DDT Vol. 7, No. 1 January 2002

Virtual screening and fast automated docking methods

Gisbert Schneider and Hans-Joachim Böhm

Recent advances in high-throughput protein structure determination and in computational chemistry have refocussed attention on virtual screening and fast automated docking methods. This review provides a brief introduction to the basic ideas and outlines computational tools currently used. We also provide several examples of where virtual screening has proved successful, highlighting the usefulness of the approach.

Gisbert Schneider* and Hans-Joachim Böhm
F. Hoffmann-La Roche
Pharmaceuticals Division
CH-4070 Basel
Switzerland
tel: +41 61 68 70696
fax: +41 61 68 87408
*e-mail:
gisbert.schneider@roche.com

▼ Virtual screening (VS) methods contribute to the drug discovery process in various ways [1,2]. Traditionally, computer-based filtering tools have been used to scrutinize individual molecules or whole compound libraries (such as a corporate-compound library), with the aim to eliminate undesired structures (negative selection). The well-known Lipinski's 'rule of 5' can be regarded as an early virtual screening method where all structures are flagged if they violate one or more of the rules. More complex drug-likeness estimators have been developed recently at various levels of sophistication; for example, sets of empirical rules and advanced computer-based decision support systems [3,4]. They can be used for both positive and negative selection, and are increasingly complemented by algorithms that predict physicochemical properties, ability to penetrate the blood–brain-barrier, toxicity flags, synthetic feasibility and cytochrome P450 liability, among others. Such systems are commonly referred to as 'general filters', that is, they can be applied for all drug discovery projects and only require the molecular structure of the candidate compound as input. These VS modules have already proved their usefulness, especially in compound library shaping and database filtering.

The required level of specificity of VS tools and their context-dependence grows with the knowledge available for a particular drug target and the underlying ligand–receptor interaction pattern or pharmacophore. If a 3D receptor structure is available, molecular docking and scoring methods can be applied to perform more target-related, fine-grained compound sieving. In the classical VS cascade, a large virtual library containing up to 10^{12} molecular structures is sequentially filtered and reduced to a small collection of 100–1000 candidates [5]. The advent of parallel and combinatorial synthesis has augmented the conventional VS scenario. In the following sections, we focus on some significant advances that have been made in the field of combinatorial molecular design tools and automated combinatorial docking procedures.

Combinatorial drug design is an optimization task

Rational drug design can be viewed as an optimization in the presence of noise in a highly complex search space with, potentially, many local optima. Robust optimization strategies are required for successful navigation in a chemical space. Three basic questions must be answered to enable a VS cycle of iterative structure generation and testing:

(1) The construction problem: how can we systematically assemble synthetically feasible novel structures?
(2) The docking problem: how does a virtual ligand interact with the receptor?
(3) The scoring problem: how can the quality of a designed structure be estimated?

During the optimization process, molecules are generated that match a given pharmacophore, that is, the spatial arrangement of relevant receptor–ligand interaction points. Typical examples of structure-based design software packages include LUDI [6] (Accelrys, San Diego, CA, USA; http://www.accelrys.com)

1359-6446/02/$ – see front matter PII: S1359-6446(02)02091-8

and BUILDER [7] (Computer Graphics Laboratory, University of California, San Francisco, CA, USA; http://www.cgl.ucsf.edu). These algorithms identify potential ligand–receptor interaction points in the receptor binding pocket and construct novel molecular entities by combinatorial or sequential assembly of atoms and molecular fragments. Several new design algorithms have been developed that rely on different scoring systems (Table 1).

If a high-resolution model of the receptor binding pocket is unavailable, which is still true for the large group of G-protein-coupled receptors (GPCRs) comprising approximately 60% of the current drug targets, then docking and/or scoring is reduced to a similarity searching problem. In this situation, various metrics can be used to rank virtual structures by similarity to an already known active molecule, for example, a patented structure or a compound described in the literature [8]. Similarity searching and docking and/or scoring complement each other and we have found it useful to follow both concepts in parallel whenever possible.

