Structure-based virtual screening: an overview

Paul D. Lyne

Enormous advances in genomics have resulted in a large increase in the number of potential therapeutic targets that are available for investigation. This growth in potential targets has increased the demand for reliable target validation, as well as technologies that can identify rapidly several quality lead candidates. Virtual screening, and in particular receptor-based virtual screening, has emerged as a reliable, inexpensive method for identifying leads. Although still an evolving method, advances in computational techniques have enabled virtual screening to have a positive impact on the discovery process. Here, the current strengths and weaknesses of the technology are discussed, and emphasis is placed on aspects of the work-flow of a virtual screening campaign, from preparation through to post-screening analysis.

Paul D. Lyne
AstraZeneca R&D
Boston, Waltham
MA 02451, USA
e-mail: paul.lyne@
astrazeneca.com

▼ The pharmaceutical industry is under ever-increasing pressure to increase its success rate in bringing drugs to the market. It is estimated that on average it can take 14 years to bring a compound from hit identification through to an approved drug [1], and the costs associated with this process are enormous, with the bulk of the expense being incurred in the development phase of the value chain. Current efforts within the industry are directed at reducing the hit-to-drug timeline, increasing the number of quality candidate drugs that make the transition from discovery to clinical development, and decreasing the attrition rate (currently 90%) of candidate drugs in the clinical stages of the value chain. As a consequence, there has been an increased investment in discovery in technologies aimed at achieving these goals. This includes investments in functional genomics [2], for improved identification and validation of therapeutic targets; in HTS [3] and combinatorial chemistry [4] for hit identification; in both experimental [5] and predictive ADME methods [6]; and in structural biology (X-ray [7] and NMR [8]) for the identification of hits and for rational-based evolution of hits to candidate drugs.

A background to virtual screening:

Virtual screening is another technology that is gaining increasing use in discovery [9]. It is seen as a complementary approach to experimental screening (HTS), and when coupled with structural biology, promises to increase the number, and enhance the success, of projects in the lead identification stage of the discovery process. There are many tools available for performing these computational analyses and broadly speaking they can be categorized as being either ligand-based or receptor-based. For ligand-based methods, the strategy is to use information provided by a compound or set of compounds that are known to bind to the desired target and to use this to identify other compounds in the corporate database or external databases with similar properties. This can be done by a variety of methods, including similarity and substructure searching [10], pharmacophore matching [11] or 3D shape matching [12].

When the structure of the target protein is known, receptor-based computational methods can be employed. These involve explicit molecular docking of each ligand into the binding site of the target, producing a predicted binding mode for each database compound, together with a measure of the quality of the fit of the compound in the target binding site. This information is then used to rank the compounds with a view to selecting and experimentally testing a small subset for biological activity. The current state-of-the-art of structure-based virtual screening is reviewed, and general approaches, successes and pitfalls associated with the technology are highlighted. This provides a general overview of structure-based screening approaches and strategies,

1359-6446/02/$ – see front matter PII: S1359-6446(02)02483-2

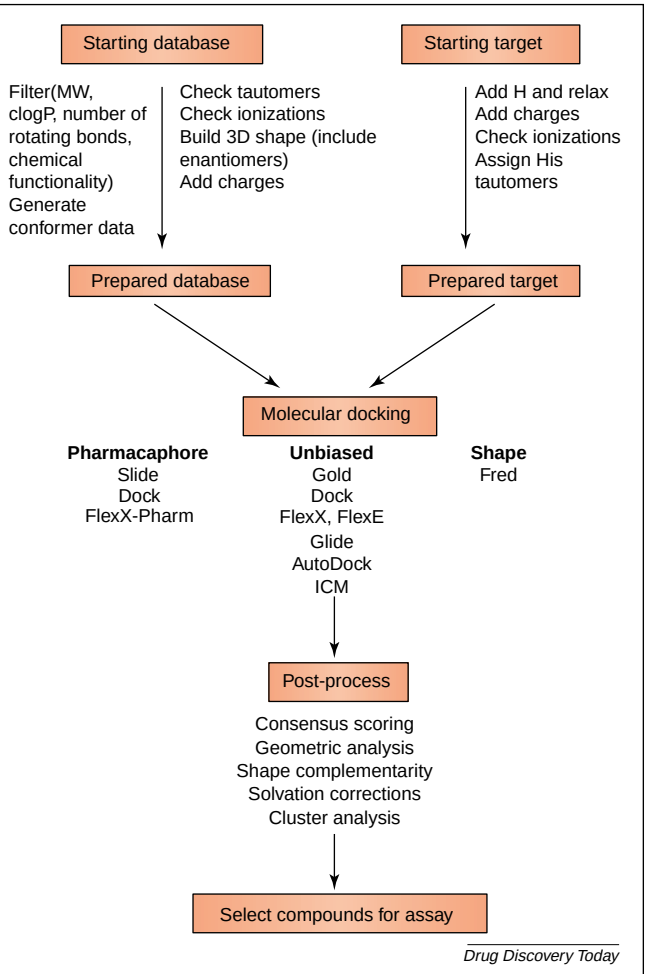

Fig. 1. A pictorial description of the workflow of a virtual screening run against a specific target. The typical workflow consists of a preparation phase for the database and the target, followed by a molecular docking phase, and concludes with the post-processing and compound selection phases. The molecular docking portion of the screen is categorized according to the approach employed. This is not an exhaustive list of molecular docking tools, but includes examples of common docking tools that could be employed. In the post-processing phase, several approaches that might be used are listed. Some or all of these might be applied in any given screen.

with relevant references to the literature if further details are required [13].

