DOCKING AND SCORING IN VIRTUAL SCREENING FOR DRUG DISCOVERY: METHODS AND APPLICATIONS

Douglas B. Kitchen, Hélène Decornez*, John R. Furr* and Jürgen Bajorath‡,§*

Abstract | Computational approaches that 'dock' small molecules into the structures of macromolecular targets and 'score' their potential complementarity to binding sites are widely used in hit identification and lead optimization. Indeed, there are now a number of drugs whose development was heavily influenced by or based on structure-based design and screening strategies, such as HIV protease inhibitors. Nevertheless, there remain significant challenges in the application of these approaches, in particular in relation to current scoring schemes. Here, we review key concepts and specific features of small-molecule–protein docking methods, highlight selected applications and discuss recent advances that aim to address the acknowledged limitations of established approaches.

**Department of Computer-Aided Drug Discovery, Albany Molecular Research, Inc. (AMRI), 21 Corporate Circle, Albany, New York 12212-5098, USA.*
‡AMRI Bothell Research Center (AMRI-BRC), 18804 North Creek Parkway, Bothell, Washington 98011, USA.
§Department of Biological Structure, University of Washington, Seattle, Washington 98195, USA.
Correspondence to J.B.
e-mail: jurgen.bajorath@albmolecular.com
doi:10.1038/nrd1549

The number of proteins with a known three-dimensional structure is increasing rapidly, and structures produced by structural genomics initiatives are beginning to become publicly available[1,2]. The increase in the number of structural targets is in part due to improvements in techniques for structure determination, such as high-throughput X-ray crystallography[3]. With large-scale structure-determination projects driven by genomics consortia, many current target proteins have been selected for their therapeutic potential.

Computational methodologies have become a crucial component of many drug discovery programmes, from hit identification to lead optimization and beyond[4–6], and approaches such as ligand-[4] or structure-based virtual screening[7] techniques are widely used in many discovery efforts. One key methodology — docking of small molecules to protein binding sites — was pioneered during the early 1980s[8], and remains a highly active area of research[7]. When only the structure of a target and its active or binding site is available, high-throughput docking is primarily used as a hit-identification tool. However, similar calculations are often also used later on during lead optimization, when modifications to known active structures can quickly be tested in computer models before compound synthesis. Furthermore, docking can also contribute to the analysis of drug metabolism using structures such as cytochrome P450 isoforms[9,10].

Here, we review basic concepts and specific features of small-molecule–protein docking methods and several selected applications, with particular emphasis on hit identification and lead optimization, but do not specifically review protein–protein docking, which is less relevant for small-molecule drug discovery. We attempt to distinguish between the problems of docking compounds into target sites and of scoring docked conformations, because the available data indicate that numerous robust and accurate docking algorithms are available, whereas imperfections of scoring functions continue to be a major limiting factor.

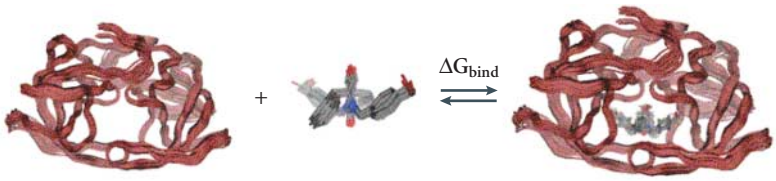
An introduction to docking

The docking process involves the prediction of ligand conformation and orientation (or posing) within a targeted binding site (BOX 1). In general, there are two aims of docking studies: accurate structural modelling and correct prediction of activity. However, the identification of molecular features that are responsible for specific biological recognition, or the prediction of compound modifications that improve potency, are

Box 1 | **Theoretical aspects of docking**

For an enzyme and inhibitor, docking aims at correct prediction of the structure of the complex [E+I] = [EI] under equilibrium conditions (see figure and equation 1).

$$[EI]_{aq} \rightleftharpoons [E]_{aq} + [I]_{aq} \quad (1)$$



$[E]_{aq}$ $[I]_{aq}$ $[E + I]_{aq}$

The figure illustrates the binding of inhibitor Dmp323 to HIV protease and is based on solution structures (PDB code: 1BVE). Multiple structures of enzyme–inhibitor complexes revealed only limited structural variations.

The free energy of binding (ΔG) is related to binding affinity by equations 2 and 3:

$$\Delta G = -RT\ln K_A \qquad K_A = K_i^{-1} = \frac{[EI]}{[E][I]} \quad (2,3)$$

Prediction of the correct structure (posing) of the [E+I] complex does not require information about K_A. However, prediction of biological activity (ranking) requires this information; scoring terms can therefore be divided in the following fashion. When considering the term [EI], the following factors are important: steric, electrostatic, hydrogen bonding, inhibitor strain (if flexible) and enzyme strain. When considering the equilibrium shown in equation 1, the following factors are also important: desolvation, rotational entropy and translational entropy.

complex issues that are often difficult to understand and — even more so — to simulate on a computer.

In view of these challenges, docking is generally devised as a multi-step process in which each step introduces one or more additional degrees of complexity[11]. The process begins with the application of docking algorithms that POSE small molecules in the active site. This in itself is challenging, as even relatively simple organic molecules can contain many conformational degrees of freedom. Sampling these degrees of freedom must be performed with sufficient accuracy to identify the conformation that best matches the receptor structure, and must be fast enough to permit the evaluation of thousands of compounds in a given docking run. Algorithms are complemented by SCORING FUNCTIONS that are designed to predict the biological activity through the evaluation of interactions between compounds and potential targets. Early scoring functions evaluated compound fits on the basis of calculations of approximate shape and electrostatic complementarities. Relatively simple scoring functions continue to be heavily used, at least during the early stages of docking simulations. Pre-selected conformers are often further evaluated using more complex scoring schemes with more detailed treatment of electrostatic and van der Waals interactions, and inclusion of at least some solvation or entropic effects[7]. It should also be noted that ligand-binding events are driven by a combination of enthalpic and entropic effects, and that either entropy or enthalpy can dominate specific interactions. This often presents a conceptual problem for contemporary scoring functions (discussed below), because most of them are much more focused on capturing energetic than entropic effects.

In addition to problems associated with scoring of compound conformations, other complications exist that make it challenging to accurately predict binding conformations and compound activity. These include, among others, limited resolution of crystallographic targets, inherent flexibility, induced fit or other conformational changes that occur on binding, and the participation of water molecules in protein–ligand interactions. Without doubt, the docking process is scientifically complex.

POSING
The process of determining whether a given conformation and orientation of a ligand fits the active site. This is usually a fuzzy procedure that returns many alternative results.

SCORING
Both posing and ranking involve scoring. The pose score is often a rough measure of the fit of a ligand into the active site. The rank score is generally more complex and might attempt to estimate binding energies.

RANKING
A more advanced process than pose scoring that typically takes several results from an initial scoring phase and re-evaluates them. This process usually attempts to estimate the free energy of binding as accurately as possible. Although the posing phase might use simple energy calculations (electrostatic and van der Waals), ranking procedures typically involve more elaborate calculations (perhaps including properties such as entropy or explicit solvation).

Molecular representations for docking

To evaluate various docking methods, it is important to consider how the protein and ligand are represented. There are three basic representations of the receptor: atomic, surface and grid[12]. Among these, atomic representation is generally only used in conjunction with a potential energy function[13] and often only during final RANKING procedures (because of the computational complexity of evaluating pair-wise atomic interactions).

Surface-based docking programs are typically, but not exclusively, used in protein–protein docking[14,15]. Connolly's early work on molecular surface representations is mainly responsible for spawning much of the research in this area[16,17]. These methods attempt to align points on surfaces by minimizing the angle between the surfaces of opposing molecules[18]. Therefore, a rigid body approximation is still the standard for many protein–protein docking techniques.

The use of potential energy grids was pioneered by Goodford[19], and various docking programs use such grid representations for energy calculations. The basic idea is to store information about the receptor's energetic contributions on grid points so that it only needs to be read during ligand scoring. In the most basic form, grid points store two types of potentials: electrostatic and van der Waals (BOX 2). FIGURE 1 shows a representative grid for capturing electrostatic potentials, and FIG. 2 illustrates the electrostatic potential of a bound inhibitor mapped on its molecular surface.

Search methods and molecular flexibility

This section focuses on algorithms used to treat ligand flexibility and, to some extent, protein flexibility. Treatment of ligand flexibility can be divided into three basic categories[11]: systematic methods (incremental construction, conformational search, databases); random or stochastic methods (Monte Carlo, genetic algorithms, tabu search); and simulation methods (molecular dynamics, energy minimization). A summary of the search approaches implemented in widely used docking programs is presented in BOX 3.

Systematic search. These algorithms try to explore all the degrees of freedom in a molecule, but ultimately face the problem of combinatorial explosion[20] (BOX 4). Therefore, ligands are often incrementally grown into active sites. A stepwise or incremental search can be accomplished in different ways — for example, by docking various molecular fragments into the active-site region and linking

Box 2 | **Standard potential energy functions**

The electrostatic potential energy is represented as a pair-wise summation of Coulombic interactions, as described in equation 1:

$$E_{coul}(r) = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (1)$$

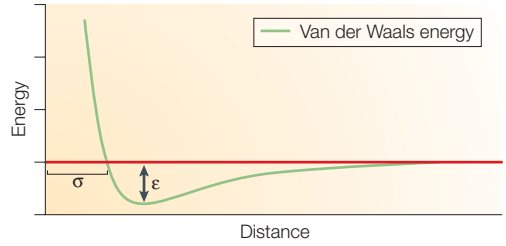
In equation 1, N is the number of atoms in molecules A and B, respectively, and q the charge on each atom.

The van der Waals potential energy for the general treatment of non-bonded interactions is often modelled by a Lennard–Jones 12–6 function, as shown in equation 2:

$$E_{vdW}(r) = \sum_{j=1}^{N} \sum_{i=1}^{N} 4\varepsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \quad (2)$$

In equation 2, ε is the well depth of the potential and σ is the collision diameter of the respective atoms *i* and *j*.

The figure shows a representation of the Lennard–Jones 12–6 function. The exp(12) term of the equation is responsible for small-distance repulsion, whereas the exp(6) provides an attractive term which approaches zero as the distance between the two atoms increases.



them covalently (which is most popular as a *de novo* ligand-design strategy) or, alternatively, by dividing docked ligands into rigid (core fragment) and flexible parts (side chains). In the latter case, once the rigid cores have been defined, they are docked into the active site. Next, flexible regions are added in an incremental fashion[21–23]. For example, DOCK 4.0[24] poses the core fragment by steric complementarity, and flexible side chains are grown one bond at a time by systematically exploring each bond's POSE SPACE. A pruning algorithm is applied to remove unfavourable conformations early on, thereby reducing the complexity of the problem[24,25]. FlexX differs from DOCK in that the placement of the rigid core fragment is based on interaction geometries between fragments and receptor groups[22,26]. Interacting groups are primarily hydrogen-bond donors and acceptors, as well as hydrophobic groups. FlexX further differs from DOCK in that it uses a pose-clustering algorithm to classify the docked poses[22,27].

The Hammerhead algorithm[28], in common with other incremental search algorithms, also divides ligands into fragments. However, Hammerhead docks each fragment and then rebuilds the ligand from fragments that have acceptable initial scores. During the fragment-growing stage, energy minimization is performed after each new addition[28].