Combinatorial docking has facilitated the construction of synthetically tractable molecules (a problem encountered by many traditional *de novo* design approaches), and has strengthened the idea of fragment-based design [9]. While atom-based techniques build up a molecule atom by atom, fragment-based methods use sets of pre-defined molecular building blocks that are connected by a virtual synthesis scheme. New virtual structures can easily be constructed from combinatorial building blocks, in an ideal situation, directly from synthons. There are two principal fragment-based design approaches: (1) sequential growth and (2) fragment-placing and linking (Fig. 1). The latter inherently reflects the idea of having an anchor fragment ('scaffold') and connecting side-chains to it via a set of linker units that are defined by the synthetic approach. Several successful docking experiments appear to support this idea (Table 2).

Despite their appeal and ease of implementation, fragment-based techniques have some limitations. Most importantly, they tend to produce relatively coarse-grained designs because fine-tuning of structures can be hampered by a limited fragment set, especially during the final optimization cycles. A chemically meaningful selection of fragments for the design process is therefore crucial for success. It can be beneficial to use different sets in the design of, for example, GPCR modulators or kinase inhibitors. An elegant concept of virtual building-block generation is to perform retro-synthetic fragmentation of 'reference libraries', such as a set of GPCR modulators [10,11]. The fragments obtained can then be used to assemble new molecules. It is hoped that such designs will have a greater chance of being GPCR modulator-like than structures that were built from a generic building-block collection.

Table 1. Examples of *de novo* design algorithms

Method	Concept	Refs
BUILDER	Recombination of docked molecules, combinatorial search	[7]
CAVEAT	Database search for fragment fitting	[31]
CONCERTS	Fragment-based, stochastic search	[32]
DLD	Atom-based, structure sampling by simulated annealing	[33]
GENSTAR	Atom-based; grows molecules *in situ* based on an enzyme contact model	[34]
GROUPBUILD	Fragment-based, sequential growth, combinatorial search	[35]
GROW	Peptide design, sequential growth	[36]
GROWMOL	Fragment-based, sequential growth, stochastic search	[37]
HOOK	Linker search for fragments placed by MCSS	[38]
LEGEND	Atom-based, stochastic search	[39]
LUDI	Fragment-based, combinatorial search	[6]
MCDNLG	Atom-based, stochastic search	[40]
MCSS	Fragment-based, stochastic sampling	[41]
MOLMAKER	Graph-theoretical 3D design	[42]
NEWLEAD	Fragment-based, builds on 3D pharmacophore-models	[43]
PRO-LIGAND	Fragment-based search	[44]
PRO-SELECT	Fragment-based, scaffold-linker approach	[45]
SKELGEN	Small-fragment based, Monte-Carlo search	[46]
SME	Peptide design, whole-molecule optimization	[47]
SMOG	Fragment-based, sequential growth, stochastic search	[48]
SPLICE	Recombination of ligands retrieved by a 3D database search	[49]
SPROUT	Fragment-based, sequential growth, combinatorial search	[50]
TOPAS	Fragment-based, evolutionary search	[11]

Structure-based molecular docking

Molecular docking methods take small-molecule structures from a database of existing compounds (or compounds

that could be made), and dock them into the protein binding site. All of the major tools currently available treat the ligand as flexible and, with very few exceptions [12–14], the protein is still treated as rigid.