The screening process

Structure-based virtual screening encompasses a variety of sequential computational phases, including target and database preparation, docking and post-docking analysis, and prioritization of compounds for testing (Fig. 1). The successful application of virtual screening depends on sound implementation of a wide range of computational techniques in each phase of the screen, which will be discussed in turn.

Database and target preparation

Database preparation

In the initial stages of a virtual screening project it is necessary to prepare the database of compounds to be screened. The source of the database typically corresponds to a corporate collection of physically available compounds, or a database of compounds available externally from chemical vendors [e.g. the MDL® Available Chemicals Directory (ACD) from MDL Information Systems; http://www.mdli.com)]. An additional source that is sometimes considered for virtual screening is an *in silico* virtual library corresponding to compounds constructed from a list of reagents and a database of known chemistries. The members of such an *in silico* library are readily synthesizable and could easily be produced once they have been selected from a virtual screen. In the case of a corporate or external collection, the initial database is reduced in size by applying several physical and chemical filters, in an attempt to have a database of compounds that have physical properties and chemical functionality consistent with the majority of known drugs. For *in silico* libraries, these factors should already have been taken into consideration during the design of the libraries. Common filtering protocols for 'drug-likeness' [6] that are applied to compound databases are variations of Lipinski's rule-of-five [14], which is an empirical set of rules based on molecular weight, lipophilicity and hydrophobicity, that provides a simple profile for orally bioavailable compounds. Other physical filters could include a limit on the number of rotatable bonds in the molecules or the polar surface area [15]. Additional filters are often applied at this stage to remove compounds containing specific chemical substructures associated with poor chemical stability or toxicity [9]. All of these filtering protocols are computationally inexpensive and can be applied rapidly to a large database (of the order of 10^5–10^6 compounds).

The essence of structure-based virtual screening is to simulate molecular recognition events. Therefore, special care needs to be made when expanding the database from its initial form [in SMILES (http://www.daylight.com) representation for example] to a database with 3D co-ordinates. Moreover, physically relevant ionization and tautomeric states need to be assigned to compounds in the database. Often compounds are registered in a database as a tautomer that is not necessarily the most relevant state of the molecule. Similarly, balanced tautomeric equilibria might exist at the pH of the experimental assay, and thus all the relevant tautomers need to be considered because there is no way of knowing *a priori* which tautomer is most likely to bind to the receptor. The stereochemistry of chiral centers are often not known before generation of the 3D database

Table 1. Common docking tools for virtual screening

Method	Ligand flexibility sampling	Scoring function	Suitability for large-scale virtual screening[a]
Dock [29]	Incremental build	Force field or contact score	High
FlexX [28]	Incremental build	Empirical score	High
Slide [30]	Conformational ensembles	Empirical score	High
Fred (Openeye Software)	Conformational ensembles	Gaussian score or empirical scores	High
Gold [25]	Genetic algorithm	Empirical score	Low
Glide (Schrodinger)	Exhaustive search	Empirical score	Low
AutoDock [26]	Genetic algorithm	Force field	Low
LigandFit (Accelrys)	Monte Carlo	Empirical score	Low
ICM [27]	Pseudo-Brownian sampling and local minimization	Mixed force field and empirical score	Low
QXP [23]	Monte Carlo	Force field	Low

[a]Qualitative assessments of the suitability of the methods to screen 100K compounds on 8 processors within few days. Methods designated as having low suitability for 100Ks of compounds should be considered suitable for 10K compounds under similar conditions, or would require considerably more computer power to be applied to 100Ks of compounds. Openeye Software, http://www.eyesopen.com; Schrodinger, http://www.schrodinger.com; Accelrys, http://www.accelrys.com.

and so it is necessary to generate enantiomers arising from chiral centers in the molecules.

There are several methods available for the generation of the 3D structure of a molecule [e.g. Corina [16], Concord and Confort (both Tripos; http://www.tripos.com) or Converter (Accelrys; htt://www.accelrys.com)]. Depending on the docking tool that will be used in the screen, it might also be necessary to assign partial charges to the compounds in the database. Typically, Gasteiger charges [17] are used, although newer, presumably more accurate charge sets, such as MMFF94 charges [18], are sometimes chosen. Other problems arise from geometric isomerism, such as asymmetrically substituted tertiary amines or amidines, inversion at tetrahedral nitrogen (treated as for chiral centers), and alicyclic ring flips.

Target preparation

The preparation of the active site is dependent on the docking tools being used. Some methods require the addition of hydrogens and care should be taken to ensure that this is done sensibly, avoiding atomic clashes. The appropriate protonation states of ionizable residues in the active site need to be determined and the correct tautomer for histidines should be assigned. It is recommended that after the addition of hydrogen atoms to the protein, the positions of the hydrogens are relaxed by energy minimization to avoid any steric clashes that have been introduced.

