers between low energy conformations. Usually, multiple steps are combined in Monte Carlo procedures. In the first step positions and conformations are sampled and the interaction energy minimized starting from an approximate pre-selected docking position. The resulting conformations are further optimized in the second step by a Metropolis Monte Carlo minimization, which optimizes each of these structures [196]. In one more Monte Carlo run following the Metropolis MC step, a pair wise atom potential function is used, and the temperature parameter is slowly lowered during the run effectively minimizing the ligand-protein complex structure (Simulated Annealing) [93]. The MC procedure can be repeated many times, potentially achieving a near exhaustive sampling of conformational possibilities for the ligand-protein complex.

## SCORING FUNCTIONS

Scoring functions are the central part of screening procedures as they direct both the posing as well as yield the final ranking of the docked compounds. There are two classes of scoring functions that are commonly used. The first type is based on energies that include terms describing electrostatic, van der Waals, H-bonding, torsional, and solvation contributions. These are referred to as force field based methods and may include all or only a subset of the listed terms. These types of scoring functions, or their derivatives, are also referred to as empirical scoring functions. The second type is referred to as knowledge-based scoring functions, which may consider geometric and chemical complementarities, requirements based on known biological functions, as well as numerical corrections based on the use of databases and the limitations in conventional scoring functions. In Table 1 examples of a number of scoring function terms representative of both types of functions used in database screening are presented.

Table 1. Example Scoring Functions Used in Virtual Database Screening

eq	typical scoring functions	Ref
1	$E_{\text{inter}} = f_{\text{vdw}} \sum_{i=1}^{\text{lig}} \sum_{j=1}^{\text{rec}} \left( \frac{A_{ij}}{r_{ij}^a} - \frac{B_{ij}}{r_{ij}^b} \right) + f_{\text{elec}} \sum_{i=1}^{\text{lig}} \sum_{j=1}^{\text{rec}} 332.0 \frac{q_i q_j}{(r_{ij})^2}$	[81]
2	$E_{\text{intra}} = \sum_{i=1}^{\text{lig}} \sum_{j>i}^{\text{lig}} F(r_{ij}^{B_{ij}}) + \sum_{k=1}^{\text{dihed}} A [1 - \cos(m_k - \theta_k)]$	[100]
3	$E_{\text{hbond}} = f_{\text{hbond}} \sum_{i=1}^{\text{lig}} \sum_{j=1}^{\text{rec}} (t) \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right)$	[86]
4	$E_{\text{hbond}} = E_{\text{hbond}}^0 e^{-[(r-r_{\text{ep}})^2/d_{\text{hbond}}^2]}$	[197]
5	$G_{\text{sol}} = f_{\text{sol}} \sum_{i=1}^{\text{lig}} \sum_{j=1}^{\text{rec}} S_i V_j e^{(-r_{ij}^2/2 - \dots)}$	[86]
6	$G_{\text{des}} = \sum_{i=1}^{\text{lig}} G_i F_i$	[177]
7	$G_{\text{des}}(\text{ACE}) = \sum_{i,j} n_{ij} e_{ij}$	[198]
8	$\text{PMF} = \sum_{i=1}^{\text{lig}} \sum_{j=1}^{\text{rec}} A_{ij}(r_{ij})$	[113]
9	$G_{\text{bind}} = G_{\text{H-bond}} f(R, \dots) + G_{\text{ionic}} f(R, \dots) + G_{\text{hydrophobic}}  A_{\text{hydrophobic}}  + G_{\text{rotor}} N_{\text{rotor}} + G_0$	[199]
10	$W = \sum_{i=1}^{\text{lig}} \sum_{j=1}^{\text{rec}} W_{ij}(r) + (1 - \dots) \times \left[ \sum_{i=1}^{\text{lig}} W_i(\text{SAS}, \text{SAS}_0) + \sum_{j=1}^{\text{rec}} W_j(\text{SAS}, \text{SAS}_0) \right]$	[200]
11	$E_{\text{pharma}} = \sum_{i=1}^{\text{lig}} \sum_{j=1}^{\text{hs}} w_j(B_{ij}) F(r_{ij}^{B_{ij}})$	[100]
12	$G_{\text{b(exp),j}} = \sum_{i=1}^n D_{ji} + \text{constant}; (j \text{ denotes ligand})$	[86]

Force field methods are typically based on the calculation of intermolecular interaction energies, although intramolecular terms related to strain energies are important and should be considered if possible. These scoring functions basically consists of the vdWs interaction expressed in terms of the Lennard-Jones potential and the electrostatic energy by Coulomb's law. Equation 1 in Table 1 gives the intermolecular interaction containing the vdW and the electrostatic terms. In eq 1,  $r_{ij}$  is the distance between atom  $i$  and  $j$ . and  $A_{ij}$  and  $B_{ij}$  are the Lennard-Jones coefficients

related to repulsion and attraction, respectively. Generally, the Lennard-Jones 6-12 potential is used, as in DOCK [81], however special potential may be employed, such as the 8-4 potential in GOLD [103].  $q_i$  and  $q_j$  are the formal charges and 332.0 is a factor converting the electrostatic energy to kilocalories per mole. The  $(r_{ij})$  is distance-dependent dielectric constant, with a value of  $\epsilon = 4$  typically used in DOCK. A distance-dependent or  $r$ -dependent dielectric is used as it is a simple way to mimic the presence of solvent due to the increase in the dielectric as the atoms move further

apart, which mimics the increase in dielectric as water moves between the atoms. A soft-core Coulombic potential, obtained by adding a constant  $c$  to the distance  $r_{ij}$ , i.e.  $(r_{ij}+c)$ , may also be used as the electrostatic term [201]. For saving computational time, the summation over receptor atoms in the equation 1 can be computed in advance on a grid overlaid on the binding site [202] such that only the summation over ligand atoms has to be computed during the conformation searching to obtain the interaction energy.

The change of the ligand intramolecular geometry leads to changes in the internal, strain or intra-molecular energy. Most docking programs ignore this term or combine this part with other empirical terms. Equation 2 gives an example of the function to calculate the intramolecular energy within a ligand.  $F(r_{ij}^{B_{ij}})$  are either hydrogen bonds or steric interactions specified by the atom type index  $B_{ij}$  between all non-hydrogen atoms. The dihedral rotation term in eq. 2 is originally from the work of Gehlhaar et al. [203]. The values of  $A$ ,  $m$ , and  $\phi_0$  are set to 3.0, 3, and  $\phi_0$  for sp<sup>3</sup>-sp<sup>3</sup> bond; and to 1.5, 6, and 0 for sp<sup>3</sup>-sp<sup>2</sup> bond. Additional parameters may be used for other types of bonds [100, 204].

Hydrogen bonding is another important factor that may be considered. In equation 3, H-bonding is described by a Lennard-Jones type formula with parameters  $C_{ij}$  and  $D_{ij}$  for repulsion and attraction, respectively.  $w(t)$  is a weight depending on the angle,  $t$ , of the H-bond [86]. In GOLD and SYBYL/ G-Score [103], the H-bond term is considered according to the type and the geometry of the donor-acceptor pairs. SYBYL/ChemScore does not distinguish neutral and ionic H-bonds, but adds an additional term for considering the interaction between ligands and metal ions residing in a binding site [205]. Alternatively, in equation 4, hydrogen bonding is calculated as a spherical Gaussian where  $E_{\text{hbond}}^0=2.5$  kcal/mole as an average of various estimates,  $d_{\text{hbond}}=1.4\text{\AA}$  as a radius of the interaction sphere, and  $r_{\text{sp}}=1.7\text{\AA}$  as the reference radius value [197].

The free energy of binding in solvent can be computed according to Hess's law of heat summation,  $G(\text{binding, solution}) = G(\text{binding, gas}) + G(\text{complex, solution}) - G(\text{receptor, solution}) - G(\text{ligand, solution})$ . The last three terms contribute to the solvation effect. As an approximation, the solvation effect can be simply evaluated as the desolvation energy of the small molecule,  $-G(\text{ligand, solution})$ , when the receptor is much larger than the ligand, i.e.  $G(\text{complex, solution}) - G(\text{receptor, solution}) \approx 0$ . Solvation effects can be computed in many different ways, including the use of an  $r$ -dielectric as discussed above. In equation 5 the solvation energy is based on the function of Stouten et al. [206]  $S$  and  $V$  denote the solvation parameter and fragmental volume, respectively. The exponential term is an envelope function with a constant of  $r_0 = 3.5\text{\AA}$ . Another approach, shown in equation 6 is based on a desolvation penalty for each ligand [177] obtained by summing over the ligand atomic contribution  $G_i$  which may be calculated using the semi-empirical program AMSOL [207, 208]. The contribution is normalized by a contacting factor  $F_i$  which has value of 1 when the atom is completely buried in the

receptor and value of zero when the atom is totally exposed to solvent molecules. Equation 7 has the desolvation free energy based on the atomic contact energy (ACE) [198] denoted by  $e_{ij}$  which is specified by a switch function  $n_{ij}$  in the range 6-10  $\text{\AA}$  in order to avoid a sharp distance cutoff [209]. By implementing the solvent accessible surface area (SASA) approximation [210, 211], accurate solvation models based on the Poisson-Boltzmann equation or Generalized Born equations, e.g. the PBSA (Poisson Boltzmann Surface Area) or GBSA (Generalized Born Surface Area) methods are also possible [212,213] although these methods are computational more expensive making them inappropriate for screening large databases.

Knowledge-based functions have attracted a lot of attention and have been shown to provide practicable predictions of the relative ordering of ligand affinity. Equation 8 is the form of a knowledge-based potential of mean force (PMF) scoring function [113, 114] which is a summation of distance-dependent interaction potentials including enthalpic and entropic effects, over all heavy atom pairs, derived based on statistical analysis of protein-ligand complexes found in the Protein Database Bank. The LUDI scoring function that yields free energies as developed by Bohm is shown in Equation 9 [199]. The neutral and ionic H-bonds are separately treated in the first two terms, each having a reference value scaled by distance- and angle-dependent factors for each interacting pair. The hydrophobic interaction is scaled by a factor based on the hydrophobic buried area. The torsional entropy loss is expressed in the fifth term. The last term is a constant based on regression analysis of available complexes for a given target protein. The F-Score used in FlexF is similar to the LUDI scoring function, but the hydrophobic term is replaced with a term for the interaction between aromatic groups and a general distance-dependent contact term [97]. Equation 10 is that used in DrugScore [200, 214], a knowledge-based potential which combines distance-dependent pair potential and solvent-accessible surface (SAS) dependent terms. Each potential is calculated as the logarithm of an occurrence frequency derived from a large set of protein-ligand complex structures. The factor  $w$  is an adjustable weight with a normal value of 0.5.

Two additional scoring functions are shown in equations 11 and 12. Equation 11 is a pharmacophore-based interaction energy between the ligand and the hot-spot atoms on the protein.  $W_j$  is the pharmacophore weight of the hot-spot atom  $j$  and will be set to zero if the interaction type of the ligand atom does not match that of the hot-spot.  $F(r_{ij}^{B_{ij}})$  is defined as in equation 2. This approach is used in GEMDOCK [100,204]. Equation 12 involves a correction [86] with consideration of quantitative structure-activity relationships (QSAR) data from statistical analysis of the multilinear regressions  $\beta_i$  for selected 2D descriptors  $D_{ji}$  labeled by index  $i$  for each ligand  $j$ .

