Top Curr Chem (2011)
DOI: 10.1007/128_2011_213

Application of NMR and Molecular Docking in Structure-Based Drug Discovery

Jaime L. Stark and Robert Powers

Abstract Drug discovery is a complex and costly endeavor, where few drugs that reach the clinical testing phase make it to market. High-throughput screening (HTS) is the primary method used by the pharmaceutical industry to identify initial lead compounds. Unfortunately, HTS has a high failure rate and is not particularly efficient at identifying viable drug leads. These shortcomings have encouraged the development of alternative methods to drive the drug discovery process. Specifically, nuclear magnetic resonance (NMR) spectroscopy and molecular docking are routinely being employed as important components of drug discovery research. Molecular docking provides an extremely rapid way to evaluate likely binders from a large chemical library with minimal cost. NMR ligand-affinity screens can directly detect a protein-ligand interaction, can measure a corresponding dissociation constant, and can reliably identify the ligand binding site and generate a co-structure. Furthermore, NMR ligand affinity screens and molecular docking are perfectly complementary techniques, where the combination of the two has the potential to improve the efficiency and success rate of drug discovery. This review will highlight the use of NMR ligand affinity screens and molecular docking in drug discovery and describe recent examples where the two techniques were combined to identify new and effective therapeutic drugs.

Keywords Drug discovery, FAST-NMR, *In silico* screening, Ligand affinity screens, Molecular docking, Nuclear magnetic resonance, Virtual screening

J.L. Stark and R. Powers (✉)
Department of Chemistry, University of Nebraska – Lincoln, 722 Hamilton Hall, Lincoln, NE 68588-0304, USA
e-mail: rpowers3@unl.edu

Contents

1 Introduction

The completion of the human genome project [1] coupled with an increase in R&D investments was widely anticipated to be the cornerstone of personalized medicine with a corresponding explosion in new pharmaceutical drugs targeting a range of diseases. Nearly a decade later, the rate at which new drugs enter clinical development and reach the market has declined dramatically despite the influx of novel therapeutic targets and R&D investments. In the past 5 years the number of new molecular entities (NMEs) receiving FDA approval has decreased by 50% from the previous 5 years [2]. There are several reasons for this decline, but most stem from the fact that drug discovery is a complex and costly endeavor. Approximately 80–90% of drugs that reach the clinical testing phase fail to make it to market [3, 4]. Efforts to reduce costs often lead pharmaceutical companies to invest their time and money in proven therapies, like "best-in-class" drugs, instead of "first-in-class" drugs that target new mechanisms of action or diseases. As a result, many diseases are "orphaned" and lack any therapeutic compounds in the discovery pipeline. Addressing these issues will require fundamental changes to create a more efficient drug discovery process.

The enormous costs and high failure rates inherent to the pharmaceutical industry are clearly contributing factors to the declining number and diversity of new therapeutics. Efforts that minimize costs without restricting research endeavors will evidently benefit the development of drugs for various human diseases. The availability of hundreds of whole-genome sequences for numerous organisms provides an invaluable data set for drug research [1, 5, 6]. Identifying a novel "druggable" protein target is a critical first step for a successful and efficient drug discovery effort. Unfortunately, bioinformatics analysis alone does not generally provide enough information to justify embarking upon an expensive drug discovery program [7, 8]. Instead, knowing the three dimensional structure of a protein greatly

enhances the value of the bioinformatics analysis. Protein structures often provide insights into the molecular basis of the protein's biological function and its relationship to a particular disease. A protein structure also provides detailed information on the sequence and structural characteristics that govern ligand binding interactions. Building a drug discovery effort based on structural information promises to help in the identification of novel therapeutic targets, in the discovery of new lead compounds, and in the optimization of drug-like properties to improve efficacy and safety. Currently, the drug discovery process within the pharmaceutical industry employs high-throughput screening (HTS) as the primary method for identifying lead compounds. However, the high false positive rate [9–12] combined with a significant cost in time and money has encouraged the development of alternative methods to drive the drug discovery process [13, 14].

Nuclear magnetic resonance (NMR) spectroscopy is uniquely qualified to assist in making the drug discovery process more efficient [15, 16]. NMR is useful for several reasons: (1) it directly detects the interaction between the ligand and protein using a variety of techniques, (2) samples are typically analyzed under native conditions, (3) hundreds of samples can be analyzed per day, and (4) information on the binding site and binding affinity can be readily obtained. These features allow NMR to be an effective tool at multiple steps in the drug discovery pathway, which includes verifying HTS and virtual screening hits [15, 17–19], screening fragment-based libraries [15, 20–22], optimizing lead compounds [15, 17, 23, 24], evaluating ADME-toxicology [25–27], and identifying and validating therapeutic targets [28, 29]. Nevertheless, there are still intrinsic costs to maintaining an NMR instrument, screening a compound library, and producing significant quantities of a protein. One way to significantly reduce experimental costs is to utilize *in silico* methodologies to supplement the lead identification and optimization steps of the drug discovery process [30].

Molecular docking is a computational tool that predicts the binding site location and conformation of a compound when bound to a protein [30–32]. This approach has been found to be fairly successful in redocking compounds into previously solved protein–ligand co-structures [33], where more than 70% of the redocked ligands reside within 2 Å root mean squared deviation (RMSD) of the actual ligand pose. During the prediction of protein–ligand co-structures, molecular docking programs calculate a binding score that allows for the selection of the best ligand pose. The binding score is typically based on a combination of geometric and energetic functions (bond lengths, dihedral angles, van der Waals forces, Lennard-Jones and electrostatic interactions, etc.) in conjunction with empirical functions unique to each specific docking program [34–39]. A large variety of docking programs are available that include AutoDock [40], DOCK [41], FlexX [42], Glide [43], HADDOCK [44], and LUDI [45, 46].

Binding energies are also routinely used to rank different ligands from a compound library after being docked to a protein target. The virtual or *in silico* screening of a library composed of thousands of theoretical compounds can be accomplished in a day with minimal cost [47–49]. Thus, a virtual screen can significantly accelerate the hit identification and optimization process while

reducing the amount of experimental effort. However, a virtual screen does have significant limitations that prevent it from completely replacing traditional HTS [50–52]. These limitations include inaccurate scoring functions, use of rigid proteins, and simplified solvation models. In essence, a virtual screen only increases the likelihood that a predicted ligand actually binds the protein target, experimental verification is essential. Despite the individual drawbacks, NMR ligand affinity screens and molecular docking are complementary techniques. This review will highlight the use of NMR ligand affinity screens and molecular docking in drug discovery and describe recent examples where the combination of the two techniques provides a powerful approach to identify new and effective therapeutic drugs.

2 NMR Ligand Affinity Screens

NMR ligand affinity screening is a versatile technique that is useful for multiple stages of the drug discovery process [15, 17, 22, 53]. This versatility arises from the ability of NMR to directly detect protein–ligand binding based on changes in several NMR parameters. A binding event is detected by the relative differences between the protein or ligand NMR spectrum in the bound and unbound states. However, the specific type of information obtained about the binding process depends on whether a ligand-based or target-based NMR experiment is used.

2.1 Ligand-Based NMR Screens

Ligand-based NMR screens typically monitor the NMR spectrum of a ligand under free and bound conditions. Distinguishing between a free ligand and a protein–ligand complex is generally based on the large molecular weight difference that affects several NMR parameters. Small molecular weight molecules have slow relaxation rates (R_2), negative NOE cross-peaks, and large translational diffusion coefficients (D_t). If a protein–ligand binding event occurs, the ligand adopts the properties of the larger molecular-weight protein, increasing R_2, producing positive NOE cross-peaks, and decreasing D_t, all of which can be observed by NMR [54]. Most ligand-based NMR screens use one-dimensional (1D) ^{1}H-NMR experiments to monitor these changes, which provide significant benefits for a high-throughput screen. 1D NMR experiments are typically fast (2–5 min) and routinely use mixtures without the need to deconvolute [55]. The deconvolution of mixtures is avoided by ensuring that NMR ligand peaks do not overlap in the NMR spectrum (Fig. 1). The application of mixtures allows for hundreds to thousands of compounds to be screened in a single day. Another advantage of ligand-based NMR methods is the minimal amount of protein required (<10 μM) for each experiment. Additionally, isotopically labeled proteins are not needed for the

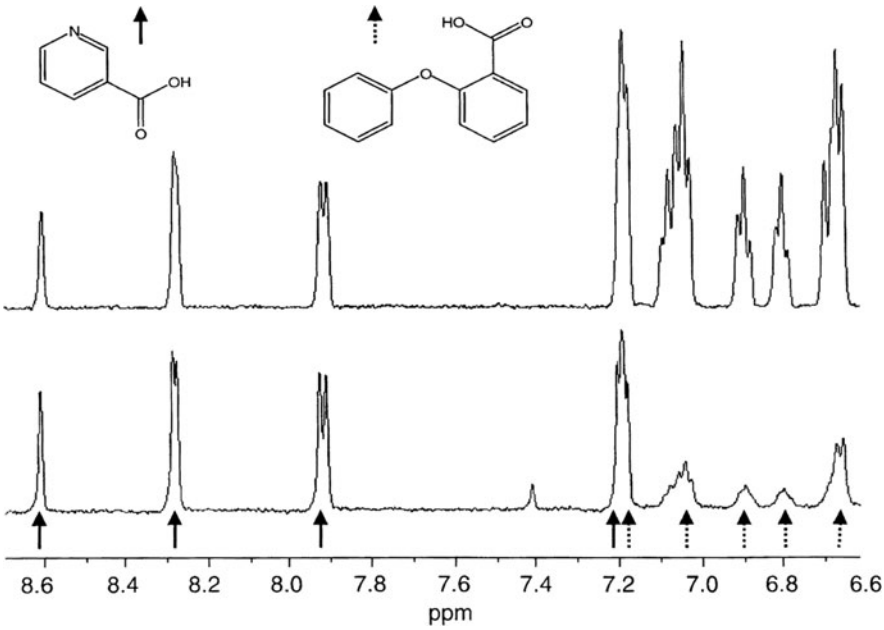

Fig. 1 An example of the use of a ligand-detect NMR experiment to observe the line broadening (increase R_2) that occurs when one compound, in a mixture of two compounds, binds a protein target. The ^{1}H-NOESY spectra of nicotinic acid (*left* structure) and 2-phenoxybenzoic acid (*right* structure) in a mixture without protein (*top* spectrum) and with the protein, p38 MAP kinase, added (*bottom* spectrum). The *solid* and *dashed arrows* represent the resonances of nicotinic acid and 2-phenoxybenzoic acid, respectively. In this case, the resonances corresponding to 2-phenoxybenzoic acid are broadened, indicating binding of this compound to the protein. (Reprinted with permission from [178], copyright 2001 by Academic Press)

NMR ligand affinity screen and protein molecular weight is not a limiting factor [21]. In fact, higher molecular-weight proteins enhance the observation of a binding event in a ligand-based NMR screen. All of these characteristics make ligand-based NMR screens a routinely used drug discovery technique.

There are several screening techniques created from ligand-based NMR experiments: line broadening [56], STD NMR [57], WaterLOGSY [58], SLAPSTIC [59], TINS [60], transferred NOEs [61], FAXS [62, 63], FABS [64, 65], and diffusion measurements [66, 67]. Each of these methods utilizes a specific NMR parameter that indicates ligand-binding, such as a change in ligand NMR peak width or diffusion, a saturation transfer from the protein or solvent to the ligand, an NOE transfer between the free and bound ligand, a spin-label induced paramagnetic relaxation, or fluorine chemical shift anisotropy. The choice of which method to use typically depends upon the protein target and the compound library being screened. In addition, line broadening and STD, among other techniques, can be used to measure dissociation constants (K_D) [68, 69]. Conversely, ligand-based NMR screens don't provide any structural information about the protein–ligand complex.