Another method of systematic search is the use of libraries of pre-generated conformations. Library conformations are typically only calculated once and the search problem is therefore reduced to a rigid body docking procedure. For example, FLOG[29] generates database conformations on the basis of distance geometry. Once acceptable conformations have been generated, the algorithm explores them in a manner similar to DOCK[11,29].

Random search. These algorithms (often called stochastic methods) operate by making random changes to either a single ligand or a population of ligands. A newly obtained ligand is evaluated on the basis of a pre-defined probability function. Two popular random approaches are Monte Carlo and genetic algorithms (BOX 5). Alternative implementations of Monte Carlo search have been reported[30,31], including a popular form in AutoDock[30]. By contrast, several other programs (including DOCK and GOLD) have implemented genetic algorithms[32–34].

The basic idea of a tabu search algorithm is to take into consideration already explored areas of conformational space[35,36]. To determine whether a molecular conformation is accepted or not, the root mean square deviation is calculated between current molecular coordinates and every molecule's previously recorded conformation. For example, PRO_LEADS makes use of a tabu search algorithm[35].

Simulation methods. Molecular dynamics is currently the most popular simulation approach. However, molecular dynamics simulations are often unable to cross high-energy barriers within feasible simulation time periods, and therefore might only accommodate ligands in local minima of the energy surface[11]. Therefore, an attempt is often made to simulate different parts of a protein–ligand system at different temperatures[37]. Another strategy for addressing the local minima problem is starting molecular dynamics calculations from different ligand positions. In contrast to molecular dynamics, energy minimization methods are rarely used as stand-alone search techniques, as only local energy minima can be reached, but often complement other search methods, including Monte Carlo[38]. DOCK performs a minimization step after each fragment addition, followed by a final minimization before scoring.

Protein flexibility. The treatment of protein flexibility is less advanced than that of ligand flexibility, but various approaches have been applied to flexibly model at least part of the target[39], including molecular dynamics and Monte Carlo calculations[31–33], rotamer libaries[40,41] and protein ensemble grids[42]. The idea behind using amino-acid side-chain rotamer libraries is to model protein conformational space on the basis of a limited number of experimentally observed and preferred side-chain conformations[40]. To reduce the number of discrete protein conformations arising from combinations of rotamers, a dead-end elimination algorithm is often used[41]. This algorithm recursively removes side-chain

POSE SPACE
All degrees of freedom involved in the process of placing one molecule relative to another. For example, for two rigid molecules the pose space simply consists of relative orientations. When one of the molecules, the ligand, is allowed to be flexible, the pose space comprises both the conformational space of the ligand and orientational space of ligand and receptor.

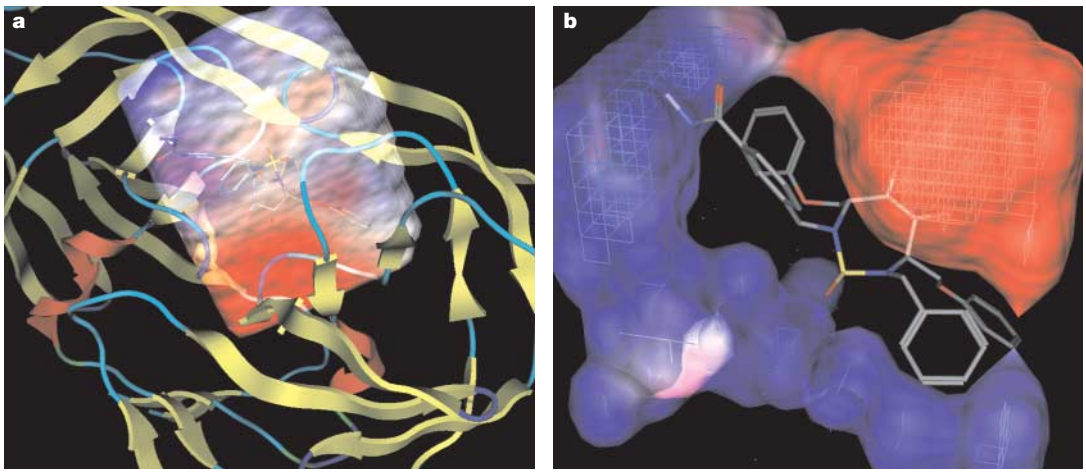


Figure 1 | **Grid representations. a** | Shown is a surface plot of a grid capturing the electrostatic potential of HIV protease (PDB code: 1BVE) around its active site (with bound inhibitor Dmp323). Red and blue indicate areas of negative and positive electrostatic potential, respectively. **b** | Shows a 'cut-away' electrostatic potential grid of the enzyme around the bound inhibitor (not included in the calculation).

conformations that do not contribute to a minimum-energy structure. Another method of treating protein flexibility is to use ensembles of protein conformations (rather than a single one) as the target for docking[42] and to map these ensembles on a grid representation. One approach generates an average potential energy grid of the ensemble, as first implemented in DOCK[42]; another maps various receptor potentials to each grid point and subsequently scores ligand conformations against each set of receptor potentials[19].

Scoring

The evaluation and ranking of predicted ligand conformations is a crucial aspect of structure-based virtual screening. Even when binding conformations are correctly predicted, the calculations ultimately do not succeed if they do not differentiate correct poses from incorrect ones, and if 'true' ligands cannot be identified. So, the design of reliable scoring functions and schemes is of fundamental importance. Free-energy simulation techniques have been developed for quantitative modelling of protein–ligand interactions and the prediction of binding affinity[43,44]. However, these expensive calculations remain impractical for the evaluation of large numbers of protein–ligand complexes and are not always accurate. Scoring functions implemented in docking programs make various assumptions and simplifications in the evaluation of modelled complexes and do not fully account for a number of physical phenomena that determine molecular recognition — for example, entropic effects. Essentially, three types or classes of scoring functions are currently applied: FORCE-FIELD-based, empirical and knowledge-based scoring functions. BOX 6 summarizes a number of currently used scoring functions, details of which will be discussed in the following section.

Force-field-based scoring. Molecular mechanics force fields usually quantify the sum of two energies, the receptor–ligand interaction energy and internal ligand energy (such as steric strain induced by binding). FIGURE 3 illustrates force-field modelling of non-bonded interactions involved in molecular recognition. Most force-field scoring functions only consider a single protein conformation, which makes it possible to omit the calculation of internal protein energy, which greatly simplifies scoring. Various force-field scoring functions are based on different force field parameter sets. For example, G-Score[26] is based on the Tripos force field[26] and AutoDock[45] on the AMBER force field[46]. However, functional forms are usually similar, as shown in supplementary information S1 (table).

Interactions between ligand and receptor are most often described by using van der Waals and electrostatic energy terms. The van der Waals energy term is given by a Lennard–Jones potential function (BOX 2). The parameters of the Lennard–Jones potential vary depending on the desired 'hardness' of the potential. Higher terms, such as a 12–6 Lennard–Jones potential of D-Score[26], result in increasingly repulsive potentials and will be less forgiving of close contacts between receptor and ligand atoms. Accordingly, lower terms, such as the 8–4 Lennard–Jones potential of G-score[26], make the potential softer. Electrostatic terms are accounted for by a Coulombic formulation with a distance-dependent dielectric function that lessens the contribution from charge–charge interactions. The functional form of the internal ligand energy is typically very similar to the protein–ligand interaction energy, and also includes van der Waals contributions and/or electrostatic terms.

Standard force-field scoring functions have major limitations, because they were originally formulated to model enthalpic gas-phase contributions to structure and energetics, and do not include solvation and entropic terms. Force-field-based scoring is further complicated by the fact that it generally requires the introduction of cut-off distances for the treatment of non-bonded interactions, which are more or less arbitrarily chosen

FORCE-FIELD
A function expressing the energy of a system as a sum of diverse molecular mechanics (or other) terms.

and complicate the accurate treatment of long-range effects involved in binding.

Recent extensions of force-field-based scoring functions include a TORSIONAL ENTROPY term for ligands in G-Score and the inclusion of explicit protein–ligand hydrogen-bonding terms in Gold[47] and AutoDock[45]. The latter terms are thought to increase the potential of specific molecular recognition. Hydrogen-bonding terms are often designed in noticeably different ways. For example, G-Score includes different hydrogen-bonding terms depending on the nature and geometry of the interaction. By contrast, AutoDock represents all of the hydrogen bonds by a 12–10 Lennard–Jones potential with a directional component, as shown in supplementary information S1 (table).

Empirical scoring functions. These scoring functions are fit to reproduce experimental data, such as binding energies and/or conformations, as a sum of several parameterized functions, as first proposed by Böhm[48]. The design of empirical scoring functions is based on the idea that binding energies can be approximated by a sum of individual uncorrelated terms. The coefficients of the various terms are obtained from REGRESSION ANALYSIS using experimentally determined binding energies and, potentially, X-ray structural information. Representative examples of empirical scoring functions are given in supplementary information S2 (table). The functional forms are often simpler than force-field scoring functions, although many of the individual contributing terms have counterparts in the force-field molecular mechanics terms. The appeal of empirical functions is that their terms are often simple to evaluate, but they are based on approximations similar to force-field functions. A disadvantage of these methods is their dependence on the molecular data sets used to perform regression analyses and fitting. This often yields different weighting factors for the various terms. As a consequence, terms from differently fitted scoring functions cannot easily be recombined into a new scoring function.

In empirical scoring functions, terms accounting for non-bonded interactions can be implemented in rather different ways. For example, in the early LUDI formulation[48], the hydrogen-bonding term is separated into neutral hydrogen bonds and ionic hydrogen bonds, whereas ChemScore[49] does not differentiate between different types of hydrogen bonds. Furthermore, the LUDI function calculates hydrophobic contributions on the basis of a representation of molecular surface area, whereas ChemScore evaluates contacts between hydrophobic atom pairs. F-Score adds an additional term to account for aromatic interactions[50].

Empirical scoring functions can include non-enthalpic contributions such as the so-called rotor term, which approximates entropy penalties on binding from a weighted sum of the number of rotatable bonds in ligands. ChemScore implements ligand rotational entropy in a more complicated form that describes the molecular environment surrounding each rotatable bond. More complex functions begin to address solvation and desolvation effects. For example, the Fresno scoring function[51], which is used for peptide docking, explicitly takes ligand desolvation into account, and desolvation energies are calculated using a continuum electrostatic model[52,53]. However, terms currently used to approximate entropy or desolvation energy provide only incomplete descriptions of these effects on protein–ligand binding.

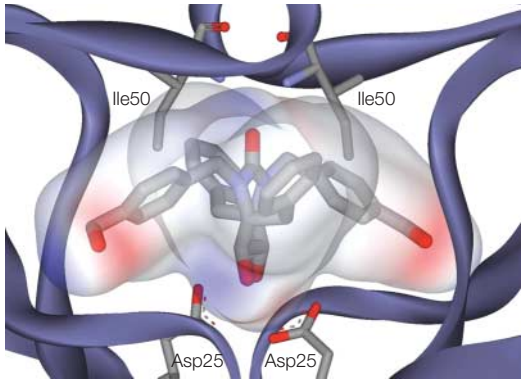


Figure 2 | **Electrostatic potential of a bound inhibitor.** Inhibitor Dmp323 is shown in complex with HIV protease (PDB code: 1BVE). The electrostatic potential of the symmetrical inhibitor in its binding conformation was mapped on its calculated molecular surface. Residues Ile50 and Asp25 from each monomer in HIV protease stabilize inhibitor binding.