Several docking programs that use different conceptual strategies have been proposed. A collection of prominent docking algorithms has been compiled by van Leeuwen (http://www.bio.vu.nl/nvtb/Docking.html) that can be used for further reference. The most widely used docking tools include FlexX [15] [Tripos, St Louis, MO, USA; and Fraunhofer-Institute for Algorithms and Scientific Computing, St Augustin, Germany (http://cartan.gmd.de/flexx/)], DOCK [16] (University of California, San Francisco, CA, USA; http://www.cmpharm.ucsf.edu/kuntz/dock.html) and GOLD [17] (Cambridge Crystallographic Data Centre, Cambridge, UK; http://www.ccdc.cam.ac.uk/prods/gold/)

FlexX can be used for protein–ligand docking following a deterministic, incremental construction algorithm. A site-point representation of the binding pocket is used in the docking process, where the ligand is flexible and the protein is treated as rigid. An interaction geometry database is used to describe intermolecular interaction patterns, and a torsion angle database serves as the basis for conformer generation. The basic idea of a FlexX search is to generate a table of all possible interaction sites, and then search the table for matching interaction points by using triangles of query points generated from ligand atoms. A ligand conformation (or 'pose') is stored and scored when a query triangle is successfully matched onto a triangle-of-receptor interaction site (Fig. 2a).

GOLD (Genetic Optimization for Ligand Docking) is a program used to predict how flexible molecules bind to proteins by using a non-deterministic sampling method. The specific features of the program are: (1) a parallel genetic algorithm (GA) for protein–ligand docking, which performs a stochastic search for preferred orientation and conformation of the ligand; (2) full-ligand and partial-protein flexibility; and (3) special energy functions that are derived, in part, from the analysis of conformation and non-bonded contacts observed in crystal structures of small molecules. In contrast to the majority of docking algorithms operating directly on real-value variables, GOLD employs a bit-string ('chromosome') representation of conformations and possible hydrogen-bonding interactions between the ligand and the receptor. The GA provides a search paradigm that enables the identification of good (though not necessarily optimal) solutions. Typically, several docking runs are required to identify most high-affinity binding modes. One advantage of this evolutionary sampling technique is its ability to find solutions in a highly complex search space, as given by flexible-ligand and protein-surface atoms.

The DOCK algorithm, which was conceived by I.D. Kuntz and colleagues (University of California, San Francisco, CA, USA) almost two decades ago, has been widely used ever since and is being continuously developed. DOCK automatically generates many possible orientations and conformations of a putative ligand within a receptor pocket either by an exhaustive search or by fragment docking. The shape of the receptor pocket is described by spheres, and the centers of the spheres are regarded as potential locations for ligand atoms (Fig. 2b). At least four ligand atoms must match individual sphere centers to count as a valid ligand match. Typically, tens of thousands of orientations are generated for each ligand candidate. Receptor–ligand complexes can be scored by accounting for steric fit, chemical complementation or pharmacophore similarity.

A common docking strategy is to dissect the ligand into rigid sub-structure fragments and then start by placing the first fragment (base-fragment) in multiple positions and orientations into the binding site. The site points are used to

