Docking, virtual high throughput screening and *in silico* fragment-based drug design

Vincent Zoete [a], Aurélien Grosdidier [a], Olivier Michielin [a, b, c]

[a] *Swiss Institute of Bioinformatics, Molecular Modeling Group, Bâtiment Génopode, Quartier Sorge, Lausanne, Switzerland*
[b] *Ludwig Institute for Cancer Research, Lausanne Branch, Ch. Des Boveresses, Epalinges, Switzerland*
[c] *Multidisciplinary Oncology Center (CePO), Rue du Bugnon, Lausanne, Switzerland*

Abstract

The drug discovery process has been profoundly changed recently by the adoption of computational methods helping the design of new drug candidates more rapidly and at lower costs. *In silico* drug design consists of a collection of tools helping to make rational decisions at the different steps of the drug discovery process, such as the identification of a biomolecular target of therapeutical interest, the selection or the design of new lead compounds and their modification to obtain better affinities, as well as pharmacokinetic and pharmacodynamic properties. Among the different tools available, a particular emphasis is placed in this review on molecular docking, virtual high-throughput screening and fragment-based ligand design.

Keywords: docking ● virtual high-throughput screening ● fragment-based drug design

Introduction

Drug discovery is an interdisciplinary, complex, time consuming and expensive process. It is widely admitted that the pharmaceutical industry now spends far more on research and development but produces fewer new molecules than 20 years ago. The PriceWaterhouseCoopers Pharma report for 2005 stressed that the pharmaceutical industry needs to find means to improve the efficiency and effectiveness of drug discovery and development. It projected that *in silico* methods will become a dominant tool to address this issue, from drug discovery to marketing. Recently, advances in computational techniques and hardware have enabled *in silico* methods to speed up lead identification and optimization. Up till now, these techniques have contributed to the design of about 50 compounds that entered clinical trials, some of which are now FDA approved [1]. As of today, *in silico* drug design should not be seen as a 'voilà' technique able to suggest directly a small number of compounds with a high affinity and selectivity for the targeted macromolecule, along with favourable pharmacokinetic and pharmacodynamic properties, and using only the three dimensional (3D) structure of the target as a starting point. It rather consists of a systematic use of a wide range of different computational tools aiming, for instance, at improving the knowledge about the target-ligand interactions (molecular docking), increasing the yield of molecules screening by focusing the search on compounds more likely to bind the target (virtual high-throughput screening

*Correspondence to: Prof Olivier Michielin,
Swiss Institute of Bioinformatics, Molecular Modeling Group,
Bâtiment Génopode, Quartier Sorge, Lausanne, Switzerland.

Tel.: +41 21 692 4053
Fax: +41 21 692 4065
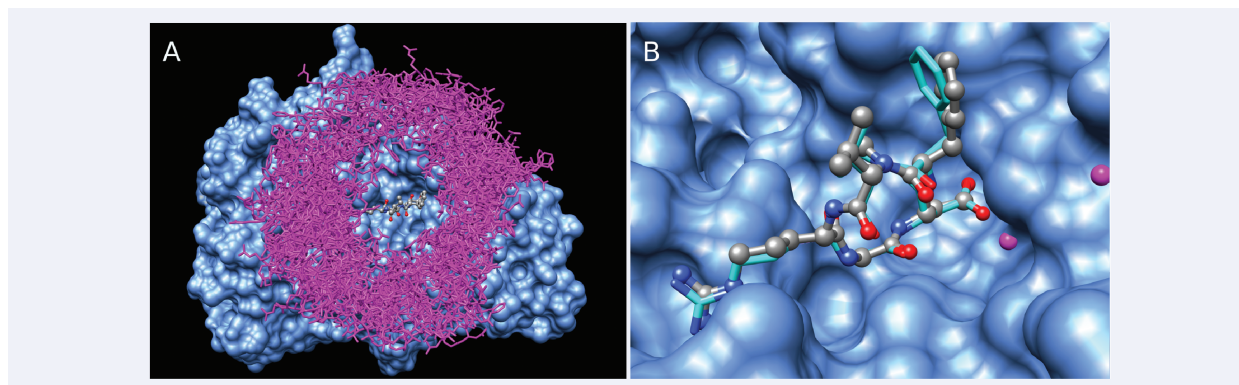E-mail: olivier.michielin@unil.ch

Fig. 1 Example of a successful docking of the Cilengitide molecule on the αVβ3 integrin surface realized with EADock [2]. (**A**) The starting population of the evolutionary process, composed of random yet plausible binding modes, is shown in magenta thick lines. These poses were generated 15 to 25 Å RMSD away from the known native binding mode (in ball and stick) to assess the sampling algorithm. (**B**) The binding mode proposed by EADock at the end of the docking process, in cyan thick lines, is compared to the native binding mode in ball and stick representation. The RMSD between the two poses is 1.2 Å. Importantly, although the native binding mode is known, this information was not used during the docking process, which employed only physical considerations.

[vHTS]) or even suggesting new potential lead compounds (fragment- fragment-based ligand design [FBD]) [1]. Those methods are detailed below.

Docking

Molecular docking tries to predict the native position, orientation and conformation (so-called native pose, or native binding mode) of a small-molecule ligand within the binding site of a targeted macromolecule. By providing the basic understanding of the interactions that are taking place between the ligand and its receptor, docking opens the door to affinity estimation prior to synthesis, as well as to ligand optimization techniques. As an example, Fig. 1 shows the successful docking of the Cilengitide molecule on the αVβ3 integrin surface realized with EADock [2]. Pioneered during the early 1980s [3], docking remains a vigorous research area, and is now among the most useful tools for *in silico* drug design and a primary component in many drug discovery programs [4–8].

Docking can be roughly described as the combination of a search algorithm that intends to suggest several possible ligand poses, and a scoring function aiming at identifying the true (native) binding mode. The number of putative binding modes for a ligand on a protein surface is virtually infinite. Hence, the search algorithm has to be fast and effective in covering the relevant conformational space, including poses very close to the native binding mode. For its part, the scoring function needs to capture the thermodynamics of the ligand–protein interaction adequately to

distinguish the true binding modes, ideally corresponding to the global minimum of the function, from all the others putative ones suggested by the search algorithm. It also has to be fast enough to treat a large number of potential solutions.

Over 30 different docking programs are available today [5]. The most widely used are AutoDock [9, 10], Genetic Optimisation for Ligand Docking (GOLD) [11, 12], FlexX [13]/FlexE [14], DOCK [3, 15] and Internal Coordinate Mechanics (ICM) [16]/ICM-flexible receptor docking algorithm (IFREDA) [17]. Table 1 gives a short description of some representative programs. Docking software differ in the way they handle the protein and ligand flexibility, their sampling algorithm and their scoring function. These aspects are detailed below.

