PROTEINS: Structure, Function, and Bioinformatics 67:1010–1025 (2007)

EADock: Docking of Small Molecules Into Protein Active Sites With a Multiobjective Evolutionary Optimization

Aurélien Grosdidier,[†] Vincent Zoete,[†] and Olivier Michielin[*]
Swiss Institute of Bioinformatics (SIB), Molecular Modeling Group, Quartier Sorges, Bâtiment Génopode, CH-1015 Lausanne, Switzerland

ABSTRACT **In recent years, protein–ligand docking has become a powerful tool for drug development. Although several approaches suitable for high throughput screening are available, there is a need for methods able to identify binding modes with high accuracy. This accuracy is essential to reliably compute the binding free energy of the ligand. Such methods are needed when the binding mode of lead compounds is not determined experimentally but is needed for structure-based lead optimization. We present here a new docking software, called EADock, that aims at this goal. It uses an hybrid evolutionary algorithm with two fitness functions, in combination with a sophisticated management of the diversity. EADock is interfaced with the CHARMM package for energy calculations and coordinate handling. A validation was carried out on 37 crystallized protein–ligand complexes featuring 11 different proteins. The search space was defined as a sphere of 15 Å around the center of mass of the ligand position in the crystal structure, and on the contrary to other benchmarks, our algorithm was fed with optimized ligand positions up to 10 Å root mean square deviation (RMSD) from the crystal structure, excluding the latter. This validation illustrates the efficiency of our sampling strategy, as correct binding modes, defined by a RMSD to the crystal structure lower than 2 Å, were identified and ranked first for 68% of the complexes. The success rate increases to 78% when considering the five best ranked clusters, and 92% when all clusters present in the last generation are taken into account. Most failures could be explained by the presence of crystal contacts in the experimental structure. Finally, the ability of EADock to accurately predict binding modes on a real application was illustrated by the successful docking of the RGD cyclic pentapeptide on the αVβ3 integrin, starting far away from the binding pocket. Proteins 2007;67:1010–1025.** © 2007 Wiley-Liss, Inc.

Key words: small ligand docking; evolutionary algorithms; rational drug design; EADock

INTRODUCTION

Structures and Drug Design

The number of experimentally resolved protein structures is growing exponentially thanks to huge efforts and improvements in crystallographic techniques. Several of these protein structures are potential targets for the pharmaceutical industry, and the importance of structure-based drug design has thus increased during the past few years. Several computational approaches based on these structures aim at rationalizing experiments by focusing on compounds more likely to have the desired activity, bioavailability, and toxicity. Presently, the most common technique to tackle this problem, called virtual high-throughput screening (VHTS), intends to rank several thousands of small molecules (usually taken from a database) according to a few properties related to the binding to a pharmaceutically relevant target. Complementary to VHTS, *in silico* rational drug design (RDD) suggests structure-based modifications of a lead compound.

The Docking Problem

Both VHTS and RDD rely heavily on the structural prediction of a complex between a ligand and a targeted receptor. Once the energetically most favorable binding mode for several ligands is identified, they can be ranked according to their estimated affinity and/or activity, in order to guide experiments. This step is extremely sensitive to the accuracy of the predicted binding mode. This article focuses on the first part of the question, known as the docking problem, which may be considered as the optimization of structural and energetic criteria described by a scoring function, given a set of degrees of freedom corresponding to the ligand and the receptor conformations and their relative positions.

Existing Approaches

An exhaustive exploration of the search space is not feasible because of its size that grows exponentially with

Grant sponsor: Swiss National Science Foundation; Grant numbers: 3232B0-103172, 3200B0-103173; Grant sponsor: Oncosuisse; Grant number: OCS 01381-08-2003; Grant sponsors: National Center of Competence in Research (NCCR); Swiss Institute of Bioinformatics.

[†]Grosdidier and V. Zoete contributed equally to this work.

*Correspondence to: Olivier Michielin, Swiss Institute of Bioinformatics (SIB), Quartier Sorges, Bâtiment Génopode, CH-1015 Lausanne, Switzerland. E-mail: olivier.michielin@isrec.unil.ch

Received 22 September 2006; Revised 22 November 2006; Accepted 12 December 2006

Published online 22 March 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21367

the number of degrees of freedom of the system. All methods primarily rely on heuristic sampling techniques to generate binding modes. Their ranking is then achieved using a scoring function. Several sampling heuristics are used.[1] They can be classified in three families: approaches derived from systematic search, molecular dynamics simulation techniques, and stochastic methods.

Systematic search methods would be too costly to be applied directly, but several approaches use them in combination with either filtering techniques[2] or with an incremental reconstruction of the ligand. The general principle of the latter is to split the ligand in rigid and flexible fragments: one[3–5] or several[6,7] rigid fragments are docked on the surface of the protein and the ligand is reconstructed.

Standard molecular mechanics simulation techniques (molecular dynamics and minimization) are appealing because of their physical foundations, but are time consuming and not effective at crossing high free energy barriers within accessible simulation time.[8] However, the reduction of van der Waals and electrostatic repulsions was found to improve the sampling by lowering conformational transition energy barriers.[9,10]

Stochastic methods (Monte Carlo, genetic algorithms, and tabu search) are general optimization techniques with a limited physical basis, and are able to explore the search space ignoring energy barriers.