DOCK [81] uses equation 1 with the option of including an additional desolvation energy term. AutoDock [86] uses equations 1 and 3 for both the inter-molecular and intra-molecular energies. In TreeDock, only a 6-12 Lennard-Jones



potential is used [119]. In PLP, a scoring system implemented in the software Cerius2 [203], all atom pairs are divided into three categories, i.e. hydrogen donor-acceptor pairs, repulsive donor-donor or acceptor-acceptor pairs, and generic dispersion of other contact pairs, each being described by the Lennard-Jones potential. In G-Score [103], the scoring function contains a Lennard-Jones 8-4 potential, a H-bond term, and an internal ligand energy term similar to the terms in the equation 2. X-Score uses the vdW interaction based on an 8-4 Lennard-Jones term, a hydrogen bonding term, a hydrophobic effect term, a torsional entropy penalty, and a regression constant [215].

Several studies have performed comparisons of some of the above scoring functions [216]. Wang and co-workers presented a comparative evaluation of 11 scoring functions [124] and pointed out that some scoring functions including PLP, F-Score, LigScore, DrugScore, LUDI, and X-Score, yield higher hit-rates. Based on an extensive evaluation of 14 scoring functions [123], they showed better correlation between the affinity and the scores calculated by X-Score, DrugScore, Sybyl/ChemScore, and Cerius2/PLP score. In most docking tools, these scoring functions can be used in combined ways, such as consensus scoring [217-220]. Users can change the weight coefficients  $f$  which dictates the contribution of each energy term to the overall consensus energy score to optimize the performance for each system. Efforts are ongoing in a number of laboratories to improve scoring functions in order to improve hit rates; however to date there appears to be no "best" scoring function that is applicable to all systems. In addition, it should be noted that the scoring functions typically do not have a direct relationship with the binding affinity of the ligands (i.e. the energy scores do not correspond to an absolute free energy of binding).

One of the general limitations with scoring functions based on some type of interaction energy is the tendency for the scoring to favor larger compounds [221]. This is simply due to the greater number of atoms leading to more favorable interactions between a larger ligand and a receptor combined with the omission of desolvation and other contributions that contribute to binding being omitted from consideration in many scoring functions, including that in the program DOCK. The latter contributions will typically be more unfavorable for the binding of larger compounds, balancing the more favorable ligand-receptor interactions. To overcome this limitation we have developed a simple normalization procedure based on the number of non-hydrogen atoms,  $N$ , in a ligand [221]. This involves taking the interaction energy score and dividing it by  $N^x$ , where  $x$  is an empirically fit term. The value of  $x$  is determined based on the resulting distribution of molecular weights of the top scoring compounds; typically a value of  $x$  is selected which yields a collection of compounds with a MW distribution between 100 and 500 with a maximum around 300 daltons. Such a distribution of MWs is desirable for lead-like compounds [33].

## DATABASES AND LIBRARIES

Virtual database screening cannot be performed without an adequate database (or library) of virtual compounds. For each compound or entry in a database, the basic information

stored includes compound name or identification number, atom types, bond types, connectivity, coordinates, formal charges, and chirality. Some databases also contain extra information including molecular properties and descriptors, as well as application fields and classification information like drug categories and reaction types. In general, databases of chemical compounds can be categorized into three kinds, with those types discussed in the remainder of this section. It should be noted that recent studies suggest that the chemical diversity of some existing commercial databases might not include the chemical space describing PPI inhibitors [222].

The first type of database involves collections of as many as possible compounds from selected sources. Most chemical companies provide such databases, which include compounds that are available in stock or can be readily synthesized. Following is a list of some online-searchable chemical databases with their approximate current size in parentheses: Asinex (129,000), Chembridge (371,000), Chemdiv (750,000), Comgenex (120,000), Life Chemicals (385,000), Maybridge (60,000), MDD (33,000), Nanosyn (47,000), NCI (250,000), Sigma\_Aldrich (90,000), Specs (232,000), Timtec (165,000) and Tripos (80,000). A longer list of databases can be found in an evaluation work by Sirois and coworkers [223]. These publicly available databases can be compiled into meta-database like "Ligand.info" [224] which contains more than 1 million low-molecular weight compounds.

The second type of database contains specific compounds selected for certain characteristics or purposes. Examples include the Derwent World Drug Index (WDI) database, which contains almost 80,000 drugs and pharmacologically active compounds, ZINC [225] containing about 728,000 purchasable compounds ready for docking, and SuperNatural [226] which contains a database of approximately 50,000 natural compounds including most drugs on the market. Several databases of ligand-protein complexes are available. SitesBase is a database about the structural similarities between known ligand binding sites found in the Protein Data Bank [227]. LigBase [228] contains approximately 50,000 ligand-protein structures also from the PDB. Ligand-Protein DataBase (LPDB) is designed to use known receptor-ligand complexes to improve docking scoring functions [229]. PDBbind database [230] (<http://www.pdbbind.org/>) provides a comprehensive collection of binding affinities for the protein-ligand complexes in the Protein Data Bank (PDB). SuperLigands [231] and PDB-Ligand [232] are both databases of about 5000 ligands derived from the Protein Data Bank. Relibase [233] is a database that provides a query tool for the analysis of receptor-ligand binding features. GLIDA [234] is a G-protein coupled receptor (GPCRs) related chemical genomic database. ChemMine [235] is a compound mining database for chemical genomics. LIGAND (<http://www.genome.ad.jp/ligand/>) is comprised of more than 7300 compounds, 5000 reactions and 3800 enzymes. A review summarizes some libraries of ligands related to human protein kinases [236].

The third kind of database includes building-blocks or molecular fragments for the purpose of further molecular design. Such libraries are provided by many companies including AKos, Chembridge(PHARMABlock), Chemdiv, Maybridge, and Key Organics Limited. This class of

database is appropriate for *de novo* design or lead optimization efforts.

Compound databases usually exist in one or more of the following three formats. One is the 2-dimensional (2D) structure database file (SDF) based on the MDL convention, which contains 2D mapping coordinates, connectivity, chirality, and bond information. In addition, it is possible to include molecular properties as well as compound identification information required for purchasing selected ligands. SDF is the format that chemical companies typically provide to users. The information in SDF files may act as the starting point for the generation of 3D structures followed by further geometry optimization [16] and are used in several database searching algorithms based on 2D molecular descriptors [237]. The second database format includes 3D structural information. Examples include the protein data bank (PDB) and Tripos MOL2 formats and the SDF format may be extended to contain 3D information. This type of format is normally used for viewing structures and is typically the starting point for virtual database screening studies and these databases also typically include the same information as the SDF database listed above. However, in many cases, in order to minimize space requirements, only the essential information is included such as atom type, bond type, coordinates, partial charges, and connectivity. In addition to their use in visualization and screening the information in 3D format databases may be used to generate 3D molecular descriptors that describe molecular similarity and diversity based, such as the GETAWAY (GEometry, Topology, and Atom-Weights Assembly) descriptors [238, 239]. The third format are searchable databases containing extensive information for each compound including that in the first two database formats as well as additional information. Accordingly, such database can be quite large and are usually stored in binary format to save disk space. Examples include the 2D database files prepared by ISIS (MDL) or other tools that are used for similarity searching, compound clustering based on the fingerprints (see below) or other approaches. These databases are often included in and used by commercially available software packages, including the molecular operating environment (MOE, Chemical Computing Group) and Sybyl (Tripos Inc).

Several points concerning databases need to be emphasized. For performing energy calculations as required for some scoring functions, it is important to have satisfactory 3D structures and high quality partial charges for each ligand. The compounds should be energy minimized to assure that proper bonds and angles geometries are obtained. However, typically 3D conformations in a database do not correspond to the global minimum, although this is not important as ligand flexibility is included in the majority of screening approaches. An alternative are 3D databases in which a large number of conformations of each compounds are stored [225]. In such cases the posing procedure will involve rigid docking of all available conformations of a compound rather than systematically sampling the ligand conformation during docking. Partial atomic charges are historically those from AMBER [143] or from the method of Gasteiger [240]. Recently, improved partial atomic charges are being used including those based on AM1 calculations, a semi-empirical quantum chemical method, combined with a

charge model like CM2 (Charge Model 2) [208] and BCC (Bond-Charge Corrections) [241]. The BCC model, which is implemented in some software packages, including OpenEye (<http://www.eyesopen.com/>), applies bond-charge corrections to the AM1 partial charges on each atom to generate the final partial charges. It should be noted that semi-empirical geometry optimization is still too expensive all the compounds in large databases, such that the charges etc. are based on geometries obtained from molecular mechanics optimization. Another recent addition to 3D databases are free energies of solvation based on the AMSOL semiempirical solvation model [207, 208]; the availability of such values may be useful for predicting desolvation penalties upon binding as well as selecting compounds with proper solubilities for experimental assay. Finally, the compounds in a database should be assigned physiologically relevant protonation states, unnecessary components should be removed (e.g. the counterions in salts) and chirality of compounds taken into account.

#### **POST-DOCK PROCESSING: RULES FOR SELECTION OF FINAL COMPOUNDS FOR EXPERIMENTAL ASSAY**

In spite of the limitations of current docking programs, docking algorithms and their associated scoring functions these approaches are helpful at selecting small molecules with a high probability of binding to the target protein. To facilitate the selection of biologically active compounds from a virtual database screen a number of additional criteria should be considered beyond the direct ranking from the docking itself. This step, as shown in Figure 1, may be referred to as post-dock processing. The factors include maximizing the chemical diversity of compounds for assay, the physical properties of the compounds and additional scoring of the compounds. The remainder of this section will focus on the first two considerations as alternate scoring considerations were discussed above.

Concerning chemical diversity, intuitively it may be considered desirable to subject compounds that cover the widest range of chemical space and are still structurally complementary to the target protein to experimental assay [16]. Indeed, work in our laboratory has shown that selection of chemical diverse compounds from the top scoring compounds leads to improved hits rates as compared to selection of compounds based on their docked scores alone [242]. To maximize the chemical diversity of compounds selected for assay, we apply the following approach. Typically, the top 1000 compounds are selected from the secondary posing and scoring in the virtual screening process based on normalized energy scoring (see Figure 1 and following paragraph). These compounds are then subjected to similarity clustering based on chemical fingerprints [243, 244]. In this approach each compound is assigned a fingerprint that is a binary string of 1024 bits with each bit representing a chemical feature (e.g. atom type, bond type, aromatic ring etc.). The fingerprint (or bit string) for each compound allows for the compounds to readily be compared based on the Tanimoto index, ultimately allowing them to be grouped into clusters of chemically similar compounds. Typically, for a 1000 compounds, approximately 100 clusters of varying size are obtained. One or two compounds

are then selected from each cluster for experimental assay, with the selection process considering physical properties discussed below. This yields a collection of 100 to 200 compounds which are all structurally complementary to the binding site and have maximum chemical diversity for assay. It should be noted that various types of fingerprints are available, including those based on the content (2D) of substructures, pharmacophore points, and biological/physical properties as well as in the spatial arrangement of these pharmacophore points [237, 245-248].

The second consideration during final compounds selection is physical properties. The classic properties considered are based on the rule-of-five proposed by Lipinski in 1997 and reprinted in 2001 [33,249,250]. The rule-of-five was derived from statistical analysis of available drugs and other active, drug-like molecules. The descriptors used in the rule are molecular weight (MW), number of H-bond acceptors (Ha), number of H-bond donors (Hd), and logarithm of the octanol/water partition coefficient estimated by Kowwin (LogP). The values for each of these terms are close to five or a factor of five, i.e.  $MW < 500$ ,  $Ha < 10$ ,  $Hd < 5$ , and  $-5 < \text{ClogP} < 5$ . This empirical rule has been further developed [251] and recently reviewed [252] and alternate rules based on polar surface area and number of rotatable bonds have been presented [253]. In addition absorption, distribution, metabolism and excretion (ADME) considerations can be taken into account when selecting compounds [254], functional groups which are problematic for medicinal chemistry during the lead optimization process can be flagged [255, 256], and checks may be made for promiscuous compounds (i.e. compounds that non specifically inhibit a wide range of proteins) [257, 258]. More detailed consideration of compound selection criteria can be found in Olah and Oprea's work [252].