Protein and ligand flexibility

During the physical binding, both the ligand and the protein adapt their conformations to each other. This phenomenon is called the *induced fit*. As a consequence, docking algorithms should handle the flexibility of both molecules. However, taking account of all these degrees of freedom (DOF) leads to a combinatorial explosion of the conformational space making the docking an even more challenging task. Therefore, almost all docking programs perform flexible ligand docking while the receptor is kept rigid. The main exceptions are GOLD, AutoDock, DOCK and EADock, which apply some flexibility to the protein during the docking through active site side chains rotations and more global minimizations, as well as FlexE and the IFREDA, which use a set of different pre-generated receptor conformations obtained experimentally or with *in silico* approaches.

Table 1 Representative docking programs

Program	Ligand flexibility	Protein flexibility	Scoring function
AutoDock 4.0 [9, 10]	EA	Flexible side chains	Force field
GOLD [11, 12]	EA	Protein side chain and backbone flexibility	Empirical score
FlexX [13] / FlexE [14]	Incremental build	Ensemble of protein structure	Empirical score
Dock 6.2[3, 15]	Incremental build	Protein side chain and backbone flexibility	Force field or contact score
Glide [18, 21, 22]	Exhaustive search	-	Empirical score
ICM [16], IFREDA [17]	Pseudo-Brownian sampling and local minimization	Flexible side chains	Force field and Empirical score
QXP [20]	MC	-	Force field
Hammerhead [19]	Incremental build	-	Empirical score
EADock	EA	Flexible side chains and backbone	Force field

MC: Monte Carlo search; EA: Evolutionary algorithm.

Sampling algorithm

Several approaches are used to sample the ligand-binding modes, and in some cases, to treat the flexibility of the protein. These sampling algorithms may be divided into three major categories: systematic search algorithms (FlexX or FlexE, DOCK, Glide [18], Hammerhead [19]), stochastic methods [AutoDock, GOLD, Quick Explore (QXP) [20], EADock] and simulation approaches.

The ideal systematic exploration of all DOF in a molecule to find its native binding mode is usually an impossible task due to the combinatorial explosion of the search space. Therefore, several methods that fall into the category of 'systematic search algorithms' use the technique of the incremental reconstruction of the ligand to compensate for this exponential dependence on the molecular size. There are basically two ways to perform incremental reconstruction. In the first one (FlexX, FlexE), the molecule is divided into a single rigid fragment and several shells of flexible extensions. The rigid fragment, selected for its ability to make the highest number of interactions with the receptor, is docked first. The flexible moieties are then reconnected incrementally. After adding one flexible component, new interactions are searched for in compliance with the torsional database, and the scoring function is used to select the best partial solutions that are used for the next extension step. In the second variant of incremental reconstruction (Hammerhead and original version of DOCK), the molecule is decomposed into various fragments that are docked independently and subsequently fused into the active site using a hinge-bending algorithm. In addition to these reconstruction algorithms, other programs approximate a complete systematic search of the binding modes space of the ligand by narrowing the latter using several filters. For instance, Glide [18, 21, 22] performs an initial rough positioning and scoring phase to narrow the search space, followed by torsionally flexible energy optimization for a few hundred surviving candidate poses. The very best candidates are further refined *via* a Monte Carlo (MC) sampling of pose conformation to improve their accuracy.

In stochastic methods, the ligand is considered as a whole, and step-by-step changes are applied to a starting pose or a population of poses. Such methods subsequently score the new poses at each step trying to enhance the interactions with the protein, leading hopefully to the native binding mode. Evolutionary algorithms (EA) and MC simulations fall into this category.

EA mimic the process of the Darwinian evolution. The starting point is a collection of poses corresponding to plausible ligand-receptor complexes, also called the starting population or seeds. An objective function assigns a score to each binding mode, so that the less likely can be replaced by new ones to form a novel generation. These new poses are generated *via* computational procedures, called operators, that mimic biological mutations and crossovers. A mutation will introduce perturbations in the binding mode, like a rotation of one dihedral angle, while a crossover combines two poses. Operators are applied on the poses selected from the fittest elements of the population, with the hope that even fitter solutions will be generated. The algorithm ends after a given number of generations or energy evaluations, or if it has converged to a solution. The best-known programs in this category are GOLD and AutoDock, but several new promising EA-based algorithms are emerging, like EADock or MolDock [23]. These programs vary in the way they handle poses, in their operators and scoring functions. The reader is referred to relevant papers for a more detailed description of these methods.

MC-based methods start from a single randomly generated pose and apply subsequent random moves, like rotation of one dihedral angle and global translation or rotation of the whole ligand. After each modification, the new pose is scored, and the Metropolis criterion [24] is applied to choose whether the new pose is retained as a starting point for the next modification, or if the algorithm continues from the previous one. The algorithm ends similarly to EA-based approaches. As an example, the QXP [20] program belongs to this category.

Simulation methods group molecular dynamics and minimization methods. These approaches 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 [5]. As a consequence, they are rarely employed as stand-alone search techniques. However, they can efficiently complement other search methods, by refining locally the poses that are suggested by one MC or EA-based step, like in AutoDock, DOCK or EADock.

Scoring functions

The scoring functions typically implemented in protein-ligand docking can be divided into three major categories [5]: knowledge-based, empirical and force-field-based scoring functions. Knowledge-based scoring functions use inter-atomic interaction potentials obtained by a reverse-Boltzmann analysis of the occurrence of different atom–atom pair contacts in known experimental complex structures [25, 26]. Empirical scoring functions are based on the idea that binding free energies can be written as a weighted sum of uncorrelated terms, such as hydrogen bonds, non-polar and aromatic contacts or entropy penalties. The weighting factors of these terms are determined by regression analysis using protein-ligand complexes with known experimental binding free energy and 3D structure [13, 27, 28]. Although easy and fast, these methods suffer from a limited description of the physical aspects of the binding process and from a dependence on the experimental dataset used for their parameterization. On the contrary, the estimation of the binding free energy by force field-based methods use unfitted, universal and physically sound energy functions, such as van der Waals and electrostatic interaction energies, and intramolecular energies [10, 12]. Recently, implicit solvation models have been introduced into docking scores to capture solvent effects upon association [2, 10, 29]. Docking programs generally approximate the exact force-field energy using a grid summation, in which the interaction energy between the protein and an atomic sample is calculated on different regularly spaced points. The binding energy of a ligand is then calculated by summing the contribution of the grid points occupied by the small molecule, taking account of the actual nature and charge of the ligand atoms. EADock is among the very few docking programs that make direct use of a universal and detailed force field such as CHARMM22 and an accurate solvation model such as Generalized Born using Molecular Volume (GB-MV2) [30, 31].