2.2 Target-Based NMR Screens

A target based screen focuses on changes in the protein (or other target) NMR spectrum to identify a binding event. Typically, chemical shift perturbations (CSPs) occur in the protein NMR spectrum upon ligand binding. The complexity and severe peak overlap in a protein 1D ^{1}H NMR spectrum makes it impractical to observe subtle CSPs for weak binding ligands. Instead, two-dimensional (2D) heteronuclear NMR [70–72] experiments are typically used for target-based NMR ligand affinity screens [73]. 2D ^{1}H-^{13}C/^{15}N HSQC/TROSY NMR experiments require a significant increase in experiment time (>10 min) due to the additional dimension and the need to collect a reference spectrum for the ligand-free protein. Also, the protein needs to be ^{15}N and/or ^{13}C isotopically labeled. Importantly, 2D ^{1}H-^{13}C/^{15}N HSQC/TROSY NMR experiments provide additional information about the ligand binding site.

A binding ligand often results in the observation of CSPs of the resonances in a 2D ^{1}H-^{15}N- or ^{1}H-^{13}C-HSQC spectrum (Fig. 2a). These CSPs are usually caused by a change in the chemical environment for residues proximal to the bound ligand or residues undergoing ligand-induced conformational changes. The availability of the protein structure and the NMR sequence assignments (correlation of an NMR resonance with a specific amino acid residue) allows for the CSPs to be mapped onto a three-dimensional (3D) representation of the protein's surface. A cluster of residues on the protein surface with observed CSPs often identifies the ligand-binding site.

The ligand binding affinity or K_D is also routinely determined from CSPs measured from a series of 2D ^{1}H-^{13}C/^{15}N HSQC/TROSY NMR experiments. The magnitude of the CSPs at varying ligand concentrations is correlated to the K_D for the protein–ligand complex using the following equation [74, 75]:

$$\mathrm{CSP}_{\mathrm{obs}} = \mathrm{CSP}_{\mathrm{max}} \frac{(K_D + [L] + [P]) - \sqrt{(K_D + [L] + [P])^2 - (4[L][P])}}{2[P]}, \qquad (1)$$

where $[P]$ is the protein concentration, $[L]$ is the ligand concentration, $\mathrm{CSP}_{\mathrm{max}}$ is the maximum CSP observed for a fully bound protein, and $\mathrm{CSP}_{\mathrm{obs}}$ is the observed CSP at a particular ligand concentration. A least squares fit of (1) to the experimental CSP data is used to calculate a K_D (Fig. 2b).

As previously mentioned, since target-based screens require the use of multidimensional NMR experiments, data collection is significantly longer relative to ligand-based NMR screens. Also target-based screens require higher protein concentrations (>50 μM compared to <10 μM). This severely limits the utility of target-based NMR screens for the high-throughput analysis of large compound libraries. Instead, the approach is typically used to validate hits from a high-throughput screen or the analysis of relatively small fragment-based libraries [76–78]. A fragment-based library consists of low molecular-weight compounds (<250–350 Da) that are fragments of known drugs or have drug-like properties

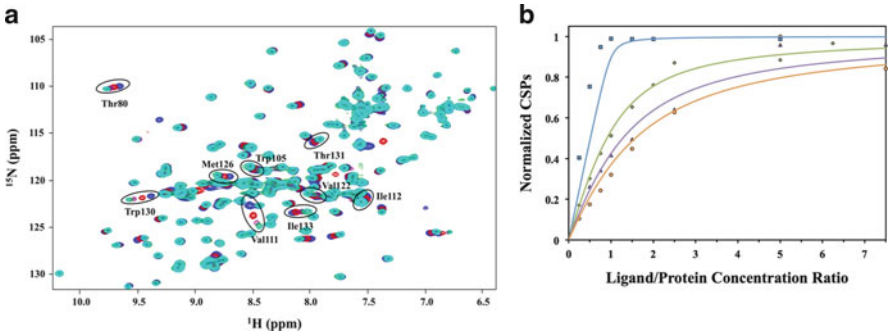

Fig. 2 (**a**) An overlay of the 2D ^{1}H-^{15}N HSQC spectra for the protein YndB titrated with increasing amounts of chalcone. The perturbed residues can be used to identify a consensus binding site. (**b**) NMR titration data for YndB bound to chalcone (*blue*), flavanone (*green*), flavone (*purple*), and flavanol (*orange*). The magnitude of the chemical shift perturbation can be used to calculate the dissociation constants for each compound. (Reprinted with permission from [112], copyright 2010 by John Wiley and Sons)

[79]. Recent advances like the SOFAST-HMQC experiment [80, 81] and the Fast-HSQC experiment [82] have decreased the time and amount of protein necessary for a target-based screen. Nevertheless, NMR ligand affinity screens are still very resource intensive, requiring a significant amount of time and material. Also, since any high-throughput screen produces a significant amount of negative data (most ligands don't bind or inhibit a protein), a more efficient approach is to screen a library of compounds with a higher probability of binding the protein target. In effect, a virtual or *in silico* screen can be used to enrich a library with likely binders.

3 Molecular Docking

An accurate prediction of the interactions between two molecules requires an in-depth understanding of the energetics that led to a stable biomolecular complex. Unfortunately, a model that correctly accounts for all the factors involved in a productive protein–ligand interaction is currently unknown. Further, the problem is exponentially more complex than just modeling the specifics of a protein–ligand interaction. A protein contains thousands of atoms that have specific interactions with each other, with the solvent, and with other ions; in addition to the bound ligand. Because of this complexity, computational efforts that attempt to model protein–ligand interactions require significant amounts of processing power and time. Many efforts that utilize molecular dynamics and distributed computing [83, 84] are generally limited to a detailed analysis of a *single* system. These methods are generally not practical for the majority of researchers interested in conducting a virtual screen of a library containing upwards of millions of compounds. To make molecular docking computationally feasible and easily accessible, many simplifications and trade-offs in the process are necessary.

Many computer programs are available to perform or assist with molecular docking. The vast number of docking programs makes it impractical to describe them all in detail within a single review (for other reviews please see [85–89]). Each docking program does have some unique features that make them particularly useful for a given situation or problem. However, nearly all the docking programs consist of two primary components: docking (or searching) and scoring [30, 31]. Docking refers to the sampling of the ligand's conformation space and its orientation relative to a receptor. Scoring is used to evaluate and rank the current pose of the ligand.

3.1 Docking

The docking process requires, at a minimum, two inputs: the three-dimensional structures of the receptor (protein) and the ligand. The most common simplification to the docking process is to keep the structure of the receptor rigid and stationary. Only the ligand is typically allowed to be flexible as it is docked to the protein. Keeping the protein rigid significantly minimizes the complexity of the calculation. Sampling the conformations and orientations of the ligand is done using systematic or stochastic methods [30, 31].

Systematic search methods attempt to sample all of the possible conformations of a ligand by incrementing the torsional angles of each rotatable bond. Unfortunately, this technique is computationally expensive due to the exponential increase in the number of possible conformations (N_{conf}) as the number of rotatable bonds increases:

$$N_{\text{conf}} = \prod_{i=1}^{N} \prod_{j=1}^{n_{\text{inc}}} \frac{360}{\theta_{i,j}}, \tag{2}$$

where N represents the number of rotatable bonds, n_{inc} is the number of incremental rotations for each rotatable bond, and $\theta_{i,j}$ is the size of the incremental rotation for each rotatable bond. As a result, purely brute force systematic approaches are generally not used. Instead, most systematic searches require the use of efficient shortcuts. As an illustration, MOLSDOCK [90] uses mutually orthogonal Latin squares (MOLS) to identify optimal ligand conformations. Latin squares are an $N \times N$ matrix, where each parameter (torsion angle value) occurs only once in each row and column. Orthogonal Latin squares are two or more superimposed $N \times N$ matrices, where each parameter still only occurs once in each row and column. MOLS are used to identify the N^2 subset of ligand conformations used to calculate binding energies. Simply, only a small subset of the possible ligand conformations is sampled to construct the potential surface and identify the minima.

Perhaps the most commonly utilized systematic search method is incremental construction, which is used by DOCK [41], FlexX [42], E-Novo [91], LUDI [45, 46], ADAM [92], and TrixX [93]. In this particular method, the ligand is

split into fragments. The most rigid fragments are often used as the core or anchor and are docked first into the receptor binding pocket. The remaining fragments are incrementally added back onto the core fragment, where each addition is systematically rotated to evaluate the most optimal conformation. Thus, incremental construction drastically reduces the number of possible conformations that need to be searched in order to identify the optimal pose.

Another systematic approach uses rigid docking in combination with a predefined library of ligand conformations, which is implemented in OMEGA [94], FLOG [95], Glide [43], and the TrixX Conformer Generator [96]. This technique generates several low energy conformers for a ligand that are clustered by RMSD. A representative conformer from each cluster is then docked into the receptor. The approach is very fast because the docking process keeps the ligand rigid, eliminating the need to spend computation time on searching torsional space. A tradeoff for this increase in speed is a potential loss in accuracy, since the binding potential for all possible conformers may not be explored. Conversely, a major benefit of the technique is the fact that the library of structural conformers only needs to be generated once. This is a significant savings in time for the pharmaceutical industry, where screening libraries may consist of millions of compounds.

Unlike systematic approaches that attempt to sample all possible ligand conformations, stochastic searches explore conformational space by making random torsional changes to a single ligand or a population of ligands. The structural changes are then evaluated using a probability function. There are three types of stochastic searches: Monte Carlo algorithms [97], genetic algorithms [98], and tabu search algorithms [99]. The most basic stochastic method is the Monte Carlo algorithm, which utilizes a Boltzmann probability function to determine whether to accept a particular ligand pose:

$$P \sim \exp\left[\frac{-(E_1 - E_0)}{K_B T}\right], \tag{3}$$

where P is the probability the conformation is accepted, E_0 and E_1 are the ligand's energy before and after the conformational change, K_B is the Boltzmann constant, and T is the temperature. The simple scoring function used by the Monte Carlo algorithms is more effective than molecular dynamics in avoiding local minima and finding the global minimum. Alternatively, genetic algorithms utilize the theory of evolution and natural selection to search ligand conformation space. In this case, the conformations, orientations, and coordinates of a ligand are encoded into variables representing a "genetic code." A population of ligands with random genetic codes is allowed to evolve using mutations, crossovers, and migrations. The new population is evaluated using a fitness function that eliminates unfavorable ligand poses. Eventually, a final population converges to ligands with the most favorable "genes" or conformations (Fig. 3). Tabu searches, like other stochastic methods, randomly modify the conformation and coordinates of a ligand, score the conformer, and then repeat the process for a new conformation. Tabu searches

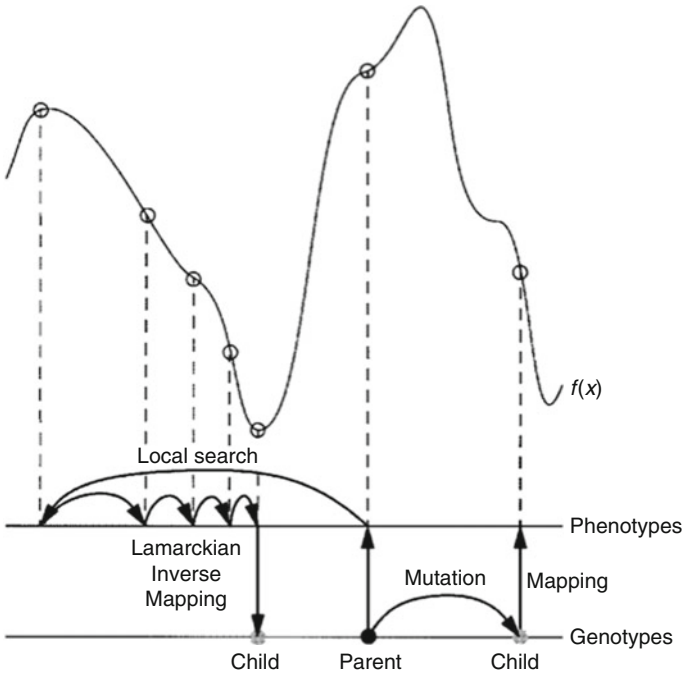

Fig. 3 An illustration of the genetic algorithm approach, where the states of the ligand (translation, orientation, and conformation relative to the protein) are interpreted as the ligand genotype and the atomic coordinates represent the phenotype. A plot of the change in the fitness function ($f(x)$) as the ligand population is allowed to mutate, crossover, and migrate. The genetic evolution of the ligand effectively samples conformational space where the best conformer is identified by a minimum in the fitness function (Reprinted with permission from [179], copyright 1998 by John Wiley and Sons)

utilize a tabu list to remember previous ligand states. A pose is immediately rejected if it is close to a prior conformation. The tabu list encourages the search to progress to unexplored regions of conformational space.