Evolutionary algorithms are iterative stochastic optimization procedures where an initial population of solutions is generated and evaluated with respect to a set of constraints, described by the fitness function. In the docking problem, the fitness function describes the interactions between the ligand and the receptor. This optimization is performed by varying the degrees of freedom related to the ligand and receptor positions, orientations and conformations. During the evolutionary cycle, worst solutions are likely to be replaced by children, created from parents selected among the fittest solutions. This process is repeated until a convergence has been reached in the population, or after a fixed number of generations. Evolutionary algorithms require a balance between diversification and selection, controlling the distribution of solutions in the search space. A high diversity combined with a slow renewal rate of the population leads to a robust and slow algorithm, roughly similar to a Monte-Carlo search. Conversely, a low diversity with a fast solution turnover is likely to cause a premature convergence.[11] This sampling bias, which can be controlled by evolutionary parameters, is a very powerful aspect of evolutionary algorithms, as they can be tuned according to the complexity of the problem to solve. Two limits can be pointed out. First, this biased sampling does not follow a Boltzmann statistic, and thus does not provide direct insights into the thermodynamical properties of a system, such as its binding free energy. Second, evolutionary algorithms extensively use stochastic elements, and consequently, finding the optimal solution is not guaranteed within a finite period of time. More efficient hybrid approaches,[12] in which problem-specific knowledge is used to drive evolution, are now widely used for docking.[13–15] Most of them uses the efficient stochastic search of evolutionary algorithms to cross energy barriers and obtain rough minima, which are subsequently refined by a local search like energy minimization.[13]

Several recent publications[16–18] introduced a two-step approach reducing the complexity of the docking problem. First, putative binding pockets are identified, in which the ligand conformation is optimized. These approaches were reported to be very efficient for VHTS.[19] However, it is not relevant for RDD, which aims at designing a ligand to achieve a high specificity for a predetermined binding region.

In addition to sampling issues, a common bottleneck of docking programs is the scoring function, responsible for driving the search, discriminating and ranking the generated binding modes.[1] The ideal score would be the free energy of the ligand–receptor association.[20,21] Since we are interested in ranking different binding modes of one given ligand on the surface of a given protein, relative free energies of association would be sufficient. Unfortunately, such calculations are currently too computationally demanding. As VHTS usually implies several thousand dockings to enrich a database, there is a need for fast scoring functions and necessary approximations. Conversely, RDD requires a very accurate but computationally expensive scoring to reliably predict the binding mode for a few tenths of complexes.

A scoring function must be both efficient and selective[10]: it must be able to drive the search as smoothly as possible, as well as able to identify the correct binding mode in a set of decoys. Scoring functions can be classified into three families: empirical scoring functions, knowledge-based, and force-field based scoring functions.

Empirical scoring function are expressed as a weighted sum of terms arising from given molecular interactions, such as hydrogen bonds, ionic, and van der Waals interactions.[22,23] The weighting factors are fitted on a database of complexes with known structures and binding free energies. Their transferability to complexes outside the training database is thought to be more limited when compared to force field-based scoring functions.

The second family is based on potential of mean force that are derived from large datasets of experimental 3D structures.[24–28]

The third family of scoring functions is based on molecular mechanics force fields, summing the interaction energy and the internal energies of both partners. If the protein is kept rigid, its internal energy does not change and can be ignored, speeding up the evaluation of a binding mode. These scoring functions are usually sensitive to atomic coordinates, limiting their applications in cross-docking experiments.[29] Softened van der Waals potentials have the advantage of being less sensitive to atomic coordinates in these cases, but also suffer from being less selective.[30] As force field-based scoring functions are not trained on a set of complexes, a good trans-

ferability to in real world applications can be expected. Two docking approaches using the CHARMM[31] package were published previously.[9,15]

Introducing EADock

Most published docking algorithms were designed for VHTS. The resulting time constraint implies that either a very fast (and thus less accurate) scoring scheme is used, or that the sampling is limited around the supposed binding pocket, or both. The docking algorithm for RDD proposed in this article, Evolutionary Algorithm for Docking (EADock) provides a unique combination of four methods, which have been presented separately.[9,15,32–34]

First, the thorough sampling procedure of EADock is inspired by evolutionary algorithms, and uses a combination of two fitness functions. The first one, which neglects solvent effects, is used to drive the search toward local minima because of its efficiency and speed. These minima are then exposed to a more selective and computationally demanding fitness function, which includes the solvation free energy. This approach thus relies on the assumption that minima of the second fitness are also minima of the first, though their rank may be different.[35]

Second, a mechanism inspired by tabu search restricts the search space as the evolution proceeds, by storing a list of previously visited unfavorable docking poses and preventing the search from revisiting these poses, thus facilitating the exploration of new conformational space. This continuous update of the search space also ensures that the evolution does not converge to complexes that do not correspond to a minimum of the second and more selective fitness.

Third, the sampling is performed with operators that combine a broad and a local search of the conformational space. Some of these operators are semi-stochastic, dealing with rotations and translations. Other operators, called "smart operators," aim at crossing energy barriers by transiently modifying the fitness landscape, in a physical and deterministic manner.