Performance

The performance of docking programs is generally assessed through re-docking calculations. First, hundreds to few thousands of experimentally determined representative ligand–protein complexes are collected, like the Ligand–Protein Database [32], the Astex/Cambridge Crystallographic Data Centre (CCDC) [33] and Astex/Diverse [34] sets or the Mother of All Databases [35]. Ligands are then removed from their binding sites, and the ability of the programs to reproduce the native binding mode is assessed. Generally, a docking is considered successful if the root mean square deviation (RMSD) between the experimental and calculated binding modes is lower than 2 Å. Although it is the current standard, this definition is arguable since it has been shown that two binding modes within 2 Å RMSD can make very different interactions with the protein [36]. Several benchmarks of different docking algorithms are available [37–39], which show that the typical success rate for re-docking ranges from 70% to 80%, depending on the authors and the test sets. It is important to note that these figures overestimate the efficiency of these programs for typical drug design studies. Indeed, the re-docking process neglects the induced-fit issue, because the protein conformer that is used for the docking of a given ligand comes from the experimental structure of the complex and is thus adapted to fit that particular compound. This is not the case when the ligand is taken from a screening database or is designed by *in silico* methods. It has been recently confirmed that docking a ligand to a non-native protein conformer, *i.e.* performing what is called a cross-docking, is a more difficult task in which the success rate of docking programs is reduced by at least 20% [40]. However, progress might be expected from methods developed to handle the protein flexibility in a fast and efficient way. Several analyses have also shown that the performance of most docking software highly depends on the particular characteristics of the binding site and ligand, so that it is hardly possible to figure out *a priori* which method, or combination of search algorithm and scoring function, is the more suited for a particular study [37, 41–43].

Virtual high throughput screening

High throughput screening (HTS) is typically used at an early stage of the drug design process in order to test a large compound collection for potential activity against the chosen target [4–7]. Unfortunately, HTS is time consuming and costly. For this reason, its computational corollary, the vHTS, has become an important tool to precede the large *in vitro* screening assays performed in pharmaceutical companies [44–46]. vHTS aims at using computational tools to estimate *a priori,* from an entire database of existing compounds (or compounds that could be made), those that are the most likely to have some affinity for the target. There are basically two approaches to this topic: ligand- and structure-based vHTS.

Ligand-based vHTS

When the structure of the target is unknown, the measured activities for some known compounds can be used to construct a pharmacophore model. The latter summarizes the positioning of key features like hydrogen-bonding and hydrophobic groups to be matched by putative ligands. Such a model can be used as a template to select the most promising candidates from the library [47, 48]. This strategy can also be used as a filter before applying a structure-based vHTS, so that only 1–10% of the initial database has finally to be docked [46].

Structure-based vHTS

Structure-based vHTS is probably the most straightforward application of docking algorithms. It consists of using a molecular docking program to determine the binding mode on the protein target for an entire database of existing or virtual compounds [44, 46, 49]. The bound conformations are used to approximate the binding free energy or the related affinity of the compound. Then, the most promising compounds are retained for further experimental testing. The most widely used docking programs for vHTS are DOCK, FlexX, Glide, GOLD and AutoDock. The size of the libraries used in such an approach ranges from hundreds of thousands to a few million compounds, limiting the time available for each docking to a few minutes or less. The size of the database is a trade off between the number of molecules that can be treated in a reasonable amount of time, and the chemical space that is desirable to cover. Despite the steady improvement of computer hardware, the conformational sampling is, therefore, very limited and vHTS suffers from a lot of false negatives. Despite the vast amount of resources invested in HTS and vHTS, and several successful studies [50–55], the outcome in terms of new compounds reaching the clinics might be seen as rather disappointing [56, 57].

In silico fragment-based drug design

Since a few years, FBD has become an attractive alternative to experimental or virtual HTS. Contrarily to HTS, where complete molecules are screened for activity, FBD aims at building new ligands piece-by-piece by connecting small and well-chosen compounds that bind into separate binding pockets, close enough to be chemically linked in their relative favourable positions [58]. When tested experimentally, hit molecular fragments exhibit generally only weak affinities, with IC_{50} in the order of 1 mM to 30 μM. However, they provide interesting starting points for follow-up strategies trying to connect several of them to give new efficient lead compounds. Fragment-based design can be performed *in silico* [59] or experimentally using nuclear magnetic resonance (NMR) or X-ray crystallography [60]. This review will focus on *in silico* approaches.

Theoretical advantages of FBD

FBD has several theoretical advantages over vHTS. First, FBD samples a higher chemical diversity than HTS. Indeed, HTS chemical libraries typically contain 10^5–10^6 individual compounds. Although it is a huge effort to handle such an amount of molecules experimentally or even *in silico*, this only covers a tiny amount of the chemical space accessible to the small drug-like molecules. Several studies have estimated this number to be around 10^{60}–10^{100} [45, 61–64], far beyond what can be tested by vHTS. Even the largest possible effort that could be imagined nowadays, using the estimated 120 million compounds available worldwide [65], only scratches the surface of the chemical space. On the contrary, FBD allows sampling of a much larger amount of the chemical diversity using a much smaller number of starting molecules. As an illustration, a chemical space of 10^6 molecules can be obtained by connecting combinatorially three fragments belonging to a 100-fragment database. But, contrarily to HTS, it only requires one virtual or experimental assay per each of the 100 fragments themselves and the few molecules that can be constructed from the most promising ones. Also, it has been calculated that the number of stable and synthetically accessible molecular fragments is around 44×10^6 [66]. This number is nearly of the same order of magnitude of what is tested with HTS, but covers a much vaster part of the chemical space.

Second, FBD leads to higher hit rates. This is illustrated by the fact that the probability of a bad ligand–protein interaction increases exponentially with the size and complexity of the molecule [67]. As a consequence, the probability that small and simple molecules bind to the protein, even with a low affinity, is much higher than for HTS-size compounds. This probability climbs up to 30% to 40% for simple fragments [67]. This supports the use of molecular fragments to anchor the drug design process rather than complex and large molecules.

Finally, FBD leads to molecules with a higher ligand efficiency. HTS chemical libraries are composed of complex molecules originally developed for other purposes than binding to the current target. As a consequence, even a HTS hit is expected to form suboptimal binding interactions with the target. On the contrary, due to its size, a high proportion of the atoms in a fragment hit are directly involved in protein-binding interaction. Their optimization has thus a better probability to lead to more efficient and therefore smaller drugs (Fig. 2), with better chances of favourable pharmacokinetic properties [57].