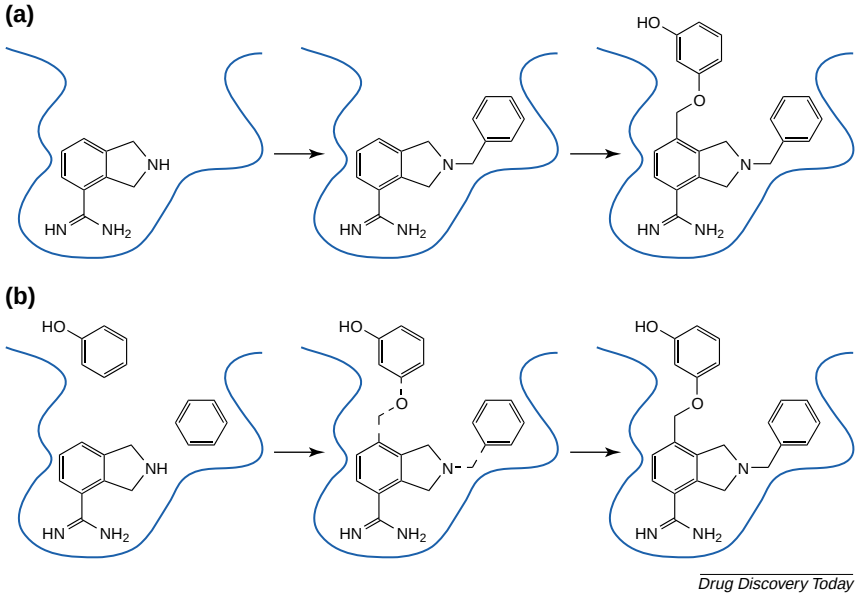
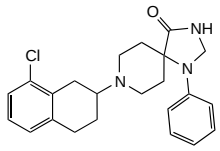
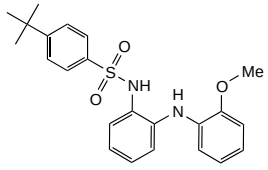
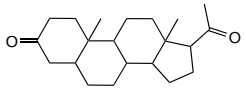
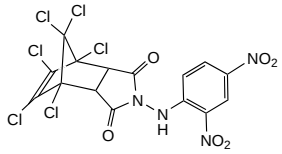
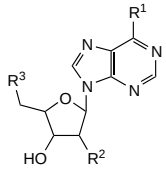
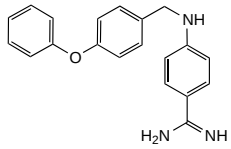
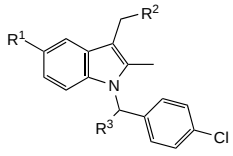
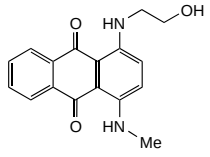


Figure 1. Two strategies for structure-based molecule assembly from fragments. The solid line represents a ligand-binding pocket on the surface of a protein. **(a)** Sequential growth technique and **(b)** fragment-placing and linking. Adapted from Ref. [58].

Table 2. Examples of bioactive molecules which were identified, optimized or designed using virtual screening methods[a]

Molecular structure	Activity	Method	Refs
	Ca^{2+} antagonist (T-channel blocker)	Pharmacophore similarity searching	[51]
	K^+ channel (kv 1.5) blocker	Fragment based evolutionary *de novo* design	[11]
	FKBP ligand	Docking and scoring	[52]
	Farnesyltransferase inhibitor	Docking and scoring	[53]
	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) inhibitor	Combinatorial docking	[54]
	Thrombin inhibitor	Combinatorial docking, *de novo* design	[55]
	Aldose reductase (AR) inhibitors	3D database searching	[56]
	HIV-1 RNA transactivation response element (TAR) inhibitor	Rigid-body docking, database searching	[57]

[a]Additional examples can be found in Ref. [58].

guide the placement of the fragment and to ensure that its position enables the formation of hydrogen bonds and lipophilic contacts. The original ligand is then reassembled within the binding site by step-wise re-attachment of the remaining fragments until the complete molecule is built (Fig. 1). This concept can also be used to generate new molecular structures by probing many different fragments in a combinatorial fashion.