In our laboratory, the Lipinski's rules are typically implemented in the following manner. The top 1000 compounds are selected based on DOCK interaction energy scores normalized to account for ligand size and subsequently clustered into approximately 100 clusters. The compounds in each cluster are then analyzed with respect to Lipinski's rules and one or two compounds that meet those rules are selected for experimental assay. Emphasis in the selection process is towards smaller compounds, as these are more lead-like versus drug like [33]. However, if none of the compounds in a cluster meet those criteria compounds are still selected from that cluster for analysis. In addition, we do not consider ADME properties at this stage as the goal is to obtain a collection of lead compounds with maximal chemical diversity rather than identifying drugs per se. Also, the possibility of promiscuous compounds is not considered, although experimental assays should be performed to determine if such compounds are identified as active in the experimental assay [257, 258]. Overall, it should be remembered that at the lead identification stage, it is lead-like compounds rather than drug-like compounds that are being selected; undesirable properties of active lead compounds may be modified via medicinal chemistry when optimizing a lead into drug candidate.

## EXAMPLES OF VIRTUAL SCREENING APPLICATIONS

As discussed in the introduction PPIs represent a class of interactions that play a significant role in almost all aspects of biology. A number of reviews on specific systems that involve PPIs were cited above allowing the reader to access information on these systems. In this section an overview of published results for 4 well studied systems is presented. These overviews are designed to allow the topics discussed above to be placed in more concrete terms, yielding a clearer understanding of how CADD may be applied to identify lead compounds that inhibit PPIs.

### BCL-2

Computational methods have been successful in identifying inhibitors of the Bcl-2 protein [285], which is involved in apoptosis (Fig. 2A) [259, 260]. Cancer cells suppress programmed cell death by upregulating Bcl-2 and preventing apoptosis, such that drugs that inhibit Bcl-2 could be used in therapies to eliminate cancer cells. The 'Bcl-2' protein family contains several proteins involved in apoptosis. Bcl-2 and Bcl-xL are anti-apoptotic proteins which inhibit the pro-apoptotic proteins Bax and Bak, and are themselves inhibited by BH3 [261, 262]. Bcl-2 is a good target for the inhibition of its PPIs because it contains a hydrophobic groove on its surface which is known to bind to alpha helices on other proteins involved in apoptosis (Fig. 2A) [263]. In addition, mutations of residues in this groove have been shown to eliminate Bcl-2's biological functions [264].

Virtual screening and homology modeling were used by Wang and co-workers to identify non-peptidic and cell-permeable inhibitors of Bcl-2 [260]. The structure of Bcl-2 was built based on the NMR structure of Bcl-xL, which has 47% sequence identity and then minimized using MSI/Insight II 98 software (Biosym Technologies, San Diego). The program DOCK 3.5 [80, 81] was then used to screen the MDL/ACD database [265], targeting the groove to which the Bak BH3 peptide binds. The top 1000 molecules were then selected and their bound orientations refined via minimization of the ligand-protein complex using Sybyl 6.2 (Tripos, St. Louis). 53 compounds with favorable binding energies, geometries that allowed hydrogen bonding and diverse structural scaffolds were chosen for assay by visual inspection. Compound HA14-1 was ultimately identified as an inhibitor in the  $\mu\text{M}$  range. This compounds allowed apoptosis to occur in human acute myeloid leukemia cells that were overexpressing the Bcl-2 protein.

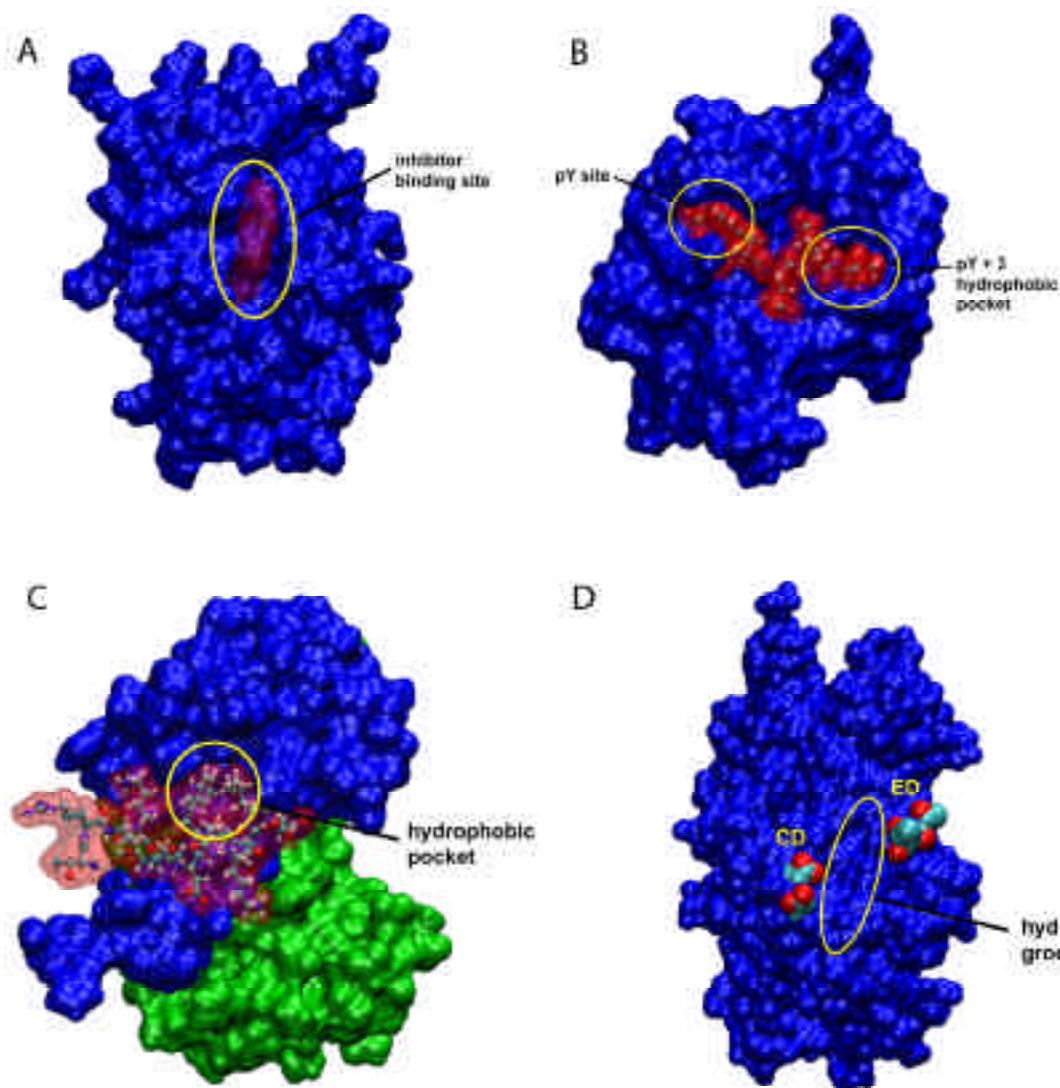
A second study also used high-throughput virtual screening to identify inhibitors of Bcl-2 [259]. As in the previous study, homology modeling was used to produce a 3D structure of Bcl-2. The sequences of Bcl-2 and Bcl-x were aligned using BLAST [136], and the 3D structures created using MODELLER [266], followed by refinement via MD simulations using CHARMM [125]. During the course of that study the experimental 3D structure of Bcl-2 became available allowing the modeled protein to be compared to the reported NMR structure (1BLX) (Petros,

2001). The fold of the simulated structure and the NMR structure were the same and the RMSD for heavy atoms in the BH3 binding site was 1.0 Å, where the binding site was defined as all residues within 8 Å of the Bak BH3 peptide. The NCI 3D database of 206,876 compounds were docked against the binding site using DOCK and ranked using the DOCK energy scoring function. Of the 35 compounds selected for biological assay seven inhibited the binding of Bak BH3 peptide to Bcl-2 and one of these had an  $IC_{50}$  of 4  $\mu$ M in the inhibition of cell growth and could induce apoptosis in cancer cells in a manner that correlated with the Bcl-2 expression level.

### P56LCK

The p56 Lymphoid T cell tyrosine kinase (Lck) protein is a tyrosine kinase from the Src family which is involved in T-

cell mediated immune responses [267, 268]. Lck is an interesting multiple domain protein that participates in both intra- and inter-molecular interactions. The protein consists of three main domains (kinase, SH2, SH3), inter-domain linkers and tails. In its inactive form, the SH2 domain is involved in intramolecular interactions and the kinase is inactive. One of the first steps in immune response is the phosphorylation of a tyrosine located in short-recognition peptides of the membrane-bound CD3 receptor referred to as ITAMs. The Lck SH2 domain has a high affinity for the phosphorylated ITAMs and binding to them leads to activation of the Lck kinase domain [269]. The activated p56lck then initiates a signal transduction cascade ultimately leading to increased interleukin-2 production and stimulation of T cell growth. Accordingly, inhibition of the protein-protein interaction between the SH2 and the ITAM peptides



**Fig. (2).** (A) The cavity of Bcl-2 (red) was based on the complex between the homologous Bcl-xL and BakBH3. The recently reported structure of Bcl-2 bound to an inhibitor (1YSW.pdb) [285] illustrates the site on the protein surface. (B) The p56lck SH2 domain bound to the pYEEI peptide (1LKK.pdb) [286] illustrates the hydrophobic pocket which was targeted in drug design. (C) Two chains of the S100B protein (blue,green) are complexed with the TRTK-12 peptide (1MWN.pdb) [287]. The hydrophobic pocket which was used to identify small molecule inhibitors is highlighted in yellow. (D) The hydrophobic groove between the CD and ED acidic residues on the ERK surface (1ERK.pdb) [288] was used as the site for docking small molecules.

could be of therapeutic value in the area of immunosuppression.

MacKerell, Hayashi and coworkers used virtual screening to identify non-peptide drug-like molecules targeting the Lck SH2 domain [270, 286]. The SH2 domain contains a hydrophobic pocket in the pY+3 binding site (pY: phosphotyrosine), which mutational studies [271] and the crystal structures of different Src SH2 domains and ITAM peptides had identified as a site imparting binding specificity. Fig. (2B) shows one of these ITAMs bound to the Lck SH2 domain and the location of the targeted hydrophobic site. A database of 2 million compounds was prepared in-house from commercially available compounds in a process that involved building the 3D structures of the compounds, assigning charges, and minimizing their structures prior to docking. A two step docking protocol similar to that shown in Figure 1 was then used. First, the compounds were placed in the targeted binding site using the anchored search method in DOCK [80,81] allowing for flexible docking of the ligands. The top 25,000 compounds from the initial screen were then selected based on normalized vdW attractive energies. Total interaction energies have an inherent bias towards highly charged compounds or very large compounds, neither of which typically would make suitable leads. In addition, compounds whose interaction with the receptor is based on electrostatics may not necessarily have high steric complementarity to the binding pocket. To eliminate these biases, compounds were ranked using the attractive van der Waals energies which were normalized according to the number of heavy atoms, as discussed above [272]. The selected compounds were subjected to a second virtual screening (secondary docking) and were then scored according to the total normalized interaction energy. The top compounds were clustered using 2D molecular fingerprints so that each cluster would have dissimilar compounds and representatives from each cluster were chosen based on Lipinski's rules, with 194 compounds obtained for biological assay from commercial vendors. Thirteen of the 194 compounds experimentally tested showed inhibitory activity in mixed lymphocyte reaction assay as well as inhibition of the binding of the ITAM with the Lck SH2 domain. Additional assays including fluorescence titrations and inhibition of Lck SH2 domain-ITAM peptide binding were used to further validate that the compounds were bound to the SH2 domain.