Knowledge-based scoring functions. In essence, knowledge-based scoring functions are designed to reproduce experimental structures rather than binding energies. In knowledge-based functions, as shown in supplementary information S3 (table), protein–ligand complexes are modelled using relatively simple atomic interaction-pair potentials. A number of atom-type interactions are defined depending on their molecular environment. So, in common with empirical methods, knowledge-based scoring functions attempt to implicitly capture binding effects that are difficult to model explicitly. Popular implementations of such functions include POTENTIAL OF MEAN FORCE (PMF)[54–56] and DrugScore[57], which also includes solvent-accessibility corrections to pair-wise potentials. SMoG[58] is another scoring function belonging to this class that utilizes pair-wise atom potentials to evaluate protein–ligand interactions. A major attraction of many knowledge-based scoring functions is their computational simplicity, which permits efficient screening of large compound databases. A disadvantage is that their derivation is essentially based on information implicitly encoded in limited sets of protein–ligand complex structures.

Consensus scoring. Given the imperfections of current scoring functions, a recent trend in this field has been the introduction of consensus scoring schemes[59]. Consensus scoring combines information from different scores to balance errors in single scores and improve the probability of identifying 'true' ligands. An exemplary implementation of consensus scoring is X-CSCORE[60], which combines GOLD-like, DOCK-like, ChemScore, PMF and FlexX scoring functions. However, the potential

TORSIONAL ENTROPY
Entropy associated with a rotatable bond in a molecule. Immobilization of a rotatable bond on binding leads to loss of its torsional (or rotational) entropy.

REGRESSION ANALYSIS
Determination of parameter values for a chosen (linear or nonlinear) function to best fit a set of observations.

POTENTIAL OF MEAN FORCE (PMF). In the context of docking and scoring, PMFs are derived from statistical analysis of experimentally observed distributions and frequencies of specific atom-pair interactions in a large collection of protein–ligand structures. Interaction potentials between each atom pair in two molecules (for example, ligand and protein) approximate the free energy of each pair-wise interaction as a function of inter-atomic distance.

Box 3 | **Flexible ligand-search methods**

Random/stochastic

- AutoDock (MC)[30]
- MOE-Dock (MC, TS)[133]
- GOLD (GA)[64]
- PRO_LEADS (TS)[35]

Systematic

- DOCK (incremental)[24]
- FlexX (incremental)[50]
- Glide (incremental)[134]
- Hammerhead (incremental)[28]
- FLOG (database)[135]

Simulation

- DOCK
- Glide
- MOE-Dock
- AutoDock
- Hammerhead

value of consensus scoring might be limited, if terms in different scoring functions are significantly correlated, which could amplify calculation errors, rather than balance them.

Evaluating scoring schemes. Perez *et al.*[61] compared force-field scoring with a combined PMF knowledge-based function. In this study, force-field terms generally performed better than PMF, but steric contributions significantly outweighed electrostatic terms. Correct ligand poses were found in nearly 80% of the cases studied when the force-field function was used, but the success rate dropped to 56% when cross-docking experiments were performed. In an effort to better understand incorrectly calculated effects, LINEAR DISCRIMINANT ANALYSIS was applied; the results indicated that better adjusted molecular volume and hydrogen bonding terms were likely to further improve force-field scoring. In the PMF function, volume and attractive dispersion effects were not considered.

Good *et al.*[62] compared several docking protocols. When PHARMACOPHORE constraints and conformational flexibility were taken into account, simple contact scores outperformed force-field treatment (even when overestimated electrostatic effects were ignored). More complex functions implemented in Prometheus[63] and GOLD[64] performed better in well-defined active sites, but simpler schemes such as the DOCK contact energy performed best in more complex and less well-understood binding sites.

Sotriffer *et al.*[65] compared DrugScore (knowledge-based) with AutoDock (force-field-based) and found that these two scoring methods had very similar abilities in predicting the correct binding modes for 158 complexes. For rigid molecules, 90–100% of cases were predicted correctly. However, as the number of rotatable bonds (and molecular flexibility) increased, the success rate dropped to 44–80%, and DrugScore only marginally improved the results over AutoDock. These findings also illustrated the difficulty in finding correct poses for highly flexible ligands.

Charifson *et al.*[59] showed that most conventional scoring functions were able to place ~50% of active (<100 nM) compounds within the top 1,000 of the score lists but that less active (~1 μM) compounds were much more difficult to identify. Chemscore, the DOCK energy and PLP scores performed well for all three binding sites that were analysed. In this study, a consensus scoring scheme was applied that combined results from each scoring function and was found to further improve prediction accuracy. In a similar study, Wang *et al.*[66] analysed more than 100 protein–ligand complexes. The best scoring conformations from PLP, F-Score, LigScore and DrugScore were within 2 Å in 70–80% of the test cases. In this study, the quality of scoring schemes was found to be more or less independent of the type of active site. However, scoring functions that relied purely on force-field or knowledge-based potentials consistently performed worse in identifying correct poses than functions that used additional terms accounting for hydrophobic effects.

The FRED program uses a Gaussian function for docking[67] to generate a smooth and easily searchable energy surface, and allows a wide latitude for errors in positioning protein atoms. Pre-generation of multiple ligand conformers is a suitable technique when this 'soft' and error-tolerant function is used, but the method has limited accuracy in ranking ligands correctly (and therefore requires more detailed follow-up scoring). In a comparison of FRED and Glide scoring schemes[68], cyclooxygenase-2, oestrogen receptor, mitogen-activated kinase, gyrase B, thrombin, gelatinase-A and neuraminidase binding sites were used as test cases. As expected, 'hard' functions such as the one in Glide performed overall better than the softer scoring function in FRED. However, FRED was found to produce accurate results for lipophilic binding sites, especially when hydrophobic effects outweighed electrostatic and hydrogen-bonding interactions. Scoring functions therefore respond differently to specific features in binding sites.

Posing versus scoring. Are calculation errors more associated with predicting binding conformations or with scoring them? Ligand flexibility has a greater effect on predicting structures correctly than size or polarity[69], which clearly relates to posing. However, the ability to discriminate docked conformations of 'fantasy' ligands from 'true' ones depends crucially on scoring. Clearly, it is often difficult to distinguish between inadequate conformational searching and flawed scoring, and relatively few studies have been designed to dissect these effects. In a parallel application of various programs[70], DOCK, FlexX and GOLD displayed a clear tendency for ligands to score better when using X-ray conformations than any incorrectly modelled conformation, whereas CDOCKER often scored correct structures worse, thereby indicating shortcomings of its scoring function. In terms of structural accuracy, GOLD and Glide produced overall satisfactory results for a set of 69 complexes. In a

LINEAR DISCRIMINANT ANALYSIS
Mathematical analysis based on two classes of data and two independent variables (a, b) that attempts to find a line that best separates the data. This line is orthogonal to the discriminant function that is a linear combination of the original variables, in this case: $F = c_a a + c_b b$ (c_a, c_b; coefficients).

PHARMACOPHORE
The spatial arrangement of atoms or groups in a molecule known or predicted to be responsible for specific biological activity.

Box 4 | **The problem of combinatorial explosion**

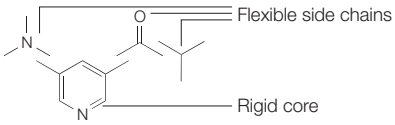
For systematic conformational search, the number of possible molecular conformations is represented by equation 1:

$$N_{Conformations} = \prod_{i=1}^{N} \prod_{j=1}^{n_{inc}} \frac{360}{\theta_{i,j}} \quad (1)$$

In equation 1, N is the number of rotatable bonds and $\theta_{i,j}$ is the size of the incremental rotational angle j for bond i. To avoid exhaustive search calculations, many conformational search algorithms use an incremental construction approach to grow a ligand within an active site that consists of three basic steps:

- Core fragment selection.
- Core fragment placement.
- Incremental ligand construction.

During the first step, the ligand is divided into a rigid core fragment and flexible side chains. Subsequently, these side chains are further divided at each new rotatable bond, as shown in the figure.



During the second and third steps, core fragments are placed and side chains are incrementally attached with rotational degrees of freedom sampled. As discussed in the text, popular algorithms often differ in the way that these processes are carried out.

recent re-parameterization of GOLD, two scoring schemes, GOLDSCORE and ChemScore, were compared and applied in a consensus scoring appproach[47]. Both functions yielded similar scoring accuracy (65–85%). Accurate prediction of relative binding affinities depended on finding correct binding conformations. As an alternative to consensus scoring, genetic-algorithm-based scoring terms were introduced to better distinguish correct ligands from 'noise' compounds[71]. On the basis of these investigations, the conclusion can be drawn that accurate modelling of binding conformations is necessary but insufficient for correct ligand scoring and ranking. Given the success rates of various efforts at structure prediction, as discussed above, it seems that imperfections in scoring functions continue to be a major limiting factor.

Improving scoring functions. How can one further improve the quality of scoring functions? As indicated earlier, a current trend in the field is to focus on the inclusion of various solvation[52,72,73] and rotational entropy[69] contributions. Scoring functions accounting for such contributions are more accurate than, for example, standard force-field functions, but are also computationally expensive, which challenges high-throughput docking. More importantly, however, the often-made observation that alternative scoring functions perform rather differently on multiple targets — for example, the GOLD validation set[74] — suggests that it might be difficult to develop scoring functions that perform equally well across many different protein families, regardless of their complexity and sophistication. Binding sites and recognition processes have unique features that ultimately render protein–ligand interactions specific (FIG. 4), which in turn might often require 'tuning' of scoring schemes on a case-by-case basis. A good example of re-evaluating scoring schemes and adjusting to differences between active sites is provided by a comparison of knowledge-based scoring of various structural classes[75]. When plotting docking scores versus the logarithm of binding constants, it was observed that slopes and intercepts significantly differed for structurally distinct active sites.

Structure-based virtual screening

General caveats. Regardless of the functions applied, scores are known to scale poorly with molecular mass and the number of rotatable bonds in compounds[76]. Large molecules can form many hypothetical interactions in binding sites and therefore have the tendency to generate better scores than smaller compounds. On the other hand, the entropy penalty for immobilization of rotatable bonds, which is frequently not taken into account, scales with the number of such bonds. As a result, if entropy penalties are included, flexible molecules tend to score lower than more rigid ones. Furthermore, the internal strain energy of a molecular pose is generally approximated using a single unbound conformation of the ligand as a reference, which has significant limitations in estimating entropy and enthalpy losses on binding. These limitations generally add to imperfections of scoring functions and make it more difficult to accurately rank test molecules on the basis of computed binding-site interactions.

The general nature and preparation of active sites also affects the quality of ligand positions and scores. Hydrophobic binding sites such as are found in, for example, HIV protease are likely to be more promising targets than sites that are more hydrophilic, or binding events involving distinct electrostatic interactions as seen, for example, in metallo-enzymes. This is mainly due to the fact that binding to hydrophobic sites can be well approximated by a calculation of shape complementarity between ligand and receptor, for which robust methodologies have existed since the early days of docking[8,77]. The calculation of shape complementarity implicitly takes hydrophobic effects into account; however, a large (and sometimes the largest) contribution to the hydrophobic effect comes from desolvation of hydrophobic ligands (such as in HIV protease), which is not adequately accounted for in docking scores and can be significantly underestimated relative to other scoring terms in some active sites. Furthermore, precise modelling and scoring of electrostatic interactions continues to be a major challenge for contemporary scoring functions. As mentioned above, simple Coulombic models are still applied for these purposes in a number of cases and have the tendency to grossly overestimate charge–charge interactions or create artificial ones.