Interestingly, the binding free energy of a molecule resulting from an optimal linking of two fragments is expected to be lower, thus more favourable, than the sum of the free energies of binding of the two isolated fragments [68] (see Fig. 3). This results from the fact that the rigid body entropic loss upon binding of a molecule is large, whereas the entropic penalty associated with freezing the rotatable bonds is small in some circumstances. The rigid body entropic loss upon binding of one molecule is due to the freezing of 6 DOF: the three rigid translations and three rigid rotations of the small molecule. 12 DOF are frozen upon binding of the two
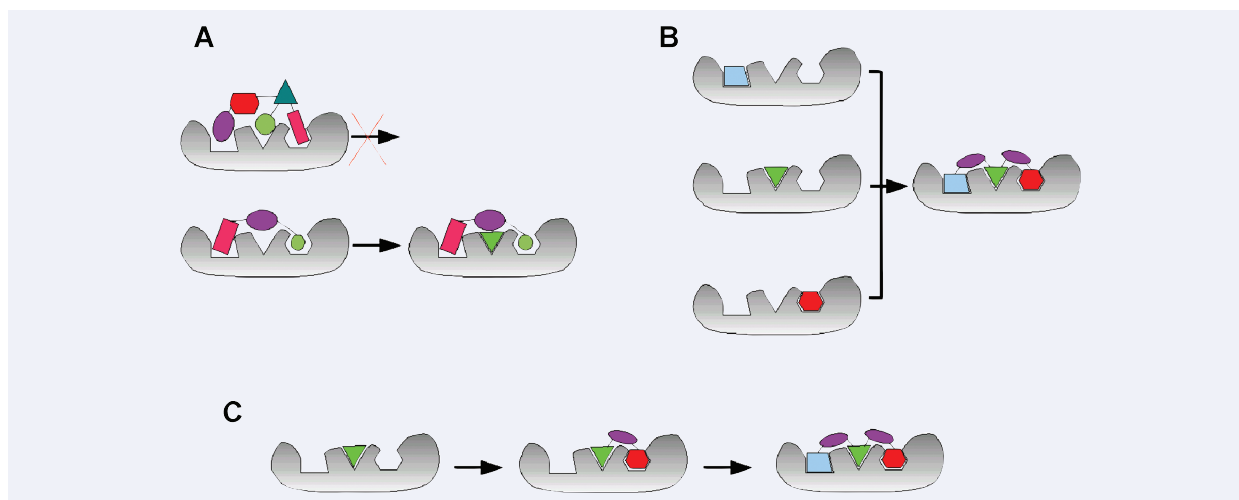
Fig. 2 HTS compared to FBD. (**A**) Typical HTS hits. The compounds found by HTS are complex and full molecules often exhibiting sub-optimal binding interactions with the target. Some can be modulated to increase the binding affinity, but inefficient compounds might still be obtained. (**B**) FBD using a linking approach. Several molecular fragments with weak affinities, each occupying a different yet close key pocket of the binding site, are connected together to provide a large, efficient and high affinity ligand. (**C**) FBD using a growing approach. Starting from a single molecular fragment, the molecule is grown piece-by-piece to give a large, efficient and high affinity ligand.

separated fragments A and B. This leads to a higher entropic penalty than when freezing the 6 DOF of the A:B joined molecule. This favourable difference in rigid body entropy prevails the conformational entropic loss of the A:B molecule, which is due to the freezing of the rotatable bonds that do not exist in the A and B fragments.

Existing FBD methods

The properties of 40 fragment hits identified experimentally against several targets indicated that they show, on average, properties consistent with a 'rule of three' [69], *i.e.* molecular weight < 300 g/mol, number of hydrogen-bond donors ≤ 3, number of hydrogen-bond acceptors ≤ 3, calculated LogP ≤ 3. In addition, it was found that the number of rotatable bonds and the polar surface area were usually lower or equal to 3 and 60 Å^2, respectively. Fragments are usually obtained using a chemoinformatics approach by breaking down biologically active compounds into a limited number of fragments. Depending on the definition of molecular fragments that is used, the chemical space of drug-like molecules reduces to some hundreds [70, 71] to thousands of fragments [72]. Several approaches are available to automatically decompose molecules into rigid fragments [73, 74].

Several methods have been developed for *in silico* FBD (see Table 2), which differ in the building blocks used to construct the ligands (atoms or fragments), the target constraints applied (ligand- or receptor-based), the strategy used to sample the chemical space (depth first [59], breadth first [59], MC, EA), the structural sampling (mainly growing, linking and random structure mutations) and the scoring function used to rank the putative ligands. Among the most

representative methods, one can find LUDI [75], Multicopy Simultaneous Search (MCSS) [76]/HOOK [77], PRO_LIGAND [78], Small Molecule Growth (SMOG) (DeWitte and Shakhnovich), LigBuilder [79], LeapFrog (Tripos Inc., Tripos, St. Louis, MO, USA), CCLD [80] and Genetic Algorithm–based de Novo Design of Inhibitors (GANDI) [81].

In ligand-based FBD, new molecules are designed based on existing ligands. From the latter, different constraints and scoring functions can be derived, like pharmacophore models, molecular similarity or Quantitative Structure Activity Relationship (QSAR) scoring functions. On the contrary, receptor-based FBD uses the 3D structure of the protein binding site to design molecules that are expected to optimize ligand–protein interactions.

Several scoring functions, called the primary constraints, can be used to rank the suggested molecules and drive the search in the chemical space. They correspond mainly to those used by docking programs, *i.e.* force field-based, empirical and knowledge-based scoring functions. In addition, several other physico-chemical parameters related to the drug-likeness of the compounds, as well as terms accounting for molecular and spatial similarity to known ligands, can be used as filters or added to the scoring functions [81, 82]. The latter are called the secondary constraints.

The linking approach (Fig. 2B) starts with the placement of building blocks at key interaction sites of the receptor. This can be done by the fragment-based design software itself, or using a dedicated software like MCSS [76], Solvation Energy for Exhaustive Docking (SEED) [83] or EADock [2]. The latter is particularly suited for the fragment-based approach since, thanks to its cluster-based sampling algorithm and its universally

Table 2 Representative FBD programs, adapted from Schneider *et al.* [59]