The positioning of hydrogen atoms on hydroxyl groups in the active site should also be examined and altered if necessary. In some instances, tightly bound crystallographic waters might need to be maintained for the virtual screen. These can be identified from the X-ray structure of the protein or with the use of analysis tools, such as Relibase+ [19], if several X-ray complexes of the target are available, to select the conserved water positions. In addition, programs such as Whatcheck can be used to identify errors in crystallographic structures arising from incorrect assignment of side-chain amide nitrogen and oxygen positions and from incorrectly oriented histidine rings [20].

Molecular docking

The next phase of the screening process involves docking each molecule in the database into the binding site of the target. The docking process consists of sampling the coordinate space of the binding site and scoring each possible ligand pose, which is then taken as the predicted binding mode for that compound. There are a large number of docking programs available for use in virtual screening and they differ in the sampling algorithms used, the handling of ligand and protein flexibility, the scoring functions they employ, and the cpu time required to dock a molecule to a given target. In Table 1, several commonly used docking methods are listed along with the types of sampling and scoring algorithms used, and an indication of their suitability for large-scale virtual screening.

Ligand flexibility

The conformation of a compound bound to a target might be different from the conformation of the unbound form in solution. In general, the free state in solution can be thought of as an ensemble of conformations of which a small subset is pertinent to the bound form. It is therefore crucial to consider the conformational flexibility of the compounds during the docking process. This can be achieved

by pre-computing a database that contains several conformers of each compound to be screened. There are several software packages [Confort, Omega (Openeye Scientific Software; http://www.eyesopen.com) and Catalyst (Accelrys)] that are available to generate large conformational databases rapidly, and then each conformer can be rigidly docked into the target binding site. Examples of methods that work with configurational ensembles [21,22] are Dock, Fred and Slide. Alternatively, the majority of docking programs explore the conformational flexibility of the compounds as the calculation proceeds, through a variety of docking algorithms. The size of conformational space for a ligand is directly related to the number of rotatable bonds present in the ligand, and efficient algorithms need to be devised to handle the combinatorics of this problem. There are several computational strategies that are used by current docking methods, and although it is not the intention to detail all the methods available, the major approaches are discussed with some common docking methods mentioned as examples (other specific methods are reviewed elsewhere in the literature [13]).

One approach to exploring conformational space of a ligand is to use Monte Carlo or simulated annealing methods [23,24], but these methods are often time-consuming and not suitable for large-scale virtual screening. An alternative approach is to use a genetic algorithm to generate conformers for a ligand. This approach has been successfully employed in the program GOLD [25], and Morris *et al.* [26] have employed a Lamarckian genetic algorithm in the most recent version of AutoDock. Genetic algorithm approaches are also time-consuming and necessitate the extensive use of parallel processing to be applied realistically in virtual screening of large databases. The ICM program [27] uses pseudo-Brownian sampling, coupled with local minimization to explore the conformational space of the ligand in the binding site. Sets of minima are stored in a history list that is used to promote efficient sampling.

Incremental growth methods are another strategy applied by docking methods [28,29] for exploring the flexibility of a ligand in a target binding site. These involve dividing the molecule into small fragments and incrementally building the molecule in the binding site. One of the more efficient docking methods currently available, FlexX [28], uses a fragment-based approach and a set of low-energy torsion angles derived from the Cambridge Structural Database (CSD; http://www.ccdc.cam.ac.uk) for each single bond. This restricts the conformational sampling needed for each rotatable bond in the molecule. In a similar manner, GOLD enables customization of the torsional energy contributions via statistical analysis of torsion angle distributions observed in the CSD.

In the docking program Glide (Schrodinger; http://www.schrodinger.com), the ligand flexibility is treated by exhaustive enumeration of the rotamer states for each rotatable bond coupled to a heuristic screen that rapidly eliminates conformations deemed unsuitable for binding to a receptor.

Target flexibility

An aspect of the docking problem that has not been discussed is the flexibility of the protein target. The majority of docking tools currently make the assumption that the protein target is held fixed in its crystal structure conformation. This is usually an inaccurate approximation but is generally a necessary one because of the increased complexity, and consequently the computational cost, that is required to accurately sample the flexibility of the binding site. There are some programs that attempt to take into account protein flexibility to a certain degree, such as Slide [30], which enables the motion and relaxation of binding-site sidechains in response to the presence of a docked ligand. Most efforts to incorporate protein flexibility try to make use of an ensemble of protein structures, which can be obtained from several high-resolution structures of the same target bound to different ligands, or in the apo form.

An alternative source of ensembles of target structures could be from an NMR structural study of the target. In the absence of experimental information, the conformational ensemble could be derived from a molecular dynamics simulation, although there are issues related to the extent of the sampling produced by simulations [31]. Several programs exist that have adopted these types of approaches [32]. A promising recent modification of the FlexX program, FlexE [33], incorporates information on the conformational flexibility of a target from several available crystal structures, using the fixed co-ordinates of structurally conserved residues and rotamer libraries for flexible sidechains in the binding site. The program AutoDock has also been used to investigate several strategies for incorporating protein mobility using an ensemble of protein structures to generate Boltzmann-weighted grids with which to generate the docking function [34].