Fourth, aside from this flexible sampling framework, coordinates handling and energy calculations are delegated to the CHARMM package, for which a Java API was developed. EADock is thus able to use the latest improvements available in CHARMM, especially sophisticated solvation models such as GB-MV2.[36,37]

The predictive ability of EADock was benchmarked on a previously used set of 37 protein complexes.[38] A successful prediction was defined by a RMSD between the predicted binding mode and the crystal structure, calculated for heavy atoms of the ligand (referred to as RMSD), lower than 2 Å. Despite challenging starting conformations, a huge search space and a very short evolution, such complexes were identified and proposed for 92% of the test complexes, and ranked first for 68% of them. Some failures may be explained by the existence of a bond between the ligand and its receptor that is out of the scope of our scoring schemes. For all remaining failures except for one, a significant interaction was found between the ligand and a neighboring complex of the crystal unit cell.

METHODS

Docking Algorithm

An overview of the algorithm is outlined in Figure 1.

EADock is initialized with parameters relative to the docking (such as reference coordinates for the targeted protein and the ligand, and a list of free dihedral angles) and for the evolutionary process itself (such as the population size and the number of generations). The crystal structure, if any, is not used in any way to introduce a bias or a driving force in the algorithm. A region of interest (ROI) is defined as a sphere. Depending on its position and its radius, this sphere can be focused around the binding site, or encompass the whole protein surface.

Seeding

The first population to be evolved, generation zero, is filled with decoys from the reference coordinates of the ligand. These decoys are referred to as seeds. Each seed is generated by random translation and rotation starting from the reference coordinates, followed by a sequential optimization of every user-defined dihedral angle. Resulting complexes are optimized by a routine called *SmartAttractor*. This procedure translates the ligand to a minimum of the interaction energy, close to the protein surface, along an axis defined by the two closest atoms. This minimum is identified by iteratively translating the ligand toward the protein with a step of 0.1 Å and minimizing its energy.

Starting from the conformations obtained from the translations/rotations, each dihedral angle specified by the user is optimized sequentially by the *OptRot* operator (see later). After this optimization, the ligand is further minimized using 50 steps of steepest descent (SD) followed by 100 steps of Adopted Basis Newton-Raphson (ABNR). Such a search guarantees that energetically unfavorable rotamers are not retained in the first generation, and that the free dihedral angles, bond lengths, and valence angles of the ligand are optimized for each initial binding mode.

Selection

Once a population has been created, two fitness functions are successively applied, providing the only two driving forces in EADock. First, complexes are ranked according to a fast and simple scoring scheme, the *SimpleFitness*. Second, *clusters of complexes* are formed and confronted to an accurate and slower scoring scheme, the *FullFitness*, and the ranks of their centers are updated.