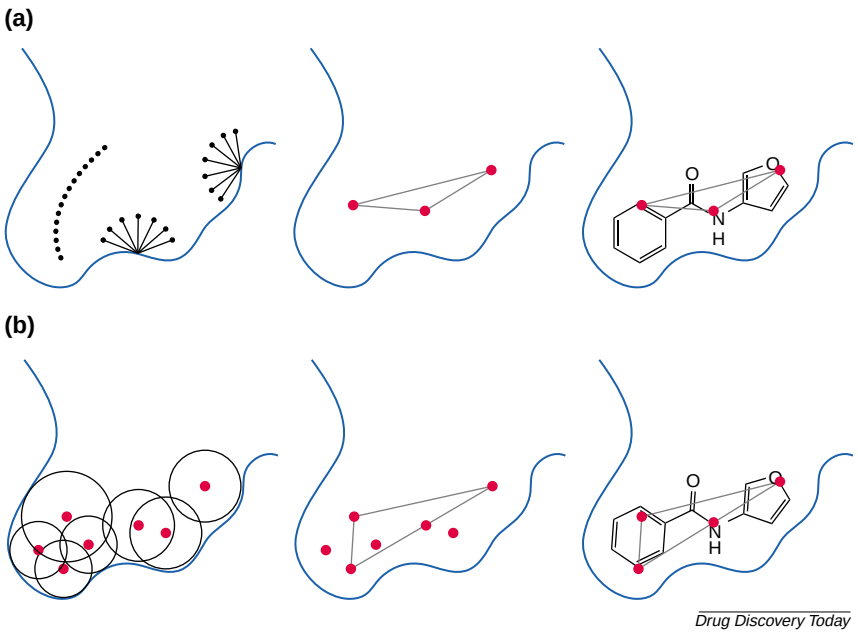


Figure 2. Schematic illustration of **(a)** the FlexX algorithm and **(b)** the DOCK algorithm. FlexX matches triangles of interaction sites onto complementary ligand atoms. The program DOCK fills the binding site with spheres, and sphere centers are then matched to the ligand atoms to determine plausible ligand–receptor complexes. The surface of the receptor pocket is drawn as a solid line. The dotted line represents potential lipophilic interaction points, and the fan-like structures indicate potential hydrogen-bonding sites. Adapted from Ref. [58]

Docking algorithms usually generate multiple docked conformations. They rely strongly on the availability of a robust and accurate scoring function that can be used to identify the correct binding mode and prioritize a large set of docked structures. Given the recent progress in hardware, it could now be feasible to dock an entire compound library of up to 1 million structures into a protein binding site in one central processing unit (CPU) day on a computer cluster with about 100 CPUs.

Scoring functions

Scoring functions are used to estimate the binding affinity of novel structures or an individual molecular fragment in a given position inside the receptor pocket [18–20]. Although fast combinatorial docking procedures have proved applicable to *de novo* design, one of the major problems remaining to be solved is the accurate prediction of binding energies. This problem has been approached in many different ways; for example, by force-field based methods, techniques based on the Poisson–Boltzmann equation, potentials of mean force, free energy perturbation and simple linear approximations. In this context, it is common to differentiate between empirical and knowledge-based scoring functions. The term 'empirical scoring function' emphasizes that these quality functions approximate the binding affinity as a sum of weighted interactions. Most empirical scoring functions are calibrated with a set of experimental binding affinities obtained from known protein–ligand complexes. Such functions usually consider individual contributions from hydrogen bonds, ionic interactions, hydrophobic interactions and binding entropy. Inconsistent calibration data can cause problems in empirical scoring.

Knowledge-based scoring functions have their foundation in the inverse formulation of Boltzmann law. This technique can be used to derive sets of atom-pair potentials (energy functions) by favoring preferred contacts and penalizing repulsive interactions. The various approaches differ in the sets of protein–ligand complexes used to obtain these potentials, the form of the energy function, the definition of protein and ligand atom types, reference states, distance cutoffs, and several other parameters.

Scoring functions provide a very active and rapidly advancing research field. Recent progress has been made by using different scoring functions simultaneously ('consensus scoring'), by the development of receptor-family specific scoring functions, and by the combination of multi-point pharmacophore searching with docking scores reflecting the constraints of the receptor binding site [21–23]. Both historical and more-recent concepts and achievements can be found in the literature [24–26].

A case study: design of novel DNA gyrase inhibitors

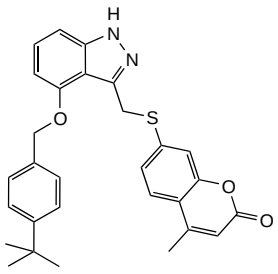
De novo design of inhibitors of the bacterial enzyme DNA gyrase is one example of a successful application of structure-based virtual screening concepts [27]. DNA gyrase is a well-established antibacterial target. It is an essential prokaryotic type II topoisomerase with no mammalian counterpart. DNA gyrase catalyzes the ATP-dependent introduction of negative supercoils into bacterial DNA as well as the decatenation and unknotting of DNA. A drug discovery project was started at Roche (Basel, Switzerland) to overcome the limitations of known DNA gyrase inhibitors.