As an alternative to using the ensemble of structures to generate a composite structure, the screening process could be conducted on each individual structure (or a subset) of the target, with a view to identifying all the ligands that bound to at least one conformational form of the protein. Naturally, this would be computationally expensive if there are many structures of the target available, but might be feasible if a small number of structures are being considered.

Scoring

Once a pose has been generated for a compound in the binding site it needs to be scored to rank the quality of the pose with respect to other poses for the compound, and with respect to other molecules in the database. There is a wide choice of scoring functions available [35], and they can be categorized as being physical-based (force-field), empirical or knowledge-based.

The physical-based scoring functions are based on atomic force fields, such as Amber [36] or CHARMM [37]. These force-fields, when employed with free energy perturbation (FEP) or thermodynamic integration (TI) methods [38], are generally accurate at estimating binding free energies. For the purposes of molecular docking, the computational time associated with FEP and TI methods is prohibitive because of the vast amount of sampling required and the inclusion of explicit solvent. Typically for docking, the physical-based scoring functions (e.g. Dock [29] and QXP [23]) employ force-fields in a minimalistic manner on a grid with no explicit solvent to obtain a single-point energy value, which is used as the score for the pose. Empirical-based scoring functions based on physicochemical properties such as hydrogen-bond counts (e.g. FlexX [28], SCORE [39], VALIDATE [40], Chemscore [41], Ludi [42] and PLP [43]), use an additive approximation to estimate the binding free energy. The overall binding free energy is composed of several free energy terms corresponding to hydrogen bonding, hydrophobic interactions, entropic changes and, in some cases, interactions with metal ions. The coefficients of each term in the sum are derived from fitting to known experimental binding energies for a variety of different protein–ligand complexes. These empirical-based scoring functions are fast and are therefore often employed by docking algorithms.

Recently, several scoring functions have been developed that use knowledge-based functions. This is an approach that has been borrowed from the field of protein folding [44]. In this case, the scoring functions are derived from the structures of protein–ligand complexes using statistical mechanics. The binding free energy of the complex is given by a sum of free energies (potentials of mean force) of interatomic contacts calculated from the frequencies of these interatomic distances in a database of experimental structures from statistical mechanics methods. Examples of knowledge-based scoring functions include PMF [45], Bleep [46], SMoG [47] and Drugscore [48]. Knowledge-based potentials are as accurate as empirical-based methods on standard test sets and are also fast to compute. Empirical-based methods rely on the availability of protein–ligand complexes with known binding affinity, and consequently they have been derived based on a relatively small set (~100) of publicly available complexes. Knowledge-based methods, by contrast, do not require binding affinity data and so are free to use the information available in protein–ligand complexes deposited in the public (~1000) or proprietary databases. Consequently, the expectation is that knowledge-based functions, when compared with empirical-based methods, are less biased and more readily transferred to systems that have not been used in the development of the scoring function. However, they do require structural data and, at present, they are limited by a paucity of suitable information, which can lead to minimal differentiation of atom types. For example, a nitrogen atom can exist in several highly different contexts – anionic, cationic, trigonal, tetrahedral and so on – but in certain implementations has been represented by a single atom type. A possible advantage of knowledge-based methods (and to some extent physical methods) is the ability to address some modes of interaction that could escape inclusion in empirical approaches, such as aromatic–Cl—O=C interactions, or cation–π interactions.

Post-analysis

A molecular docking screen of a large compound database against a receptor yields a vast amount of data comprising the predicted binding pose for each compound, along with the predicted binding affinity of that ligand for the target. It is conceivable that one could choose a list of compounds to be tested based upon the rank ordering of these compounds without further analysis; however, there are several reasons why this might be unwise. First, it is well known that current scoring functions used in virtual screening campaigns are often inadequate at predicting the true binding affinity of a ligand for a receptor. A recent study [49] of the performance of several scoring functions of varying complexity, for predicting the binding affinities of several ligands to p38 MAP kinase, highlighted the inadequacies of some common scoring functions (PLP, Chemscore and Dock), with little or no correlation found between the predicted and experimental binding affinities. These findings have been corroborated for similar studies on several in-house projects at AstraZeneca using Chemscore, Gold and PMF. Although there have been reports of progress in the area of scoring-function development [48,50], the accurate or reliable prediction of binding affinities coupled with the computational speed characteristics required remain elusive. Nonetheless, there are several post-analysis strategies that can be employed to minimize the number of false positives in the selection list and to propagate the true hits to the top of the list.

Additional scoring

One popular strategy is to use the concept of consensus scoring [51]. In this approach, a given docking function is used to generate the top-ranked poses for the compounds in the target receptor, and then multiple scoring functions are used to score the top-ranked pose. Only the top-ranked compounds common to each scoring function (the consensus score) are chosen for biological testing. This approach has been shown to improve the enrichment of true hit outputs from a virtual screen [51–53]. Nonetheless, care still needs to be taken when applying the consensus scoring approach. The performance of individual scoring functions should be assessed when possible so that the most appropriate combination of scoring functions are chosen for each specific system. Additionally, there are several ways that the consensus can be applied [54]. An evaluation of several docking methods found that invariably the methods were able to identify the correct binding pose but usually could not rank this pose as the top binding pose for the ligand [55]. It is therefore more prudent to perform consensus scoring, not only on the top pose, but on a larger set of poses for each ligand. This approach was recently used by Stahl and Rarey in an evaluation of the performance of various scoring functions for virtual screening [53].