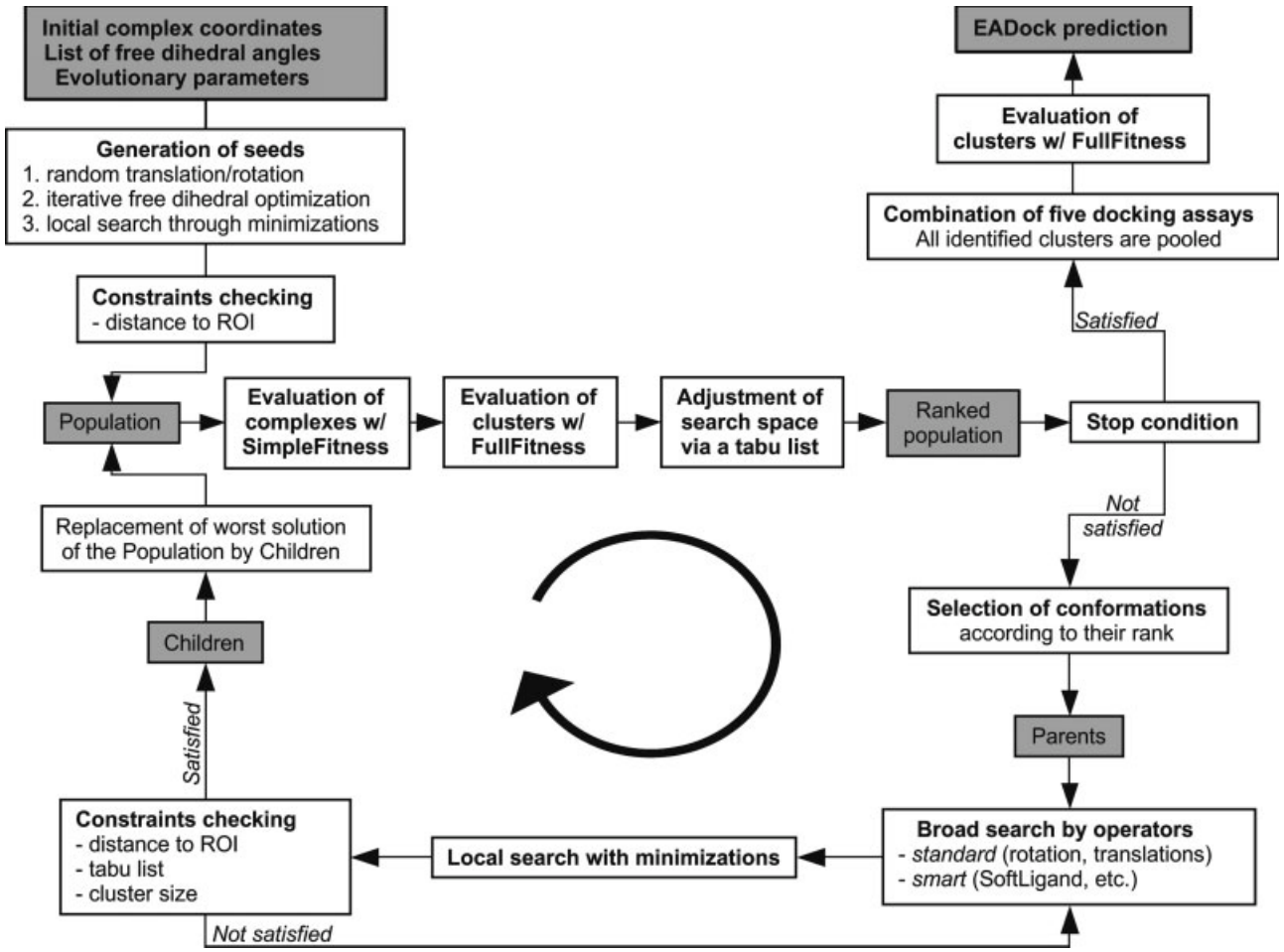


Fig. 1. Main steps of the docking algorithm implemented in EADock. Typical parameters are a population size of 250 complexes, 400 generations, a clustering cutoff of 2 Å, and a maximum cluster size of 8 elements.

The *SimpleFitness* is equal to the total energy of the system calculated with the CHARMM22 molecular mechanics force field, with a dielectric constant of 1 and no cutoff:

$$SimpleFitness = E_{\text{intra}}^{\text{ligand}} + E_{\text{intra}}^{\text{recept}} + E_{\text{inter}}$$

$E_{\text{intra}}^{\text{ligand}}$ and $E_{\text{intra}}^{\text{recept}}$ are the internal energy of the ligand and the receptor, respectively. They are equal to the sum of the internal bonded (bonds, angles, etc.) and nonbonded (electrostatic and van der Waals interactions) terms. When the receptor is fixed, its internal energy, $E_{\text{intra}}^{\text{recept}}$, is constant. E_{inter} is the interaction energy between the ligand and the receptor, and is equal to the sum of the van der Waals and electrostatic interaction energies. The *SimpleFitness* is fast, but neglects the effect of solvent known to have an important contribution to the binding free energy. It is nevertheless likely to focus on reasonable solutions.[35]

The *FullFitness* is subsequently used to evaluate clusters that are identified using the RMSD matrix between all complexes in the population. The most favorably ranked complex is chosen as center for the first cluster. Its neighboring complexes in the population, defined with a RMSD threshold, are assigned to this first cluster. The next most favorably ranked complex is chosen as a the center for the second cluster, and its neighbors are assigned to this second cluster. This procedure continues until all complexes of the population have been assigned to a cluster. When at least three clusters have reached their maximum number of members (typically 8), their *FullFitness* is computed.

The *FullFitness* of a cluster is calculated by averaging the 30% most favorable effective energies of its elements, in order to limit the risk of a few complexes penalizing the whole cluster. This effective energy is written as the sum of the total energy of the system and a solvation term. Neglecting the solute entropic contribution, we can write:

$$G_{\text{eff}} = E_{\text{intra}}^{\text{ligand}} + E_{\text{intra}}^{\text{recept}} + E_{\text{inter}} + \Delta G_{\text{elec.solv}} + \sigma \times \text{SASA}$$

where $E_{\text{intra}}^{\text{ligand}}$, $E_{\text{intra}}^{\text{recept}}$, and E_{inter} are calculated as described above. The solvation energy is composed of the

electrostatic, $\Delta G_{elec,solv}$, and the nonpolar contributions. The latter can be considered as the sum of a cavity term and a solute–solvent van der Waals term, and is assumed to be proportional to the solvent accessible surface area, SASA.[39,40] We use a value of 0.0072 kcal/(mol Å^2) for the parameter σ[41–43] and the SASA was calculated analytically in CHARMM.

$\Delta G_{elec,solv}$ is calculated using the analytical Generalized Born Molecular Volume (GB-MV2) model implemented in CHARMM that is about 20 times faster than solving the Poisson equation. Recent results showed that the deviation between the desolvation energies calculated with the GB-MV2 model and with the Poisson-Boltzmann (PB) model is constant for a series of different conformations of a given complex, which means that the use of GB-MV2 does not alter the ranking of binding modes.[44]

Each fitness function modifies the rank of complexes in the population, and this rank is used to select parents for the next generation; the centers of clusters with the most favorable *FullFitness* are ranked on top of all other complexes in the current population. Conversely, centers corresponding to clusters with a less favorable *FullFitness* are removed from the population and added to the tabu list.

Parent complexes are selected from the top ranked half of the population, according to their internal rank in the cluster to which they belong, then by the rank of this cluster among other clusters. Isolated binding modes are considered as mono-element clusters. The best member of each cluster is selected first, then the second best member of each cluster, and so on. This selection continues with elements of the following ranks until enough parents have been collected. Members of small clusters can be selected several times.