3.2 Scoring

While docking algorithms are generally efficient at generating the correct ligand pose, it is important for the docking program to actually select the correct ligand conformation from an ensemble of similar conformers. In essence, the scoring function should be able to distinguish between the true or optimal binding conformation and all the other poses. The scoring function is also used to rank the relative binding affinities for each compound in the library. Ideally, the scoring function should be able to calculate the free energy ($\Delta G_{\text{binding}}$) of the protein–ligand binding interaction, which is directly related to the K_{D}:

$$\Delta G_{\text{binding}} = -RT\ln\frac{1}{K_{\text{D}}}. \tag{4}$$

Unfortunately, accurately calculating the binding free energy is very challenging due to the many forces that influence binding. In molecular docking, there are five primary types of scoring functions: force field-based, empirical, knowledge-based, shape-based, and consensus [100–102].

Force field-based scoring functions [30, 31] are used to calculate the free energy of binding by combining the receptor–ligand interaction energy and the change in internal energies of the ligand based on its bound conformation (Fig. 4). The internal energy of the receptor is usually ignored since the receptor is kept rigid in most docking programs. The protein–ligand binding energies are typically defined by van der Waal forces, hydrogen bonding energies, and electrostatic energy terms. The van der Waals and hydrogen bonding terms often utilize a Lennard-Jones potential function, while the electrostatic terms are described by a coulombic function. Unfortunately, these interaction energies were originally derived from measuring enthalpic interactions in the gas phase. Of course, receptor–ligand binding interactions actually occur in an aqueous solution, which introduces additional interactions between the solvent molecules, the receptor, and the ligand. Protein–ligand binding energies are also dependent on the entropic changes that occur upon binding, which include torsional, vibrational, rotational, and translational entropies. Most entropy and solvation-based energy terms can't be calculated using force field-based scoring functions. As a result, force field-based scoring functions are incomplete and inaccurate.

Empirical scoring functions [103–106] are similar to force field-based scoring functions since they use a summation of individual energy terms. But empirical scoring functions also attempt to include solvation and entropic terms. This is typically achieved by using experimentally determined binding energies of known ligand–receptor interactions to train the scoring system using regression analysis. Empirical scoring functions are fast, but the accuracy is completely dependent upon the experimental data set used to train the scoring function. In general, empirical scoring functions are reliable for ligand–receptor complexes that are similar to the training set.

Knowledge-based scoring functions [107–109] are fundamentally different from force field-based and empirical scoring functions. Knowledge-based scoring functions don't attempt to calculate the free energy of binding. Instead, these scoring functions utilize a sum of protein–ligand atom pair interaction potentials to calculate a binding affinity. The atom pair interaction potentials are generated based upon a probability distribution of interatomic distances found in known protein–ligand structures. The probability distributions are then converted into distance-dependent interaction energies. In this manner, knowledge-based scoring functions allow for the modeling of binding interactions that are not well understood. The approach is also very simple, which is useful for screening large compound libraries. Unfortunately, knowledge-based scoring functions are designed

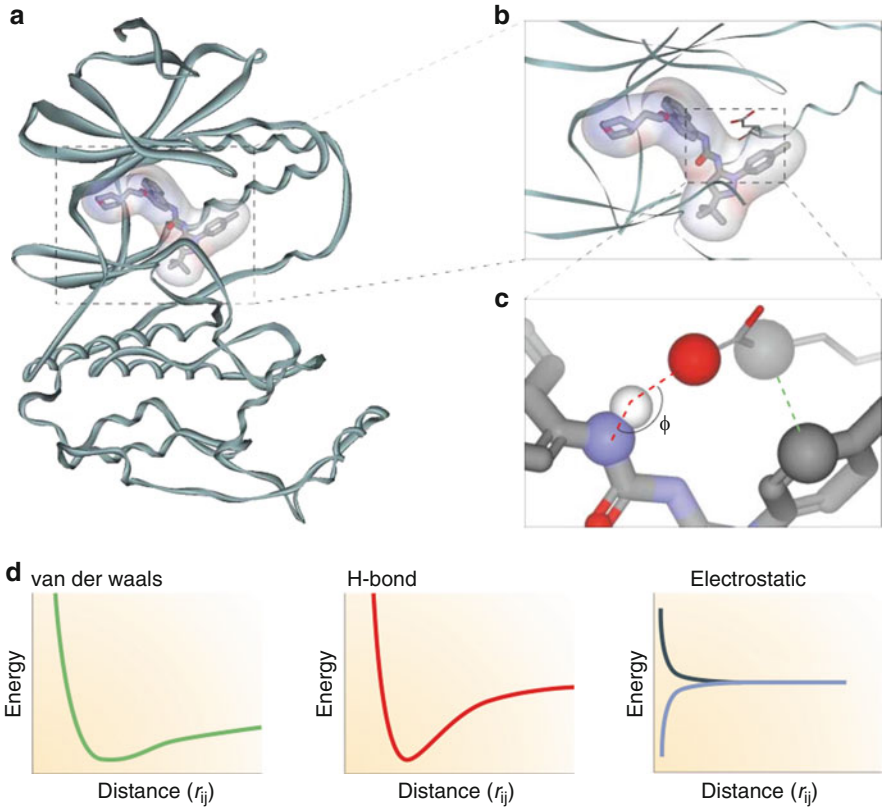

Fig. 4 (**a**) A representation of p38 mitogen-activated protein kinase structure bound to BIRB796 and (**b**) an expanded view of the binding site. (**c**) A representation of the hydrogen-bonding (*red*) and electrostatic interactions (*green*) between the atoms of the protein and the atoms of the ligand. (**d**) A representation of three force-field energy terms (van der Waals, hydrogen-bonding, and electrostatic) as distance between the interacting atom pairs change. (Reprinted with permission from [30], copyright 2004 by the Nature Publishing Group)

to reproduce known experimental structures, and the binding score generated has little relevance to an actual binding affinity. This is an issue similar to empirical scoring functions; the accuracy of the scoring function is strongly dependent on the similarity of the protein–ligand complex to the training data set.

As implied, shape-based scoring functions are based on a shape match between the ligand and the ligand binding site [110]. These scoring functions are typically used as prefilters to eliminate compounds that are unable to fit into the ligand binding site [111, 112]. Shape-based scoring functions are very fast, but are limited relative to more accurate scoring functions that calculate binding affinities. Shape-based scoring functions typically generate smooth energy surfaces using Gaussian functions [111], which are more tolerant to atomic variations and make protein

clash interactions "softer." This essentially helps minimize the effect of small structural variations that may occur during ligand binding.

While the above scoring methods are generally useful in describing protein–ligand interactions, the simplifications used in each approach limits the overall accuracy in predicting the correct docked ligand pose [113, 114]. The major weakness of most docking programs has been shown to be the scoring function. One approach to compensate for this deficiency is to use a consensus score from a combination of scoring functions to rescore a docked pose. Consensus scoring [31, 115] has been shown in several examples to improve docking results compared to a single scoring function. However, like individual scoring functions, the improvement is not consistent and the proper choice of scoring functions to calculate a consensus score is typically based on trial and error.

3.3 Protein Flexibility

Proteins are inherently flexible and undergo a range of motions over different time scales, and thus the use of rigid protein structures by molecular docking is problematic [116, 117]. This is especially troublesome for therapeutic targets where only an apo-structure is available. Conformational changes upon ligand-binding may range from small perturbations in side chain conformation at the site of ligand binding to large rearrangements of the entire protein structure. Not accounting for such structural changes during ligand docking can drastically alter the ability to identify reliable protein–ligand models correctly [118–122]. Conversely, attempting to dock a large library of flexible ligands to a completely flexible protein structure using molecular dynamics is too computationally expensive to be practical.

Several approaches to "solve" the protein flexibility problem have been explored. The first generally applicable approach utilized soft docking in the scoring function, which reduces the van der Waals repulsion terms in the empirical scoring function [123, 124]. This allows for some overlap between ligand and protein atoms. While this approach is simple and fast, it can only accommodate very small changes in side chain conformations. Other approaches attempt to implement protein structural changes into the docking process. For example, a library of side chain rotamers for residues only in the ligand binding site is routinely used [40, 125]. This dramatically reduces the number of active rotatable bonds during the docking process and has a lower computational cost compared to molecular dynamics. However, the inclusion of a library of rotamers in the docking protocol is significantly slower than rigid protein docking. Furthermore, the approach is limited to local side chain conformational changes.

The most common docking technique that attempts to account for protein flexibility uses multiple protein structures. The ensemble of structures is expected to represent the range of conformations sampled by the protein and has the benefit of being able to evaluate both small and large conformational changes. The molecular docking is repeated for each individual protein conformation, which

results in a proportional increase in computational time. Also, the results may be ambiguous, since there may be several equally valid ligand poses for each different protein conformation. This is especially apparent in virtual screening approaches where enrichment factors suffer when docking to multiple structures (please see Sect. 3.4). This is likely due to an increase in the number of false positives among the top hits [126]. Ensemble docking is an alternative to docking multiple structures that removes the ambiguity [118]. All the protein structures from the ensemble are superimposed in order to generate an average structure or an average receptor grid. The docking is then performed against the average structure or average receptor grid (Fig. 5). The ensemble docking approach allows for a single docking at a significantly lower computational cost; however, it may suffer from accuracy problems if the ensemble is biased towards the unbound form of the protein. Effectively, a biased ensemble may negate the goal of incorporating protein flexibility if it represents a single conformation.

3.4 Virtual Screening and Assessment

Using molecular docking to identify lead candidates is an attractive approach for the pharmaceutical industry; it allows for the rapid evaluation of millions of chemical compounds while using minimal resources compared to traditional HTS. The process by which molecular docking is used to rank compounds within a library based on a predicted binding affinity is known as virtual screening [127, 128]. The potential benefit to drug discovery has inspired the development and evaluation of numerous virtual screening methodologies. A virtual screen requires a balance between optimizing speed and maximizing accuracy. Specifically, the goal of a drug discovery virtual screen is the rapid and efficient separation of a small subset of active compounds from a relatively large random library of inactive compounds. Unfortunately, determining the effectiveness of a specific virtual screening process is challenging, where independent evaluators routinely generate inconsistent results [87, 129–131].

The ambiguous nature of the results from a virtual screen requires additional methods to evaluate its success. Typically, a virtual screening process is evaluated against a protein target with a set of known binders. Assessing the performance of a virtual screen is primarily based on the accuracy of the predicted ligand pose and binding affinity. The correct binding pose is often evaluated by calculating the RMSD between the docked and experimental ligand structures. The evaluation of binding affinity is typically based on the accurate ranking of known binders instead of the absolute scores because of the known limitations with calculating a binding energy. Other modes of performance assessment involve evaluating enrichment and generating diverse hit lists.