A subsequent study has been used to validate several of the initial actives as lead compounds suitable for optimization studies [273]. This involved identifying structurally similar compounds for 12 of the 13 active compounds via chemical similarity fingerprint analysis. The new compounds were then experimentally assayed. Results showed a majority of the similar compounds to be active for the more active lead compounds, validating them as viable compounds for optimization. In contrast, for some of the less active compounds the similar compounds all had low or no activity, indicating that additional optimization of these was not feasible. However, for two of the lowest activity lead compounds, the majority of similar compounds had good activity with some of the similar compounds being more active than the original lead. Thus, these compounds were "rescued" as viable leads for future development. The results

for these studies were also used to develop structure-activity relationship for the inhibitors as well as identify common binding motifs of the ligands.

### S100B

S100B is a calcium binding protein which binds to the C-terminus of p53 and inhibits its tumor suppressor function [274, 275]. When S100B binds calcium, it undergoes a conformational change that exposes a hydrophobic patch of residues that interact with p53 as seen in Fig. (2C). The presence of this well-defined 'hot spot' deduced from high resolution crystal and NMR structures [287] made S100B an attractive target for the development of PPI inhibitors [276]. Virtual screening of 640,000 compounds from an in-house [16, 277] database of small molecules was done using DOCK [69, 278]. The binding site was defined based on spheres created by SPHGEN that were within 14 Å of the amino acids involved in the S100B-p53 interaction. Compound scoring was based on the attractive vdW interaction energy in order to ensure that the compounds had good steric complementarity with the binding pocket. Two step docking was performed and compounds were chosen from the second step for biological assay based on their drug-like properties, their similarity to drug scaffolds and their chemical diversity. From 60 purchased compounds, seven were soluble and determined to bind to the S100B site by fluorescence spectroscopy with  $K_D$  values between 1 and 120  $\mu$ M and by heteronuclear single quantum coherence NMR titrations. NOEs obtained for one of the compounds showed that the intended hydrophobic area was being targeted. Five compounds inhibited growth of melanoma cancer cells at micromolar concentrations and are currently being subjected to lead optimization studies.

### ERK

The extracellular signal-regulated kinases (ERK1/2) are part of a signaling pathway involved in cell proliferation. ERK1/2, c-Jun-N-terminal kinase (JNK) and p38 kinase are three known mitogen activated (MAP) protein kinases [288] whose unregulated activation has been linked to cancer and inflammation [279-281]. In addition, it is thought that inhibition of ERK may arrest cancer cell proliferation [279-281]. In the ERK1/2 signaling pathway, only two proteins are known to activate ERK1/2 through phosphorylation of residues Thr183 and Tyr185, MAP kinase-1 and 2 (MEK1/2). Upon activation, ERK undergoes a conformational change [282] and phosphorylates its many substrate proteins. In vitro, dozens of proteins can be phosphorylated by ERK. Some are known cognate ERK substrates, like RSK-1 and ELK-1, but others are also involved in other signaling pathways. Thus, ERK1/2 represents an interesting system as it is involved in multiple PPIs, with those different interactions involved in different signal transduction pathways. Accordingly, it would be desirable to selectively inhibit the interaction of ERK1/2 with only one (or a subset) of their substrate proteins.

In a study by Hancock *et al*, virtual screening was used to identify drug-like molecules that could specifically target ERK phosphorylation of some substrates and not others [76, 283]. To attain this goal, rather than inhibition of the active (i.e. catalytic or kinase) site, which is likely to block the

phosphorylation of many substrate proteins, a site in the protein involved only in PPIs was targeted as shown in Fig. (2D). The selected site had been characterized through mutation studies, showing it to be involved in interactions with MEK1/2 and substrates. In the crystal structure of ERK [288] this site was shown to consist of two sets of protruding polar amino acids surrounding a groove in the protein [284] that is located in the hinge region of the kinase opposite to the active site. Of these two sites one is common to other MAP kinases (the CD domain) while the second site (ED domain) affect the specificity of the interactions with substrate proteins. Therefore, this region of the protein was selected as the putative binding site, such that bound inhibitors would only inhibit downstream phosphorylation of ERK substrate proteins.

A database of 800,000 compounds was screened using the DOCK software in a procedure similar to the one described above for the Lck protein. Bias towards highly polar compounds was eliminated through the use of a two step docking algorithm that initially emphasized the vdW attractive interaction energy and ligand size normalization for the primary docking run. 20,000 compounds selected from the primary docking were then subjected to a more rigorous secondary docking with 500 compounds selected based on normalized DOCK total interaction energy scores. The top 500 compounds were then clustered using Jarvis-Patrick and MACC-BITS 2D fingerprints into dissimilar subsets of compounds. One or two compounds with desirable properties based on Lipinski's rules were then selected from each group. 80 compounds were obtained and tested for biological activity. Several compounds inhibited the phosphorylation of two ERK substrates, RSK-1 and ELK-1 and direct binding to the protein was confirmed using fluorescence spectroscopy. Compounds also showed dose-dependent inhibition of proliferation in several cancer cell lines. More recently, inhibitors of ERK targeting the phosphorylated form of the protein have been identified (A.D. MacKerell, Jr. and P. Shapiro, Work in progress) and additional characterization of all the ERK inhibitors identified to date is ongoing.

## SUMMARY

In the present review an overview of various CADD approaches used in the identification of inhibitors of PPIs has been given. Emphasis was placed on issues of particular relevance to PPIs, such as identification of the appropriate binding site to be targeted during virtual database screening. However, the majority of the approaches used for PPIs are relevant to any target protein, making the present overview of more general utility. In addition, considerations associated with the final selection of compounds for experimental assays, referred to as post-dock processing, including the use of similarity clustering to obtain chemical diverse structures, were emphasized. This last step is important to maximize success (i.e. the hit rate) in virtual screen studies as the methods used in the docking approaches including a variety of approximations as required due to computational considerations. Despite the limitations, low-molecular weight inhibitors of PPIs have been identified for a number of different systems indicating the viability of the CADD approaches outlined in this review. Such compounds offer

potential as tools in the field of chemical biology and chemical genetics and have the potential to be developed into novel therapeutic agents for a wide variety of disease states.

## ACKNOWLEDGEMENTS

Support from the University of Maryland, School of Pharmacy Computer-Aided Drug Design Center and NIH grants CA107331 and CA120215 is acknowledged.

## REFERENCES

- [1] Xenarios, I.; Eisenberg, D. Protein interaction databases. *Curr. Opin. Biotechnol.* **2001**, *12*, 334-339.
- [2] Archakov, A. I.; Govorun, V. M.; Dubanov, A. V.; Ivanov, Y. D.; Veselovsky, A. V.; Lewi, P.; Janssen, P. Protein-protein interactions as a target for drugs in proteomics. *Proteomics* **2003**, *3*, 380-391.
- [3] Pagel, P.; Kovac, S.; Oesterheld, M.; Brauner, B.; Dunger-Kaltenbach, I.; Frishman, G.; Montrone, C.; Mark, P.; Stumpflen, V.; Mewes, H. W.; Ruepp, A.; Frishman, D. The MIPS mammalian protein-protein interaction database. *Bioinformatics* **2005**, *21*, 832-834.
- [4] Beuming, T.; Skrabanek, L.; Niv, M. Y.; Mukherjee, P.; Weinstein, H. PDZBase: a protein-protein interaction database for PDZ domains. *Bioinformatics* **2005**, *21*, 827-828.
- [5] Ryan, D. P.; Matthews, J. M. Protein-protein interactions in human disease. *Curr. Opin. Struct. Biol.* **2005**, *15*, 441-446.
- [6] Lampson, M. A.; Kapoor, T. M. Unraveling cell division mechanisms with small-molecule inhibitors. *Nat. Chem. Biol.* **2006**, *2*, 19-27.
- [7] Walsh, C. T. Natural insights for chemical biologists. *Nat. Chem Biol* **2005**, *1*, 122-124.
- [8] Roskoski, R., Jr. Structure and regulation of Kit protein-tyrosine kinase--the stem cell factor receptor. *Biochem. Biophys. Res. Commun.* **2005**, *338*, 1307-1315.
- [9] Hopfner, K. P.; Karcher, A.; Shin, D. S.; Craig, L.; Arthur, L. M.; Carney, J. P.; Tainer, J. A. Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* **2000**, *101*, 789-800.
- [10] He, M. M.; Smith, A. S.; Oslob, J. D.; Flanagan, W. M.; Braisted, A. C.; Whitty, A.; Cancilla, M. T.; Wang, J.; Lugovskoy, A. A.; Yoburn, J. C.; Fung, A. D.; Farrington, G.; Eldredge, J. K.; Day, E. S.; Cruz, L. A.; Cacherero, T. G.; Miller, S. K.; Friedman, J. E.; Choong, I. C.; Cunningham, B. C. Small-molecule inhibition of TNF-alpha. *Science* **2005**, *310*, 1022-1025.
- [11] Park, S. H.; Raines, R. T. Genetic selection for dissociative inhibitors of designated protein-protein interactions. *Nat. Biotechnol.* **2000**, *18*, 847-851.
- [12] Smukste, I.; Stockwell, B. R. Advances in chemical genetics. *Annu. Rev. Genomics Hum. Genet.* **2005**, *6*, 261-286.
- [13] Spring, D. R. Chemical genetics to chemical genomics: small molecules offer big insights. *Chem. Soc. Rev.* **2005**, *34*, 472-482.
- [14] Lokey, R. S. Forward chemical genetics: progress and obstacles on the path to a new pharmacopoeia. *Curr. Opin. Chem. Biol.* **2003**, *7*, 91-96.
- [15] Barril, X.; Hubbard, R. E.; Morley, S. D. Virtual screening in structure-based drug discovery. *Mini Rev. Med. Chem.* **2004**, *4*, 779-791.
- [16] Huang, N.; Nagarsekar, A.; Xia, G.; Hayashi, J.; MacKerell, A. D., Jr. Identification of non-phosphate-containing small molecular weight inhibitors of the tyrosine kinase p56 Lck SH2 domain via *in silico* screening against the pY + 3 binding site. *J. Med. Chem.* **2004**, *47*, 3502-3511.
- [17] Cole, J. C.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Comparing protein-ligand docking programs is difficult. *Proteins* **2005**, *60*, 325-332.
- [18] Khandelwal, A.; Lukacova, V.; Comez, D.; Kroll, D. M.; Raha, S.; Balaz, S. A combination of docking, QM/MM methods, and MD simulation for binding affinity estimation of metalloprotein ligands. *J. Med. Chem.* **2005**, *48*, 5437-5447.
- [19] Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **2005**, *26*, 915-931.