The deficiencies of current scoring functions can be traced back to an inadequate description of the physics of the interaction between a ligand and the target. The incorporation of solvation energies and more rigorous descriptions of the electrostatic interactions of the system are sometimes used as a post-processing tool to aid in re-ranking the list from a molecular docking screen. The importance of accurate descriptions of electrostatic interactions for the accurate determination of ligand–protein binding affinities is well established [56]. Most rigorous approaches to describing the electrostatics of protein–ligand complexes solve the Poisson-Boltzmann equation for the system, which unfortunately, because of the computational cost of the calculations, is impractical for the number of ligand–protein poses scored in a typical virtual-screening experiment. An elegant solution to the Poisson–Boltzmann equation based on Gaussian functions has been proposed and implemented in the program ZAP (Openeye Scientific Software) [57]. This is an extremely rapid and accurate method for solving the Poisson–Boltzmann equation and thus it is feasible to use ZAP as a post-processing tool for the output from a molecular docking screen. Alternative approaches to include solvation corrections involve precomputing desolvation estimates for the ligands in the database. This type of approach has been implemented in the program Dock, which can be modified to include a desolvation penalty to the scoring function based on precalculated Amsol energies; this has been reported to improve the quality of the ranking compared to the original scoring function used in Dock [58].

Geometric analysis

Geometric analysis can also be used in the post-processing phase [59]. This can include calculations to determine the surface complementarity of the ligand and protein for each docked molecule; the percentage of ligand surface area that remains exposed in the predicted complex (with particular attention paid to exposed hydrophobic regions); and the total amount of surface area of the ligand and protein that are buried in the complex. Another form of geometric post-analysis relies on an interrogation of the structural aspects of the docked protein–ligand complexes, such as the requirement of a specific hydrogen-bonding interaction between the ligand and the protein. If particular interactions between potential ligands and the protein are known to be desirable (for example a hydrogen-bond donor interaction with the catalytic aspartates of aspartyl proteases [60]) and a docking method was employed that does not make use of biased sampling, then an analysis of all the docked poses retaining only those that enable certain key interactions can also be useful in reducing the number of false positives in the final selected list. Finally, before making a selection for experimental testing, the compounds are inspected visually by the modeller, making sure that the interactions between the ligand and the protein make sense, and that the bound conformation of the ligand is reasonable. This can be identified visually or by comparison with known conformations from the CSD. There is a lack of commercially available software that is suitable for browsing through large lists of data from a virtual screen that enables simultaneous visual inspection and interrogation of various aspects of the data (i.e. the score and various ligand properties), which result in several in-house solutions. An exception to this is the molecular browser package VIDA (Openeye Scientific Software), which enables simultaneous visual and graphical analysis of large compound lists and data.

In all the post-analysis cases discussed here it is strongly advisable to use, when possible, a set of known ligands (preferably with known binding modes and affinities) to calibrate the virtual screening methodology.

Strategies for virtual screening

Ultimately, structure-based virtual screening is knowledge-driven and thus there are several rational approaches that can be applied to any specific screening project. Certainly it is feasible to take a prepared 3D database and dock each

member to the receptor, and follow this up with careful post-analysis to make a final selection of compounds to be tested. Often, however, there is additional information available that could be used rapidly to reduce the number of compounds before they are screened with a time-consuming docking method. This tiered approach to virtual screening, beyond the normal filtering applied in the preparation of the database, is particularly useful when the size of the prepared screening database is large (up to a million or more compounds). Several computational and structural strategies are shown in Fig. 2.

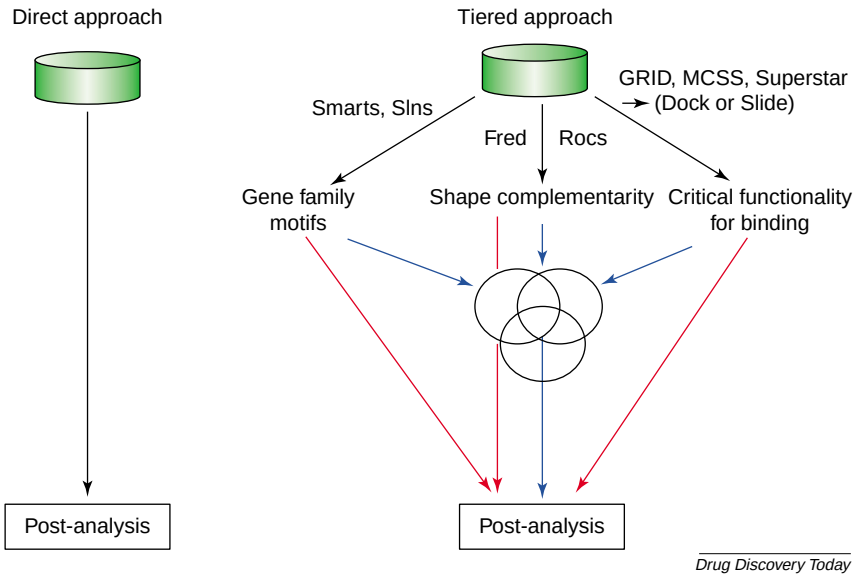

Fig. 2. In any virtual screening campaign, several strategies might be employed. Shown here are two contrasting approaches to such a campaign. One strategy is to take a direct approach to the screen, whereby the prepared database is docked into the target using a certain molecular docking tool. An alternative strategy is to take a tiered approach, whereby several fast computational techniques are initially applied to the database to produce an enriched subset of compounds that are then docked by one or several molecular docking methods. In the case of several docking methods being used, the analysis of the poses could be over all the hits produced by each of the methods (red arrows), or over only those 'hits' that have been found by all the methods (the intersection in the Venn diagram – blue arrows).