In a virtual screening protocol, every compound in a library (N_{tot}) is docked to the protein and a corresponding binding score is calculated. The binding score for the ligand's best docked pose is used to rank the ligand relative to the entire library.

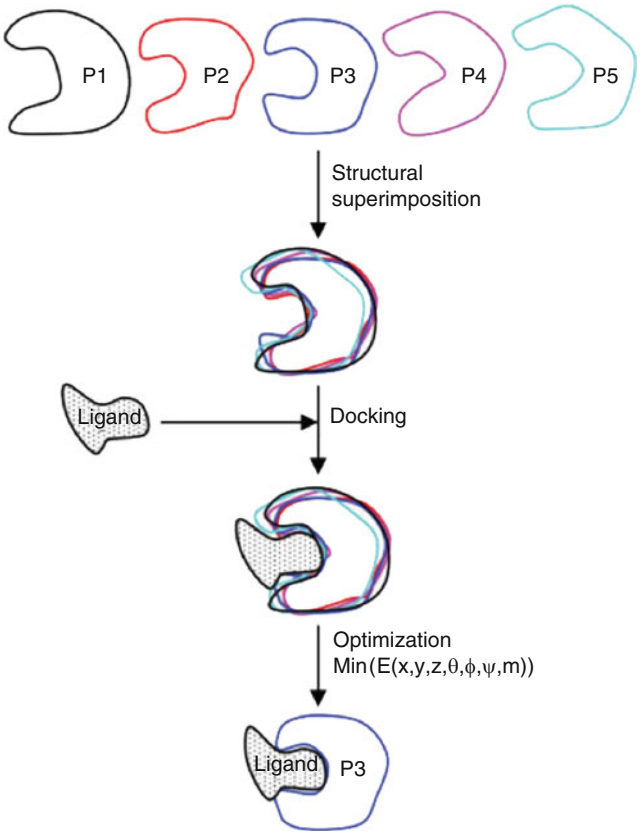

Fig. 5 A cartoon illustration of ensemble docking, where five individual protein structures are superimposed to create a single scoring parameter for the docked ligand. Ensemble docking minimizes the computational effort since a single docking occurs to select the best conformer instead of five separate molecular docking simulations. (Reprinted with permission from [118], copyright 2007 by John Wiley and Sons)

A virtual screen never results in all the truly active compounds being top ranked. Instead, most virtual screening protocols set a binding score or ranking threshold to identify the predicted active compounds or "hits." In general, top ranked compounds are expected to be enriched with active compounds compared to a random selection (Fig. 6a). A high enrichment factor ($EF > 10$) is considered the benchmark of success for a virtual screening [132]. Enrichment is dependent on sensitivity (Se) and specificity (Sp). Sensitivity represents the true positive rate, which is the ratio of true positives (TP) found by the virtual screening vs the total number of actives (A) in the library. The number of actives corresponds to both true positive (TP) and false negative (FN):

$$Se = \frac{TP}{TP + FN}. \tag{5}$$

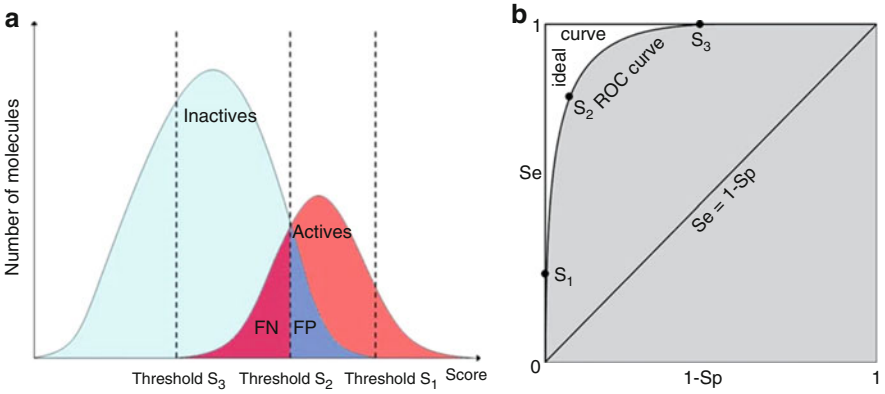

Fig. 6 (**a**) A theoretical distribution of compounds in a virtual screen based upon the docking score. The overlap between active and inactive compounds indicates that the scoring threshold used to identify a hit by virtual screening is critical. (**b**) A ROC curve is used to evaluate the enrichment of a virtual screen and select a scoring threshold. A ROC curve that approaches $Se = 1$ and 1-$Sp = 0$ represents perfect enrichment. The area under the ROC curve (AUC) represents the probability that a true active is identified. (Reprinted with permission from [131], copyright 2008 by Springer)

Specificity is the measure of the true negative rate, which represents the ratio of true negatives (*TN*) to the total number of inactive compounds. The number of inactive compounds corresponds to both true negatives (*TN*) and false positives (*FP*):

$$Sp = \frac{TN}{TN + FP}. \tag{6}$$

The enrichment factor is a common method for evaluating the enrichment capabilities of a virtual screen:

$$EF = \frac{\left(\frac{TP}{TP+FP}\right)}{\left(\frac{TP+FN}{N_{\text{tot}}}\right)}. \tag{7}$$

The enrichment factor is dependent upon the ratio of active compounds to the total number of compounds in the library. As a result, enrichment scores are difficult to compare between virtual screens with different libraries. Also, the enrichment factor does not distinguish between high and low ranking compounds.

Perhaps the more popular approach for evaluating enrichment is to generate a receiver operating characteristic (ROC) curve [133]. The ROC curve is a plot of the true positive rate (*Se*) against the false positive rate ($1-Sp$) at varying thresholds for determining a hit. A ROC curve allows for the evaluation of a virtual screening method without using an arbitrary scoring threshold. Enrichment occurs when the resulting data point at a particular threshold resides above the diagonal ($Se = 1-Sp$), which corresponds to a random selection of compounds. In a perfect virtual

screen where every active compound is identified as a hit and every inactive compound falls below the threshold, the ROC curve approaches the top left corner ($Se = 1$ and $1-Sp = 0$) (Fig. 6b).

Hit list diversity is also an important consideration for the success of a virtual screen since there is more value in identifying a few unique compounds instead of many compounds all based on the same chemical scaffold. One way that diversity can be determined is by comparing the structural similarities of hits from a virtual screen by using the Tanimoto index [134] and then clustering the results. Basically, a Tanimoto index is calculated based on the fraction of similar chemical substructures present in two structures. Generally, 1,365 chemical substructures are used to describe a structure. The substructures include individual elements, two-atom substructures, single rings, condensed rings, aromatic rings, other rings, chains, branches, and functional groups:

$$TI = \frac{C}{A + B + C}, \tag{8}$$

where A represents the substructural features present in the first structure, B represents the substructural features present in the second structure, and C represents the substructural features common to both structures. Identical structures have a TI score of 1, where completely dissimilar structures have a TI value of 0.

4 Combining Molecular Docking with NMR Ligand Affinity Screens

The vast majority of initial leads in drug discovery are identified from HTS [13, 135, 136]. Pharmaceutical companies have invested heavily in developing and maintaining large chemical libraries (>1,000,000 compounds), which are screened using automated, biological assays intended to monitor a specific response or biological effect [136]. Unfortunately, HTS is extremely inefficient due to the high cost of developing, maintaining, and screening such large libraries of compounds. Furthermore, the random search for an effective drug in the vastness of chemical space (~10^{60} compounds) [137] is almost guaranteed to fail. Thus, HTS hit rates are typically very low, where <0.5% of compounds exhibit any inhibitor activity in an assay [138]. Correspondingly, HTS assays are highly inefficient since most of the screening effort is spent on the analysis of negative data. Additionally, HTS assays, by nature, are mechanistic "black boxes," and a response does not provide any information on the mechanism of inhibition. This often leads to numerous false positives from undesirable interactions [11, 12, 139] that may lead the drug discovery project astray. Improving the efficiency of drug discovery requires the implementation of advanced techniques that better guide the selection of lead candidates without sacrificing speed.

Ideally, an entirely *in silico* approach to screening a large compound library would significantly improve efficiency and reduce costs [140, 141]. However, several assessments of virtual screens have concluded that, without prior in-depth analysis of the protein's ligand binding site, only a marginal improvement in finding successful leads is observed relative to standard HTS [32]. NMR can complement a virtual screen by providing rapid experimental validation of lead compounds. NMR allows for a ligand-binding event to be directly observed instead of relying on false-positive prone activity assays. Also, NMR provides detailed structural information about the ligand binding site and the orientation of the bound ligand. An NMR ligand affinity screen can be used to validate upwards of thousands of predicted hits from a virtual screen [142]. Thus, combining NMR with virtual screens may provide a more efficient approach to lead identification and drug discovery.

4.1 Identification of New Therapeutic Targets

The functional assignment of unannotated proteins is essential to the drug discovery process. Greater than 40% of protein sequences encoded in eukaryotic genomes consist of proteins of unknown function and represent an important opportunity to identify new therapeutic targets [143]. Assigning a function to an uncharacterized protein is an arduous and time-consuming task. The process often requires detailed biochemical studies that may include analyzing cell phenotypes through knockout libraries, monitoring of gene expression levels, or utilizing pull-down assays [144–147].

Since the interactions of proteins with other biomolecules or small molecules is the basis of a functional definition or classification, identifying the functional ligand, the functional epitope or ligand binding site, and the 3D structure of the protein–ligand complex are invaluable for a functional annotation. A functional epitope or ligand binding site is evolutionarily conserved relative to the rest of the protein structure in order for the protein to maintain its biological function. Therefore, proteins that share similar binding site structures are expected to be functional homologs and bind a similar set of ligands [28, 29]. Correspondingly, numerous *in silico* approaches attempt to infer a function for an uncharacterized protein by predicting ligand binding sites using geometry-based, information-based, and energy-based algorithms [148–150]. Unfortunately, unambiguously identifying the ligand binding site on a protein can be challenging without experimental evidence, especially for proteins with no known function.