In addition, the placement of water molecules that are either structurally important or directly involved in binding interactions, and assumed rigidity of side-chain

Box 5 | **Search techniques**

Monte Carlo algorithm in its basic form:

- Generate an initial configuration of a ligand in an active site consisting of a random conformation, translation and rotation.
- Score the initial configuration.
- Generate a new configuration and score it.
- Use a Metropolis criterion (see below) to determine whether the new configuration is retained.
- Repeat previous steps until the desired number of configurations is obtained.

Metropolis criterion

If a new solution scores better than the previous one, it is immediately accepted. If the configuration is not a new minimum, a Boltzmann-based probability function is applied. If the solution passes the probability function test, it is accepted; if not, the configuration is rejected.

Molecular dynamics

Molecular dynamics is a simulation technique that solves Newton's equation of motion for an atomic system: $F_i = m_i a_i$, in which F is force, m is mass and a is acceleration. The force on each atom is calculated from a change in potential energy (usually based on molecular mechanics terms) between current and new positions: $F_i = -(dE/r_i)$, in which r is distance. Atomic forces and masses are then used to determine atomic positions over series of very small time steps: $F_i = m_i (d^2 r_i/dt^2)$, in which t is time. This provides a trajectory of changes in atomic positions over time. Practically, it is easier to determine time-dependent atomic positions by first calculating accelerations a_i from forces and masses, then velocities v_i from $a_i = dv_i/dt$ and, ultimately, positions from velocities $v_i = dr_i/dt$.

Genetic algorithms

Genetic algorithms are a class of computational problem-solving approaches that adapt the principles of biological competition and population dynamics. Model parameters are encoded in a 'chromosome' and stochastically varied. Chromosomes yield possible solutions to a given problem and are evaluated by a fitness function. The chromosomes that correspond to the best intermediate solutions are subjected to crossover and mutation operations analogous to gene recombination and mutation to produce the next generation. For docking applications, the genetic algorithm solution is an ensemble of possible ligand conformations.

Tabu search algorithm in its basic form:

- Make n small random changes to the current conformation.
- Rank each change according to the value of the chosen fitness function.
- Determine which changes are 'tabu' (that is, previously rejected conformations).
- If the best modification has a lower value than any other accepted so far, accept it, even if it is in the 'tabu'; otherwise, accept the best 'non-tabu' change.
- Add the accepted change to the 'tabu' list and record its score.
- Go to the first step.

conformations within a binding site, can dramatically influence posing of test compounds[78]. Clearly, whenever conformational changes are involved in binding, rigidly defined binding sites are limited in their predictive potential. Finally, it has been observed that structure-based virtual screening often selects compounds that are biologically promiscuous and are therefore termed 'frequent hitters'[79,80]. The fairly unspecific inhibition by such compounds can, at least in part, be attributed to dominating hydrophobic character and aggregation effects that tend to favour their detection in both docking simulations and screening assays (albeit for different reasons).

Selection strategies. Considering the many approximations and limitations involved in system set-up, posing and scoring, one might ask the question of why structure-based virtual screening actually 'works'. The principal reasons are that computational screening is an enrichment process; that accurately calculated energies and scores are not necessarily required for meaningful compound selection; and that appropriate selection strategies compensate for some methodological shortcomings. For example, in a typical docking study, a large compound database will probably be reduced to a shortlist of preferred candidates, perhaps ~100 or so. To enrich this selection with compounds that have a high probability of being active, de-selection of inappropriate compounds (which most of the database compounds are) is as important as finding the most promising candidates. Importantly, de-selection of inappropriate compounds is more easily achieved than selection within the accuracy limitations of the calculations. Also, some binding events, such as those dominated by shape complementarity, can be treated well given the approximations of posing and scoring. Furthermore, as long as active compounds are found in the shortlist, their relative ranking becomes less important. Simply put, an active compound within the top-five scoring compounds will be as good as one within the top 50, as long as these compounds are tested, which further compensates for limitations of scoring. In addition, it is also a rather common practice to subject a reasonably small number of pre-selected candidates (for example, 100–500) to visual inspection, which adds another dimension to the selection process (that is, chemical intuition, knowledge and experience)[81]. So, although virtual screening inevitably produces false-positives and -negatives, rationalizing the analysis as an enrichment process helps to explain its successes.

Even very fast docking and scoring methods typically require several to tens of seconds per compound for a fully flexible search and therefore become prohibitive in the presence of millions of database compounds. As a result, complex posing and scoring schemes are often carried out only after the source database has been significantly reduced in size by the application of de-selection or filtering methods.

Structures of target sites. The choice and preparation of the structural model of a targeted binding site are important variables. Experimentally determined (X-ray or nuclear magnetic resonance) structures are generally preferred. However, as the number of proteins of pharmaceutical interest has grown faster than the number whose structures have been determined, homology modelling has risen in popularity. A recent study compared the quality of docking results when either crystal structures of HOLO- or APO-ENZYMES or homology models were used as templates[82]. Perhaps surprisingly, homology models yielded enrichments factors of ten or better in eight of ten test cases studied, and apo-enzymes and homology modelled structures performed comparably well. However, by far the best performance was observed when ligand-bound protein conformations were used as starting points. The study demonstrated that even subtle protein conformational changes that result from ligand binding were sufficient to significantly influence the quality of docking results. Nevertheless, homology

HOLO-, APO-ENZYME
Holo-: ligand-bound form of an enzyme; apo-: uncomplexed form. The original definitions referred to enzymes and cofactors, rather than ligands, but ligands and cofactors are often synonymously used.

Box 6 | **Types of scoring functions**

Force-field-based

- D-Score[26]
- G-Score[26]
- GOLD[47]
- AutoDock[45]
- DOCK[24]

Empirical

- LUDI[92,93]
- F-Score[50]
- ChemScore[49]
- SCORE[131,132]
- Fresno[51]
- X-SCORE[60]

Knowledge-based

- PMF[54–56]
- DrugScore[57]
- SMoG[58]

models built in the presence of high sequence similarity provided reasonable docking templates. A similar study was undertaken to examine the value of homology modelling within the protein kinase family[76]. Here, the quality of calculated poses was found to correlate well with ligand enrichment factors. Again, crystal structures of ligand-bound target sites provided the best results when used as templates. The importance of protein conformation was recently also demonstrated by cross-docking studies on trypsin, thrombin and HIV-1 protease[69]. Error rates in docking of ligands to apo-binding sites correlated with the magnitude of protein structural change observed as a result of binding.

Pre-screening: three-dimensional filtering. In addition to conventional one/two-dimensional filters such as the rule-of-five[83], three-dimensional filter functions have been implemented to efficiently pre-screen very large databases and reduce the final number of docking and scoring steps. For example, shape similarity methods can be applied for filtering. The heuristic is based on identifying similar molecular shapes on the basis of signatures, triplets, quartets or higher-order groups of atoms[84–86]. However, these shape filters are usually limited to pre-screening of databases that contain single molecular conformations, which can be a source of false-negatives. In addition, pharmacophore-based screening can be carried out where pre-defined chemical and geometric features in compounds are matched[87]. Signatures and bit strings derived from triplets of distances[84] and surface triplets and histograms[85] have been be used to identify preferred candidates. Recently, a ray-tracing-based approach[86] has been applied to calculate shape signatures of molecules for database searching. These types of descriptors are also highly conformation-dependent and therefore limited in their predictive value when only a single (hypothetical) molecular conformation is used.

Hit identification. The ultimate measure of success for the methods discussed above is their ability to produce significant hit rates while reducing the number of compounds that need to be tested. A number of successful case studies have been published over the years that demonstrate the potential of the approach (one also needs to take into account that many successful applications in pharmaceutical settings are unlikely to be disclosed). TABLE 1 summarizes recent case studies on a wide variety of targets that have produced some impressive results. It is interesting to note that nearly all groups performed pre-filtering and used two-dimensional similarity methods and shape or drug-like filters to reduce the number of database compounds for the time-consuming steps of flexible docking, elaborate scoring and visual analysis. Hits with at least low-micromolar potency were usually found, often without biasing search calculations towards previously identified hits. The results also mirror the general trend that hits in the micromolar range are much more frequently identified than nanomolar hits in these calculations (this is similar to the situation in biological screening). Major reasons for this are that newly identified active compounds are rarely optimized for potency against a given target and that nanomolar potency is typically only obtained after chemical modification in the course of hit-to-lead transition and lead optimization.

Structure-based lead optimization

In addition to hit identification, docking techniques are increasingly used to support lead optimization efforts. Here, the scenario changes: to facilitate a hit-to-lead transition, the compound potency typically has to be increased by two to three orders of magnitude and relatively small chemical modifications can lead to significant changes in binding. The requirement to estimate the effects of relatively small chemical changes further complicates the calculations and therefore distinguishing a micromolar compound from a nanomolar analogue often requires much greater accuracy than typical docking and scoring can provide. However, once hits or leads have been co-crystallized with their targets and exact binding conformations have been established, docking of analogues can be facilitated by the application of algorithms such as 'anchored search'[24] that model compound modifications on pre-defined core fragments of leads. These 'conservatively' predicted complexes usually involve only a limited number of analogues, and so alternative and consensus scoring schemes can be easily explored. Typical structure-based analogue design is illustrated in FIG. 5. At the very least, automated analogue design and evaluation makes it possible to quickly eliminate molecules that are too large or do not satisfy binding constraints, and shifts focus towards more promising synthetic candidates. For example, a series of caspase-3 inhibitors was optimized starting from a co-crystal structure with salicylic acid[88]. Modelling of analogues resulted in a compound with 20-nM potency whose modelled structure was experimentally confirmed. Going beyond a one-by-one evaluation of analogues, combining docking and design of analogue libraries