- [20] Schneider, G.; Fechner, U. Computer-based *de novo* design of drug-like molecules. *Nat. Rev. Drug Discov.* **2005**, *4*, 649-663.
- [21] Stahura, F. L.; Bajorath, J. New methodologies for ligand-based virtual screening. *Curr. Pharm. Des.* **2005**, *11*, 1189-1202.
- [22] Lawrence, D. S. Signaling protein inhibitors via the combinatorial modification of peptide scaffolds. *Biochim. Biophys. Acta* **2005**, *1754*, 50-57.
- [23] Eichler, J. Synthetic peptide arrays and peptide combinatorial libraries for the exploration of protein-ligand interactions and the design of protein inhibitors. *Comb. Chem. High Throughput Screen.* **2005**, *8*, 135-143.
- [24] Deng, Z.; Chuaqui, C.; Singh, J. Knowledge-based design of target-focused libraries using protein-ligand interaction constraints. *J. Med. Chem.* **2006**, *49*, 490-500.
- [25] Macias, A. T.; Mia, M. Y.; Xia, G.; Hayashi, J.; MacKerell, A. D., Jr. Lead validation and SAR development via chemical similarity searching: application to compounds targeting the pY+3 site of the SH2 domain of p56lck. *J. Chem. Inf. Model.* **2005**, *45*, 1759-1766.
- [26] Hall, L. H.; Hall, L. M. QSAR modeling based on structure-information for properties of interest in human health. *SAR QSAR Environ. Res.* **2005**, *16*, 13-41.
- [27] Guner, O.; Clement, O.; Kurogi, Y. Pharmacophore modeling and three dimensional database searching for drug design using catalyst: recent advances. *Curr. Med. Chem.* **2004**, *11*, 2991-3005.
- [28] Bernard, D.; Coop, A.; MacKerell, A. D., Jr. 2D conformationally sampled pharmacophore: a ligand-based pharmacophore to differentiate delta opioid agonists from antagonists. *J. Am. Chem. Soc.* **2003**, *125*, 3101-3107.
- [29] Bernard, D.; Coop, A.; MacKerell Jr, A. D. Conformationally sampled pharmacophore for peptidic delta opioid ligands. *J. Med. Chem.* **2005**, *48*, 7773-7780.
- [30] Bernard, D.; Coop, A.; Mackerell, A. D. Computer-Aided Drug Design: Structure-Activity Relationships of Delta Opioid Ligands. *Drug Design Rev.* **2005**, *2*, 277-291.
- [31] Dean, P. M.; Lloyd, D. G.; Todorov, N. P. *De novo* drug design: integration of structure-based and ligand-based methods. *Curr. Opin. Drug Discov. Devel.* **2004**, *7*, 347-353.
- [32] Honma, T. Recent advances in *de novo* design strategy for practical lead identification. *Med. Res. Rev.* **2003**, *23*, 606-632.
- [33] Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1308-1315.
- [34] Doman, T. N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly, D. T.; Shoichet, B. K. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem.* **2002**, *45*, 2213-2221.
- [35] May, A.; Zacharias, M. Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochim. Biophys. Acta* **2005**, *1754*, 225-231.
- [36] van Dijk, A. D.; Boelens, R.; Bonvin, A. M. Data-driven docking for the study of biomolecular complexes. *FEBS J.* **2005**, *272*, 293-312.
- [37] Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P.; DesJarlais, R. L. Docking: successes and challenges. *Curr. Pharm. Des.* **2005**, *11*, 323-333.
- [38] Fradera, X.; Mestres, J. Guided docking approaches to structure-based design and screening. *Curr. Top. Med. Chem.* **2004**, *4*, 687-700.
- [39] Alvarez, J. C. High-throughput docking as a source of novel drug leads. *Curr. Opin. Chem. Biol.* **2004**, *8*, 365-370.
- [40] Schneidman-Duhovny, D.; Nussinov, R.; Wolfson, H. J. Predicting molecular interactions *in silico*: II. Protein-protein and protein-drug docking. *Curr. Med. Chem.* **2004**, *11*, 91-107.
- [41] Wodak, S. J.; Mendez, R. Prediction of protein-protein interactions: the CAPRI experiment, its evaluation and implications. *Curr. Opin. Struct. Biol.* **2004**, *14*, 242-249.
- [42] Hou, T.; Xu, X. Recent development and application of virtual screening in drug discovery: an overview. *Curr. Pharm. Des.* **2004**, *10*, 1011-1033.
- [43] Jansen, J. M.; Martin, E. J. Target-biased scoring approaches and expert systems in structure-based virtual screening. *Curr. Opin. Chem. Biol.* **2004**, *8*, 359-364.
- [44] Jain, A. N. Virtual screening in lead discovery and optimization. *Curr. Opin. Drug Discov. Devel.* **2004**, *7*, 396-403.
- [45] Stahura, F. L.; Bajorath, J. Virtual screening methods that complement HTS. *Comb. Chem. High Throughput Screen.* **2004**, *7*, 259-269.
- [46] Campbell, S. J.; Gold, N. D.; Jackson, R. M.; Westhead, D. R. Ligand binding: functional site location, similarity and docking. *Curr. Opin. Struct. Biol.* **2003**, *13*, 389-395.
- [47] Kroemer, R. T. Molecular modelling probes: docking and scoring. *Biochem. Soc. Trans.* **2003**, *31*, 980-984.
- [48] Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335-373.
- [49] Krumrine, J.; Raubacher, F.; Brooijmans, N.; Kuntz, I. Principles and methods of docking and ligand design. *Methods Biochem. Anal.* **2003**, *44*, 443-476.
- [50] Veselovsky, A. V.; Ivanov, A. S. Strategy of computer-aided drug design. *Curr. Drug Targets Infect. Disord.* **2003**, *3*, 33-40.
- [51] Toledo-Sherman, L. M.; Chen, D. High-throughput virtual screening for drug discovery in parallel. *Curr. Opin. Drug Discov. Devel.* **2002**, *5*, 414-421.
- [52] Sotriffer, C.; Klebe, G. Identification and mapping of small-molecule binding sites in proteins: computational tools for structure-based drug design. *Farmacologia* **2002**, *57*, 243-251.
- [53] Shoichet, B. K.; McGovern, S. L.; Wei, B.; Irwin, J. J. Lead discovery using molecular docking. *Curr. Opin. Chem. Biol.* **2002**, *6*, 439-446.
- [54] Smith, G. R.; Sternberg, M. J. Prediction of protein-protein interactions by docking methods. *Curr. Opin. Struct. Biol.* **2002**, *12*, 28-35.
- [55] Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* **2002**, *47*, 409-443.
- [56] Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein-small molecule docking methods. *J. Comput. Aided Mol. Des.* **2002**, *16*, 151-166.
- [57] Waszkowycz, B. Structure-based approaches to drug design and virtual screening. *Curr. Opin. Drug Discov. Devel.* **2002**, *5*, 407-413.
- [58] Schneider, G.; Bohm, H. J. Virtual screening and fast automated docking methods. *Drug Discov. Today* **2002**, *7*, 64-70.
- [59] Abagyan, R.; Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **2001**, *5*, 375-382.
- [60] Langer, T.; Hoffmann, R. D. Virtual screening: an effective tool for lead structure discovery? *Curr. Pharm. Des.* **2001**, *7*, 509-527.
- [61] Xu, D.; Xu, Y.; Uberbacher, E. C. Computational tools for protein modeling. *Curr. Protein Pept. Sci.* **2000**, *1*, 1-21.
- [62] Zeng, J. Mini-review: computational structure-based design of inhibitors that target protein surfaces. *Comb. Chem. High Throughput Screen.* **2000**, *3*, 355-362.
- [63] Feller, S. M.; Lewitzky, M. Potential Disease Targets for Drugs that Disrupt Protein-Protein Interactions of Grb2 and Crk Family Adaptors. *Curr. Pharm. Des.* **2006**, *12*, 529-548.
- [64] Geyer, J. A.; Prigge, S. T.; Waters, N. C. Targeting malaria with specific CDK inhibitors. *Biochim. Biophys. Acta* **2005**, *1754*, 160-170.
- [65] Inestrosa, N. C.; Alvarez, A.; Dinamarca, M. C.; Perez-Acle, T.; Colombes, M. Acetylcholinesterase-amyloid-beta-peptide interaction: effect of Congo Red and the role of the Wnt pathway. *Curr. Alzheimer Res.* **2005**, *2*, 301-306.
- [66] Savarino, A. Expanding the frontiers of existing antiviral drugs: possible effects of HIV-1 protease inhibitors against SARS and avian influenza. *J. Clin. Virol.* **2005**, *34*, 170-178.
- [67] Zhu, S.; Tytgat, J. Evolutionary epitopes of Hsp90 and p23: implications for their interaction. *FASEB J.* **2004**, *18*, 940-947.
- [68] Markowitz, J.; Chen, I.; Gitti, R.; Baldisseri, D. M.; Pan, Y.; Udan, R.; Carrier, F.; MacKerell, A. D., Jr.; Weber, D. J. Identification and characterization of small molecule inhibitors of the calcium-dependent S100B-p53 tumor suppressor interaction. *J. Med. Chem.* **2004**, *47*, 5085-5093.
- [69] Markowitz, J.; Mackerell, A. D., Jr.; Carrier, F.; Charpentier, T. H.; Weber, D. J. Design of Inhibitors for S100B. *Curr. Top. Med. Chem.* **2005**, *5*, 1093-1108.
- [70] Archakov, A. I.; Govorun, V. M.; Dubanov, A. V.; Ivanov, Y. D.; Veselovsky, A. V.; Lewi, P.; Janssen, P. Protein-protein interactions as a target for drugs in proteomics. *Proteomics* **2003**, *3*, 380-391.