Searching for novel inhibitors of DNA gyrase by HTS of the Roche Compound Inventory (RCI) yielded no suitable lead structures. Therefore, a rational VS approach was used

to generate lead structures. The idea was to use as a starting point the detailed 3D structural information of the ATP binding site located on gyrase subunit B. At the beginning of the project, x-ray structures were available showing the DNA gyrase subunit B complexed with the substrate analog ADPNP, novobiocin and cyclothialidine. In the inner part of the receptor pocket, these ligands share a common binding motif: each of them donates a hydrogen bond to an aspartic acid side chain (Asp73) and accepts a hydrogen bond from a conserved water molecule. It was thought that a novel inhibitor would require an ability to form these essential two hydrogen bonds. In addition, the molecule should contain a lipophilic part to pick up some lipophilic interactions with the enzyme.

A computational search of the Available Chemicals Directory (ACD; Molecular Design, San Leandro, CA, USA) and the RCI using the programs LUDI and CATALYST (Accelrys) was carried out to identify molecules with a low molecular weight that met the criteria described previously. Relying on the results of the *in silico* screening, only 600 compounds were tested initially. In the next step, close analogs of the first hits were assayed. Overall, 3000 compounds were tested in the biased screen, providing 150 hits that clustered into 14 different chemical classes. Seven of these classes were subsequently validated as true, novel DNA gyrase inhibitors that act by binding to the ATP binding site located on the B subunit. Their activity was in the range of 5–64 μg ml^{-1}, that is, two to three orders-of-magnitude higher than the activity of novobiocin or cyclothialidine.

Subsequent structure-based optimization of the hits led to compounds with potencies equal or up to 10 times better than novobiocin. Figure 3 (MNEC < 0.03 μg ml^{-1}) is an example of a novel potent inhibitor of DNA gyrase B resulting from structure-based VS.



Drug Discovery Today

Figure 3. A novel potent inhibitor of DNA gyrase resulting from structure-based virtual screening.

Conclusion and outlook

VS methods support the decision-making process in drug discovery by the evaluation of large virtual libraries and *in silico* compound filtering. Automated docking procedures have been successfully applied to database screening, *de novo* design and the analysis of binding modes of individual molecules. The most crucial aspect is still the calibration and use of appropriate scoring functions. Different scoring functions work better for different protein–ligand complexes. Two recent publications discuss this problem and begin to provide some insight into the differences [26,28]. It is worth noting that all docking programs and scoring functions have a tendency to generate a significant number of false positives. New developments indicate that progress can be made by combining pharmacophore filtering and docking methods, as well as by using several scoring functions in parallel [28,29] [Good, A.C. *et al.*, Putting the horse before the cart: analysis and optimization of structure-based virtual screening protocols. *220th National Meeting of the American Chemical Society*, 20–24 August 2000, Washington, DC, USA]. The problem of 'frequent hitters' or 'promiscuous binders' is a recurring problem in both HTS and VS campaigns. Such compounds show up as hits in many different VS results, docking results, and biological assays covering a wide range of targets. This happens for two main reasons: (1) the activity of the compound is not specific for the target; or (2) the compound perturbs the assay or detection method. In both cases, such molecules are usually poor starting points for lead optimization programs, which leads to loss of time and money, without any benefit. Occasionally, medicinal chemists are able to identify frequent hitters by obvious undesired structural features or properties. Rationalizing these characteristics and automating the frequent-hitter identification process can increase efficiency and thus assist in the selection of promising hit or lead candidates [30]. Library design by VS and automated combinatorial docking already plays a significant role in current drug discovery projects, and we expect that they will become an indispensable part of future medicinal chemistry.