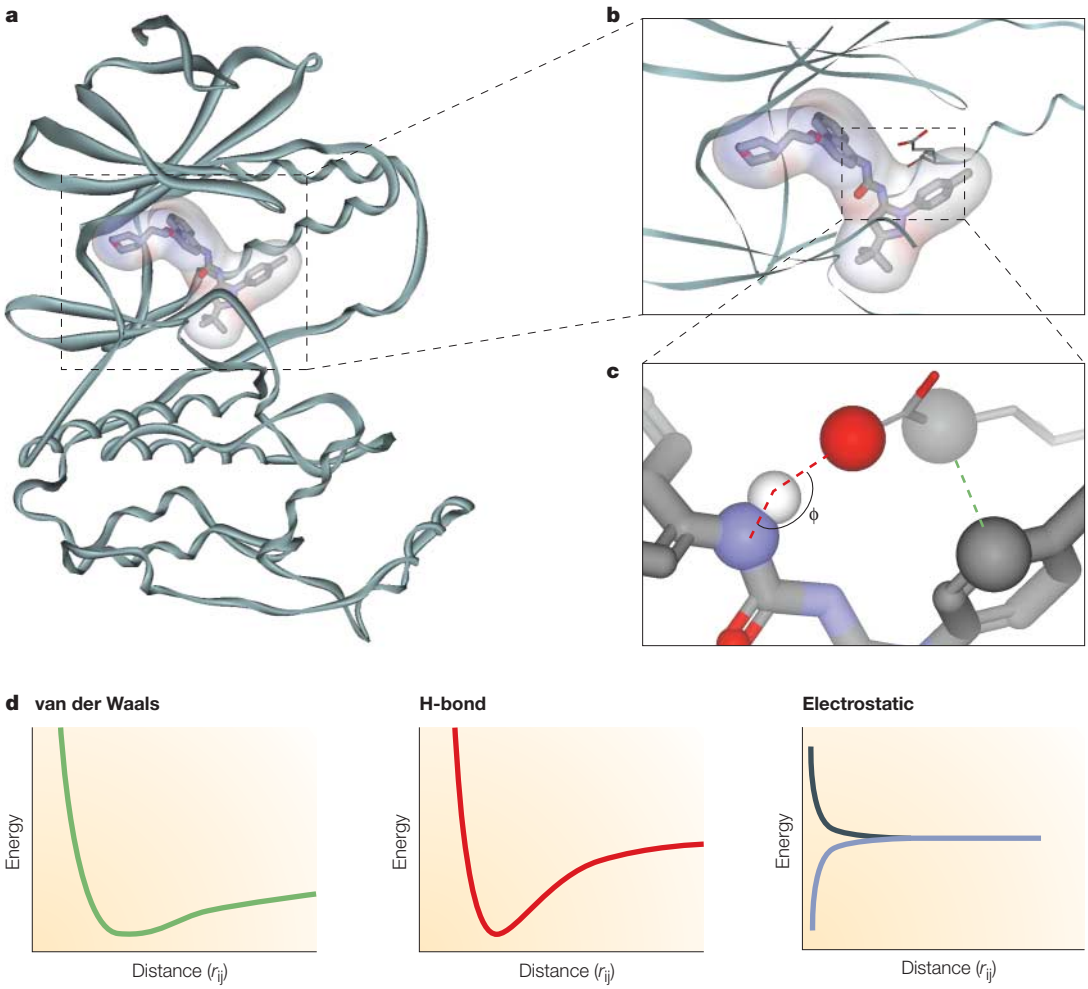


Figure 3 | **Modelling molecular recognition. a** | Structure of p38 mitogen-activated protein kinase with bound inhibitor BIRB796 (PDB code: 1KV2). The inhibitor is shown with its electrostatic potential surface. **b** | Enlarged view of the active site. **c** | Close-up view of the interaction between residue Glu71 and BIRB796. Hydrogen bonding (H-bond) and van der Waals interactions are colour-coded red and green, respectively. **d** | Schematic representation of functions used to model pair-wise interactions that contribute to binding. Interactions are calculated as a function of the distance (r_{ij}) between two atoms *i* and *j*. Left of part **d**: van der Waals interaction given by a 12–6 Lennard–Jones potential (note the smoother attractive part of the potential compared to hydrogen bond term). Middle of part **d**: hydrogen-bond potential given by a 'harder' 12–10 Lennard–Jones potential (see also BOX 2). This term is angle-dependent (as indicated in **c**). Right of part **d**: electrostatic potential for two like (blue) or opposite (black) charges of same magnitude calculated using a distance-dependent dielectric constant of *4r*.

provides a particularly promising route to lead optimization. This has been well demonstrated by a study on cathepsin D that produced low-nM inhibitors by iterative anchored docking calculations and targeted library design[89]. This approach is attractive because docking and scoring are no longer challenged to predict a 'home run' modification to a lead, but rather to prioritize sites and groups for experimental modification (which is well within the accuracy limits of the calculations).

De novo *design.* An early method for the *de novo* design of compounds in active sites is multiple-copy simultaneous search (MCSS)[90,91]. Many small fragments are docked and simultaneously minimized within an active site. After scoring and sorting, preferred fragments are combined into larger molecules. Much like LUDI[92,93], results of MCSS can provide a map of likely sub-sites for binding of selected functional groups. Other algorithms that grow and score compounds within binding sites are implemented in programs such as Groupbuild[94], GenStar[95], Grow[96] and Growmol[97]. A known limitation of such approaches is the difficulty of computationally estimating the synthetic accessibility of 'designer' molecules. However, the SYNOPSIS program represents a recent effort to couple *de novo* and synthetic design[98]. For example, a mini-library containing only 28 compounds was selected from 373 million possible candidate molecules in a study targeting HIV reverse transcriptase (HIV-RT). This was achieved by use of a genetic algorithm to simultaneously evaluate conformational and synthetic parameters embedded in a fitness function. Of the 28 selected compounds, 18 could be synthesized and of these molecules, 10 were found to be active below 100 μM. In another study, carbonic anhydrase inhibitors

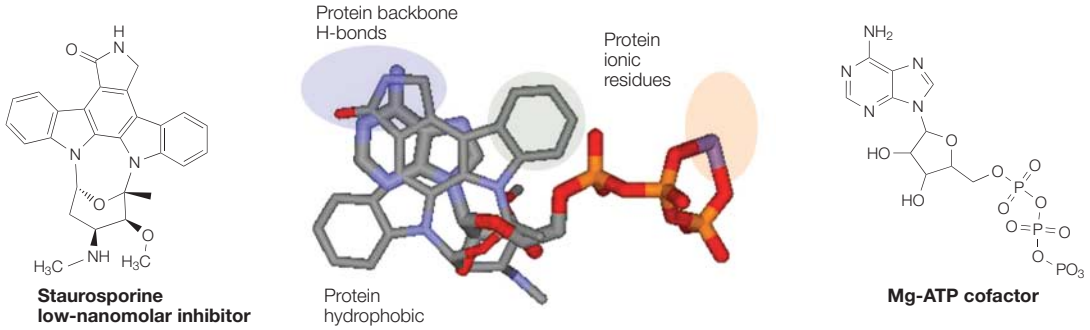


Figure 4 | **Complexity of protein–ligand interactions.** The figure shows a schematic illustration of various interaction components that need to be considered to predict the structure and binding energetics of two compounds within the same active site. In this case, the natural cofactor of cyclic-AMP-dependent kinase (PDB code: 1atp), Mg ATP, is compared with the ATP-binding site-directed inhibitor staurosporine. To correctly predict staurosporine as an inhibitor in a docking study, relative weights in energy functions for the treatment of hydrophobic (indoles), hydrogen bonding (lactam ring) and ionic (aliphatic amine salt bridge, Mg-phosphate + protein chelation) interactions must be appropriately adjusted to balance their effects. Finding preferred scoring conditions for a specific target is a non-trivial process and often involves many trial-and-error runs.

were constructed using a Monte Carlo combinatorial growth algorithm and a knowledge-based scoring scheme[75]. From ~100,000 theoretical candidates, only two compounds were selected for synthesis but both exhibited sub-nM potency.

Simulations. Free-energy simulations are applicable to evaluate limited numbers of analogues; for example, a series of thrombin inhibitors[99]. Various approximations have been proposed for reducing the complexity of perturbation calculations for these purposes. For example, the OwFeg method performs a free-energy simulation over bound and unbound states of ligands but maps energy changes to a grid[100], which greatly simplifies calculations for transforming one functional group into another. Grid points that are energetically relevant for various chemical modifications can be monitored during analogue design. Moreover, linear response approximations that utilize ligand-interaction energies with the protein and solvent environment are now more commonly applied in lead optimization[101]. These methods require the availability of at least a few experimental data points across the range of activities considered to be significant. A quantitative structure–activity relationship (QSAR) is applied to combine non-bonded interactions that occur within the simulated system. Linear response methods have been shown to provide some promising results in analogue design studies on β-secretase[102,103], HIV-RT[104–107], factor Xa[108] and the oestrogen receptor[109].

Molecular mechanics Poisson–Boltzmann surface area (MM/PBSA)[110] calculations are another molecular-dynamics-based simulation technique involving both force-field and solvation terms that are important for binding. Solvation effects are estimated using a continuum Poisson–Boltzmann model[52]. A major difference between MM/PBSA and linear response methods is the treatment of the ligand in its unbound state: MM/PBSA uses NORMAL MODE analysis to calculate enthalpic and entropic contributions to the ligand free energy. The methodology was recently applied in analyses of neuraminidase[111] and cathepsin D[53] inhibitors.

NORMAL MODE
An oscillation in which all particles of a system move with the same frequency and phase.

Active-site analysis. Graphical computational analysis of binding sites has greatly contributed to structure-based drug design since its early days. Docking and simulation techniques have also been applied to analyse features of the active site, including various hydrophobic and hydrophilic molecular fields that can identify promising areas for ligand docking and/or *de novo* design[112]. Surface maps and molecular fields are mostly stored on grids that are used to semi-quantitatively compare active sites in homologous enzymes to explore differences in specificity[112]. The evaluation of potential interactions in active sites can complement docking analyses. Another recent approach generates structural interaction fingerprints (SIFts) that allow pre-screening for potential ligands in databases prior to docking[113]. In an exploration of the active site of trypanothione reductase, 44 diverse inhibitors were initially docked and the resulting conformations were sampled and used to train a scoring function for this enzyme[114]. Then 2,500 novel compounds were docked into the active site and evaluated using this scoring scheme; 13 compounds were selected for testing and 9 were found to be active.

Active-site analysis techniques can also be applied in combination with quantitative methods. For example, in a study of factor Xa, QSAR models were established for a series of amidino inhibitors[115]. A total of 120 analogues were then docked into active sites using four crystal structures and the resulting molecular alignments were used to calculate molecular fields. The resulting fields guided the design of analogue libraries. In a similar vein, 133 known factor Xa inhibitors were docked, scored and subjected to regression analysis[116]. Scoring terms were fitted to experimental binding energies to develop a factor Xa-specific scoring scheme. Applying this scheme, 80% of known inhibitors could be retrieved from a compound library with only a 15% false-positive rate[116].

Absorption, distribution, metabolism and excretion properties. Docking techniques are currently also applied to aid in structure-based absorption, distribution, metabolism and excretion (ADME) evaluation. Cytochrome

Table 1 | **Comparisons of docking/scoring methods and virtual screening results**

Method	Number of structures	Proteins and families
Selected docking studies focusing on structural accuracy		
Comparison of 13 docking scores[61]	34 PDB structures; 100–300 compounds per activity range	Kinases, dehydrogenases, HIV protease
Comparison of 11 scoring functions[66]	100 protein–ligand complexes	
AutoDock, DrugScore[65]	41 protein–ligand complexes	
LibDock, PLP2[76]	1,000 kinase inhibitors	Kinases
Examples of successful structure-based virtual screens		
DOCK[136]	Crystal structure; 200,000 compounds; 1,000 top-scoring clustered; rule-of-five; 15 compounds tested; 3 hits (0.4 μM)	BCR-ABL tyrosine kinase
DOCK[137]	Homology model; 250,000 NCI compounds; rigid docking (50,000 orientations per ligand); 444 ligands flexibly docked; 13 tested; 3 hits (<100 μM)	Thymidine phosphorylase
FlexX[138]	800,000 ACD compounds reduced to 856; 9 hits; (0.25–256 μM)	tRNA-guanine transglycosylase
DOCK[81]	1,700-fold enrichment over actual HTS; 1.7 μM inhibitor through VS	Phosphatase-1B
FlexX[139]	ACD search, 1 hit (43 μM)	IGF1/IGF-BP-5
DOCK[140]	Homology model; 400,000 compounds; 12 test; one hit (80 nM)	Protein kinase CK2
Catalyst/DOCK[87]	4,000 ACD compounds; pharmacophore search; 24 tested; 12 in μM range across 6 mutants	*Plasmodium falciparum* DHFR
DOCK[141]	Homology model; 200,000 NCI compounds; 35 tested; 7 hits (1.6–14 μM)	BCL2/BCL-XL

ACD, Available Chemicals Directory; BCL, B-cell lymphoma protein; DHFR, dihydrofolate reductase; HTS, high-throughput screening; IGF, insulin-like growth factor; NCI, National Cancer Institute; VS, virtual screening.