One approach is to screen the large database initially with a fast docking tool such as Fred (Openeye Scientific Software), Slide (in biased site point mode) or Dock (using critical clusters). In the case where the binding site is small or predominantly hydrophobic in nature (where shape complementarity is expected to be especially important), Fred (a shape-based method) can rapidly eliminate compounds in the database that do not have shapes that are complementary to the binding site under consideration. The resultant reduced database can then be processed through atom-based molecular docking screens, using higher fidelity, but substantially slower, throughput methods.

The chemical landscape of the binding-site can be used to derive functional-group maps or hot-spots for protein–ligand interaction, which can then be used as binding-site pharmacophores. The programs Dock and Slide can make use of these binding-site pharmacophores in the target to screen out rapidly compounds in the database that do not have appropriate chemical functionality to satisfy a percentage of the hot-spots that are present in the protein binding site. The binding site pharmacophore could be derived from an analysis of X-ray complexes of known inhibitors bound to the protein, or from computational tools such as GRID [61], MCSS [62] or Superstar [63].

Another source for pharmacophoric screening is not based directly on the 3D structure of the target but capitalizes on in-house knowledge of the chemical motifs that are useful for a specific target class, such as kinases or metalloenzymes. These motifs can be encoded in SMARTS or SLNs and used to screen the initial database for a subset of compounds to be docked. An example could be to include

Table 2. Recent successes of structure-based virtual screening approaches

Target	Method	Ref.
Carbonic anhydrase II	FlexX	[64]
Ptp1b	DOCK	[65]
Estrogen receptor	PRO_LEADS	[66]
Thrombin	PRO_LEADS	[66]
Factor Xa	PRO_LEADS	[66]
Thymidylate synthase	DOCK	[67]
Retinoic acid receptor	ICM	[68]
Farnesyl transferase	EUDOC	[69]
Kinesin	DOCK	[70]
Hypoxanthine-guanine-xanthine phosphoribosyl tansferase	DOCK	[71]
DNA gyrase	LUDI	[72]
HIV-1 RNA transactivation response element	ICM	[73]

compounds that have specific zinc-binding groups when performing a virtual screen on a zinc containing metalloenzyme, which was a strategy successfully used in a screen for inhibitors of carbonic anhydrase II [64].

If an inhibitor or set of inhibitors for the protein target is known, a powerful shape-based tool, ROCS (Openeye Scientific Software), can be used to rapidly reduce the database to compounds that have similar 3D shapes to the known inhibitors. This approach could be combined with the hot-spot analysis described above to produce a pharmacophore-shape based filter for the initial database. Another factor influencing the strategy adopted is the nature and scope of the experimental validation of the output. For example, NMR shift-mapping requires the identification of small candidates of restricted functionality that will potentially bind only weakly (10^{-3} to 10^{-5}M), and consequently the choice of scoring method is important in attempt to minimize false negatives.

Although any of these approaches could be applied, the most appropriate one for the problem in-hand should be chosen. It is often useful to apply many of these approaches to the same problem and then to consider only the subset of the database that satisfied all of the binding-site based filters. All of these tiers of the virtual screening process are aimed at enriching the final database that is docked before post-analysis.

Outlook

In recent years there have been several published successes for structure-based virtual screening and these are listed in Table 2. The technology is maturing into a viable method for the identification of hits and is being considered as an essential part of the armory available for enriching the lead identification phase of the value chain for the pharmaceutical industry. This review emphasizes that virtual screening is more than just a molecular docking exercise but is also dependent upon the care taken in the preparation of the input and the analysis of the output. The main issue that continues to impede the progress of the field concerns the quality of the scoring functions employed. It could be argued that for the purposes of virtual screening, a scoring function does not necessarily have to rank true hits correctly, but it should be able to discriminate true hits from non-binders. Although there has been undoubted success with virtual screening, the false-positive rate remains high. This shortcoming could be a tolerable consequence of being able to reduce the number of compounds that need to be tested from hundreds of thousands to just hundreds. However, in the instance of screening massive virtual libraries of the order of millions or billions of compounds, the current false-positive ratio implies that a much larger number of compounds would need to be tested (10000s) to enable a comfortable margin for error. Irrespective of the wisdom of attempting a virtual screen of millions or billions of compounds (because only a small area of chemical space is relevant for a given target), efforts to improve the quality of the scoring function and/or the techniques used in the post-screening analysis are needed to progress this technology further.

Acknowledgements

I am grateful to Andy Maynard, Dave Timms and Diane Joseph-McCarthy for helpful comments.