Method	Building block	Structure sampling	Target constraints	Search strategy	Scoring function
HSITE [94]	Fr	Fitting and clipping of planar skeleton	Rec	BFS	Steric constraints, HB
Legend [95]	At	Growing	Rec	Random	Force field
LUDI [75, 96]	Fr	Growing, Linking	Rec	BFS	Emp
SPROUT [97]	Fr	Growing, Linking	Rec	DFS, BFS	Emp
MCSS/HOOK [76, 77]	Fr	Linking	Rec	BFS	Force field
DLD [98, 99]	At	Stochastic	Rec	MC	Force field
PRO_LIGAND [78]	Fr	Growing, Linking	Rec, Lig	DFS	Emp
SMoG [100]	Fr	Growing		MC	K-B
BUILDER [101]	At	Lattice	Rec	BFS	Steric constraints
CONCERTS [102]	Fr	MD	Rec	MC	Force field
PRO_SELECT [103, 104]	Fr	Growing	Rec	BFS	Emp
Skelgen [105]	Fr	Stochastic	Rec, Lig	MC	Geom. and chem. constraints
LigBuilder [79]	Fr	Growing, Linking	Rec	EA	Emp
TOPAS [106]	Fr	Stochastic	Lig	EA	Mol Sim
ADAPT [107]	Fr	Stochastic	Rec	EA	Emp using DOCK score
SYNOPSIS [88]	Fr	Stochastic	Rec	EA	Emp
CoG [108]	At, Fr	Stochastic	Lig	EA	Mol Sim
BREED [87]	Fr	Linking	Lig	Ex	No scoring
LEA3D [82]	Fr	Stochastic	Rec	EA	Emp using FLExX score
Gandi [81]	Fr	Linking	Rec	EA	Force field and Sim3D

Fr: fragment; At: atom; MD: Molecular Dynamics; Rec: receptor-based; Lig: ligand-based; BFS: Breadth-first search; DFS: depth-first search; MC: Monte Carlo search; EA: Evolutionary algorithm; Ex: exhaustive enumeration; HB: hydrogen bonds; Emp: empirical scoring function; K-B: knowledge-based scoring function; Mol Sim: molecular similarity; Sim3D: spatial similarity.

applicable scoring function, it is able to both map fragments favourable positions and dock complete molecules [2]. The positioned fragments are then automatically connected to each other using linkers, resulting in several complete molecules that satisfy all key interaction sites. On the contrary, the growing procedure (Fig. 2C) starts from a single fragment located at one of the key interaction site of the target. This fragment can be chosen by the user or by the program. The structure is then grown from this first fragment iteratively, piece-by-piece. Each addition is made so as to yield favourable interactions between the target and the new fragments, while keeping those already shown by the starting molecule. Connection rules are derived from the existence of certain bonds in organic compounds, or from organic synthesis reactions. Both growing and linking strategies have strengths and weaknesses [59]. Growing might run into difficulties if the active site contains several distinct pockets separated by a large gap in which the interactions between a ligand and the protein are limited. When using a linking approach, slightly misplaced fragments or fragments with loosely defined spatial orientation (like a phenyl ring with no preferred orientation in a large lipophilic binding pocket) can lead to the construction of a suboptimal molecule.

We should not expect *ab initio* FBD to yield nanomolar compounds in the first instance. Rather, the methods will probably
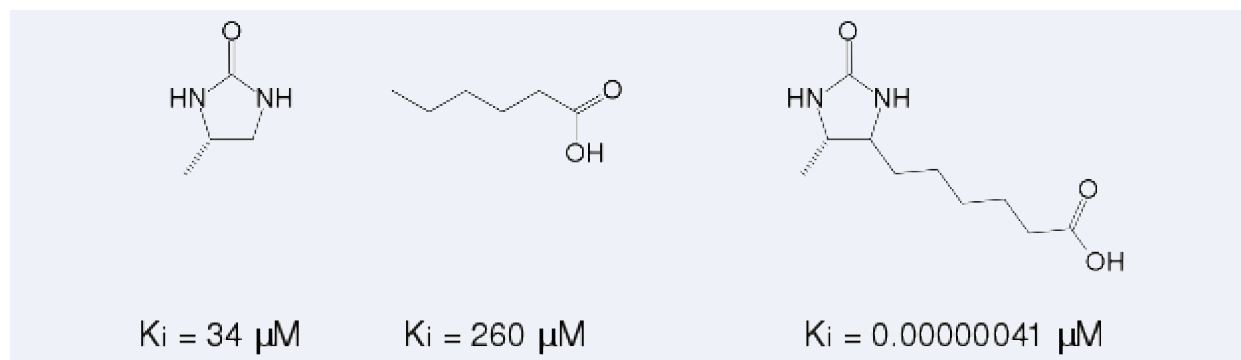
Fig. 3 Influence of fragment linking on the experimental affinity in a FBD study targeting avidin [109].

design new perspective lead compounds of medium affinity, which will be the starting point of further optimization [59]. However, FBD techniques already contributed to generate an impressive number of high affinity ligands [84–90] and drug leads for clinical trials, although they were only recently adopted in the drug discovery pipeline. FBD represents a very promising technique to address tomorrow's challenges of drug discovery.

Synthetic accessibility of molecules proposed by FBD

One critical aspect of *in silico* FBD is the synthetic accessibility of the proposed compounds. Obviously, experimental HTS hits are known to be synthesizable, since they have already been synthesized to be present in the tested molecules collection. It can also be expected that their derivatives are accessible using an approach similar to that used for the parental compound. On the contrary, all molecules assembled on the screen of a computer using *in silico* FBD are not insured to be easily synthesized. However, several strategies can be designed to optimize this aspect. First, drug design studies often aim at deriving new elements of a known class of drugs (the so-called '*me too*' approach). In this case, the synthetic issue might be limited thanks to the knowledge already available for such families of molecules. Second, the fragments that are used *in silico* can be selected to involve organic reactions that are in the core competence of the in house pharmacochemist, or a set of other virtual organic reaction schemes, like in the Retrosynthetic Combinatorial Analysis Procedure (RECAP) [91] or Synthesize and Optimize System in Silico (SYNOPSIS) [88] approaches. Once a few fragments have been successfully assembled in the active site, another option is to screen databases like Zinc (http://zinc.docking.org/) for compounds containing this motif. The results of this search will provide commercially available molecules, thus for which the synthesis has likely been described and optimized. It is also possible to assess the synthetic accessibility of the candidate compounds by an additional

software attempting to define synthetic routes and select potential precursors from databases of available compounds [89, 92]. Similarly, scoring functions have been established recently that try to mimic the intuition of the organic chemist and estimate the synthetic feasibility of molecules by examining their chemical structures, without suggesting any retro-synthesis [93].

Conclusion

A brief outline on the most common types of *in silico* tools has been presented, emphasizing the great progress that *in silico* drug design has made great changes over the past years, making it a valuable and efficient tool for drug discovery.

Despite the numerous successful studies and the very positive picture that is often drawn, the docking problem is far from being solved [5]. Molecular docking still holds several limitations, like the lack of a universally applicable scoring function, able to efficiently combine accuracy and speed. Several directions of improvements are being investigated, like the use of implicit solvent models and entropic terms. In addition, although ligands are commonly handled with full flexibility, the protein flexibility is still only partially considered, at best. Further studies are still necessary to tackle this issue and address the induced-fit problem. Also, the dynamic inclusion of water molecules during the docking process, to take account of eventually important water-mediated hydrogen bond bridges between the ligand and the protein, could increase the efficiency of the approach. As of today, the results of a docking experiment should be taken with care, and be seen as a good starting point for more involved studies [5].