In summary, EADock uses two fitness functions on two different levels: complexes are ranked according to the *SimpleFitness* (fast, efficient) to guarantee reasonable electrostatic and van der Waals interactions. Clusters of complexes, corresponding to binding modes, are then evaluated by the *FullFitness* (slow, selective) taking into account the solvent effect, and the search space is adjusted consequently. Both fitness functions modify the rank of complexes in the population, as this rank is used to select parents for the next generation. After a user defined number of generations, the evolution is stopped.

Evolutionary algorithms require a balance between selection and generation of diversity, the latter being embodied in new complexes created from parent complexes. In EADock, parents are chosen to refine and fill identified clusters, so that these clusters can compete with respect to their *FullFitness* as frequently as possible, limiting the risk of discarding interesting complexes due to the poor selectivity of the *SimpleFitness*.

Diversity

To generate a child, one or two parent complexes are selected according to their rank in the collection described above, and modified by an operator. Operators combine a broad search procedure (see below) followed by a local search through energy minimization. The latter was shown to speed up convergence and improve prediction quality[13] by resolving simple steric problems that might be introduced by the former, as well as adjusting valence angles and bond lengths. Once a child belonging to the search space has been generated and confronted to the tabu list, it is included in the population. If it belongs to a cluster that is already full, it replaces the member with the least favorable *SimpleFitness*. Otherwise, the worst ranked solution of the whole population is replaced. This prevents the premature convergence of the whole population to a single minimum.

Several operators are available to generate new complexes. Four operators optimize the position and orientation of the ligand relative to the protein: two consist of random rotations around a random axis and two of random translations along a random axis. For each kind of movement, two sets of parameters are used, either focusing on a local (rotations up to 40° and translations up to 2.5 Å) or on a long-range exploration (free rotations and translations up to 10 Å). The latter are referred to as long-range operators, the former and all other operators being short-range operators. As long-range operators are likely to deeply modify the ligand pose, the *SmartAttractor* procedure described previously is applied to the newly generated pose to avoid steric clashes or complexes with little or no interaction between the ligand and its receptor.

The *OptRot* operator optimizes the free dihedral angles of the ligand. One of these dihedral angles is randomly chosen and optimized as follows; two groups of atoms are identified, one on each side of the bond defining the rotation axis. The first group of atoms is held fixed while the second is rotated by 60° steps. Scanned poses are minimized using 50 steps of SD followed by 100 steps of ABNR, and assigned a score with the *SimpleFitness*. The conformation of the second group of atoms with the lowest score is kept with a Metropolis-like criterion. The angle scan is then repeated swapping the two groups of atoms in order to rotate the first one, and the rotamer with the most favorable score is retained.

Three operators, *ElectrostaticOptimizer*, *VanDerWaalsOptimizer*, and *SoftLigand*, were designed to modify the ligand-binding mode of the parent thanks to a minimization in a transiently altered force field. To some extend, this can be related to the limitation of the repulsive van der Waals term, which has been described in Refs. 9 and 10 that also smooths the energy landscape. They all follow the same principle: first, the relative contribution of a specific energy term is artificially increased or decreased. Second, an energy minimization is performed in this altered force field. Third, initial contributions of energetic terms are restored, and fourth, an additional energy minimization is performed to relax the ligand in the original force field.

ElectrostaticOptimizer and *VanDerWaalsOptimizer* transiently increase by a factor of five the electrostatic interaction or the van der Waals interaction energies, respectively. Both the sampling and relaxation minimizations follow the same scheme consisting of 50 steps of SD, followed by 100 steps of ABNR. *SoftLigand* transiently decreases the self energy of the ligand by a factor of four. This alteration of the force field allows the ligand to be transiently distorted to cross energy barriers and to improve its interaction with the surface of the protein during 150 steps of ABNR minimization. The relaxation minimization consists of 500 steps of ABNR.

The last operator, *Interpolator*, uses two parent complexes provided that the RMSD between them ranges from 0.2 to 5 Å. A set of interpolated conformations are generated and optimized by the *SmartAttractor* procedure (see Seeding), and the conformation with the lowest interaction energy is retained.

Once a parent has been randomly chosen, an operator is selected based on its probability to be applied, which is increased automatically according to its contribution to the fitness improvement over the last five generations. This procedure, called automatic operator scheduling, has been described previously.[12] In brief, each time a child is created, the operator that was applied is credited with the *SimpleFitness* difference between this child and its parent. Every fifth generations, the probability of operators are adjusted according to this credit. In addition, a bias is introduced depending on the size of the cluster to which the parent belongs. If this cluster has reached its maximum allowed size, long-range operators are more likely to be selected to escape from this identified minimum. Conversely, short-range operators are more likely to be applied if the cluster is almost empty, to refine it and increase its number of members, in order to evaluate it as soon as possible with the *FullFitness*.

Postprocessing

The reliability of each docking experiment is enhanced by combining several (typically five) independent evolutions. Complexes for which the effective energy have been calculated are merged into a single optimized population, which is then reclustered. These new clusters are ranked according to their *FullFitness*, which is calculated as described previously.

A 2 Å RMSD threshold between the top-ranked cluster and the crystal structure defines an successful prediction. If such a conformation was never sampled, a sampling failure is reported. If a successful complex was generated but lost before its evaluation by the *FullFitness*, a *SimpleFitness* failure is reported. If an acceptable cluster was evaluated by the *FullFitness*, but lost afterwards because of its poor score, a *FullFitness* failure is reported.