References

1 Myers, S. and Baker, A. (2001) Drug discovery – an operating model for a new era. *Nat. Biotechnol.* 19, 727–730
2 Dean, N.M. (2001) Functional genomics and target validation approaches using antisense oligonucleotide technology. *Curr. Opin. Biotechnol.* 12, 622–625
3 Hertzberg, R.P. and Pope, A.J. (2000) High-throughput screening: new technology for the 21st century. *Curr. Opin. Chem. Biol.* 4, 445–451
4 Otto, S. *et al.* (2002) Dynamic combinatorial chemistry. *Drug Discov. Today* 7, 117–125
5 Li, A.P. (2001) Screening for human ADME/Tox drug properties in drug discovery. *Drug Discov. Today* 6, 357–366
6 Walters, W.P. and Murcko, M.A. (2002) Prediction of 'drug-likeness'. *Adv. Drug Deliv. Rev.* 54, 255–271
7 Carr, R. and Jhoti, H. (2002) Structure-based screening of low-affinity compounds. *Drug Discov. Today* 7, 522–527
8 Diercks, T.C.M. *et al.* (2001) Applications of NMR in drug discovery. *Curr. Opin. Chem. Biol.* 5, 285–291
9 Walters, W.P. *et al.* (1998) Virtual screening – an overview. *Drug Discov. Today* 3, 160–178
10 Mestres, J. and Knegtel, R.M.A. (2000) Similarity versus docking in 3D virtual screening. *Perspect. Drug Des. Discovery* 20, 191–207
11 Mason, J.S. *et al.* (2001) 3-D pharmacophores in drug discovery. *Curr. Pharm. Des.* 7, 567–597
12 Srinivasan, J. *et al.* (2002) Evaluation of a novel shape-based computational filter for lead evolution: Application to thrombin inhibitors. *J. Med. Chem.* 45, 2494–2500
13 Halperin, I. *et al.* (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins Struct. Funct. Genet.* 47, 409–443
14 Lipinski, C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Discov. Res.* 23, 3–25
15 Palm, K. *et al.* (1996) Correlation of drug absorption with molecular surface properties. *J. Pharm. Sci.* 85, 32–39
16 Sadowski, J. (1997) A hybrid approach for ring flexibility in 3D database searching. *J. Comput.-Aided Mol. Design* 11, 53–60
17 Gasteiger, J. and Marsili, M. (1980) Iterative partial equalization of orbital electronegativity – rapid access to atomic charges. *Tetrahedron* 36, 3219–3288
18 Halgren, T.A. (1996) Merck molecular force field. 2. MMFF94 van der Waals and electrostatic parameters for intermolecular interations. *J. Comput. Chem.* 17, 520–552
19 Hendlich, M. (1998) Databases for protein–ligand complexes. *Acta Crystallogr. D* 54, 1178–1182
20 Hooft, R.W.W. *et al.* (1996) Errors in protein structures. *Nature* 381, 272
21 Kearsley, S.K. *et al.* (1994) Flexibases – a way to enhance the use of molecular docking methods. *J. Comput.-Aided Mol. Design* 8, 565–582
22 Lomber, D.M. and Shoichet, B.K. (1998) Flexible ligand docking using configurational ensembles. *Proteins Struct. Funct. Genet.* 7, 938–950

23 McMartin, C. and Bohacek, R.S. (1997) QXP: powerful, rapid computer algorithms for structure-based drug design. *J. Comput. Aided Mol. Des.* 11, 333–344

24 Liu, M. and Wang, S. (1999) MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. *J. Comput.-Aided Mol. Design* 13, 435–451

25 Jones, G. *et al.* (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* 245, 43–53

26 Morris, G.M. *et al.* (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639–1662

27 Abagyan, R.A. *et al.* (1994) ICM – a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* 15, 488–506

28 Rarey, M. *et al.* (1996) A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* 261, 470–489

29 Ewing, T.J.A. *et al.* (2001) DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Design* 15, 411–428

30 Schnecke, V. and Kuhn, L.A. (2000) Virtual screening with solvation and ligand-induced complementarity. *Perspect. Drug Des. Discov.* 20, 171–190

31 Philippopoulos, M. and Lim, C. (1999) Exploring the dynamic information content of a protein NMR structure: Comparison of a molecular dynamics simulation with NMR and X-ray structures of *E. coli* ribonuclease H1. *Proteins Struct. Funct. Genet.* 36, 87–110

32 Knegtel, R.M.A. *et al.* (1997) Molecular docking to ensembles of protein structures. *J. Mol. Biol.* 266, 424–440

33 ClauBen, H. *et al.* (2001) Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* 308, 377–395

34 Osterberg, F. *et al.* (2002) Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins Struct. Funct. Genet.* 46, 34–40

35 Ajay, A. and Murcko, M.A. (1995) Computational methods to predict binding free energy in ligand-receptor complexes. *J. Med. Chem.* 38, 4953–4967

36 Cornell, W.D. *et al.* (1995) A second generation force-field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* 117, 5179–5197

37 MacKerell, A.D. *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616

38 Kollman, P.A. (1993) Free energy calculations: application to chemical and biochemical phenomena. *Chem. Rev.* 93, 2395–2417

39 Bohm, H.J. (1994) The development of a simple empirical scoring function to estimate the binding constant of a protein-ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Design* 8, 243–256