- [71] Arkin, M. R.; Wells, J. A. Small-molecule inhibitors of protein-protein interactions: Progressing towards the dream. *Nat. Rev. Drug Discov.* **2004**, *3*, 301-317.
- [72] Bogan, A. A.; Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **1998**, *280*, 1-9.
- [73] Toogood, P. L. Inhibition of protein-protein association by small molecules: Approaches and progress. *J. Med. Chem.* **2002**, *45*, 1543-1558.
- [74] Li, X.; Keskin, O.; Ma, B.; Nussinov, R.; Liang, J. Protein-protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking. *J. Mol. Biol.* **2004**, *344*, 781-795.
- [75] Halperin, I.; Wolfson, H.; Nussinov, R. Protein-protein interactions: coupling of structurally conserved residues and of hot spots across interfaces. Implications for docking. *Structure* **2004**, *12*, 1027-1038.
- [76] Hancock, C. N.; Macias, A.; MacKerell, J. A. D.; Shapiro, P. Mitogen Activated Protein (MAP) Kinases: Development of ATP and Non-ATP Dependent Inhibitors. *Med. Chem.* **2006**, *2*, 213-222.
- [77] Pieper, U.; Eswar, N.; Davis, F. P.; Braberg, H.; Madhusudhan, M. S.; Rossi, A.; Marti-Renom, M.; Karchin, R.; Webb, B. M.; Eramian, D.; Shen, M. Y.; Kelly, L.; Melo, F.; Sali, A. MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* **2006**, *34*, D291-295.
- [78] John, B.; Sali, A. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Res.* **2003**, *31*, 3982-3992.
- [79] Canutescu, A. A.; Dunbrack, R. L., Jr. MolIDE: a homology modeling framework you can click with. *Bioinformatics* **2005**, *21*, 2914-2916.
- [80] Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269-288.
- [81] Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput Aided Mol. Des.* **2001**, *15*, 411-428.
- [82] Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100-5109.
- [83] Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **1997**, *272*, 106-120.
- [84] Moont, G.; Gabb, H. A.; Sternberg, M. J. Use of pair potentials across protein interfaces in screening predicted docked complexes. *Proteins* **1999**, *35*, 364-373.
- [85] Jackson, R. M.; Gabb, H. A.; Sternberg, M. J. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J. Mol. Biol.* **1998**, *276*, 265-285.
- [86] Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639-1662.
- [87] Palma, P. N.; Krippahl, L.; Wampler, J. E.; Moura, J. J. BiGGER: a new (soft) docking algorithm for predicting protein interactions. *Proteins* **2000**, *39*, 372-384.
- [88] Wu, G.; Robertson, D. H.; Brooks, C. L., 3rd; Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER-A CHARMM-based MD docking algorithm. *J. Comput. Chem.* **2003**, *24*, 1549-1562.
- [89] Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* **1997**, *11*, 425-445.
- [90] Comeau, S. R.; Gatchell, D. W.; Vajda, S.; Camacho, C. J. ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* **2004**, *32*, W96-99.
- [91] Murcia, M.; Ortiz, A. R. Virtual screening with flexible docking and COMBINE-based models. Application to a series of factor Xa inhibitors. *J. Med. Chem.* **2004**, *47*, 805-820.
- [92] de Groot, B. L.; van Aalten, D. M.; Scheek, R. M.; Amadei, A.; Vriend, G.; Berendsen, H. J. Prediction of protein conformational freedom from distance constraints. *Proteins* **1997**, *29*, 240-251.
- [93] Hart, T. N.; Read, R. J. A multiple-start Monte Carlo docking method. *Proteins* **1992**, *13*, 206-222.
- [94] Law, D. S.; Ten Eyck, L. F.; Katzenelson, O.; Tsigelny, I.; Roberts, V. A.; Pique, M. E.; Mitchell, J. C. Finding needles in haystacks: Reranking DOT results by using shape complementarity, cluster analysis, and biological information. *Proteins* **2003**, *52*, 33-40.
- [95] Trosset, J. Y.; Scheraga, H. A. Reaching the global minimum in docking simulations: a Monte Carlo energy minimization approach using Bezier splines. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 8011-8015.
- [96] Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. FDS: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. *J. Comput. Chem.* **2003**, *24*, 1637-1656.
- [97] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470-489.
- [98] Gardiner, E. J.; Willett, P.; Artymiuk, P. J. GAPDOCK: a Genetic Algorithm Approach to Protein Docking in CAPRI round 1. *Proteins* **2003**, *52*, 10-14.
- [99] Li, H.; Li, C.; Gui, C.; Luo, X.; Chen, K.; Shen, J.; Wang, X.; Jiang, H. GASDock: a new approach for rapid flexible docking based on an improved multi-population genetic algorithm. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4671-4676.
- [100] Yang, J. M.; Chen, C. C. GEMDOCK: a generic evolutionary method for molecular docking. *Proteins* **2004**, *55*, 288-304.
- [101] Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750-1759.
- [102] Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739-1749.
- [103] Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727-748.
- [104] Vakser, I. A. Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. *Proteins* **1997**, *Suppl 1*, 226-230.
- [105] Dominguez, C.; Boelens, R.; Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731-1737.
- [106] Fernandez-Recio, J.; Totrov, M.; Abagyan, R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins* **2003**, *52*, 113-117.
- [107] Abagyan, R. A.; Totrov, M. M.; Kuznetsov, D. N. ICM - a new method for protein modelling and design. Applications to docking and structure prediction from the distorted native conformation. *J. Comp. Chem.* **1994**, *15*, 488-506.
- [108] Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337*, 209-225.
- [109] Diller, D. J.; Merz, K. M., Jr. High throughput docking for library design and library prioritization. *Proteins* **2001**, *43*, 113-124.
- [110] Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **1995**, *8*, 127-134.
- [111] Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graph. Model.* **2003**, *21*, 289-307.
- [112] Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **2005**, *33*, W363-367.
- [113] Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791-804.
- [114] Muegge, I. Effect of ligand volume correction on PMF scoring. *J. Comput. Chem.* **2001**, *22*, 418-425.
- [115] Pei, J.; Wang, Q.; Liu, Z.; Li, Q.; Yang, K.; Lai, L. PSI-DOCK: towards highly efficient and accurate flexible ligand docking. *Proteins* **2006**, *62*, 934-946.
- [116] Li, L.; Chen, R.; Weng, Z. RDOCK: refinement of rigid-body protein docking predictions. *Proteins* **2003**, *53*, 693-707.



- [117] Wriggers, W.; Milligan, R. A.; McCammon, J. A. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **1999**, *125*, 185-195.
- [118] Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46*, 499-511.
- [119] Fahmy, A.; Wagner, G. TreeDock: a tool for protein docking based on minimizing van der Waals energies. *J. Am. Chem. Soc.* **2002**, *124*, 1241-1250.
- [120] Chen, R.; Li, L.; Weng, Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins* **2003**, *52*, 80-87.
- [121] Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004**, *47*, 45-55.
- [122] Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks, C. L., 3rd. Comparative study of several algorithms for flexible ligand docking. *J. Comput. Aided Mol. Des.* **2003**, *17*, 755-763.
- [123] Wang, R.; Lu, Y.; Fang, X.; Wang, S. An Extensive Test of 14 Scoring Functions Using the PDBbind Refined Set of 800 Protein-Ligand Complexes. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2114 - 2125.
- [124] Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287-2303.
- [125] Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187-217.
- [126] Mackerell, A. D., Jr. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* **2004**, *25*, 1584-1604.
- [127] Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* **1999**, *37*, 228-241.
- [128] Ferrari, A. M.; Wei, B. Q. Q.; Costantino, L.; Shoichet, B. K. Soft docking and multiple receptor conformations in virtual screening. *J. Med. Chem.* **2004**, *47*, 5076-5084.
- [129] Janin, J.; Seraphin, B. Genome-wide studies of protein-protein interaction. *Curr. Opin. Struct. Biol.* **2003**, *13*, 383-388.
- [130] Ritchie, D. W. Evaluation of protein docking predictions using Hex 3.1 in CAPRI rounds 1 and 2. *Proteins* **2003**, *52*, 98-106.
- [131] Mendez, R.; Leplae, R.; Lensink, M. F.; Wodak, S. J. Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures. *Proteins* **2005**, *60*, 150-169.
- [132] Mendez, R.; Leplae, R.; De Maria, L.; Wodak, S. J. Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins* **2003**, *52*, 51-67.
- [133] Janin, J.; Henrick, K.; Moul, J.; Eyck, L. T.; Sternberg, M. J.; Vajda, S.; Vakser, I.; Wodak, S. J. CAPRI: a Critical Assessment of PRredicted Interactions. *Proteins* **2003**, *52*, 2-9.
- [134] Dixon, J. S. Evaluation of the CASP2 docking section. *Proteins* **1997**, *Suppl 1*, 198-204.
- [135] Dunbrack, R. L., Jr.; Gerloff, D. L.; Bower, M.; Chen, X.; Lichtarge, O.; Cohen, F. E. Meeting review: the Second meeting on the Critical Assessment of Techniques for Protein Structure Prediction (CASP2), Asilomar, California, December 13-16, 1996. *Fold. Des.* **1997**, *2*, R27-42.
- [136] Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389-3402.
- [137] Jones, D. T.; Swindells, M. B. Getting the most from PSI-BLAST. *Trends Biochem. Sci.* **2002**, *27*, 161-164.
- [138] Schaffer, A. A.; Aravind, L.; Madden, T. L.; Shavirin, S.; Spouge, J. L.; Wolf, Y. I.; Koonin, E. V.; Altschul, S. F. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* **2001**, *29*, 2994-3005.
- [139] Higgins, D. G. CLUSTAL V: multiple alignment of DNA and protein sequences. *Methods Mol. Biol.* **1994**, *25*, 307-318.
- [140] Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673-4680.
- [141] Huang, X. On global sequence alignment. *Comput. Appl. Biosci.* **1994**, *10*, 227-235.
- [142] Sakakibara, Y.; Brown, M.; Hughey, R.; Mian, I. S.; Sjolander, K.; Underwood, R. C.; Haussler, D. Stochastic context-free grammars for tRNA modeling. *Nucleic Acids Res.* **1994**, *22*, 5112-5120.
- [143] 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. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668-1688.
- [144] Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **2005**, *26*, 1719-1751.
- [145] Lee, B.; Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379-400.
- [146] Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. The MIDAS display system. *J. Mol. Graphics* **1988**, *6*, 13-27.
- [147] Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605-1612.
- [148] Connolly, M. Analytical molecular surface calculation. *J. Appl. Cryst.* **1983**, *16*, 548-558.
- [149] Connolly, M. L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983**, *221*, 709-713.
- [150] Bhat, S.; Purisima, E. O. Molecular surface generation using a variable-radius solvent probe. *Proteins* **2006**, *62*, 244-261.
- [151] Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem.* **2002**, *23*, 128-137.
- [152] Pedretti, A.; Villa, L.; Vistoli, G. VEGA: a versatile program to convert, handle and visualize molecular structure on Windows-based PCs. *J. Mol. Graph. Model.* **2002**, *21*, 47-49.
- [153] Bystroff, C. MASKER: improved solvent-excluded molecular surface area estimations using Boolean masks. *Protein Eng.* **2002**, *15*, 959-965.
- [154] Monteiro, M. A.; Chan, K. H.; Rasko, D. A.; Taylor, D. E.; Zheng, P. Y.; Appelmeik, B. J.; Wirth, H. P.; Yang, M.; Blaser, M. J.; Hynes, S. O.; Moran, A. P.; Perry, M. B. Simultaneous expression of type 1 and type 2 Lewis blood group antigens by *Helicobacter pylori* lipopolysaccharides. Molecular mimicry between *h. pylori* lipopolysaccharides and human gastric epithelial cell surface glycoforms. *J. Biol. Chem.* **1998**, *273*, 11533-11543.
- [155] Vorobjev, Y. N.; Hermans, J. SIMS: computation of a smooth invariant molecular surface. *Biophys. J.* **1997**, *73*, 722-732.
- [156] Sanner, M. F.; Olson, A. J.; Spehner, J. C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **1996**, *38*, 305-320.
- [157] Totrov, M.; Abagyan, R. The contour-buildup algorithm to calculate the analytical molecular surface. *J. Struct. Biol.* **1996**, *116*, 138-143.
- [158] Zauhar, R. J. SMART: a solvent-accessible triangulated surface generator for molecular graphics and boundary element applications. *J. Comput. Aided Mol. Des.* **1995**, *9*, 149-159.
- [159] Pascual-Ahuir, J. L.; Silla, E.; Tunon, I. GEPOL: an improved description of molecular surface. III. A new algorithm for the computation of a solvent-excluding surface. *J. Comput. Chem.* **1994**, *15*, 1127-1138.
- [160] Connolly, M. L. The molecular surface package. *J. Mol. Graph* **1993**, *11*, 139-141.
- [161] Roth, C. M.; Neal, B. L.; Lenhoff, A. M. Van der Waals interactions involving proteins. *Biophys. J.* **1996**, *70*, 977-987.
- [162] Berchanski, A.; Shapira, B.; Eisenstein, M. Hydrophobic complementarity in protein-protein docking. *Proteins* **2004**, *56*, 130-142.
- [163] Zhong, S.; MacKerell Jr., A. D. Novel scoring functions for *in silico* database screening: Binding response and pose-based scaling. *230th ACS National Meeting, in Washington, DC* **2005**, Aug 28 - Sept 1.
- [164] Vigers, G. P.; Rizzi, J. P. Multiple active site corrections for docking and virtual screening. *J. Med. Chem.* **2004**, *47*, 80-89.
- [165] Fukunishi, Y.; Mikami, Y.; Nakamura, H. Similarities among receptor pockets and among compounds: analysis and application