Functional Annotation Screening Technology using NMR (FAST-NMR) [28, 29] is one approach that combines HTS by NMR with molecular docking and bioinformatics analysis in order to assign a function to a protein (Fig. 7). In this process, a compound library that contains approximately 430 biologically relevant compounds [151] is screened by NMR using a multistep approach [152]. First, a ligand-based screen using 1D NMR ^{1}H line-broadening experiments identifies

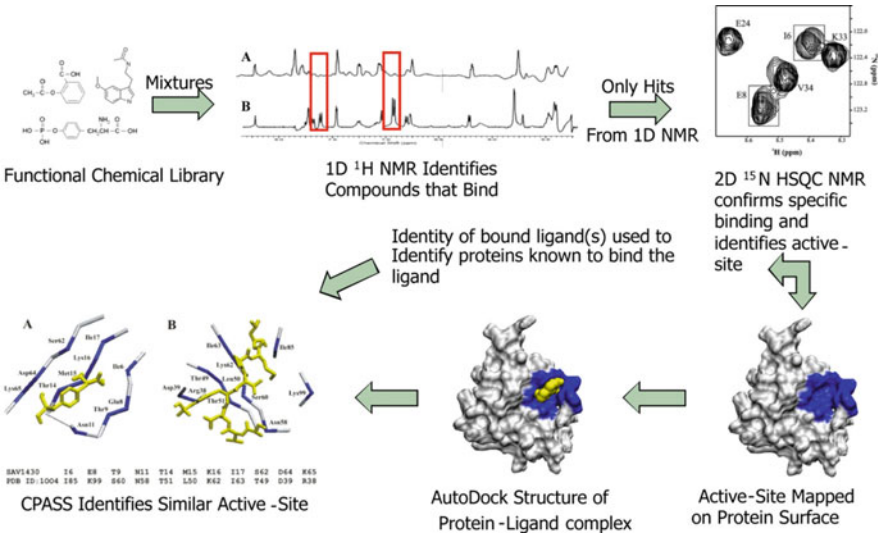

Fig. 7 A flow diagram of the FAST-NMR process. Mixtures of biologically active compounds are first assayed in a ligand-based 1D line broadening screen against the protein of interest. Compounds that are identified as hits are then verified using CSPs from a 2D ^{1}H-^{15}N HSQC experiment that define a binding site on the protein surface. The CSPs are used to guide and filter an AutoDock molecular docking calculation to generate a protein–ligand co-structure. The ligand binding site defined by the co-structure is then compared to other experimental binding sites in the PDB using CPASS. (Reprinted with permission from [28], copyright 2008 by Elsevier)

potential binders. These hits are then verified in a target-based screen using a 2D ^{1}H-^{15}N HSQC experiment, where the occurrence of CSPs allows for the identification of the ligand binding site. Molecular docking is used to generate a rapid protein–ligand co-structure [121] that serves as input for the Comparison of Protein Active-Site Structures (CPASS) program [153]. CPASS compares the sequence and structure of this NMR modeled ligand binding site to ~36,000 unique experimental ligand binding sites from the RCSB Protein Databank [143]. Thus, a protein of unknown function can be annotated from a protein with a known function that shares a similar ligand binding site [154]. The FAST-NMR and CPASS approach has been used for the successful annotation of two hypothetical proteins, SAV1430 from *S. aureus* [29] and PA1324 from *P. aeruginosa* [155]. It has also been used to identify a structural and functional similarity between the bacterial type III secretion system and eukaryotic apoptosis [156].

The FAST-NMR approach was recently applied to protein YndB from *Bacillus subtilis* to generate a functional annotation [112]. FAST-NMR was augmented by the inclusion of a virtual screen using the Nature Lipidomics Gateway library that contains ~22,000 lipids. Eight major categories of lipids are represented in the library (fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids, and polyketides), which are further divided into a total of 538 distinct subclasses. The initial goal was to identify lipid scaffolds that

preferentially bound YndB to infer the natural ligand. OMEGA [94] was used to generate a database of ~10,000,000 conformers from the lipid library. The program FRED was then used to dock the lipid conformer library to YndB. FRED [111] used rigid docking based on shape complementarity and a consensus scoring system to rank the ligands. The relative enrichment for each lipid class was calculated at different thresholds. Only one lipid category, the polyketides, had a positive relative enrichment, where all of the polyketides identified belonged to the flavonoid class of lipids. Within the flavonoids, three subclasses emerged as favorable hits from the virtual screen, where chalcones/hydroxychalcones, flavanones, and flavones/flavonols accounted for 44.9%, 28.6%, and 14.3% of the top 50 hits, respectively. *trans*-Chalcone, flavanone, flavone, and flavonol were selected to represent each class. The compounds were titrated into YndB to confirm binding and to measure K_D. The titrations were followed using a series of 2D ^{1}H-^{15}N HSQC NMR experiments, where CSPs were measured to calculate K_Ds (Fig. 2). *trans*-Chalcone (K_D <1 μM), flavanone (K_D 32 ±3 μM), flavone (K_D 62 ± 9 μM), and flavonol (K_D 86 ± 16 μM) were all shown to bind YndB in the same ligand binding site with K_Ds that mimicked the virtual screen ranking. Chalcones and flavonoids have not been identified among the natural products of *Bacillus* organisms, but are important precursors to plant antibiotics. The screening results are consistent with the symbiotic relationship between *B. subtilis* and plants. *B. subtilis* YndB is proposed to be part of a stress-response network that senses chalcone-like molecules during a plant's response to a pathogen infection. The stress-response may induce *B. subtilis* sporulation or the production of antibiotics to assist in combating the plant pathogens.

4.2 Rapid Protein–Ligand Structure Determination

A protein–ligand complex is instrumental to a structure-based approach to drug discovery. A new protein–ligand structure is required for each iteration of the lead modification process, until the compound has been evolved into a drug candidate. As a result, rapid protein–ligand structure determination benefits the drug discovery process. There are several methods that utilize NMR CSPs from a protein–ligand binding interaction with molecular docking to generate a corresponding co-structure. Some recent techniques include the McCoy and Wyss method [157], LIGDOCK [158], NMRScore [159], AutoDockFilter [121], QCSP-Steered Docking [160], and HADDOCK [44]. Basically, the CSPs are used to guide the docking process qualitatively and then to steer or filter the docking quantitatively. The docked model is validated by an agreement with the experimental CSPs.

AutoDockFilter (ADF) utilizes a post-filtering approach for rapidly (~35–45 min) generating a co-structure. First, CSPs from the 2D ^{1}H-^{15}N HSQC spectrum are mapped onto the protein surface to define the AutoDock 4.0 3D search grid. A 100 docked ligand poses are generated within the CSP defined search grid. Second, the CSPs are used to filter the ligand conformers and select the best pose

with the AutoDockFilter (ADF) program. ADF calculates a pseudodistance (d_{CSP}) based on the magnitude of the CSPs and compares it to the shortest distance (d_S) between any atom in the residue that incurred the CSP with any atom in the docked ligand pose. A violation energy is attributed to each protein residue that is further from the docked ligand pose then predicted by the CSP pseudodistance. The sum of these violation energies generates an overall NMR energy (E_{NMR}) for the docked ligand conformer:

$$E_{NMR} = k\sum_{i=1}^{n}(\Delta_{Dist})^2\Delta_{Dist} = \begin{cases} d_{CSP} - d_S & d_{CSP} < d_S \\ 0 & d_S \leq d_{CSP} \end{cases}. \tag{9}$$

The conformer with the lowest NMR energy corresponds to the best protein-ligand co-structure based on a consistency with the experimental CSPs. The NMR energy also provides a qualitative way to evaluate the reliability of the co-structure, with high NMR energies correlating to unreliable co-structures (Fig. 8).

NMRScore [159] is very similar to ADF. NMRScore uses poses generated by AutoDock and seven other docking programs. CSPs are calculated for each pose using DivCon, where a CSP RMSD is determined between the calculated and experimental CSPs. The best pose corresponds to the conformer with the lowest CSP RMSD. The McCoy and Wyss method [157] also uses simulated chemical shift changes. But, unlike the NMRScore approach, the docked ligand is replaced by a number of randomly placed amino-acid probes within the ligand binding site. Proton chemical shifts, primarily from ring-current effects, are calculated for the protein with and without the docked amino-acid probes. The proton chemical shifts are calculated using the SHIFTS program [161], where CSPs are determined based on the difference between the two sets of calculated proton chemical shifts. The best pose for the amino-acid probe is chosen based on a minimal difference between the experimental and calculated proton CSPs. The ligand is then docked to the protein by aligning the ligand with the amino-acid probes.

Instead of simulated chemical shifts, the HADDOCK [44] and LIGDOCK [158] programs use CSPs to define ambiguous interaction restraints (AIRs) [162]. AIRs are an intermolecular distance restraint between all atoms of the residue with the CSP and all atoms of the ligand. Importantly, other experimental information (STDs, mutational data, etc.) can also be used to define AIRS. HADDOCK and LIGDOCK employ a three-tiered approach to refining the protein–ligand complex. First, the ligand is docked to a rigid protein structure. Next, the protein–ligand structure is refined with simulated annealing in torsional space [163]. Finally, the structure is optimized with explicit solvent to remove any remaining structural problems. HADDOCK and LIGDOCK are particularly beneficial since the protein–ligand co-structure is directly refined against the experimental CSPs. The methods do suffer from long computation times and potential difficulties with proper parameterization of the ligand. HADDOCK was initially developed to dock protein–protein interactions and was later modified to accommodate ligands, whereas LIGDOCK was specifically designed to generate protein–ligand co-structures.

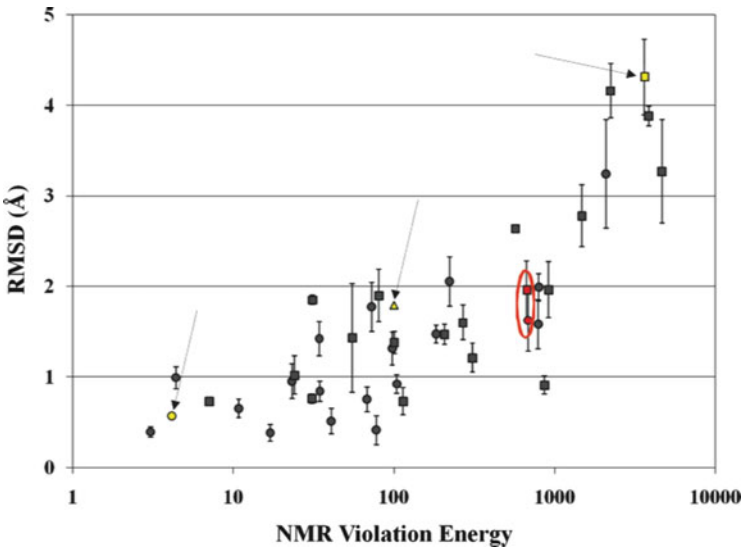

Fig. 8 A comparison of the NMR docking energy from AutoDockFilter to the rmsds between the best docked ligand conformers and the experimental protein–ligand co-structure. An improved correlation is observed for the docking of ligands to the bound form of the protein (*circles*) compared to the apo-protein structure (*squares*). The *red* data points correspond to AutoDockFilter docking results using experimental CSPs for staphylococcal nuclease (PDB-ID: 1EY0, 1SNC) [180–182]. The *yellow* data points correspond to a docking to the apo-structure of acetylcholinesterase (PDB-ID:1ACJ, 1QIF) that resulted in a high rmsd. However, the inclusion of side chain flexibility for residues in the ligand binding site resulted in an improved docking and lower rmsd. (Reprinted with permission from [121], copyright 2008 by the American Chemical Society)

Gonzalez-Ruiz and Gohlke describe a conceptual hybrid (QCSP-Steered Docking) of the AutoDockFilter and the HADDOCK/LIGDOCK procedures, effectively combining the best features of both methods [160]. AutoDock 3.0.5 was modified to incorporate a new hybrid scoring scheme utilizing the DrugScore target function [164] with an amended CSP energy function. Basically, AutoDock is used to generate poses similar to AutoDockFilter, but when an energetically acceptable pose is obtained, CSPs are calculated for the pose. The calculated CSPs are based only on ring current effects [165] from aromatic rings in the ligand. A comparison between the calculated and experimental CSPs is used to calculate an energy violation. Instead of an absolute difference, a Kendall's rank correlation coefficient is used to account for magnitude differences between the experimental and calculated CSP values. The pose with the lowest DrugScore and CSP energy is chosen. Thus, QCSP-Steered Docking is as fast as AutoDockFilter, but allows for direct refinement against the experimental CSPs like HADDOCK/LIGDOCK.

4.3 Lead Identification

Several recent approaches have investigated the combination of NMR and molecular docking for identifying inhibitors for specific proteins. Typically, these approaches apply one of two methodologies: (1) a virtual screen of a large compound library followed by validation of potential binders by NMR or (2) a fragment-based screen using NMR followed by the use of molecular docking to generate a protein–ligand co-structure for optimization.