40 Head, R.D. *et al.* (1996) VALIDATE: a new method for the receptor-based prediction of binding affinities of novel ligands. *J. Am. Chem. Soc.* 118, 3959–3969

41 Eldridge, M.D. *et al.* (1997) Empirical scoring functions. 1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Design* 11, 425–445

42 Bohm, H.J. (1992) The computer program LUDI: a new method for the *de novo* design of enzyme inhibitors. *J. Comput.-Aided Mol. Design* 6, 61–78

43 Gehlhaar, D.K. *et al.* (1995) Molecular recognition of the inhibitor AG-1343 by HIV-1 protease – conformationally flexible docking by evolutionary programming. *Chem. Biol.* 2, 317–324

44 Sippl, M. (1995) Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.* 5, 229–235

45 Muegge, I. (2000) A knowledge-based scoring function for protein–ligand interactions: probing the reference state. *Perspect. Drug Des. Discov.* 20, 99–114

46 Mitchell, J.B.O. *et al.* (1999) Bleep – potential of mean force describing protein–ligand interactions: II. Calculation of binding energies and comparison with experimental data. *J. Comp. Chem.* 20, 1177–1185

47 Ishchenko, A.V. and Shakhnovich, E.I. (2002) Small molecule growth 2001 (SMoG2001): An improved knowledge-based scoring function for protein-ligand interactions. *J. Med. Chem.* 45, 2770–2780

48 Gohlke, H. *et al.* (2000) Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* 295, 337–356

49 Pearlman, D.A. and Charifson, P.S. (2001) Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. *J. Med. Chem.* 44, 3417–3423

50 Sotriffer, C.A. *et al.* (2002) Docking into knowledge-based potential fields: a comparitive evaluation of DrugScore. *J. Med. Chem.* 45, 1967–1970

51 Charifson, P.S. *et al.* (1999) Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* 42, 5100–5109

52 Bissantz, C. *et al.* (2000) Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* 43, 4759–4767

53 Stahl, M. and Rarey, M. (2001) Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* 44, 1035–1042

54 Wang, R.X. and Wang, S.M. (2001) How does consensus scoring work for virtual library screening? An idealized computer experiment. *J. Chem. Inf. Comput. Sci.* 41, 1422–1426

55 Dixon, J.S. (1997) Evaluation of the CASP2 docking section. *Proteins Struct. Funct. Genet.* (Suppl. 1), 198–204

56 Honig, B. and Nicholls, A. (1995) Classical electrostatics in biology and chemistry. *Science* 268, 1144–1149

57 Grant, J.A. *et al.* (2001) A smooth permitivity function for Poisson-Boltzman solvation methods. *J. Comp. Chem.* 22, 608–640

58 Shoichet, B.K. *et al.* (1999) Ligand solvation in molecular docking. *Proteins Struct. Funct. Genet.* 34, 4–16

59 Stahl, M. and Bohm, H.J. (1998) Development of filter functions for protein–ligand docking. *J. Mol. Graph. Model.* 16, 121–132

60 Lam, P.Y.S. *et al.* (1994) Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* 263, 380–384

61 Goodford, P.J. (1985) A computational procedure for determining energetically favourable binding sites on biologically important molecules. *J. Med. Chem.* 28, 849–857

62 Miranker, A. and Karplus, M. (1991) Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins Struct. Funct. Genet.* 11, 29–34

63 Verdonk, M.L. *et al.* (1999) Superstar: a knowledge-based approach for identifying interaction sites on proteins. *J. Mol. Biol.* 289, 1093–1108

64 Gruneberg, S. *et al.* (2001) Subnanomolar inhibitors from computer screening: a model study using human carbonic anhydrase II. *Angew. Chem., Int. Ed. Engl.* 40, 389–393

65 Doman, T.N. *et al.* (2002) Molecular docking and high throughput screening for novel inhibitors of protein tyrosine phosphatase 1B. *J. Med. Chem.* 45, 2213–2221

66 Baxter, C.A. *et al.* (2000) New approach to molecular docking and its application to virtual screening of chemical databases. *J. Chem. Inf. Comput. Sci.* 40, 254–262

67 Tondi, D. *et al.* (1999) Structure-based discovery and in-parallel optimization of novel competitive inhibitors of thymidylate synthase. *Chem. Biol.* 6, 319–331

68 Schapira, M. *et al.* (2000) Rational discovery of novel nuclear hormone receptor antagonists. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1008–1013

69 Perola, E. *et al.* (2000) Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *J. Med. Chem.* 43, 401–408

70 Hopkins, S.C. *et al.* (2000) Inhibitors of kinesin activity from structure-based computer screening. *Biochemistry* 39, 2805–2814

71 Aronov, A.M. *et al.* (2000) Rational design of selective submicromolar inhibitors of Tritrichmonas foetus hypoxanthine-guanine-xanthine phosphoribosyltransferase. *Biochemistry* 39, 4684–4691

72 Boehm, H-J. *et al.* (2000) Novel inhibitors of DNA gyrase: 3D structure based biased needle screening. Hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J. Med. Chem.* 43, 2664–2674

73 Filikov, A.V. *et al.* (2000) Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. *J. Comput.-Aided Mol. Design* 14, 593–610