Several studies have illustrated the ability of vHTS to suggest putative lead compounds, and help its experimental counterpart by reducing drastically the number of molecules that will be effectively tested. However, despite the large efforts that have been deployed, the outcome in terms of new compounds reaching the clinical trials might be seen as rather disappointing [56, 57].

Structure-based FBD could also benefit from a better treatment of the flexibility of the target protein and improvement in binding free energy estimation methods. However, automated *de novo* design, and in particular FBD, has already proven its value for hit and lead-structure identification [59]. *In silico* designed molecules can provide the medicinal chemist with rational support to guide his ideas about valuable new chemical entities, and thus help the development of novel and patentable leads.

Acknowledgements

References

1. **Jorgensen WL.** The many roles of computation in drug discovery. *Science*. 2004; 303: 1813–8.
2. **Grosdidier A, Zoete V, Michielin O.** EADock: docking of small molecules into protein active sites with a multiobjective evolutionary optimization. *Proteins*. 2007; 67: 1010–25.
3. **Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE.** A geometric approach to macromolecule-ligand interactions. *J Mol Biol*. 1982; 161: 269–88.
4. **Kontoyianni M, Madhav P, Suchanek E, Seibel W.** Theoretical and practical considerations in virtual screening: a beaten field? *Curr Med Chem*. 2008; 15: 107–16.
5. **Sousa SF, Fernandes PA, Ramos MJo.** Protein-ligand docking: current status and future challenges. *Proteins*. 2006; 65: 15–26.
6. **Kitchen DB, Decornez H, Furr JR, Bajorath J.** Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*. 2004; 3: 935–49.
7. **Brooijmans N, Kuntz ID.** Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct*. 2003; 32: 335–73.
8. **Taylor RD, Jewsbury PJ, Essex JW.** A review of protein-small molecule docking methods. *J Comput Aided Mol Des*. 2002; 16: 151–66.
9. **Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ.** Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*. 1998; 19: 1639–62.
10. **Huey R, Morris GM, Olson AJ, Goodsell DS.** A semiempirical free energy force field with charge-based desolvation. *J Comput Chem*. 2007; 28: 1145–52.
11. **Jones G, Willett P, Glen RC, Leach AR, Taylor R.** Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*. 1997; 267: 727–48.

12. **Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD.** Improved protein-ligand docking using GOLD. *Proteins*. 2003; 52: 609–23.
13. **Rarey M, Kramer B, Lengauer T, Klebe G.** A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*. 1996; 261: 470–89.
14. **Claussen H, Buning C, Rarey M, Lengauer T.** FlexE: efficient molecular docking considering protein structure variations. *J Mol Biol*. 2001; 308: 377–95.
15. **Ewing TJ, Makino S, Skillman AG, Kuntz ID.** DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des*. 2001; 15: 411–28.
16. **Abagyan R, Totrov M, Kuznetsov D.** ICM†-†A new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *Journal of Computational Chemistry*. 1994; 15: 488–506.
17. **Cavasotto CN, Abagyan RA.** Protein flexibility in ligand docking and virtual screening to protein kinases. *J Mol Biol*. 2004; 337: 209–25.
18. **Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS.** Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*. 2004; 47: 1739–49.
19. **Welch W, Ruppert J, Jain AN.** Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem Biol*. 1996; 3: 449–62.
20. **McMartin C, Bohacek RS.** QXP: powerful, rapid computer algorithms for structure-based drug design. *J Comput Aided Mol Des*. 1997; 11: 333–44.
21. **Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL.** Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment fac-

tors in database screening. *J Med Chem*. 2004; 47: 1750–9.
22. **Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA.** Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem*. 2006; 49: 6177–96.
23. **Thomsen R, Christensen MH.** MolDock: a new technique for high-accuracy molecular docking. *J Med Chem*. 2006; 49: 3315–21.
24. **Metropolis N, Rosenbluth AW, Rosenbluth AH, Teller AH, Teller E.** Equations of State Calculations by Fast Computing Machines. *J Chem Phys*. 1953; 21: 1087–92.
25. **Muegge I, Martin YC.** A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J Med Chem*. 1999; 42: 791–804.
26. **Gohlke H, Hendlich M, Klebe G.** Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol*. 2000; 295: 337–56.
27. **Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP.** Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J Comput Aided Mol Des*. 1998; 12: 309–23.
28. **Eldridge MD.** Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des*. 1997; 11: 425–45.
29. **Zou X, Yaxiong S, Kuntz ID.** Inclusion of solvation in ligand binding free energy calculations using the generalized-born model. *J. Am. Chem. Soc*. 1999; 121: 8033–43.
30. **Lee MS, Salsbury FR, Brooks CL.** Novel generalized Born methods. *Journal of Chemical Physics*. 2002; 116: 10606–14.
31. **Lee MS, Feig M, Salsbury FR, Brooks CL.** New analytic approximation to the standard

molecular volume definition and its application to generalized born calculations. *Journal of Computational Chemistry*. 2003; 24: 1348–56.

32. **Roche O, Kiyama R, Brooks CL.** Ligand-protein database: linking protein-ligand complex structures to binding data. *J Med Chem*. 2001; 44: 3592–8.

33. **Nissink JWM, Murray C, Hartshorn M, Verdonk ML, Cole JC, Taylor R.** A new test set for validating predictions of protein-ligand interaction. *Proteins*. 2002; 49: 457–71.

34. **Hartshorn MJ, Verdonk ML, Chessari G, Brewerton SC, Mooij WTM, Mortenson PN, Murray CW.** Diverse, high-quality test set for the validation of protein-ligand docking performance. *J Med Chem*. 2007; 50: 726–41.

35. **Hu L, Benson ML, Smith RD, Lerner MG, Carlson HA.** Binding MOAD (Mother Of All Databases). *Proteins*. 2005; 60: 333–40.

36. **Cole JC, Murray CW, Nissink JWM, Taylor RD, Taylor R.** Comparing protein-ligand docking programs is difficult. *Proteins: Struct Funct Bioinform*. 2005; 60: 325–32.

37. **Kellenberger E, Rodrigo J, Muller P, Rognan D.** Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins*. 2004; 57: 225–42.

38. **Westhead DR, Clark DE, Murray CW.** A comparison of heuristic search algorithms for molecular docking. *J Comput Aided Mol Des*. 1997; 11: 209–28.

39. **Bursulaya BD, Totrov M, Abagyan R, Brooks CL.** Comparative study of several algorithms for flexible ligand docking. *J Comput Aided Mol Des*. 2003; 17: 755–63.