Virtual screening followed by NMR validation is perhaps the most commonly used combination of these two techniques. Several recent studies have highlighted the use of this approach [166–169]. Branson et al. [166] used a virtual screen with NMR to identify inhibitors of lupindiadenosine 5′,5‴-P^1,P^4-tetraphosphate (Ap_4A) hydrolase. These proteins are found in eukaryotes, prokaryotes, and archaea and have been proposed to be involved in several biological functions, ranging from apoptosis, DNA repair, to gene expression. In bacteria, it has also been shown to be involved in pathogenesis, which makes this a potential target for developing antimicrobial agents. There is also a significant difference in sequence between the bacterial and animal forms of the protein, which makes this even more attractive as a drug target. In this study, a virtual screen using DOCK 4 [41] was performed on Ap_4A hydrolase from *Lupinusangustifolius* with a database of ~120,000 compounds. The docked poses from DOCK were reranked according to consensus scoring using six different scoring functions, where the top 100 ranked ligands were selected and then filtered again to remove all compounds with a logP of 3 or greater in order to select for compounds likely to be water soluble. The result was seven compounds, of which six were commercially available. These six compounds were then subjected to isothermal titration calorimetry to identify any inhibition of hydrolase activity. From that analysis, one compound (NSC51531), which contains a 1,4-diaminoanthracene-9,10-dione core, showed significant binding affinity (~1 μM K_D) and was chosen for analysis by 2D ^1H-^{15}N HSQC. The NMR analysis showed CSPs consistent with the ATP binding site of the protein. In addition, introducing NSC51531 to the human Ap_4A hydrolase showed non-specific binding and had no apparent toxic effects against human fibroblasts. This is likely due to structural differences between the binding sites of the lupin and human forms of Ap_4A hydrolase. Potentially, a scaffold based upon NSC51531 could result in an inhibitor with specificity towards the bacterial form of the protein leading to an effective microbial agent (Fig. 9a).

Veldcamp and coworkers [169] utilized a similar method that targeted the chemokine CXCL12, which activates the CXCR4 receptor shown to be involved with cancer progression. In this approach, nearly 1.5 million compounds from the ZINC database [170] were screened using DOCK 3.5 [171] against the region of CXCL12 that interacts with CXCR4. Specifically, a sulfotyrosine (sY21) was targeted since it was anticipated to be an important residue for the CXCL12-CXCR4 interaction. The top 1,000 hits were manually inspected to identify five compounds with a favorable interaction with sY21. These five compounds were

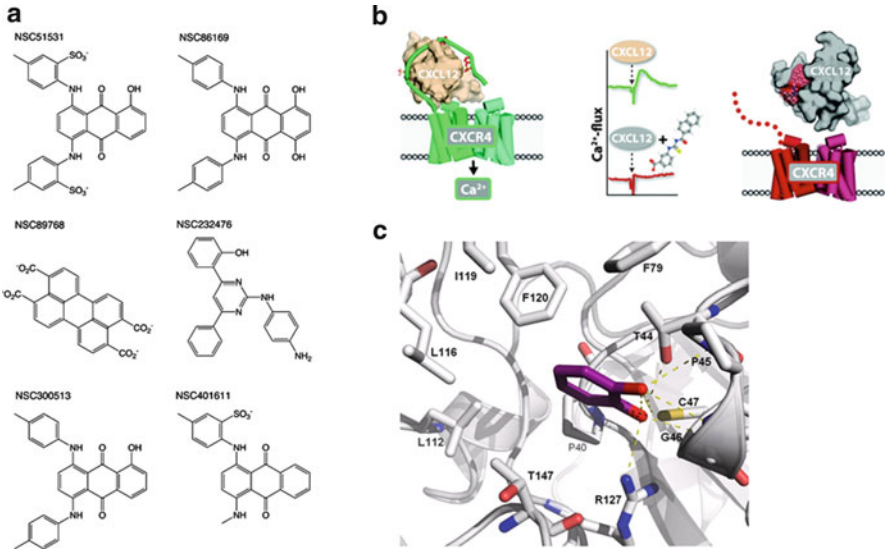

Fig. 9 (**a**) The inhibitors to lupin Ap_4A hydrolase, where NSC51531, NSC232476, and NSC89768 were identified by the virtual screen and NSC86169, NSC300513, and NSC401611 were structural analogs of NSC51531. (Reprinted with permission from [166], copyright 2009 by the American Chemical Society). (**b**) A representation of the interaction between the three sulfotyrosine groups of chemokine CXCL12 and the N-terminal region of the G-protein-coupled receptor CXCR4. Virtual screening and NMR identified 3-(naphthalene-2-carbonylthiocarbomoylamino)benzoic acid (ZINC 310454) as a possible inhibitor of the binding between CXCL12 and CXCR4, which was verified with a calcium flux assay. (Reprinted with permission from [169], copyright 2010 by the American Chemical Society). (**c**) The docked pose of fragment F152 (*magenta*) in the active site of human peroxiredoxin 5 with the hydroxyl groups oriented towards catalytic cysteine (C47). (Reprinted with permission from [174], copyright 2010 by PLoS)

then screened using 2D ^{1}H-^{15}N HSQCs, which showed that four of the compounds bound weakly, but specifically, to CXCL12 in the region of interest. The strongest binder, ZINC 310454, had a K_D of ~64 μM. Additional NMR screens with analogs to ZINC 310454 showed the importance of the carboxylic acid and naphthyl group, since analogs lacking these features showed no binding in the 2D ^{1}H-^{15}N HSQC experiments. Furthermore, a calcium flux assay demonstrated that 100 μM ZINC 310454 inhibited CXCL12-mediated signaling. Correspondingly, ZINC 310454 may be a useful scaffold for drug development (Fig. 9b). The results also reinforced the validity of chemokines as a target for drug discovery.

Using molecular docking to screen a large compound library does reduce the time and resources relative to an HTS assay, but it still suffers from an unfocused approach. In general, virtual screens or HTS assays don't efficiently sample chemical space or improve the diversity of hits. Molecular modeling also requires a priori knowledge of the binding site to guide the virtual screen, which may be difficult when dealing with new potential therapeutic targets. One approach to these problems may be to utilize NMR as the primary screening tool and molecular

docking to generate protein–ligand co-structures. Since it is not practical to use NMR to screen the large library of compounds typically utilized by HTS or virtual screening, a more focused approach with a smaller compound library is employed.

Fragment-based screening utilizes a significantly smaller library consisting of simple, low molecular-weight (<250–350 Da) molecules [15, 20–22]. These fragment-like molecules typically have weaker binding affinities (millimolar range) compared to hits found in high-throughput screens (micromolar range), but NMR is sensitive enough to detect these weak protein–ligand interactions. Importantly, fragment-based libraries are more efficient in covering chemical space. Simply, the number of possible compounds decreases drastically as the number of atoms is reduced. Thus, a smaller chemical library actually covers a larger percentage of chemical space. An even greater structural diversity can be achieved by chemically linking multiple fragments. This also results in an additive improvement in binding affinity. Evolving a drug from smaller fragments in this manner has the added benefit of improving ligand efficiency, which typically results in a more bioavailable compound that minimizes non-specific and unfavorable interactions [172, 173].

A recent study [174] by Barelier and colleagues utilized fragment-based screening by NMR and molecular docking in the investigation of the human peroxiredoxin 5 (PRDX5) ligands. Peroxiredoxins are important enzymes that catalyze the reduction of hydroperoxides through a conserved cysteine. However, very few ligands have been identified that bind these proteins despite the availability of crystal structures for PRDX5 bound with benzoate (PDB ID: 1HD2, 1H40) [175]. A compound library of 200 fragment compounds was screened by NMR using STD and WaterLOGSY experiments, where six fragments were identified as binders. STD experiments were also used to calculate the binding affinities for the six fragment molecules, which were in the 1–5 mM range. Since the 1D experiments did not provide information about the location of the binding site, AutoDock 4 [40] was used to dock the fragments to the PRDX5 protein structure. The docking was done against the entire protein structure; a grid search focusing on the benzoate ligand binding site was not used. Not surprisingly, ambiguous results were obtained. The molecular fragments bound to several locations on the PRDX5 structure that were indistinguishable based on binding energies.

Of necessity, the NMR backbone assignments for PRDX5 were obtained to enable the identification of the ligand binding site by monitoring CSPs in 2D ^{1}H-^{15}N HSQC experiments. All the fragments were shown to generate a similar set of CSPs consistent with a binding site that included the proposed catalytically active cysteine. The docked binding conformation was also further confirmed from CSPs for derivatives of these fragments. Analysis of the PRDX5 structure with the docked fragments identified the presence of a potentially important hydroxyl functional group that was pointed towards the catalytic cysteine (Fig. 9c). Interestingly, the benzoate compound found in the PRDX5 crystal structure did not show binding by NMR. But, derivatives of benzoate that included a hydroxyl functional group showed improved affinity, further indicating the importance of this hydroxyl group in ligand binding to PRDX5. These results provide further validation of the

value of combining fragment-based NMR screens with molecular docking to generate chemical leads.

While fragment-based screens have been shown to be an effective approach to drug discovery, NMR ligand affinity screens require more time and material than a virtual screen. However, fragment-based screens are extremely helpful for new therapeutic targets with unknown binding sites. Also, the approach has the added benefit of providing information about the druggability of the protein target. There is a correlation between the hit rate of a fragment-based NMR screen and the ability of the protein target to bind drug-like compounds with high affinity [176, 177].

5 Concluding Remarks

Significant advances continue to be made in the fields of molecular docking and NMR ligand affinity screens that are benefiting drug discovery. Molecular docking provides an extremely rapid way to evaluate likely binders from a large chemical library with minimal cost. Unfortunately, limitations in the accurate ranking of true binders by molecular docking programs require further experimental validation. Conversely, NMR ligand-affinity screens can directly detect a protein–ligand interaction, can measure a corresponding K_D, and can reliably identify the ligand binding site. However, NMR-ligand affinity screens are resource intensive and are generally limited to relatively small chemical libraries. Thus, the strengths and weakness of virtual screens and NMR ligand affinity screens are perfectly complementary. Combining the two screening techniques has the potential of significantly improving the efficiency of drug discovery. The combination of NMR and molecular modeling techniques has been shown to enable the rapid determination of reliable protein–ligand co-structures, the identification of new therapeutic targets, and the successful discovery of new drug leads.