- to *in silico* ligand screening. *J. Mol. Graph. Model.* **2005**, *24*, 34-45.
- [166] Smith, G. R.; Fitzjohn, P. W.; Page, C. S.; Bates, P. A. Incorporation of flexibility into rigid-body docking: applications in rounds 3-5 of CAPRI. *Proteins* **2005**, *60*, 263-268.
- [167] Barril, X.; Morley, S. D. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J. Med. Chem.* **2005**, *48*, 4432-4443.
- [168] Bastard, K.; Prevost, C.; Zacharias, M. Accounting for loop flexibility during protein-protein docking. *Proteins* **2006**, *62*, 956-969.
- [169] Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A. Representing receptor flexibility in ligand docking through relevant normal modes. *J. Am. Chem. Soc.* **2005**, *127*, 9632-9640.
- [170] Zacharias, M. Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci.* **2003**, *12*, 1271-1282.
- [171] Liu, Z.; Dominy, B. N.; Shakhnovich, E. I. Structural mining: self-consistent design on flexible protein-peptide docking and transferable binding affinity potential. *J. Am. Chem. Soc.* **2004**, *126*, 8515-8528.
- [172] Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. Geometry-based flexible and symmetric protein docking. *Proteins* **2005**, *60*, 224-231.
- [173] Ehrlich, L. P.; Nilges, M.; Wade, R. C. The impact of protein flexibility on protein-protein docking. *Proteins* **2005**, *58*, 126-133.
- [174] Alberts, I. L.; Todorov, N. P.; Dean, P. M. Receptor flexibility in *de novo* ligand design and docking. *J. Med. Chem.* **2005**, *48*, 6585-6596.
- [175] Zavodszky, M. I.; Kuhn, L. A. Side-chain flexibility in protein-ligand binding: the minimal rotation hypothesis. *Protein Sci.* **2005**, *14*, 1104-1114.
- [176] Tatsumi, R.; Fukunishi, Y.; Nakamura, H. A hybrid method of molecular dynamics and harmonic dynamics for docking of flexible ligand to flexible receptor. *J. Comput. Chem.* **2004**, *25*, 1995-2005.
- [177] Wei, B. Q.; Weaver, L. H.; Ferrari, A. M.; Matthews, B. W.; Shoichet, B. K. Testing a flexible-receptor docking algorithm in a model binding site. *J. Mol. Biol.* **2004**, *337*, 1161-1182.
- [178] Lin, J. H.; Peryman, A. L.; Schames, J. R.; McCammon, J. A. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers* **2003**, *68*, 47-62.
- [179] Althaus, E.; Kohlbacher, O.; Lenhof, H. P.; Muller, P. A combinatorial approach to protein docking with flexible side chains. *J. Comput. Biol.* **2002**, *9*, 597-612.
- [180] Huang, P. S.; Love, J. J.; Mayo, S. L. Adaptation of a fast Fourier transform-based docking algorithm for protein design. *J. Comput. Chem.* **2005**, *26*, 1222-1232.
- [181] Kovacs, J. A.; Chacon, P.; Cong, Y.; Metwally, E.; Wriggers, W. Fast rotational matching of rigid bodies by fast Fourier transform acceleration of five degrees of freedom. *Acta Crystallogr. D Biol. Crystallogr.* **2003**, *59*, 1371-1376.
- [182] Jackson, R. M. Q-fit: a probabilistic method for docking molecular fragments by sampling low energy conformational space. *J. Comput. Aided Mol. Des.* **2002**, *16*, 43-57.
- [183] Sandak, B.; Nussinov, R.; Wolfson, H. J. A method for biomolecular structural recognition and docking allowing conformational flexibility. *J. Comput. Biol.* **1998**, *5*, 631-654.
- [184] Cecchini, M.; Kolb, P.; Majeux, N.; Cafilisch, A. Automated docking of highly flexible ligands by genetic algorithms: a critical assessment. *J. Comput. Chem.* **2004**, *25*, 412-422.
- [185] Pegg, S. C.; Haresco, J. J.; Kuntz, I. D. A genetic algorithm for structure-based *de novo* design. *J. Comput. Aided Mol. Des.* **2001**, *15*, 911-933.
- [186] Gardiner, E. J.; Willett, P.; Artymiuk, P. J. Protein docking using a genetic algorithm. *Proteins* **2001**, *44*, 44-56.
- [187] Politowska, E.; Drabik, P.; Kazmierkiewicz, R.; Ciarkowski, J. Docking ligands to vasopressin and oxytocin receptors via genetic algorithm. *J. Recept. Signal Transduct. Res.* **2002**, *22*, 393-409.
- [188] Hou, T.; Wang, J.; Chen, L.; Xu, X. Automated docking of peptides and proteins by using a genetic algorithm combined with a tabu search. *Protein Eng.* **1999**, *12*, 639-648.
- [189] Verkhivker, G. M.; Rejto, P. A.; Gehlhaar, D. K.; Freer, S. T. Exploring the energy landscapes of molecular recognition by a genetic algorithm: analysis of the requirements for robust docking of HIV-1 protease and FKBP-12 complexes. *Proteins* **1996**, *25*, 342-353.
- [190] Oshiro, C. M.; Kuntz, I. D.; Dixon, J. S. Flexible ligand docking using a genetic algorithm. *J. Comput. Aided Mol. Des.* **1995**, *9*, 113-130.
- [191] Willett, P. Genetic algorithms in molecular recognition and design. *Trends Biotechnol.* **1995**, *13*, 516-521.
- [192] Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43-53.
- [193] Rohs, R.; Bloch, I.; Sklenar, H.; Shakked, Z. Molecular flexibility in *ab initio* drug docking to DNA: binding-site and binding-mode transitions in all-atom Monte Carlo simulations. *Nucleic Acids Res.* **2005**, *33*, 7048-7057.
- [194] Liu, M.; Wang, S. MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. *J. Comput. Aided Mol. Des.* **1999**, *13*, 435-451.
- [195] Knechtel, R. M.; Antoon, J.; Rullmann, C.; Boelens, R.; Kaptein, R. MONTY: a Monte Carlo approach to protein-DNA recognition. *J. Mol. Biol.* **1994**, *235*, 318-324.
- [196] Cafilisch, A.; Niederer, P.; Anliker, M. Monte Carlo docking of oligopeptides to proteins. *Proteins* **1992**, *13*, 223-230.
- [197] Fernandez-Recio, J.; Totrov, M.; Abagyan, R. Soft protein-protein docking in internal coordinates. *Protein Sci.* **2002**, *11*, 280-291.
- [198] Zhang, C.; Vasmatazis, G.; Cornette, J. L.; DeLisi, C. Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.* **1997**, *267*, 707-726.
- [199] Bohm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* **1994**, *8*, 243-256.
- [200] Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337-356.
- [201] Li, C. H.; Ma, X. H.; Chen, W. Z.; Wang, C. X. A protein-protein docking algorithm dependent on the type of complexes. *Protein Eng.* **2003**, *16*, 265-269.
- [202] Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849-857.
- [203] Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* **1995**, *2*, 317-324.
- [204] Yang, J. M.; Shen, T. W. A pharmacophore-based evolutionary approach for screening selective estrogen receptor modulators. *Proteins* **2005**, *59*, 205-220.
- [205] Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425-445.
- [206] Stouten, P. F. W.; Frommel, C.; Nakamura, H.; Sander, C. An Effective Solvation Term Based on Atomic Occupancies for Use in Protein Simulations. *Molecular Simulation* **1993**, *10*, 97-108.
- [207] Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Model for aqueous solvation based on class IV atomic charges and first solvation shell effects. *J. Phys. Chem.* **1996**, *100*, 16385-16398.
- [208] Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. New class IV charges model for extracting accurate partial charges from wave functions. *J. Phys. Chem. ser. A* **1998**, *102*, 1820-1831.
- [209] Zhang, C.; Chen, J.; DeLisi, C. Protein-protein recognition: exploring the energy funnels near the binding sites. *Proteins* **1999**, *34*, 255-267.
- [210] Guvench, O.; Brooks, C. L., 3rd Efficient approximate all-atom solvent accessible surface area method parameterized for folded and denatured protein conformations. *J. Comput. Chem.* **2004**, *25*, 1005-1014.
- [211] Wagoner, J.; Baker, N. A. Solvation forces on biomolecular structures: a comparison of explicit solvent and Poisson-Boltzmann models. *J. Comput. Chem.* **2004**, *25*, 1623-1629.
- [212] Feig, M.; Im, W.; Brooks, C. L., 3rd Implicit solvation based on generalized Born theory in different dielectric environments. *J. Chem. Phys.* **2004**, *120*, 903-911.

- [213] Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., 3rd Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **2004**, *25*, 265-284.
- [214] Velec, H. F.; Gohlke, H.; Klebe, G. DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.* **2005**, *48*, 6296-6303.
- [215] Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* **2002**, *16*, 11-26.
- [216] Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., 3rd Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* **2004**, *47*, 3032-3047.
- [217] Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759-4767.
- [218] Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100-5109.
- [219] Terp, G. E.; Johansen, B. N.; Christensen, I. T.; Jorgensen, F. S. A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein-ligand binding affinities. *J. Med. Chem.* **2001**, *44*, 2333-2343.
- [220] Wang, R.; Wang, S. How does consensus scoring work for virtual library screening? An idealized computer experiment. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1422-1426.
- [221] Pan, Y.; Huang, N.; Cho, S.; MacKerell, A. D., Jr. Consideration of molecular weight during compound selection in virtual target-based database screening. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 267-272.
- [222] Pagliaro, L.; Felding, J.; Audouze, K.; Nielsen, S. J.; Terry, R. B.; Krog-Jensen, C.; Butcher, S. Emerging classes of protein-protein interaction inhibitors and new tools for their development. *Curr. Opin. Chem. Biol.* **2004**, *8*, 442-449.
- [223] Sirois, S.; Hatzakis, G.; Wei, D.; Du, Q.; Chou, K. C. Assessment of chemical libraries for their druggability. *Comput. Biol. Chem.* **2005**, *29*, 55-67.
- [224] von Grothuss, M.; Koczyk, G.; Pas, J.; Wyrwicz, L. S.; Rychlewski, L. Ligand.Info small-molecule Meta-Database. *Comb. Chem. High Throughput Screen.* **2004**, *7*, 757-761.
- [225] Irwin, J. J.; Shoichet, B. K. ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177-182.
- [226] Dunkel, M.; Fullbeck, M.; Neumann, S.; Preissner, R. SuperNatural: a searchable database of available natural compounds. *Nucleic Acids Res.* **2006**, *34*, D678-683.
- [227] Gold, N. D.; Jackson, R. M. SitesBase: a database for structure-based protein-ligand binding site comparisons. *Nucleic Acids Res.* **2006**, *34*, D231-234.
- [228] Stuart, A. C.; Ilyin, V. A.; Sali, A. LigBase: a database of families of aligned ligand binding sites in known protein sequences and structures. *Bioinformatics* **2002**, *18*, 200-201.
- [229] Roche, O.; Kiyama, R.; Brooks, C. L., 3rd Ligand-protein database: linking protein-ligand complex structures to binding data. *J. Med. Chem.* **2001**, *44*, 3592-3598.
- [230] Wang, R.; Fang, X.; Lu, Y.; Yang, C. Y.; Wang, S. The PDBbind database: methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111-4119.
- [231] Michalsky, E.; Dunkel, M.; Goede, A.; Preissner, R. SuperLigands - a database of ligand structures derived from the Protein Data Bank. *BMC Bioinformatics* **2005**, *6*, 122.
- [232] Shin, J. M.; Cho, D. H. PDB-Ligand: a ligand database based on PDB for the automated and customized classification of ligand-binding structures. *Nucleic Acids Res.* **2005**, *33*, D238-241.
- [233] Hendlich, M.; Bergner, A.; Gunther, J.; Klebe, G. Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. *J. Mol. Biol.* **2003**, *326*, 607-620.
- [234] Okuno, Y.; Yang, J.; Taneishi, K.; Yabuuchi, H.; Tsujimoto, G. GLIDA: GPCR-ligand database for chemical genomic drug discovery. *Nucleic Acids Res.* **2006**, *34*, D673-677.
- [235] Girke, T.; Cheng, L. C.; Raikhel, N. ChemMine. A compound mining database for chemical genomics. *Plant Physiol.* **2005**, *138*, 573-577.
- [236] Kimmich, R. D.; Park, W. K. Chemical libraries towards protein kinase inhibitors. *Comb. Chem. High Throughput Screen.* **2003**, *6*, 661-672.
- [237] Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors, In Methods and Principles in Medicinal Chemistry*. Wiley-VCH: New York **2000**.
- [238] Consonni, V.; Todeschini, R.; Pavan, M.; Gramatica, P. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 693-705.
- [239] Consonni, V.; Todeschini, R.; Pavan, M. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 682-692.
- [240] Gasteiger, J.; Saller, H. Calculation of the Charge Distribution in Conjugated Systems by a Quantification of the Resonance Concept. *Angew. Chem. Int. Ed. Engl.* **1985**, *24*, 687-689.
- [241] Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623-1641.
- [242] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3-26.
- [243] Butina, D. Unsupervised Data Base Clustering on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way to Cluster Small and Large Data Sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 747-750.
- [244] Godden, J. W.; Stahura, F. L.; Bajorath, J. Anatomy of fingerprint search calculations on structurally diverse sets of active compounds. *J. Chem. Inf. Model.* **2005**, *45*, 1812-1819.
- [245] Matter, H. Selecting optimally diverse compounds from structure databases: A validation study of two-dimensional and three-dimensional molecular descriptors. *J. Med. Chem.* **1997**, *40*, 1219-1229.
- [246] Bajorath, J. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J. Chem. Inform. Comput. Sci.* **2001**, *41*, 233-245.
- [247] Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzouzi, K.; Jacoby, E.; Schuffenhauer, A. Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. *Org. Biomol. Chem.* **2004**, *2*, 3256-3266.
- [248] Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Molecular similarity searching using atom environments, information-based feature selection, and a naive Bayesian classifier. *J. Chem. Inform. Comput. Sci.* **2004**, *44*, 170-178.
- [249] Oprea, T. I. Current trends in lead discovery: are we looking for the appropriate properties? *J. Comput. Aided Mol. Des.* **2002**, *16*, 325-334.
- [250] Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput. Aided Mol. Des.* **2000**, *14*, 251-264.
- [251] Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today: Technol.* **2004**, *1*, 337-341.
- [252] Olah, M. M.; Bologna, C. G.; Oprea, T. I. Strategies for compound selection. *Curr. Drug Discov. Technol.* **2004**, *1*, 211-220.
- [253] Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45*, 2615-2623.
- [254] Ekins, S.; Boulanger, B.; Swaan, P. W.; Hupcey, M. A. Towards a new age of virtual ADME/TOX and multidimensional drug discovery. *J. Comput. Aided Mol. Des.* **2002**, *16*, 381-401.
- [255] Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening - an overview. *Drug Discov. Today* **1998**, *3*, 160-178.
- [256] Baurin, N.; Baker, R.; Richardson, C.; Chen, I.; Foloppe, N.; Potter, A.; Jordan, A.; Roughley, S.; Parratt, M.; Greaney, P.; Morley, D.; Hubbard, R. E. Drug-like annotation and duplicate analysis of a 23-supplier chemical database totalling 2.7 million compounds. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 643-651.
- [257] McGovern, S. L.; Helfand, B. T.; Feng, B.; Shoichet, B. K. A specific mechanism of nonspecific inhibition. *J. Med. Chem.* **2003**, *46*, 4265-4272.