40. **Verdonk ML, Mortenson PN, Hall RJ, Hartshorn MJ, Murray CW.** Protein-ligand docking against non-native protein conformers. *J Chem Inf Model*. 2008.

41. **Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS.** A critical assessment of docking programs and scoring functions. *J Med Chem*. 2006; 49: 5912–31.

42. **Schulz-Gasch T, Stahl M.** Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J Mol Model*. 2003; 9: 47–57.

43. **Wang R, Lu Y, Wang S.** Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem*. 2003; 46: 2287–303.

44. **McInnes C.** Virtual screening strategies in drug discovery. *Curr Opin Chem Biol*. 2007; 11: 494–502.

45. **Schneider G.** Trends in virtual combinatorial library design. *Curr Med Chem*. 2002; 9: 2095–101.

46. **Klebe G.** Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today*. 2006; 11: 580–94.

47. **Kurogi Y, Güner OF.** Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Curr Med Chem*. 2001; 8: 1035–55.

48. **Rarey M, Dixon JS.** Feature trees: a new molecular similarity measure based on tree matching. *J Comput Aided Mol Des*. 1998; 12: 471–90.

49. **Alvarez JC.** High-throughput docking as a source of novel drug leads. *Curr Opin Chem Biol*. 2004; 8: 365–70.

50. **Ghosh S, Nie A, An J, Huang Z.** Structure-based virtual screening of chemical libraries for drug discovery. *Curr Opin Chem Biol*. 2006; 10: 194–202.

51. **Kolb P, Huang D, Dey F, Caflisch A.** Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. *J Med Chem*. 2008; 51: 1179–88.

52. **Scarsi M, Podvinec M, Roth A, Hug H, Kersten S, Albrecht H, Schwede T, Meyer UA, Rücker C.** Sulfonylureas and glinides exhibit peroxisome proliferator-activated receptor gamma activity: a combined virtual screening and biological assay approach. *Mol Pharmacol*. 2007; 71: 398–406.

53. **Kraemer O, Hazemann I, Podjarny AD, Klebe G.** Virtual screening for inhibitors of human aldose reductase. *Proteins*. 2004; 55: 814–23.

54. **Lu IL, Mahindroo N, Liang PH, Peng YH, Kuo CJ, Tsai K-C, Hsieh H-P, Chao Y-S, Wu S-Y.** Structure-based drug design and structural biology study of novel nonpeptide inhibitors of severe acute respiratory syndrome coronavirus main protease. *J Med Chem*. 2006; 49: 5154–61.

55. **Lu Y, Nikolovska-Coleska Z, Fang X, Gao W, Shangary S, Qiu S, Qin D, Wang S.** Discovery of a nanomolar inhibitor of the human murine double minute 2 (MDM2)-p53 interaction through an integrated, virtual database screening strategy. *J Med Chem*. 2006; 49: 3759–62.

56. **Lahana R.** How many leads from HTS? *Drug Discov Today*. 1999; 4: 447–8.

57. **Carr RAE, Congreve M, Murray CW, Rees DC.** Fragment-based lead discovery: leads by design. *Drug Discov Today*. 2005; 10: 987–92.

58. **Erlanson DA, Jahnke W.** Fragment-based approches in drug discovery. In: Jahnke W, Erlanson DA, editors. Wiley-VCH, Weinheim, Germany, 2006.

59. **Schneider G, Fechner U.** Computer-based de novo design of drug-like molecules. *Nat Rev Drug Discov*. 2005; 4: 649–63.

60. **Congreve M, Chessari G, Tisi D, Woodhead AJ.** Recent developments in fragment-based drug discovery. *J Med Chem*. 2008; 51: 3661–80.

61. **Bohacek RS, McMartin C, Guida WC.** The art and practice of structure-based drug design: a molecular modeling perspective. *Med Res Rev*. 1996; 16: 3–50.

62. **Dobson CM.** Chemical space and biology. *Nature*. 2004; 432: 824–8.

63. **Lipinski C, Hopkins A.** Navigating chemical space for biology and medicine. *Nature*. 2004; 432: 855–61.

64. **Walters WP, Stahl MT, Murcko MA.** Virtual screening – an overview. *Drug Discov Today*. 1998; 3: 160–78.

65. **Hann MM, Oprea TI.** Pursuing the leadlikeness concept in pharmaceutical research. *Curr Opin Chem Biol*. 2004; 8: 255–63.

66. **Fink T, Bruggesser H, Reymond J-L.** Virtual exploration of the small-molecule chemical universe below 160 Daltons. *Angew Chem Int Ed Engl*. 2005; 44: 1504–8.

67. **Hann MM, Leach AR, Harper G.** Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci*. 2001; 41: 856–64.

68. **Murray CW, Verdonk ML.** Fragment-based approaches in drug discovery. In: Jahnke W, Erlanson DA, editors. Wiley-VCH, Weinheim, Germany, 2006.

69. **Congreve M, Carr R, Murray C, Jhoti H.** A 'rule of three' for fragment-based lead discovery? *Drug Discov Today*. 2003; 8: 876–77.

70. **Bemis GW, Murcko MA.** The properties of known drugs. 1. Molecular frameworks. *J Med Chem*. 1996; 39: 2887–93.

71. **Bemis GW, Murcko MA.** Properties of known drugs. 2. Side chains. *J Med Chem*. 1999; 42: 5095–9.

72. **Mauser H, Stahl M.** Chemical fragment spaces for *de novo* design. *J Chem Inf Model*. 2007; 47: 318–24.

73. **Vieth M, Siegel MG, Higgs RE, Watson IA, Robertson DH, Savin KA, Durst GL, Hipskind PA.** Characteristic physical properties and structural fragments of marketed oral drugs. *J Med Chem*. 2004; 47: 224–32.

74. **Kolb, Caflisch A.** Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based

high-throughput docking. *J Med Chem.* 2006; 49: 7384–92.

75. **Böhm HJ.** The computer program LUDI: a new method for the *de novo* design of enzyme inhibitors. *J Comput Aided Mol Des.* 1992; 6: 61–78.

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

77. **Eisen MB, Wiley DC, Karplus M, Hubbard RE.** HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins.* 1994; 19: 199–221.

78. **Clark DE, Frenkel D, Levy SA, Li J, Murray CW, Robson B, Waszkowycz B, Westhead DR.** PRO-LIGAND: an approach to de novo molecular design. 1. Application to the design of organic molecules. *J Comput Aided Mol Des.* 1995; 9: 13–32.

79. **Wang R, Gao Y, Lai L.** LigBuilder: A multipurpose program for structure-based drug design. *J Mol Model.* 2000; 6: 498–516.

80. **Caflisch A.** Computational combinatorial ligand design: application to human alpha-thrombin. *J Comput Aided Mol Des.* 1996; 10: 372–96.