References

1 Bailey, D. and Brown, D. (2001) High-throughput chemistry and structure-based design: survival of the smartest. *Drug Discov. Today* 6, 57–59
2 Böhm, H-J. and Schneider, G., eds (2000) *Virtual Screening for Bioactive Molecules*, Wiley–VCH Weinheim
3 Clark, D.E. and Pickett, S.D. (2000) Computational methods for the prediction of 'drug-likeness'. *Drug Discov. Today* 5, 49–58
4 Muegge, I. *et al.* (2001) Simple criteria for drug-like chemical matter. *J. Med. Chem.* 44, 1841–1846
5 Walters, W.P. *et al.* (1998) Virtual screening: an overview. *Drug Discov. Today* 3, 160–178
6 Böhm, H-J. (1992) The computer program LUDI: a new method for the *de novo* design of enzyme inhibitors. *J. Comput.-Aided Mol. Des.* 6, 61–78
7 Roe, D.C. and Kuntz, I.D. (1995) BUILDER v.2: improving the chemistry of a *de novo* design strategy. *J. Comput.-Aided Mol. Des.* 9, 269–282
8 Willett, P. *et al.* (1998) Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* 38, 983–996

9 Leach, A.R. *et al.* (2000) Synergy between combinatorial chemistry and *de novo* design. *J. Mol. Graph. Model.* 18, 358–367

10 Lewell, X.Q. *et al.* (1998) RECAP – retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* 38, 511–522

11 Schneider, G. *et al.* (2000) *De novo* design of molecular architectures by evolutionary assembly of drug-derived building blocks. *J. Comput.-Aided Mol. Des.* 14, 487–494

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

13 Schnecke, V. *et al.* (1998) Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. *Proteins* 33, 74–87

14 Claussen, H. *et al.* (2001) Molecular docking into the flexible active site of aldosereductase using FlexE. In *Rational Approaches to Drug Design* (Höltje, H. and Sippl, W., eds), pp. 324–333, Prous Science

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

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

17 Jones, G. *et al.* (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 267, 727–748

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

19 Hirst, J.D. (1998) Predicting ligand binding energies. *Curr. Opin. Drug Des. Dev.* 1, 28–33

20 Böhm, H-J. and Stahl, M. (1999) Rapid empirical scoring functions in virtual screening applications. *Med. Chem. Res.* 9, 445–462

21 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

22 Mason, J.S. and Cheney, D.L. (2000) Library design and virtual screening using multiple 4-point pharmacophore fingerprints. *Pac. Symp. Biocomput. 2000*, pp. 576–587, World Scientific Press

23 Koehler, R.T. and Villar, H.O. (2000) Statistical relationships among docking scores for different protein binding sites. *J. Comput.-Aided Mol. Des.* 14, 23–37

24 Joseph-McCarthy, D. (1999) Computational approaches to structure-based ligand design. *Pharmacol. Ther.* 84, 179–191

25 Gane, P.J. and Dean, P.M. (2000) Recent advances in structure-based rational drug design. *Curr. Opin. Struct. Biol.* 10, 401–404

26 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

27 Böhm, 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

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

29 Thomas, B.E. *et al.* (2000) Pharmacophore-based molecular docking. In *Pharmacophore Perception, Development, and Use in Drug Design* (Guner, O.F., ed.), pp. 351–367, International University Line

30 Roche, O. *et al.* Development of a virtual screening method for identification of 'frequent hitters' in compound libraries. *J. Med. Chem.* (in press)

31 Lauri, G. and Bartlett, P.A. (1994) CAVEAT: a program to facilitate the design of organic molecules. *J. Comput.-Aided Mol. Des.* 8, 51–66

32 Pearlman, D.A. and Murcko, M.A. (1996) CONCERTS: dynamic connection of fragments as an approach to *de novo* ligand design. *J. Med. Chem.* 39, 1651–1663

33 Miranker, A. and Karplus, M. (1995) An automated method for dynamic ligand design. *Proteins* 23, 472–490

34 Rotstein, S.H. and Murcko, M.A. (1993) GenStar: a method for *de novo* drug design. *J. Comput.-Aided Mol. Des.* 7, 23–43