- [258] Feng, B. Y.; Shelat, A.; Doman, T. N.; Guy, R. K.; Shoichet, B. K. High-throughput assays for promiscuous inhibitors. *Nat. Chem. Biol.* **2005**, *1*, 146-148.
- [259] Enyedy, I. J.; Ling, Y.; Nacro, K.; Tomita, Y.; Wu, X. H.; Cao, Y. Y.; Guo, R. B.; Li, B. H.; Zhu, X. F.; Huang, Y.; Long, Y. Q.; Roller, P. P.; Yang, D. J.; Wang, S. M. Discovery of small-molecule inhibitors of bcl-2 through structure-based computer screening. *J. Med. Chem.* **2001**, *44*, 4313-4324.
- [260] Wang, J. L.; Liu, D. X.; Zhang, Z. J.; Shan, S. M.; Han, X. B.; Srinivasula, S. M.; Croce, C. M.; Alnemri, E. S.; Huang, Z. W. Structure-based discovery of an organic compound that binds Bcl-2 protein and induces apoptosis of tumor cells. *Proc. Nat. Acad. Sci. USA* **2000**, *97*, 7124-7129.
- [261] Beuparlant, P.; Shore, G. C. Therapeutic activation of caspases in cancer; a question of selectivity. *Curr. Opin. Drug Discov. Devel.* **2003**, *6*, 179-187.
- [262] O'Neill, J.; Manion, M.; Schwartz, P.; Hockenbery, D. M. Promises and challenges of targeting Bcl-2 anti-apoptotic proteins for cancer therapy. *Biochim. Biophys. Acta* **2004**, *1705*, 43-51.
- [263] Petros, A. M.; Olejniczak, E. T.; Fesik, S. W. Structural biology of the Bcl-2 family of proteins. *Biochim. Biophys. Acta* **2004**, *1644*, 83-94.
- [264] Yin, X. M.; Oltvai, Z. N.; Korsmeyer, S. J. *Nature* **1994**, *369*, 321-323.
- [265] Sprous, D. G.; Lewis, D. R.; Leonard, J. M.; Heritage, T.; Burkett, S. N.; Baker, D. S.; Clark, R. D. OptiDock: virtual HTS of combinatorial libraries by efficient sampling of binding modes in product space. *J. Comb. Chem.* **2004**, *6*, 530-539.
- [266] Sali, A.; Potterton, L.; Yuan, F.; van Vlijmen, H.; Karplus, M. Evaluation of comparative protein modeling by MODELLER. *Proteins* **1995**, *23*, 318-326.
- [267] Lawrence, D. S.; Niu, J. Protein kinase inhibitors: the tyrosine-specific protein kinases. *Pharmacolo. Ther.* **1998**, *77*, 81-114.
- [268] Neel, B. G.; Tonks, N. K. Protein tyrosine phosphatases in signal transduction. *Curr. Opin. Cell Biol.* **1997**, *9*, 193-204.
- [269] Moarefi, I.; LaFevre-Bernt, M.; Sicheri, F.; Huse, M.; Lee, C. H.; Kuriyan, J.; Miller, W. T. Activation of the Src-family tyrosine kinase Hck by SH3 domain displacement. *Nature* **1997**, *385*, 650-653.
- [270] Huang, N.; Nagarsekar, A.; Xia, G. J.; Hayashi, J.; MacKerell, A. D. Identification of non-phosphate-containing small molecular weight inhibitors of the tyrosine kinase p56 Lck SH2 domain via *in silico* screening against the pY+3 binding site. *J. Med. Chem.* **2004**, *47*, 3502-3511.
- [271] Songyang, Z.; Cantley, L. C. Recognition and specificity in protein tyrosine kinase-mediated signalling. *Trends Biochem. Sci.* **1995**, *20*, 470-475.
- [272] Pan, Y. P.; Huang, N.; Cho, S.; MacKerell, A. D. Consideration of molecular weight during compound selection in virtual target-based database screening. *J. Chem. Inform. Comput. Sci.* **2003**, *43*, 267-272.
- [273] Macias, A. T.; Mia, M. Y.; Xia, G. J.; Hayashi, J.; MacKerell, A. D. Lead validation and SAR development via chemical similarity searching; Application to compounds targeting the pY+3 site of the SH2 domain of p56(lck). *J. Chem. Inf. Model.* **2005**, *45*, 1759-1766.
- [274] Rustandi, R. R.; Baldisseri, D. M.; Weber, D. J. Structure of the negative regulatory domain of p53 bound to S100B(beta beta). *Nat. Struct. Biol.* **2000**, *7*, 570-574.
- [275] Weber, D. J.; Rustandi, R. R.; Carrier, F.; Zimmer, D. B. Interaction of dimeric S100B(BB) with the tumor suppressor protein: A model for Ca-dependent S100-target protein interactions; Kluwer Academic Publishers: Dordrecht, The Netherlands **2000**, pp. 469-487.
- [276] Fry, D. C.; Vassilev, L. T. Targeting protein-protein interactions for cancer therapy. *J. Mol. Med.* **2005**, *83*, 955-963.
- [277] Chen, I. J.; Neamati, N.; MacKerell, A. D., Jr. Structure-based inhibitor design targeting HIV-1 integrase. *Curr. Drug Targets Infect. Disord.* **2002**, *2*, 217-234.
- [278] Markowitz, J.; Chen, J.; Gitti, R.; Baldisseri, D. M.; Pan, Y. P.; Udan, R.; Carrier, F.; MacKerell, A. D.; Weber, D. J. Identification and characterization of small molecule inhibitors of the calcium-dependent S100B-p53 tumor suppressor interaction. *J. Med. Chem.* **2004**, *47*, 5085-5093.
- [279] Reuter, C. W.; Morgan, M. A.; Bergmann, L. Targeting the Ras signaling pathway: a rational, mechanism-based treatment for hematologic malignancies? *Blood* **2000**, *96*, 1655-1669.
- [280] Kyriakis, J. M.; Abruch, J. Protein kinase cascades activated by stress and inflammatory cytokines. *Bioessays* **1996**, *18*, 567-577.
- [281] Duesbery, N. S.; Webb, C. P.; Vande Woude, G. F. MEK wars, a new front in the battle against cancer. *Nat. Med.* **1999**, *5*, 736-737.
- [282] Zhang, J.; Zhang, F.; Ebert, D.; Cobb, M. H.; Goldsmith, E. J. Activity of the MAP kinase ERK2 is controlled by a flexible surface loop. *Structure* **1995**, *3*, 299-307.
- [283] Hancock, C. N.; Macias, A.; Lee, E. K.; Yu, S. Y.; Mackerell, A. D., Jr.; Shapiro, P. Identification of novel extracellular signal-regulated kinase docking domain inhibitors. *J. Med. Chem.* **2005**, *48*, 4586-4595.
- [284] Tanoue, T.; Maeda, R.; Adachi, M.; Nishida, E. Identification of a dock groove on ERK and p38 MAP kinases that regulates the specificity of docking interactions. *EMBO J.* **2001**, *20*, 466-479.
- [285] Oltsdorf, T.; Elmore, S. W.; Shoemaker, A. R.; Armstrong, R. C.; Augeri, D. J.; Belli, B. A.; Bruncko, M.; Deckwerth, T. L.; Dinges, J.; Hajduk, P. J.; Joseph, M. K.; Kitada, S.; Korsmeyer, S. J.; Kunzer, A. R.; Letai, A.; Li, C.; Mitten, M. J.; Nettesheim, D. G.; Ng, S.; Nimmer, P. M.; O'Connor, J. M.; Aleksijew, A.; Petros, A. M.; Reed, J. C.; Shen, W.; Tahir, S. K.; Thompson, C. B.; Tomaselli, K. J.; Wang, B.; Wendt, M. D.; Zhang, H.; Fesik, S. W.; Rosenberg, S. H. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* **2005**, *435*, 677-681.
- [286] Tong, L.; Warren, T. C.; King, J.; Betageri, R.; Rose, J.; Jakes, S. Crystal structures of the human p56(lck) SH2 domain in complex with two short phosphotyrosyl peptides at 1.0 angstrom and 1.8 angstrom resolution. *J. Mol. Biol.* **1996**, *256*, 601-610.
- [287] Inman, K. G.; Yang, R.; Rustandi, R. R.; Miller, K. E.; Baldisseri, D. M.; Weber, D. J. Solution NMR structure of S100B bound to the high-affinity target peptide TRTK-12. *J. Mol. Biol.* **2002**, *324*, 1003-1014.
- [288] Zhang, F.; Strand, A.; Robbins, D.; Cobb, M. H.; Goldsmith, E. J. Atomic structure of the MAP kinase ERK2 at 2.3 A resolution. *Nature* **1994**, *367*, 704-711.