81. **Dey F, Caflisch A.** Fragment-based de novo ligand design by multiobjective evolutionary optimization. *J Chem Inf Model.* 2008; 48: 679–90.

82. **Douguet D, Munier-Lehmann H, Labesse G, Pochet S.** LEA3D: a computer-aided ligand design for structure-based drug design. *J Med Chem.* 2005; 48: 2457–68.

83. **Majeux N, Scarsi M, Caflisch A.** Efficient electrostatic solvation model for protein-fragment docking. *Prot-Struct Funct Genet.* 2001; 42: 256–68.

84. **Alig L, Alsenz J, Andjelkovic M, Bendels S, Bénardeau A, Bleicher K, Bourson A, David-Pierson P, Guba W, Hildbrand S, Kube D, Lübbers T, Mayweg AV, Narquizian R, Neidhart W, Nettekoven M, Plancher J-M, Rocha C, Rogers-Evans M, Röver S, Schneider G, Taylor S, Waldmeier P.** Benzodioxoles: novel cannabinoid-1 receptor inverse agonists for the treatment of obesity. *J Med Chem.* 2008; 51: 2115–27.

85. **Firth-Clark S, Willems HMG, Williams A, Harris W.** Generation and selection of novel estrogen receptor ligands using the de novo structure-based design tool, SkelGen. *J Chem Inf Model.* 2006; 46: 642–47.

86. **Jorgensen WL, Ruiz-Caro J, Tirado-Rives J, Basavapathruni A, Anderson KS, Hamilton AD.** Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg Med Chem Lett.* 2006; 16: 663–7.

87. **Pierce AC, Rao G, Bemis GW.** BREED: Generating novel inhibitors through hybridization of known ligands. Application to CDK2, p38, and HIV protease. *J Med Chem.* 2004; 47: 2768–75.

88. **Vinkers HM, de Jonge MR, Daeyaert FFD, Heeres J, Koymans LMH, van Lenthe JH, Lewi PJ, Timmerman H, Aken KV, Janssen PAJ.** Synopsis: synthesize and optimize system in silico. *J Med Chem.* 2003; 46: 2765–73.

89. **Honma T, Hayashi K, Aoyama T, Hashimoto N, Machida T, Fukasawa K, Iwama T, Ikeura C, Ikuta M, Suzuki-Takahashi I, Iwasawa Y, Hayama T, Nishimura S, Morishima H.** Structure-based generation of a new class of potent Cdk4 inhibitors: new de novo design strategy and library design. *J Med Chem.* 2001; 44: 4615–27.

90. **Böhm HJ, Banner DW, Weber L.** Combinatorial docking and combinatorial chemistry: design of potent non-peptide thrombin inhibitors. *J Comput Aided Mol Des.* 1999; 13: 51–6.

91. **Lewell XQ, Judd DB, Watson SP, Hann MM.** 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.* 1998; 38: 511–22.

92. **Gillet V, Myatt G, Zsoldos Z, Johnson A.** SPROUT, HIPPO and CAESA: Tools for de novo structure generation and estimation of synthetic accessibility. *Perspect Drug Discov Design.* 1995; 3: 34–50.

93. **Boda K, SeidelT, Gasteiger J.** Structure and reaction based evaluation of synthetic accessibility. *J Comput Aided Mol Des.* 2007; 21: 311–25.

94. **Danziger DJ, Dean PM.** Automated site-directed drug design: the prediction and observation of ligand point positions at hydrogen-bonding regions on protein surfaces. *Proc R Soc Lond B Biol Sci.* 1989; 236: 115–24.

95. **Nishibata Y, Itai A.** Confirmation of usefulness of a structure construction program based on three-dimensional receptor structure for rational lead generation. *J Med Chem.* 1993; 36: 2921–8.

96. **Böhm HJ.** On the use of LUDI to search the Fine Chemicals Directory for ligands of proteins of known three-dimensional structure. *J Comput Aided Mol Des.* 1994; 8: 623–32.

97. **Gillet V, Johnson AP, Mata P, Sike S, Williams P.** SPROUT: A program for structure generation. *J Comput Aided Mol Des.* 1993; 7: 127–53.

98. **Miranker A, Karplus M.** An automated method for dynamic ligand design. *Proteins.* 1995; 23: 472–90.

99. **Stultz CM, Karplus M.** Dynamic ligand design and combinatorial optimization: designing inhibitors to endothiapepsin. *Proteins.* 2000; 40: 258–89.

100. **DeWitte RS, Shakhnovich EI.** SMoG: De novo design method based on simple, fast, and accurate free energy estimates. 1. Methodology and Supporting Evidence. *J. Am. Chem. Soc.* 1996; 118: 11733–44.

101. **Roe DC, Kuntz ID.** BUILDER v.2: improving the chemistry of a de novo design strategy. *J Comput Aided Mol Des.* 1995; 9: 269–82.

102. **Pearlman DA, Murcko MA.** CONCERTS: dynamic connection of fragments as an approach to de novo ligand design. *J Med Chem.* 1996; 39: 1651–63.

103. **Liebeschuetz JW, Jones SD, Morgan PJ, Murray CW, Rimmer AD, Roscoe JME, Waszkowycz B, Welsh PM, Wylie WA, Young SC, Martin H, Mahler J, Brady L, Wilkinson K.** PRO_SELECT: combining structure-based drug design and array-based chemistry for rapid lead discovery. 2. The development of a series of highly potent and selective factor Xa inhibitors. *J Med Chem.* 2002; 45: 1221–32.

104. **Murray CW, Clark DE, Auton TR, Firth MA, Li J, Sykes RA, Waszkowycz B, Westhead DR, Young SC.** PRO_SELECT: combining structure-based drug design and combinatorial chemistry for rapid lead discovery. 1. Technology. *J Comput Aided Mol Des.* 1997; 11: 193–207.

105. **Todorov N, Dean PM.** Evaluation of a method for controlling molecular scaffold diversity in de novo ligand design. *J Comput Aided Mol Des.* 1997; 11: 175–92.

106. **Schneider G, Lee ML, Stahl M, Schneider PJ.** De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks. *J Comput Aided Mol Des.* 2000; 14: 487–94.

107. **Pegg SC, Haresco JJ, Kuntz ID.** A genetic algorithm for structure-based de novo design. *J Comput Aided Mol Des.* 2001; 15: 911–33.

108. **Brown N, McKay B, Gilardoni F, Gasteiger J.** A graph-based genetic algorithm and its application to the multiobjective evolution of median molecules. *J Chem Inf Comput Sci.* 2004; 44: 1079–87.

109. **Green NM.** Avidin. *Adv Protein Chem.* 1975; 29: 85–133.