References

1. Venter JC et al (2001) The sequence of the human genome. Science 291(5507):1304–1351
2. Paul SM et al (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov 9(3):203–214
3. Kola I, Landis J (2004) Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 3(8):711–715
4. Cuatrecasas P (2006) Drug discovery in jeopardy. J Clin Invest 116(11):2837–2842
5. Bernal A, Ear U, Kyrpides N (2001) Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. Nucleic Acids Res 29(1):126–127
6. Frishman D et al (2003) The PEDANT genome database. Nucleic Acids Res 31(1):207–211
7. White RH (2006) The difficult road from sequence to function. J Bacteriol 188(10): 3431–3432
8. Gerlt JA, Babbitt PC (2000) Can sequence determine function? Genome Biol 1(5): REVIEWS0005

9. Rishton GM (1997) Reactive compounds and in vitro false positives in HTS. Drug Discov Today 2(9):382–384
10. Seidler J et al (2003) Identification and prediction of promiscuous aggregating inhibitors among known drugs. J Med Chem 46(21):4477–4486
11. McGovern SL et al (2003) A specific mechanism of nonspecific inhibition. J Med Chem 46(20):4265–4272
12. McGovern SL et al (2002) A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. J Med Chem 45(8):1712–1722
13. Kenny BA et al (1998) The application of high-throughput screening to novel lead discovery. Prog Drug Res 51:245–269
14. Macarron R (2006) Critical review of the role of HTS in drug discovery. Drug Discov Today 11(7–8):277–279
15. Powers R (2009) Advances in nuclear magnetic resonance for drug discovery. Expert Opin Drug Discov 4(10):1077–1098
16. Pellecchia M et al (2008) Perspectives on NMR in drug discovery: a technique comes of age. Nat Rev Drug Discov 7(9):738–745
17. Roberts GCK (2000) Applications of NMR in drug discovery. Drug Discov Today 5(6): 230–240
18. Huth JR et al (2004) ALARM NMR: a rapid and robust experimental method to detect reactive false positives in biochemical screens. J Am Chem Soc 127(1):217–224
19. Dalvit C et al (2006) NMR-based quality control approach for the identification of false positives and false negatives in high throughput screening. Curr Drug Discov Technol 3(2): 115–124
20. Schade M (2007) Fragment-based lead discovery by NMR. Front Drug Des Discov 3:105–119
21. Zartler ER, Mo H (2007) Practical aspects of NMR-based fragment discovery. Curr Top Med Chem 7(16):1592–1599
22. Dalvit C (2009) NMR methods in fragment screening: theory and a comparison with other biophysical techniques. Drug Discov Today 14(21/22):1051–1057
23. Fesik SW (1993) NMR structure-based drug design. J Biomol NMR 3(3):261–269
24. Kubinyi H (1998) Structure-based design of enzyme inhibitors and receptor ligands. Curr Opin Drug Discov Devel 1(1):4–15
25. Ishihara K et al (2009) Identification of urinary biomarkers useful for distinguishing a difference in mechanism of toxicity in rat model of cholestasis. Basic Clin Pharmacol Toxicol 105(3):156–166
26. Ott K-H, Aranibar N (2007) Nuclear magnetic resonance metabonomics: methods for drug discovery and development. Methods Mol Biol 358:247–271
27. Powers R (2009) NMR metabolomics and drug discovery. Magn Reson Chem 47(S1): S2–S11
28. Powers R, Mercier KA, Copeland JC (2008) The application of FAST-NMR for the identification of novel drug discovery targets. Drug Discov Today 13(3–4):172–179
29. Mercier KA et al (2006) FAST-NMR: functional annotation screening technology using NMR spectroscopy. J Am Chem Soc 128(47):15292–15299
30. Kitchen DB et al (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3(11):935–949
31. Halperin I et al (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 47(4):409–443
32. Warren GL et al (2006) A critical assessment of docking programs and scoring functions. J Med Chem 49(20):5912–5931
33. Hartshorn MJ et al (2007) Diverse, high-quality test set for the validation of protein-ligand docking performance. J Med Chem 50(4):726–741
34. Jacobsson M et al (2003) Improving structure-based virtual screening by multivariate analysis of scoring data. J Med Chem 46(26):5781–5789

35. Loving K, Salam NK, Sherman W (2009) Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. J Comput Aided Mol Des 23(8):541–554
36. Salaniwal S et al (2007) Critical evaluation of methods to incorporate entropy loss upon binding in high-throughput docking. Proteins 66(2):422–435
37. Vasilyev V, Bliznyuk A (2004) Application of semiempirical quantum chemical methods as a scoring function in docking. Theor Chem Acc 112(4):313–317
38. Wei D et al (2010) Binding energy landscape analysis helps to discriminate true hits from high-scoring decoys in virtual screening. J Chem Inf Model 50(10):1855–1864
39. Zavodszky MI et al (2009) Scoring ligand similarity in structure-based virtual screening. J Mol Recognit 22(4):280–292
40. Morris GM et al (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30(16):2785–2791
41. Ewing TJ et al (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des 15(5):411–428
42. Rarey M et al (1996) A fast flexible docking method using an incremental construction algorithm. J Mol Biol 261(3):470–489
43. Friesner RA et al (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47(7):1739–1749
44. Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125(7): 1731–1737
45. Bohm HJ (1992) LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. J Comput Aided Mol Des 6(6):593–606
46. Bohm HJ (1992) The computer program LUDI: a new method for the de novo design of enzyme inhibitors. J Comput Aided Mol Des 6(1):61–78
47. Cerqueira NMFSA et al. (2010) Virtual screening of compound libraries. Methods Mol Biol 572:57–70 (Ligand-Macromolecular Interactions in Drug Discovery)
48. Ripphausen P et al (2010) Quo vadis, virtual screening? A comprehensive survey of prospective applications. J Med Chem 53(24):8461–8467
49. Sousa SF et al (2010) Virtual screening in drug design and development. Comb Chem High Throughput Screen 13(5):442–453
50. Eckert H, Bajorath J (2007) Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. Drug Discov Today 12(5&6):225–233
51. Merz KM Jr (2010) Limits of free energy computation for protein-ligand interactions. J Chem Theory Comput 6(5):1769–1776
52. Proschak E et al (2007) Shapelets: possibilities and limitations of shape-based virtual screening. J Comput Chem 29(1):108–114
53. Wyss DF, McCoy MA, Senior MM (2002) NMR-based approaches for lead discovery. Curr Opin Drug Discov Devel 5(4):630–647
54. Lepre CA, Moore JM, Peng JW (2004) Theory and applications of NMR-based screening in pharmaceutical research. Chem Rev 104(8):3641–3676
55. Mercier KA, Powers R (2005) Determining the optimal size of small molecule mixtures for high throughput NMR screening. J Biomol NMR 31(3):243–258
56. Hajduk PJ, Olejniczak ET, Fesik SW (1997) One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules. J Am Chem Soc 119:12257–12261
57. Mayer M, Meyer B (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. Angew Chem Int Ed 38(12):1784–1788
58. Dalvit C et al (2000) Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. J Biomol NMR 18(1):65–68
59. Jahnke W, Rudisser S, Zurini M (2001) Spin label enhanced NMR screening. J Am Chem Soc 123(13):3149–3150

60. Vanwetswinkel S et al (2005) TINS, target immobilized NMR screening: an efficient and sensitive method for ligand discovery. Chem Biol 12(2):207–216
61. Fejzo J et al (1999) The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. Chem Biol 6(10):755–769
62. Dalvit C et al (2003) Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability. J Am Chem Soc 125(25): 7696–7703
63. Dalvit C et al (2002) Fluorine-NMR competition binding experiments for high-throughput screening of large compound mixtures. Comb Chem High Throughput Screen 5(8):605–611
64. Dalvit C et al (2003) A general NMR method for rapid, efficient, and reliable biochemical screening. J Am Chem Soc 125(47):14620–14625
65. Dalvit C et al (2004) Reliable high-throughput functional screening with 3-FABS. Drug Discov Today 9(14):595–602
66. Price SW (1997) Pulsed-field gradient nuclear magnetic resonance as a tool for studying translational diffusion: part 1. Basic theory. Concepts Magn Reson 9:299–336
67. Price SW (1998) Pulsed-field gradient nuclear magnetic resonance as a tool for studying translational diffusion: part II. Experimental aspects. Concepts Magn Reson 10:197–237
68. Shortridge MD et al (2008) Estimating protein-ligand binding affinity using high-throughput screening by NMR. J Comb Chem 10(6):948–958
69. Ji Z, Yao Z, Liu M (2009) Saturation transfer difference nuclear magnetic resonance study on the specific binding of ligand to protein. Anal Biochem 385(2):380–382
70. Muhandiram DR et al (1993) A gradient 13C NOESY-HSQC experiment for recording NOESY spectra of 13C-labeled proteins dissolved in H2O. J Magn Reson B 102(3):317–321
71. Sklenar V et al (1993) Gradient-tailored water suppression for proton-nitrogen-15 HSQC experiments optimized to retain full sensitivity. J Magn Reson A 102(2):241–245
72. Per VK et al (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci USA 94(23):12366–12371
73. Shuker SB et al (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science 274(5292):1531–1534
74. Fielding L (2007) NMR methods for the determination of protein-ligand dissociation constants. Prog Nucl Magn Reson Spectrosc 51:219–242
75. Morton CJ et al (1996) Solution structure and peptide binding of the SH3 domain from human Fyn. Structure 4(6):705–714
76. Stoll F (2003) Library design. Chimia 57(5):224–228
77. Erlanson DA, McDowell RS, O'Brien T (2004) Fragment-based drug discovery. J Med Chem 47(14):3463–3482
78. Siegal G, Ab E, Schultz J (2007) Integration of fragment screening and library design. Drug Discov Today 12(23&24):1032–1039
79. Lipinski CA (2004) Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol 1(4):337–341
80. Schanda P, Kupce E, Brutscher B (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. J Biomol NMR 33(4):199–211
81. Schanda P, Brutscher B (2006) Hadamard frequency-encoded SOFAST-HMQC for ultrafast two-dimensional protein NMR. J Magn Reson 178(2):334–339
82. Mori S et al (1995) Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. J Magn Reson B 108(1):94–98
83. Taufer M et al (2005) Study of an accurate and fast protein-ligand docking algorithm based on molecular dynamics. Concurr Comput 17(14):1627–1641
84. Garcia-Sosa AT, Sild S, Maran U (2009) Docking and virtual screening using distributed grid technology. QSAR Comb Sci 28:815–821

85. Kuntz ID, Meng EC, Shoichet BK (1994) Structure-based molecular design. Acc Chem Res 27(5):117–123
86. Krovat EM, Steindl T, Langer T (2005) Recent advances in docking and scoring. Curr Comput Aided Drug Des 1(1):93–102
87. Cole JC et al (2005) Comparing protein-ligand docking programs is difficult. Proteins 60 (3):325–332
88. Wandzik I (2006) Current molecular docking tools and comparisons thereof. MATCH 55 (2):271–278
89. Dias R, de Azevedo WF Jr (2008) Molecular docking algorithms. Curr Drug Targets 9 (12):1040–1047
90. Viji SN, Prasad PA, Gautham N (2009) Protein-ligand docking using mutually orthogonal Latin squares (MOLSDOCK). J Chem Inf Model 49(12):2687–2694
91. Pearce BC et al (2009) E-novo: an automated workflow for efficient structure-based lead optimization. J Chem Inf Model 49(7):1797–1809
92. Mizutani MY, Tomioka N, Itai A (1994) Rational automatic search method for stable docking models of protein and ligand. J Mol Biol 243(2):310–326
93. Schlosser J, Rarey M (2009) Beyond the virtual screening paradigm: structure-based searching for new lead compounds. J Chem Inf Model 49(4):800–809
94. Bostrom J, Greenwood JR, Gottfries J (2003) Assessing the performance of OMEGA with respect to retrieving bioactive conformations. J Mol Graph Model 21(5):449–462
95. Miller MD et al (1994) FLOG: a system to select quasi-flexible ligands complementary to a receptor of known three-dimensional structure. J Comput Aided Mol Des 8(2):153–174
96. Griewel A et al (2009) Conformational sampling for large-scale virtual screening: accuracy versus ensemble size. J Chem Inf Model 49(10):2303–2311
97. Hart TN, Read RJ (1994) Multiple-start Monte Carlo docking of flexible ligands. Birkhaeuser, Boston
98. Fuhrmann J et al (2010) A new Lamarckian genetic algorithm for flexible ligand-receptor docking. J Comput Chem 31(9):1911–1918
99. Cao T, Li T (2004) A combination of numeric genetic algorithm and tabu search can be applied to molecular docking. Comput Biol Chem 28(4):303–312
100. Huang S-Y, Zou X (2010) Advances and challenges in protein-ligand docking. Int J Mol Sci 11:3016–3034
101. Huang S-Y, Grinter SZ, Zou X (2010) Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. Phys Chem Chem Phys 12(40): 12899–12908
102. Huang S-Y, Zou X (2010) Mean-force scoring functions for protein-ligand binding. Annu Rep Comput Chem 6:281–296
103. Bohme A et al (1998) Piperacillin/tazobactam versus cefepime as initial empirical antimicrobial therapy in febrile neutropenic patients: a prospective randomized pilot study. Eur J Med Res 3(7):324–330
104. Eldridge MD et al (1997) 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 11(5):425–445
105. Tao P, Lai L (2001) Protein ligand docking based on empirical method for binding affinity estimation. J Comput Aided Mol Des 15(5):429–446
106. Wang R, Lai L, Wang S (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J Comput Aided Mol Des 16(1): 11–26
107. Muegge I, Martin YC (1999) A general and fast scoring function for protein-ligand interactions: a simplified potential approach. J Med Chem 42(5):791–804
108. Gohlke H, Hendlich M, Klebe G (2000) Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol 295(2):337–356