Dataset

To ensure unbiased benchmarks,[45] complexes used for the validation of EADock were taken from a previous study.[38] They are presented in Table I.

As can be seen, featured ligands are diverse in terms of charge, molecular weight, number of internal degrees of freedom, number of hydrogen-bond donors and acceptors, and logP.

The binding sites can be classified according to their accessibility. Binding sites are considered buried if the fraction of the ligand surface buried upon binding is greater than 95% for all ligands. This is the case for cytochrome, L-arabinose, and intestinal FABP. Nonburied binding sites can be either easily (trypsin, neuraminidase, ribonuclease, carbonic anhydrase), or poorly accessible (ε-thrombin, thermolysin, penicillopepsin, carbocypeptidase), depending on the shape of the binding pocket.

Titrable groups were considered to be in their standard protonation state at neutral pH. The protonation state of histidine residues was defined based on inspection of their environment. The proteins and ions were modeled using the all-atom CHARMM22[46] force field. Missing hydrogens in the crystal structure were added using the HBUILD[47] procedure of CHARMM. Missing parameters for the ligand, for use in conjunction with CHARMM22, were derived from the Merck Molecular Force Field (MMFF)[48–52] by taking the dihedral angle term as is, and the quadratic part of the bond and angle energy terms. The partial charges and van der Waals parameters of the ligand atoms were taken from the MMFF. The ligands were modeled with all hydrogens.

Before starting the docking process, the crystal structures were minimized using 100 steps of SD with the GB-MV2 solvation model. No cutoff was used. This short minimization was used to remove clashes arising from the crystal structure and hydrogen atoms placement without affecting the protein conformation. The RMSD between the starting and final conformations, calculated for all heavy atoms, was always lower than 0.15 Å. The ligand was removed before starting the docking process.

Algorithm Assessment and Benchmark

First, the scoring strategy and the sampling strategy were assessed. Then, the predictive ability of EADock was benchmarked in conditions similar to a real application.

For each test case, a hundred decoys were generated using the seeding procedure described earlier. The *SimpleFitness* and the effective energy of these decoys were calculated, and plotted against the RMSD to the crystal structure. The convergence of the evolutionary process relies on the correspondence between minima of the two fitness functions. Such a correspondence implies that if a complex has been first minimized in the *SimpleFitness* force field, a subsequent minimization in the *FullFitness* force field has only a limited impact on its coordinates. This was confirmed for each test case, using 250 decoys generated with the seeding procedure described above, except that no minimizations were performed (conformations A). These conformations were minimized in the *SimpleFitness* force field (50 steps of SD followed by 100

TABLE I. Complexes Used for the Sampling Assessment and for the Benchmark of the Algorithm

Experiment	Binding site accessibility	Complex	q	DoF	Hb A.	Hb D.	Mass	%B.Sur.
28 test cases:	Accessible	*Carbonic anhydrase*						
algorithm assessment		1cil	1	3	6	2	323.4	85.1
and benchmark		1okl	0	2	4	1	249.3	87.7
		1cnx	0	10	6	3	331.4	74.2
		Neuraminidase						
		1nsc	1	4	9	6	308.3	92.0
		1nsd	1	4	8	5	290.3	92.6
		1nnb	1	4	8	5	290.3	89.7
		Ribonuclease						
		1gsp	0	2	9	3	360.3	80.2
		1rhl	2	3	10	4	361.2	78.1
		1rls	2	3	10	4	361.2	79.2
		Trypsin						
		3ptb	1	1	0	2	121.2	94.6
		1tng	1	1	0	1	114.2	91.6
		1tnj	1	2	0	1	122.2	92.4
		1tnk	1	3	0	1	136.2	91.0
		1tni	1	4	0	1	150.2	85.6
		1tnl	1	1	0	1	134.2	92.7
		1tpp	0	2	3	2	206.2	86.9
		1pph	1	7	3	3	429.6	69.9
	Poorly accessible	*Carbocypeptidase*						
		1cbx	1	3	4	1	207.2	98.2
		3cpa	0	4	4	3	238.2	97.7
		6cpa	1	9	8	2	477.4	82.3
		Penicillopepsin						
		1apt	1	17	6	5	501.7	85.9
		1apu	0	15	6	4	485.7	85.0
		Thermolysin						
		3tmn	0	5	3	3	303.4	73.0
		5tln	1	7	5	3	320.3	79.8
		6tmn	1	11	8	3	471.5	73.2
		ε-Thrombin						
		1etr	0	7	6	4	504.6	87.9
		1ets	1	7	4	4	552.7	88.3
		1ett	1	5	3	3	429.6	88.2
9 test cases: algorithm	Buried	*Cytochrome P-450cam*						
benchmark only		1phf	0	1	1	1	144.2	100.0
		1phg	0	3	3	0	226.3	100.0
		2cpp	0	0	1	0	152.2	100.0
		Intestinal FABP						
		1icm	1	11	2	0	227.4	95.6
		1icn	0	14	2	1	282.5	96.0
		L-Arabinose						
		1abe	0	0	5	4	150.1	100.0
		1abf	0	0	5	4	164.2	100.0
		5abp	0	1	6	5	180.2	100.0