P450 isoforms are major drug-metabolizing enzymes and have become focal points in the study of rapid metabolism and drug–drug interactions[117,118]. Several groups have therefore developed structure-based approaches for the prediction of compounds that would be metabolized by or inhibit P450s, and various homology models of human P450 isoforms have been generated for these purposes as templates for docking to predict drug metabolism[9,119–122]. Recently, a crystal structure was determined of a human P450 isoform in complex with warfarin[10]. The inhibitor binds proximally to the iron–porphyrin system in the enzyme but had no direct interaction with the cofactor. These structural insights should help to further refine docking studies on human P450s and increase their predictive value.

Perspectives

Many of the examples and applications discussed in this review indicate that the scoring and reliable ranking of test compounds continue to be major bottlenecks in structure-based virtual screening and lead optimization. Despite a plethora of already available scoring functions, further progress will be required to better account for and balance entropic effects and electrostatic interactions. Many current limitations are the result of the assumption that implemented solvation or entropic and electrostatic terms are generally applicable and transferable to different protein systems. However, structure-based screening calculations have produced impressive results and many novel hits. These successes are at least in part due to the fact that virtual-screening campaigns mostly aim at the enrichment of active compounds, rather than, for example, accurate calculation of binding energies. For efficient compound selection, relying solely on computed scores is currently not sufficient; experience and intuition are often still a key to success. Taking this into account, further progress can be made in establishing more advanced scoring schemes, even if it is not possible to develop conceptually novel scoring functions in the near future. Importantly, scoring schemes can be advanced by modifying molecular systems used for benchmarking, calibrating selected functions for specific applications, or determining the most relevant scoring ranges. Scientific foundations for such efforts have already been laid, as described in the following section.

The statistical analysis of score distributions resulting from docking of large compound databases into different target sites has enabled scoring ranges to be determined that are most likely to reflect 'nonspecific' binding events[123]. Similarly, docking of compound collections into arbitrarily selected (or random) targets can provide information about background or 'noise' scoring levels, regardless of the scoring functions that are applied. This type of strategy has its roots in earlier investigations designed to determine similarity measures for ligands on the basis of docking against panels of at least partly irrelevant receptor sites[124]. Compound ranking has also been improved by the classification of databases into groups of similar molecules prior to docking and final selection of only the best scoring representative of each group[125]. Another knowledge-based approach is the use of three-dimensional similarity information from co-crystallized ligands as an additional constraint or scoring term[126].

Scoring schemes can also be improved by tailoring them to a specific target site, to designed sites or to multiple related receptors. For example, altered binding

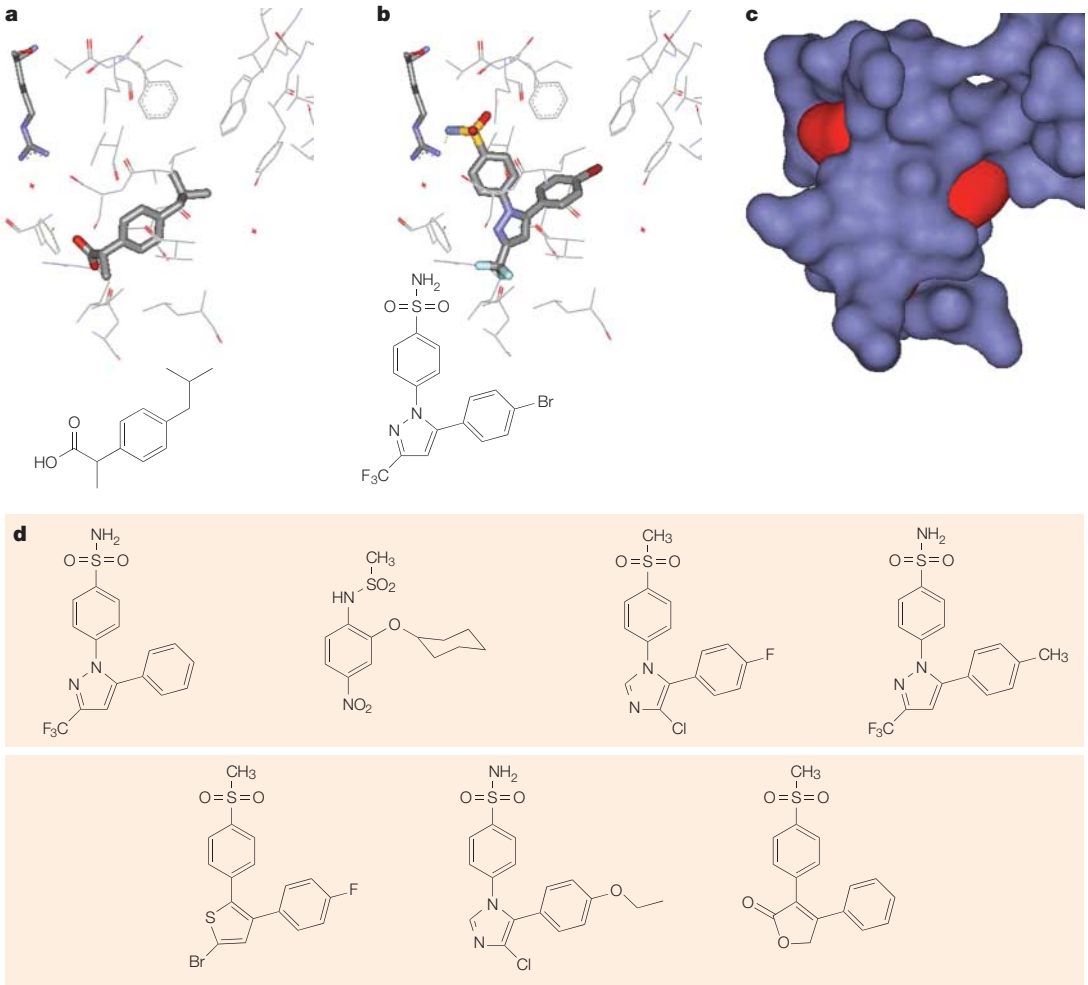


Figure 5 | **Design of specific inhibitors.** The active site of cyclooxygenase-2 (COX2) (PDB code: 1cx2) is shown in complex with ibuprofen, a non-selective COX inhibitor (**a**), and a selective COX2 inhibitor (**b**); **c** shows a space-filling representation of the active site. **d** | Several other potent COX2 inhibitors are shown. These COX2 lead compounds have different scaffolds and functional groups that can be experimented with in the environment of the active site using docking techniques taking crystallographic information into account.

sites that emphasize distinct chemical features can be applied to specifically analyse or calibrate electrostatic contributions, hydrophobic interactions or solvation energies of scoring functions, as has been demonstrated in docking studies on structures of mutated T4 lysozyme active sites[127]. Moreover, relatively simple scoring terms might be selected and refined for a specific type of binding site[128]. Such tailored scoring terms can produce accurate results but are, of course, not transferable. Finally, docking against protein families is likely to improve the predictive value of calculations focusing on single targets[129] and help identify specific inhibitors. This has been illustrated, for example, by the design of novel antiparasitic agents by combining virtual screening against a target family with structure-based compound library design[130].

Concluding remarks

Docking calculations have been applied in pharmaceutical research for nearly two decades. Virtual screening on protein templates, which differs from molecular similarity- and ligand-based virtual screening methods, provides an opportunity for the *de novo* identification of active compounds, without bias towards known hits or leads. From an algorithmic point of view, contemporary posing and scoring methodologies are rather diverse. The interplay between docking and scoring functions is fairly complex, but it is often easier to produce reliable models of bound ligands than to distinguish 'true' ligands from false-positives. As also discussed in this article, further improvement of scoring and compound ranking schemes does not necessarily depend on the development of novel scoring functions. Furthermore, compound filter functions, two- or three-dimensional similarity-based methods and pharmacophore models are frequently combined with docking to reduce the number of candidate compounds for fairly complex scoring calculations. Although docking and scoring relies on many approximations, the application of these techniques during lead optimization, often in concert with other computational methods, already extends more traditional approaches to structure-based design.