35 Rotstein, S.H. and Murcko, M.A. (1993) GroupBuild: a fragment-based method for *de novo* drug design. *J. Med. Chem.* 36, 1700–1710

36 Moon, J. and Howe, W. (1991) Computer design of bioactive molecules: a method for receptor-based *de novo* ligand design. *Proteins* 11, 314–328

37 Bohacek, R.S. and McMartin, C. (1994) 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

38 Eisen, M.B. *et al.* (1994) HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins* 19, 199–221

39 Nishibata, Y. and Itai, A. (1991) Automatic creation of drug candidate structures based on receptor structure. Starting point for artificial lead generation. *Tetrahedron* 41, 8985–8990

40 Gehlhaar, D.K. *et al.* (1995) *De novo* design of enzyme inhibitors by Monte Carlo ligand generation. *J. Med. Chem.* 38, 466–472

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

42 Clark, D.E. *et al.* (1996) MOLMAKER: *de novo* generation of 3D databases for use in drug design. *J. Chem. Inf. Comput. Sci.* 36, 137–145

43 Tschinke, V. and Cohen, N.C. (1993) The NEWLEAD program: a new method for the design of candidate structures from pharmacophoric hypotheses. *J. Med. Chem.* 36, 3863–3870

44 Clark, D.E. *et al.* (1995) PRO-LIGAND: an approach to *de novo* molecular design. 1. Application to the design of organic molecules. *J. Comput.-Aided Mol. Des.* 9, 13–32

45 Murray, C.W. *et al.* (1997) PRO_SELECT: combining structure-based drug design and combinatorial chemistry for rapid lead discovery. 1. Technology. *J. Comput.-Aided Mol. Des.* 11, 193–207

46 Todorov, N.P. and Dean, P.M. (1997) Evaluation of a method for controlling molecular scaffold diversity in *de novo* ligand design. *J. Comput.-Aided Mol. Des.* 11, 175–192

47 Schneider, G. *et al.* (1998) Peptide design by artificial neural networks and computer-based evolutionary search. *Proc. Natl. Acad. Sci. U. S. A.* 95, 12179–12184

48 DeWitte, R.S. *et al.* (1997) SMoG: *de novo* design method based on simple, fast, and accurate free energy estimates. 2. Case studies in molecular design. *J. Am. Chem. Soc.* 119, 4608–4617

49 Ho, C.M. and Marshall, G.R. (1993) FOUNDATION: a program to retrieve all possible structures containing a user-defined minimum number of matching query elements from three-dimensional databases. *J. Comput.-Aided Mol. Des.* 7, 623–647

50 Gillet, V. *et al.* (1993) SPROUT: a program for structure generation. *J. Comput.-Aided Mol. Des.* 7, 127–153

51 Schneider, G. *et al.* (2000) Virtual screening for bioactive molecules by evolutionary *de novo* design. *Angew. Chem., Int. Ed. Engl.* 39, 4130–4133

52 Burkhard, P. *et al.* (1999) The discovery of steroids and other novel FKBP inhibitors using a molecular docking program. *J. Mol. Biol.* 287, 853–858

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

54 Bressi, J.C. *et al.* (2001) Adenosine analogues as selective inhibitors of glyceraldehyde-3-phosphate dehydrogenase of Trypanosomatidae via structure-based drug design. *J. Med. Chem.* 44, 2080–2093

55 Böhm, H-J. *et al.* (1999) Combinatorial docking and combinatorial chemistry: design of potent non-peptide thrombin inhibitors. *J. Comput.-Aided Mol. Des.* 13, 51–56

56 Iwata, Y. *et al.* (2001) Discovery of novel aldose reductase inhibitors using a protein structure-based approach: 3D-database search followed by design and synthesis. *J. Med. Chem.* 44, 1718–1728

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

58 Stahl, M. (2000) Structure-based library design. In *Virtual Screening for Bioactive Molecules* (Böhm, H-J. and Schneider, G., eds), pp. 229–264, Wiley-VCH