Our approach was tested on 37 complexes. This distinction is based on the accessibility of the binding site. This table lists the PDB code for each complex, as well as the charge of the ligand (q), the number of internal degrees of freedom of the ligand (DoF), the number of hydrogen-bond donors and acceptors (Hb. A. and Hb. D.), the mass of the ligand (Mass) and the percentage of surface of the ligand buried upon complexation. (% B. Sur.).

steps of ABNR) giving conformations B. The latter were further minimized in the *FullFitness* force field (100 steps of SD) giving conformations C. The RMSDs between corresponding poses in B and C were calculated ($RMSD_{BC}$). In view of the hypothesis that the two fitness share the same minima, the distribution of $RMSD_{BC}$ is expected to be close to zero. To provide a reference, conformations A were also minimized directly in the *FullFitness* force field (100 steps of SD) giving conformations D. The RMSDs between corresponding poses in A and D were also calculated ($RMSD_{AD}$). The $RMSD_{BC}$ and $RMSD_{AD}$ distributions were compared, to verify that $RMSD_{BC}$ is closer to zero than $RMSD_{AD}$.

The quality of the sampling of a docking algorithm can be measured by its ability to converge on the crystal structure when starting from remote seeds. Docking

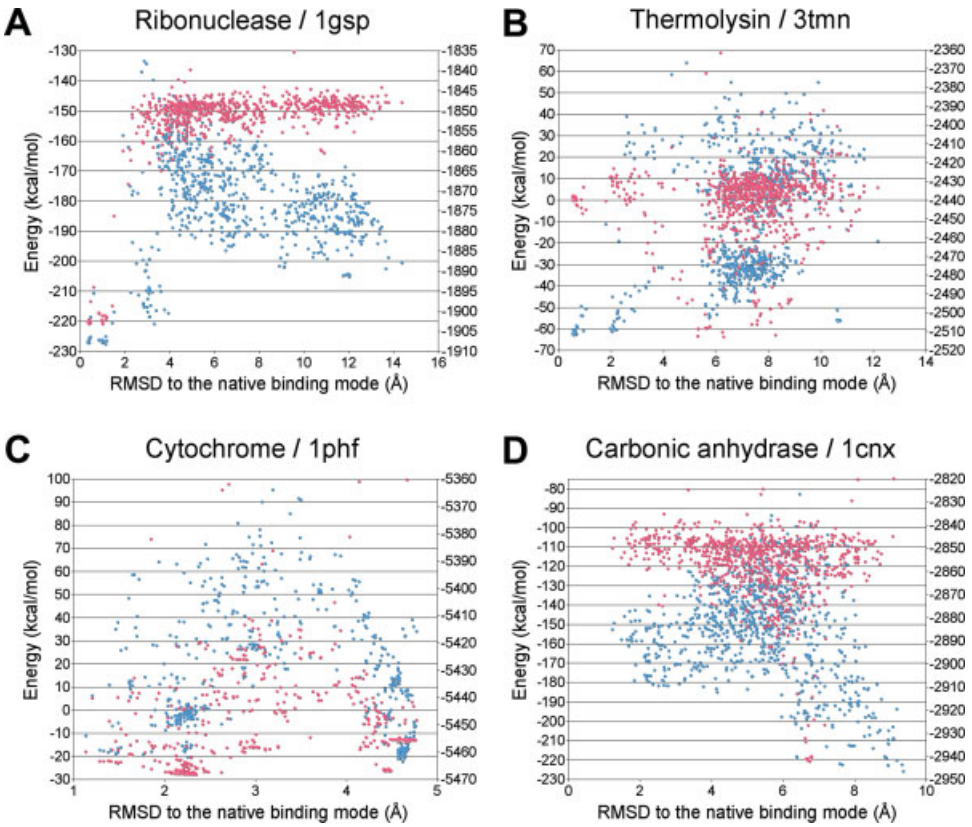


Fig. 2. Correlation between the RMSD and the *SimpleFitness* (pink, left *y*-axis) and the *FullFitness* (blue, right *y*-axis). The four plots correspond to four representative test cases. For each of them, a set of 1000 decoys were generated using the seeding procedure (see Material and Method). **A**: Both fitness are able to identify a cluster very close to the crystal structure. **B**: The *SimpleFitness* fails at ranking the cluster corresponding to the crystal structure correctly, but point it out as a local minimum which is ranked correctly by the *FullFitness*. **C**: A correct cluster was ranked first by the *SimpleFitness*, but not by the *FullFitness*. In this case, a bond exists between the ligand and the receptor. **D**: Both fitness fail at identifying an acceptable cluster. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

assays were thus performed with increasingly challenging seedings for the 28 test cases with an accessible or poorly accessible binding pocket. Test cases with buried binding pockets were excluded; since their binding pockets are a priori unambiguously identified, they would have led to an overestimation of the efficiency of our sampling. A total of 140 docking assays were performed: for each of the 28 test cases, five groups of seeds were generated as described above, with a RMSD to the crystal structure varying between 0–3 Å (easiest), 2–5 Å, 4–7 Å, 6–9 Å, and 8–11 Å (most difficult). For all docking runs, 25 out of the 250 docking poses in the population were renewed at each generation. Children were generated from parent conformations selected out of the top-ranked half of the population. All operators shared the same base probability (0.2), and their maximum adaptive probability was set to 0.1. A clustering distance cutoff of 2 Å was used together with a maximum of eight conformations per cluster. Identified clusters were evaluated by averaging the three most favorable effective energies of their members.