109. Velec HF, Gohlke H, Klebe G (2005) 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 48(20):6296–6303
110. Kortagere S, Krasowski MD, Ekins S (2009) The importance of discerning shape in molecular pharmacology. Trends Pharmacol Sci 30(3):138–147
111. McGann MR et al (2003) Gaussian docking functions. Biopolymers 68(1):76–90
112. Stark JL et al (2010) Solution structure and function of YndB, an AHSA1 protein from Bacillus subtilis. Proteins 78(16):3328–3340
113. Merlitz H, Herges T, Wenzel W (2004) Fluctuation analysis and accuracy of a large-scale in silico screen. J Comput Chem 25(13):1568–1575
114. Tirado-Rives J, Jorgensen WL (2006) Contribution of conformer focusing to the uncertainty in predicting free energies for protein-ligand binding. J Med Chem 49(20):5880–5884
115. Charifson PS et al (1999) Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. J Med Chem 42(25): 5100–5109
116. Peng JW (2009) Communication breakdown: protein dynamics and drug design. Structure 17(3):319–320
117. Hayward S, de Groot BL (2008) Normal modes and essential dynamics. Methods Mol Biol 443:89–106 (Molecular Modeling of Proteins)
118. Huang SY, Zou X (2007) Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. Proteins 66(2):399–421
119. Erickson JA et al (2004) Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. J Med Chem 47(1):45–55
120. Sherman W et al (2006) Novel procedure for modeling ligand/receptor induced fit effects. J Med Chem 49(2):534–553
121. Stark J, Powers R (2008) Rapid protein-ligand costructures using chemical shift perturbations. J Am Chem Soc 130(2):535–545
122. B-Rao C, Subramanian J, Sharma SD (2009) Managing protein flexibility in docking and its applications. Drug Discov Today 14(7–8):394–400
123. Jiang F, Kim SH (1991) "Soft docking": matching of molecular surface cubes. J Mol Biol 219 (1):79–102
124. Claussen H et al (2001) FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 308(2):377–395
125. Alberts IL, Todorov NP, Dean PM (2005) Receptor flexibility in de novo ligand design and docking. J Med Chem 48(21):6585–6596
126. Barril X, Morley SD (2005) Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. J Med Chem 48(13):4432–4443
127. Klebe G (2006) Virtual ligand screening: strategies, perspectives and limitations. Drug Discov Today 11(13–14):580–594
128. Schneider G, Bohm HJ (2002) Virtual screening and fast automated docking methods. Drug Discov Today 7(1):64–70
129. Chen H et al (2006) On evaluating molecular-docking methods for pose prediction and enrichment factors. J Chem Inf Model 46(1):401–415
130. Kontoyianni M, McClellan LM, Sokol GS (2004) Evaluation of docking performance: comparative data on docking algorithms. J Med Chem 47(3):558–565
131. Kirchmair J et al (2008) Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection–what can we learn from earlier mistakes? J Comput Aided Mol Des 22(3–4):213–228
132. Bender A, Glen RC (2005) A discussion of measures of enrichment in virtual screening: comparing the information content of descriptors with increasing levels of sophistication. J Chem Inf Model 45(5):1369–1375
133. Truchon J-F, Bayly CI (2007) Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. J Chem Inf Model 47(2):488–508

134. Scsibrany H et al (2003) Clustering and similarity of chemical structures represented by binary substructure descriptors. Chemom Intell Lab Syst 67(2):95–108
135. Davis AM et al (2005) Components of successful lead generation. Curr Top Med Chem 5(4):421–439
136. Sams-Dodd F (2006) Drug discovery: selecting the optimal approach. Drug Discov Today 11(9–10):465–472
137. Fink T, Reymond JL (2007) Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. J Chem Inf Model 47(2):342–353
138. Lahana R (1999) How many leads from HTS? Drug Discov Today 4(10):447–448
139. Goode DR et al (2008) Identification of promiscuous small molecule activators in high-throughput enzyme activation screens. J Med Chem 51(8):2346–2349
140. Foloppe N et al (2006) Identification of chemically diverse Chk1 inhibitors by receptor-based virtual screening. Bioorg Med Chem 14(14):4792–4802
141. Richardson CM et al (2007) Discovery of a potent CDK2 inhibitor with a novel binding mode, using virtual screening and initial, structure-guided lead scoping. Bioorg Med Chem Lett 17(14):3880–3885
142. Pellecchia M et al (2004) NMR-based techniques in the hit identification and optimisation processes. Expert Opin Ther Targets 8(6):597–611
143. Galperin MY, Koonin EV (2010) From complete genome sequence to 'complete' understanding? Trends Biotechnol 28(8):398–406
144. Tucker CL (2002) High-throughput cell-based assays in yeast. Drug Discov Today 7(18 Suppl):S125–S130
145. Lee YH et al (2005) Gene knockdown by large circular antisense for high-throughput functional genomics. Nat Biotechnol 23(5):591–599
146. Joshi T et al (2004) Genome-scale gene function prediction using multiple sources of high-throughput data in yeast Saccharomyces cerevisiae. OMICS 8(4):322–333
147. del Val C et al (2004) High-throughput protein analysis integrating bioinformatics and experimental assays. Nucleic Acids Res 32(2):742–748
148. Laurie AT, Jackson RM (2006) Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. Curr Protein Pept Sci 7(5):395–406
149. Blundell TL et al (2006) Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. Philos Trans R Soc Lond B Biol Sci 361(1467):413–423
150. Vajda S, Guarnieri F (2006) Characterization of protein-ligand interaction sites using experimental and computational methods. Curr Opin Drug Discov Devel 9(3):354–362
151. Mercier KA, Germer K, Powers R (2006) Design and characterization of a functional library for NMR screening against novel protein targets. Comb Chem High Throughput Screen 9(7): 515–534
152. Mercier KA, Shortridge MD, Powers R (2009) A multi-step NMR screen for the identification and evaluation of chemical leads for drug discovery. Comb Chem High Throughput Screen 12(3):285–295
153. Powers R et al (2006) Comparison of protein active site structures for functional annotation of proteins and drug design. Proteins 65(1):124–135
154. Park K, Kim D (2008) Binding similarity network of ligand. Proteins 71(2):960–971
155. Mercier KA et al (2009) Structure and function of Pseudomonas aeruginosa protein PA1324 (21-170). Protein Sci 18(3):606–618
156. Shortridge MD, Powers R (2009) Structural and functional similarity between the bacterial type III secretion system needle protein PrgI and the eukaryotic apoptosis Bcl-2 proteins. PLoS One 4(10):e7442
157. McCoy MA, Wyss DF (2000) Alignment of weakly interacting molecules to protein surfaces using simulations of chemical shift perturbations. J Biomol NMR 18(3):189–198

158. Schieborr U et al (2005) How much NMR data is required to determine a protein-ligand complex structure? Chembiochem 6(10):1891–1898
159. Wang B, Westerhoff LM, Merz KM Jr (2007) A critical assessment of the performance of protein-ligand scoring functions based on NMR chemical shift perturbations. J Med Chem 50(21):5128–5134
160. Gonzalez-Ruiz D, Gohlke H (2009) Steering protein-ligand docking with quantitative NMR chemical shift perturbations. J Chem Inf Model 49(10):2260–2271
161. Xu X-P, Case DA (2001) Automated prediction of 15N, 13CÎ±, 13CÎ and 13C′ chemical shifts in proteins using a density functional database. J Biomol NMR 21(4):321–333
162. Nilges M (1995) Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. J Mol Biol 245(5):645–660
163. Guntert P, Wuthrich K (2001) Sampling of conformation space in torsion angle dynamics calculations. Comput Phys Commun 138(2):155–169
164. Gohlke H, Hendlich M, Klebe G (2000) Predicting binding modes, binding affinities and "hot spots" for protein-ligand complexes using a knowledge-based scoring function. Perspect Drug Discov Des 20:115–144
165. Osapay K, Case DA (1991) A new analysis of proton chemical shifts in proteins. J Am Chem Soc 113(25):9436–9444
166. Branson KM et al (2009) Discovery of inhibitors of lupin diadenosine 5′,5‴-P(1), P(4)-tetraphosphate hydrolase by virtual screening. Biochemistry 48(32):7614–7620
167. Jacobsson M et al (2008) Identification of Plasmodium falciparum spermidine synthase active site binders through structure-based virtual screening. J Med Chem 51(9):2777–2786
168. Lee Y et al (2009) Identification of compounds exhibiting inhibitory activity toward the Pseudomonas tolaasii toxin tolaasin I using in silico docking calculations, NMR binding assays, and in vitro hemolytic activity assays. Bioorg Med Chem Lett 19(15):4321–4324
169. Veldkamp CT et al (2010) Targeting SDF-1/CXCL12 with a ligand that prevents activation of CXCR4 through structure-based drug design. J Am Chem Soc 132(21):7242–7243
170. Irwin JJ, Shoichet BK (2005) ZINC–a free database of commercially available compounds for virtual screening. J Chem Inf Model 45(1):177–182
171. Lorber DM, Shoichet BK (2005) Hierarchical docking of databases of multiple ligand conformations. Curr Top Med Chem 5(8):739–749
172. Bembenek SD, Tounge BA, Reynolds CH (2009) Ligand efficiency and fragment-based drug discovery. Drug Discov Today 14(5–6):278–283
173. Reynolds CH, Tounge BA, Bembenek SD (2008) Ligand binding efficiency: trends, physical basis, and implications. J Med Chem 51(8):2432–2438
174. Barelier S et al (2010) Discovery of fragment molecules that bind the human peroxiredoxin 5 active site. PLoS One 5(3):e9744
175. Declercq JP et al (2001) Crystal structure of human peroxiredoxin 5, a novel type of mammalian peroxiredoxin at 1.5 A resolution. J Mol Biol 311(4):751–759
176. Hajduk PJ, Huth JR, Fesik SW (2005) Druggability indices for protein targets derived from NMR-based screening data. J Med Chem 48(7):2518–2525
177. Hajduk PJ, Huth JR, Tse C (2005) Predicting protein druggability. Drug Discov Today 10(23–24):1675–1682
178. Peng JW et al (2001) Nuclear magnetic resonance-based approaches for lead generation in drug discovery. Methods Enzymol 338:202–230
179. Morris GM et al (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19(14):1639–1662
180. Wang JF et al (1992) Solution studies of staphylococcal nuclease H124L. 2. 1H, 13C, and 15N chemical shift assignments for the unligated enzyme and analysis of chemical shift changes that accompany formation of the nuclease-thymidine 3′,5′-bisphosphate-calcium ternary complex. Biochemistry 31(3):921–936

181. Wang JF et al (1990) Two-dimensional NMR studies of staphylococcal nuclease. 2. Sequence-specific assignments of carbon-13 and nitrogen-15 signals from the nuclease H124L-thymidine 3′,5′-bisphosphate-Ca2+ ternary complex. Biochemistry 29(1):102–113
182. Wang JF, LeMaster DM, Markley JL (1990) Two-dimensional NMR studies of staphylococcal nuclease. 1. Sequence-specific assignments of hydrogen-1 signals and solution structure of the nuclease H124L-thymidine 3′,5′-bisphosphate-Ca2+ ternary complex. Biochemistry 29(1):88–101