1. Berman, H. M. *et al*. The protein data bank and the challenge of structural genomics. *Nature Struct. Biol.* **7**, 957–959 (2000).
2. Westbrook, J., Feng, Z., Chen, L., Yang, H. & Berman, H. M. The protein data bank and structural genomics. *Nucleic Acid Res.* **31**, 489–491 (2003).
3. Blundell, T. L., Jhoti, H. & Abell, C. High-throughput crystallography for lead discovery in drug design. *Nature Rev. Drug Discov.* **1**, 45–54 (2002).
4. Bajorath, J. Integration of virtual and high-throughput screening. *Nature Rev. Drug Discov*. **1**, 882–894 (2002).
5. Walters, W. P., Stahl, M. T. & Murcko, M. A. Virtual screening — an overview. *Drug Discov. Today* **3**, 160–178 (1998).
6. Langer, T. & Hoffmann, R. D. Virtual screening: an effective tool for lead structure discovery. *Curr. Pharm. Design* **7**, 509–527 (2001).
7. Gohlke, H. & Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem. Int. Ed.* **41**, 2644–2676 (2002).
A very extensive and informative review with emphasis on quantitative analysis of protein–ligand interactions.
8. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. A geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.* **161**, 269–288 (1982).
9. Venhorst, J. *et al*. Homology modeling of rat and human cytochrome P450 2D (CYP2D) isoforms and computational rationalization of experimental ligand-binding specificities. *J. Med. Chem.* **46**, 74–86 (2003).
10. Williams, P. A. *et al*. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **424**, 464–468 (2003).
11. Brooijmans, N. & Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biolmol. Struct.* **32**, 335–373 (2003).
Excellent review of research in the docking arena that contains an instructive section on the conceptually different processes involved in ligand–protein docking.
12. Halperin, I., Ma, B., Wolfson, H. & Nussinov, R. Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins* **47**, 409–443 (2002).
13. Burnett, R. M. & Taylor, J. S. DARWIN: a program for docking flexible molecules. *Proteins* **41**, 173–191 (2000).
14. Norel, R., Lin, S. L., Wolfson, H. & Nussinov, R. Shape complementarity at protein–protein interfaces. *Biopolymers* **34**, 933–940 (1994).
15. Norel, R., Petrey, D., Wolfson, H. & Nussinov, R. Examination of shape complementarity in docking of unbound proteins. *Proteins* **35**, 403–419 (1999).
16. Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Cryst.* **16**, 548–558 (1983).
17. Connolly, M. Solvent-accessible surface of proteins and nucleic acids. *Science* **221**, 709–713 (1983).
References 16 and 17 outline the theoretical foundation of molecular surface calculations that have also become a crucial component of many shape-based docking algorithms.
18. Norel, R., Wolfson, H. & Nussinov, R. Small molecular recognition: solid angles surface representation and shape complementarity. *Comb. Chem. High Throughput Screen* **2**, 177–191 (1999).
19. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28**, 849–857 (1985).
This seminal paper introduced the idea of potential energy grids and its application to understanding protein–ligand interactions. This concept has been applied and extended in many contemporary docking programs.
20. Leach, A. R. *Molecular Modelling: Principles and Applications* (Addison Wesley Longman Limited, Harlow, 1996).
21. DesJarlais, R. L. Docking flexible ligands to macromolecular receptors by shape. *J. Med Chem.* **29**, 2149–2153 (1986).
22. Klebe, G. & Rarey, M. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **261**, 470–489 (1996).
23. Kuntz, I. D. & Leach, A. R. Conformational analysis of flexible ligands in macromolecular receptor sites. *J. Comput. Chem.* **13**, 730–748 (1992).
24. Ewing, T. J. A., Makino, S., Skillman, A. G. & Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.* **15**, 411–428 (2001).
25. Conformation search [online], <http://dock.compbio.ucsf.edu/dock4/html/Manual.f.html> (1998).
26. Kramer, B., Rarey, M., Lengauer, T. Evaluation of the FlexX incremental construction algorithm for protein–ligand docking. *Proteins* **37**, 228–241 (1999).
27. Linnainmaa, S., Harwood, D. & Davis, L. S. Pose determination of a three-dimensional object using triangle pairs. *IEEE Trans. Comput. Anal. Machine Intelligence* **10**, 634–646 (1988).
An in-depth study of a computer vision technique (pose clustering) that is utilized, for example, in FlexX.
28. Welch, W., Ruppert, J. & Jain, A. N. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* **3**, 449–462 (1996).
29. Kearsly, S. K., Underwood, D. J., Sheridan, R. P. & Miller, M. D. Flexibase: a way to enhance the use of molecular docking methods. *J. Comput. Aided Mol. Des.* **8**, 565–582 (1994).
30. Olson, A. J. & Goodsell, D. S. Automated docking in crystallography: analysis of the substrates of aconitase. *Proteins* **17**, 1–10 (1993).
31. Read, R. J. & Hart, T. N. A multiple-start Monte Carlo docking method. *Proteins* **13**, 206–222 (1992).
32. Dixon, J. S. & Oshiro, C. M. Flexible ligand docking using a genetic algorithm. *J. Comput. Aided Mol. Des.* **9**, 113–130 (1995).
33. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R. & Hart, W. E. Automated docking using a Lamarckian genetic algorithm and an empirical free energy function. *J. Comput. Chem.* **19**, 1639–1662 (1998).
34. Jones, G., Willet, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748 (1997).
35. Westhead, D. R., Clark, D. E. & Murray, C. W. A comparison of heuristic search algorithms for molecular docking. *J. Comput. Aided Mol. Des.* **11**, 209–228 (1997).
36. Baxter, C. A., Murray, C. W., Clark, D. E., Westhead, D. R. & Eldridge, M. D. Flexible docking using tabu search and an empirical estimate of binding affinity. *Proteins* **33**, 367–382 (1998).
37. Di Nola, A., Berendsen, H. J. C. & Roccatano, D. Molecular dynamics simulation of the docking of substrates to proteins. *Proteins* **19**, 174–182 (1994).
38. Trosset, J.-Y. & Scheraga, H. A. Reaching the global minimum in docking simulations: a Monte Carlo energy minimization approach using Bezier Splines. *Proc. Natl Acad. Sci. USA* **95**, 8011–8015 (1995).
39. Carlson, H. A. & McGammon, J. A. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **57**, 213–218 (2000).
Informative review of approaches to treat protein flexibility in the computational study of protein–ligand interactions.
40. Leach, A. R. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* **235**, 245–356 (1994).
41. Desmet, J., Maeyer, M. D., Hazes, B. & Lasters, I. The dead end elimination theorem and its use in protein side-chain positioning. *Nature* **356**, 539–542 (1992).
42. Knegtel, R. M. A., Kuntz, I. D. & Oshiro, C. M. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **266**, 242–440 (1997).
43. Kollman, P. A. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **93**, 2395–2417 (1993).
Review of the theory of free-energy calculations and their areas of application, including ligand binding.
44. Simonson, T., Archontis, G. & Karplus, M. Free energy simulations come of age: protein–ligand recognition. *Acc. Chem. Res.***35**, 430–437 (2002).
45. Morris, G. M. *et al*. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662 (1998).
46. Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. An all-atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **7**, 252 (1986).
47. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein–ligand docking using GOLD. *Proteins* **52**, 609–623 (2003).
48. Böhm, H.-J. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J. Comput. Aided Mol. Des.* **6**, 593–606 (1992).
49. Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V. & Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* **11**, 425–445 (1997).
50. Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **261**, 470–489 (1996).
51. Rognan, D., Lauemoller, S. L., Holm, A., Buus, S. & Tschinke, V. Predicting binding affinities of protein ligands from three-dimensional models: application to peptide binding to class I major histocompatibility proteins. *J. Med. Chem.* **42**, 4650–4658 (1999).
52. Sitkoff, D. F., Sharp, K. A. & Honig, B. Accurate calculation of hydration free energies using macroscopic continuum models. *J. Phys. Chem.* **98**, 1978–1983 (1998).
53. Huo, S., Wang, J., Cieplak, P., Kollman, P. A. & Kuntz, I. D. Molecular dynamics and free energy analyses of cathepsin D–inhibitor interactions: insight into structure-based ligand design. *J. Med. Chem.* **45**, 1412–1419 (2002).
54. Muegge, I. A knowledge-based scoring function for protein–ligand interactions: probing the reference state. *Perspect. Drug Discov. Des.* **20**, 99–114 (2000).
55. Muegge, I. Effect of ligand volume correction on PMF scoring. *J. Comput. Chem.* **22**, 418–425 (2001).
56. Muegge, I. & Martin, Y. C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **42**, 791–804 (1999).
57. Gohlke, H., Hendlich, M. & Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **295**, 337–356 (2000).
58. DeWitte, R. S. & Shakhnovich, E. I. SMoG: *de novo* design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. *J. Am. Chem. Soc.* **118**, 11733–11744 (1996).
59. Charifson, P. S., Corkery, J. J., Murcko, M. A. & Walters, W. P. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **42**, 5100–5109 (1999).
This study introduced the concept of consensus scoring as an approach to balance imperfections of single scoring functions and improve prediction accuracy.
60. Wang, R., Lai, L. & Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* **16**, 11–26 (2002).
61. Perez, C. & Ortiz, A. R. Evaluation of docking functions for protein–ligand docking. *J. Med. Chem.* **44**, 3768–3785 (2001).
62. Good, A. C. *et al*. Analysis and optimization of structure-based virtual screening protocols 2. Examination of docked ligand orientations sampling methodology: mapping a pharmacophore for success. *J. Mol. Graph. Model.* **22**, 31–40 (2003).
63. Baxter, C. A. *et al*. New approach to molecular docking and its application to virtual screening of chemical databases. *J. Chem. Inf. Comput. Sci.* **40**, 254–262 (2000).
64. GOLD Version 1.2. [online], <http://www.ccdc.cam.ac.uk/products/life_sciences/gold/> (2003).
65. Sotriffer, C. A., Gohlke, H. & Klebe, G. Docking into knowledge-based potential fields: a comparative evaluation of DrugScore. *J. Med. Chem.* **45**, 1967–1970 (2002).
66. Wang, R., Lu, Y. & Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **46**, 2287–2303 (2003).
67. McGann, M. R., Almond, H. R., Nicholls, A., Grant, J. A. & Brown, F. K. Gaussian docking functions. *Biopolymers* **68**, 76–90 (2003).
68. Schultz-Gasch, T. & Stahl, M. Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J. Mol. Model* **9**, 47–57 (2003).
69. Erickson, J. A., Jalaie, M., Robertson, D. H., Lewis, R. A. & Vieth, M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem*. **47**, 45–55 (2004).
70. Kontoyianni, M., McClellan, L. M. & Sokol, G. S. Evaluation of docking performance: comparative data on docking algorithms. *J. Med. Chem.* **47**, 558–565 (2004).
71. Smith, R., Hubbard, R. E., Gschwend, D. A., Leach, A. R. & Good, A. C. Analysis and optimization of structure-based virtual screening protocols 3. New Methods and old problems in scoring function design. *J. Mol. Graph. Model.* **22**, 41–53 (2003).
72. Still, W. C., Tempczyk, A., Hawley, R. C. & Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**, 6127–6129 (1990).
73. Ghosh, A., Rapp, C. S. & Friesner, R. A. A generalized Born model based on a surface integral formulation. *J. Phys. Chem. B* **102**, 10983–10990 (1998).
74. Nissink, J. W. M. *et al*. A new test set for validating predictions of protein–ligand interaction. *Proteins* **49**, 457–471 (2002).
75. Grzybowski, B. A., Ishchenko, A. V., Shimada, J. & Shakhnovich, E. I. From knowledge-based potentials to combinatorial lead design *in silico*. *Acc. Chem. Res.* **35**, 261–269 (2002).
76. Diller, D. J. & Li, Y. Kinases, homology models, and high throughput docking. *J. Med. Chem.* **46**, 4638–4647 (2003).
77. DesJarlais, R. L. *et al*. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J. Med. Chem.* **31**, 722–729 (1988).

78. Dean, P. M. & Poornima, C. S. Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein–ligand interactions. *J. Comput. Aided Mol. Des.* **9**, 500–512 (1995).
79. McGovern, S. L., Caselli, E., Grigorieff, N. & Shoichet, B. K. A common mechanism underlying promiscous inhibitors from virtual and high-throughput screening. *J. Med. Chem.* **45**, 1712–1722 (2002).
80. Roche, O. *et al*. Development of a virtual screening method for identification of 'frequent hitters' in compound libraries. *J. Med. Chem.* **45**, 137–142 (2002).
81. Doman, T. N. *et al*. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem.* **45**, 2213–2221 (2002).
An impressive example of the performance of structure-based virtual screening.
82. McGovern, S. L. & Shoichet, B. K. Information decay in molecular docking screens against holo, apo and modeled conformations of enzymes. *J. Med Chem.* **46**, 2895–2907 (2003).
Informative analysis of the influence of chosen protein-structure templates on the quality of docking and scoring.
83. Lipinski, C. A. & Christopher, A. L. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23**, 3–25 (1997).
84. Nilakantan, R., Bauman, N. & Venkataraghavan, R. New method for rapid characterization of molecular shapes: applications in drug design. *J. Chem. Inf. Comput. Sci.* **33**, 79–85 (1993).
85. Good, A. C., Ewing, T. J. A., Gschwend, D. A. & Kuntz, I. D. New molecular shape descriptors: application in database screening. *J. Comput. Aided Mol. Des.* **9**, 1–12 (1995).
86. Zauhar, R. J., Moyna, G., Tian, L., Li, Z. & Welsh, W. J. Shape signatures: a new approach to computer-aided ligand-and receptor-based drug design. *J. Med. Chem.* **46**, 5674–5690 (2003).
87. Rastelli, G. *et al*. Docking and database screening reveal new classes of *Plasmodium falciparum* dihydrofolate reductase inhibitors. *J. Med. Chem.* **46**, 2834–2845 (2003).
88. Choong, I. C. *et al*. Identification of potent and selective small-molecule inhibitors of caspase-3 through the use of extended tethering and structure-based drug design. *J. Med. Chem.* **45**, 5005–5022 (2002).
89. Kick, E. K. *et al*. Structure-based design and combinatorial chemistry yield low nanomolar inhibitors of cathepsin D. *Chem. Biol.* **4**, 297–307 (1997).
An instructive study highlighting the potential of interfacing docking analysis and targeted library design.
90. Karplus, M. & Miranker, A. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* **11**, 29–34 (1991).
91. Caflisch, A. Computational combinatorial ligand design: application to human α-thrombin. *J. Comput. Aided Mol. Des.* **10**, 372–396 (1996).
92. Böhm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* **8**, 243–256 (1994).
Pioneering development of an empirical scoring function using multiple linear regression to calculate coefficients for the most important terms contributing to ligand binding.
93. Böhm, H. J. Prediction of binding constants of protein ligands: a fast method for the polarization of hits obtained from *de novo* design on 3D database search programs. *J. Comput. Aided Mol. Des.* **12**, 309–323 (1998).
94. Murcko, M. A. & Rotstein, S. H. GroupBuild: a fragment-based method for *de novo* drug design. *J. Med. Chem.* **36**, 1700–1710 (1993).
95. Murcko, M. A. & Rotstein, S. H. GenStar: a method for *de novo* drug design. *J. Comput. Aided Mol. Des.* **7**, 23–43 (1993).
96. Howe, W. J. & Moon, J. B. 3D database searching and *de novo* construction methods in molecular design. *Comput. Meth.* **3**, 697–711 (1990).
97. Bohacek, R. S. & McMartin, C. Multiple highly diverse structures complementary to enzyme binding sites: results of extensive application of a *de novo* design method incorporating combinatorial growth. *J. Am. Chem. Soc.* **116**, 5560–5571 (1994).
98. Vinkers, H. M. *et al*. SYNOPSIS: SYNthesize and OPtimize system *in silico*. *J. Med. Chem.* **46**, 2765–2773 (2003).
99. Guimaraes, C. R. W. & de Alencastro, R. B. Thrombin inhibition by novel benzamidine derivatives: a free-energy perturbation study. *J. Med. Chem.* **45**, 4995–5004 (2003).
100. Pearlman, D. A. & Charifson, P. S. Improved scoring of ligand–protein interactions using OWFEG free energy grids. *J. Med. Chem.* **44**, 502–511 (2001).
101. Aquist, J., Medina, C. & Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **7**, 385–391 (1994).
This article presents an early formulation and use of linear response and linear interaction approximations in estimating binding affinity of protein ligands.
102. Tounge, B. A. & Reynolds, C. H. Calculation of the binding affinity of β-secretase inhibitors using the linear interaction energy method. *J. Med. Chem.* **46**, 2074–2082 (2003).
103. Rizzo, R. C., Wang, D.-P., Tirado-Rives, J. & Jorgensen, W. L. Validation of a model for the complex of HIV-1 reverse transcriptase with sustiva through computation of resistance profiles. *J. Am. Chem. Soc.* **122**, 12898–12900 (2003).
104. Rizzo, R. C., Tirado-Rives, J. & Jorgensen, W. L. Estimation of binding affinities for HEPT and nevirapine analogues with HIV-1 reverse transcriptase via Monte Carlo simulations. *J. Med. Chem.* **44**, 145–154 (2003).
105. Kroeger-Smith, M. B. *et al*. Molecular modeling calculations of HIV-1 reverse transcriptase nonnucleoside inhibitors: correlation of binding energy with biological activity for novel 2-aryl-substituted benzimidazole analogues. *J. Med. Chem.* **46**, 1940–1947 (2003).
106. Udier-Blagovic, M., Tirado-Rives, J. & Jorgensen, W. L. Validation of a model for the complex of HIV-1 reverse transcriptase with nonnucleoside inhibitor TMC125. *J. Am. Chem. Soc.* **125**, 6016–6017 (2003).
107. Rizzo, R. C. *et al*. Prediction of activity for nonnucleoside inhibitors with HIV-1 reverse transcriptase based on Monte Carlo simulations. *J. Med. Chem.* **45**, 2970–2987 (2002).
108. Ostrovsky, D., Udier-Blagovic, M. & Jorgensen, W. L. Analyses of activity for Factor Xa inhibitors based on Monte Carlo simulations. *J. Med. Chem.* **46**, 5691–5699 (2003).
109. van Lipzig, M. M. *et al*. Prediction of ligand binding affinity and orientation of xenoestrogens to the estrogen receptor by molecular dynamics simulations and the linear interaction energy method. *J. Med. Chem.* **47**, 1030 (2004).
This work provides a good example of linear interaction methods applied to binding energies ranging over many orders of magnitude.
110. Kollman, P. A. *et al*. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897 (2000).
111. Masukawa, K. M., Kollman, P. A. & Kuntz, I. D. Investigation of neuraminidase-substrate recognition using molecular dynamics and free energy calculations. *J. Med. Chem.* **46**, 5628–5637 (2003).
112. Sheridan, R., Holloway, M. K., McGaughey, G. B., Mosley, R. T. & Singh, S. B. A simple method for visualizing the differences between related receptor sites. *J. Mol. Graph. Model.* **21**, 71–79 (2002).
113. Deng, Z., Chuaqui, C. & Singh, J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein–ligand binding interactions. *J. Med. Chem.* **47**, 337–344 (2004).
114. Horvath, D. A virtual screening approach applied to the search for trypanothione reductase inhibitors. *J. Med. Chem.* **40**, 2412–2423 (1997).
The study details many possible scoring terms for protein–ligand complexes and is a good example of the value of refitting parameters for a particular protein class and series of ligands.
115. Matter, H. *et al*. Design and quantitative structure–activity relationship of 3-amidinobenzyl-1H-indole-2-carboxamides as potent, nonchiral and selective inhibitors of blood coagulation factor Xa. *J. Med. Chem.* **45**, 2749–2769 (2002).
116. Murcia, M. & Ortiz, A. R. Virtual screening with flexible docking and COMBINE-based models. Application to a series of factor Xa inhibitors. *J. Med. Chem.* **47**, 805–820 (2004).
117. van de Waterbeemd, H. & Gifford, E. ADMET *in silico* modelling: towards prediction paradise? *Nature Rev. Drug Discov.* **2**, 192–204 (2003).
118. Omiecinski, C. J. Concise review of the cytochrome P450s and their roles in toxicology. *Toxicol. Sci.* **48**, 151–156 (1999).
119. de Groot, M. J., Ackland, M. J., Horne, V. A., Alex, A. A. & Jones, B. C. Novel approach to predicting P450-mediated drug metabolism: development of a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* **42**, 1515–1524 (1999).
120. de Groot, M. J., Ackland, M. J., Horne, V. A., Alex, A. A. & Jones, B. C. A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed n-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* **42**, 4062–4070 (1999).
121. de Groot, M. J. Development of a combined protein and pharmacophore model for cytochrome P450 2C9. *J. Med. Chem.* **45**, 1983–1993 (2002).
122. Park, J.-Y. & Harris, D. Construction and assessment of models of CYP2E1: Predictions of metabolism from docking, molecular dynamics, and density functional theoretical calculations. *J. Med. Chem.* **46**, 1645–1660 (2003).
123. Godden, J. W., Stahura, F. L. & Bajorath, J. Statistical analysis of computational docking of large compound databases to distinct protein binding sites. *J. Comput. Chem.* **20**, 1634–1643 (1999).
124. Briem, H. & Kuntz, I. D. Molecular similarity based on DOCK-generated fingerprints. *J. Med. Chem.* **39**, 3401–3408 (1996).
125. Su, A. I. *et al*. Docking molecules by families to increase the diversity of hits in database screens: computational strategy and experimental evaluation. *Proteins* **42**, 279–293 (2001).
126. Rognan, D., Lauemoller, S. L., Holm, A., Buus, S., Tschinke V. Predicting binding affinities of protein ligands from three-dimensional models: application to peptide binding to class I major histocompatibility proteins. *J. Med. Chem.* **42**, 4650–4658 (1999).
127. Wei, B. Q., Baase, W. A., Weaver, L. H., Matthews, B. W. & Shoichet, B. K. A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol.* **322**, 339–355 (2002).
128. Fradera, X., Knegtel, M. A., Mestres, J. Similarity-driven flexible ligand docking. *Proteins* **40**, 623–626 (2000).
129. Lamb, M. L. *et al*. Design, docking, and evaluation of multiple libraries against multiple targets. *Proteins* **42**, 296–318 (2001).
130. Aronov, A. M., Munagala, N. R., Kuntz, I. D. & Wang, C. C. Virtual screening of combinatorial libraries across a gene family in search of inhibitors of Giardia lamblia guanine phosphoribosyltransferase. *Antimicrob. Agents Chemother.* **45**, 2571–2576 (2001).
131. Wang, R., Liu, L., Lai, L. & Tang,Y. SCORE: a new empirical method for estimating the binding affinity of a protein-ligand complex. *J. Mol. Model* 4, 379–394 (1998).
132. Tao, P. & Lai, L. Protein ligand docking based on empirical method for binding affinity estimation. *J. Comput. Aided Mol. Des.* **15**, 429–446 (2001).
133. Chemical Computing Group. MOE. 2003. Montreal, Quebec, Canada.
134. Friesner, R. A. *et al*. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739–1749 (2004).
135. Kearsley, S. K., Underwood, D. J., Sheridan, R. P. & Miller, M. D. Flexibases: a way to enhance the use of molecular docking methods. *J. Comput. Aided Mol. Des.* **8**, 565–582 (1994).
136. Peng, H. *et al*. Identification of novel inhibitors of BCR-ABL tyrosine kinase via virtual screening. *Bioorg. Med. Chem. Lett.* **13**, 3693–3699 (2003).
137. McNally, V. A. *et al*. Identification of a novel class of inhibitor of human and Escherichia coli thymidine phosphorylase by *in silico* screening. *Bio. Med. Chem. Lett.* **13**, 3705–3709 (2003).
138. Brenk, R. *et al*. Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. *J. Med. Chem.* **46**, 1133–1143 (2003).
139. Kamionka, M. *et al*. *In silico* and NMR identification of inhibitors of the IGF-I and IGF-Binding protein-5 interaction. *J. Med. Chem.* **45**, 5655–5660 (2002).
140. Vangrevelinghe, E. *et al*. Discovery of a potent and selective protein kinase CK2 inhibitor by high-througput docking. *J. Med. Chem.* **46**, 2656–2662 (2003).
141. Enyedy, I. J. *et al*. Discovery of small-molecule inhibitors of Bcl-2 through structure-based computer screening. *J. Med. Chem.* **44**, 4313–4324 (2001).

Acknowledgements

H.D. and J.R.F. contributed equally to this paper. This manuscript is dedicated to Wolfram Saenger, Free University Berlin, on the occasion of his sixty-fifth birthday.

Competing interests statement

The authors declare no competing financial interests.

Online links

FURTHER INFORMATION
Protein Structure Prediction Center: http://predictioncenter.llnl.gov/
Research Collaboratory for Structural Biology Protein Data Bank: http://www.rcsb.org/pdb
Biomolecular Interaction Network Database: http://www.bind.ca/
Drug Design Resources: http://www.drugdesign.org
Access to this interactive links box is free online.