Finally, EADock was benchmarked in real conditions on the same test set: a docking assay was realized for the 37 test cases, using the parameters described above, combined with a less restricted seeding. Seeds were generated between 3 and 10 Å RMSD to the crystal structure for test cases with accessible or poorly accessible binding pockets (seeds too close to the crystal structure were explicitly excluded, see discussion). The ROI was limited to a radial distance of 15 Å around the center of mass of the crystal structure. For buried test cases, seeds were generated between 3 and 5 Å RMSD while the ROI was limited to 5 Å to prevent sampling outside of the binding pocket.

RESULTS

First, the scoring and sampling component of the algorithm were assessed by investigating the relationship between the *SimpleFitness* or the *FullFitness* with the RMSD to the crystal structure using a set of decoys, and by starting the evolution from unfavorably biased seeds.

Second, the predictive ability of the algorithm was benchmarked in realistic conditions, excluding seeds with a RMSD to the crystal structure lower than 3 Å RMSD, in order not to introduce a favorable bias in our results (see Discussion). Both evaluations were performed on the same data set of 37 ligands.[38]

Algorithm Assessment

Four representative examples of the relationships that we observed between the RMSD and both the *SimpleFitness* and *FullFitness* are shown in Figure 2: a successful identification of an acceptable cluster by both fitness (A), a *SimpleFitness* failure (B), a *FullFitness* failure (C), and a failure of both fitness (D). Compared to the *SimpleFitness*, the *FullFitness* is generally noisier, and its driving force is present only near the global minimum.

The correspondence between minima of both fitness functions was assessed by plotting histograms for $RMSD_{BC}$ and $RMSD_{AD}$ (see Material and Methods). A representative example, ribonuclease/1gsp, is shown in Figure 3. As can be seen, the $RMSD_{BC}$ is lower and its distribution is much tighter. This supports the hypothesis that minima of the *SimpleFitness* are also minima of the *FullFitness*. As their ranking may be different, the ROI is dynamically updated to prevent the sampling to be focused around minima of the *SimpleFitness* that are not relevant according to the *FullFitness*.

Each seeding was assessed according to the randomization of the ligand position and to its conformation prior to starting the docking procedure, since both have a significant impact on the results.[45] To estimate the conformational diversity among seeds, each one was fitted to the ligand crystal structure conformation (used as a reference), and the corresponding *fitted* RMSD ($RMSD_{fit}$) was calculated. It is important to note that the RMSD between seeds and the crystal structure reflect a combination of the randomization of the ligand position and conformation, while the $RMSD_{fit}$ between seeds and the crystal structures reflect the randomization of the ligand conformation alone. For the former, the distinction was made between test cases with an accessible or poorly accessible binding site, and test cases with a buried binding site [Fig. 4(A)]. For the latter, a representative example, trypsin/1pph, is shown in Figure 4(B). As can be seen, the native ligand conformation was poorly represented, showing the good randomization of seeds dihedral angles. This thorough randomization makes the docking assay more realistic and difficult.[53]

In our approach, clusters are used to establish the connection between the *SimpleFitness* and the *FullFitness*. An example of the structural variability within clusters is shown in Figure 5(A). All members of the depicted clusters correspond to a well defined binding mode.

During evolution, clusters are evaluated by the *FullFitness*, by averaging over the 30% most favorable effective energies of their members. This implies that for each cluster, the distribution of the effective energies of these members is tight enough for their average to be relevant. The standard deviation of these distributions, σ_{eff}, was measured for all clusters that reached their maximum size [Fig. 5(B)]. For 50, 80, and 90% of the clusters, σ_{eff} is below 1.29, 3.93, and 6.36 kcal/mol, respectively. This indicates that the distribution of the effective energies of elements belonging to a cluster are narrow, and that the *FullFitness* calculated for clusters are relevant.

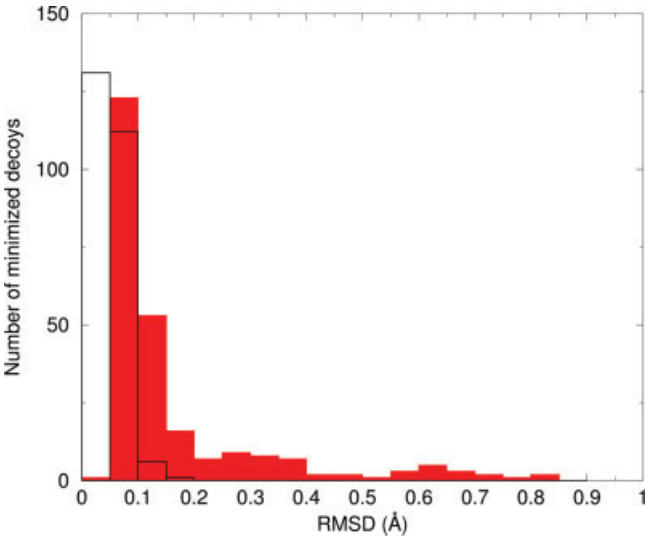
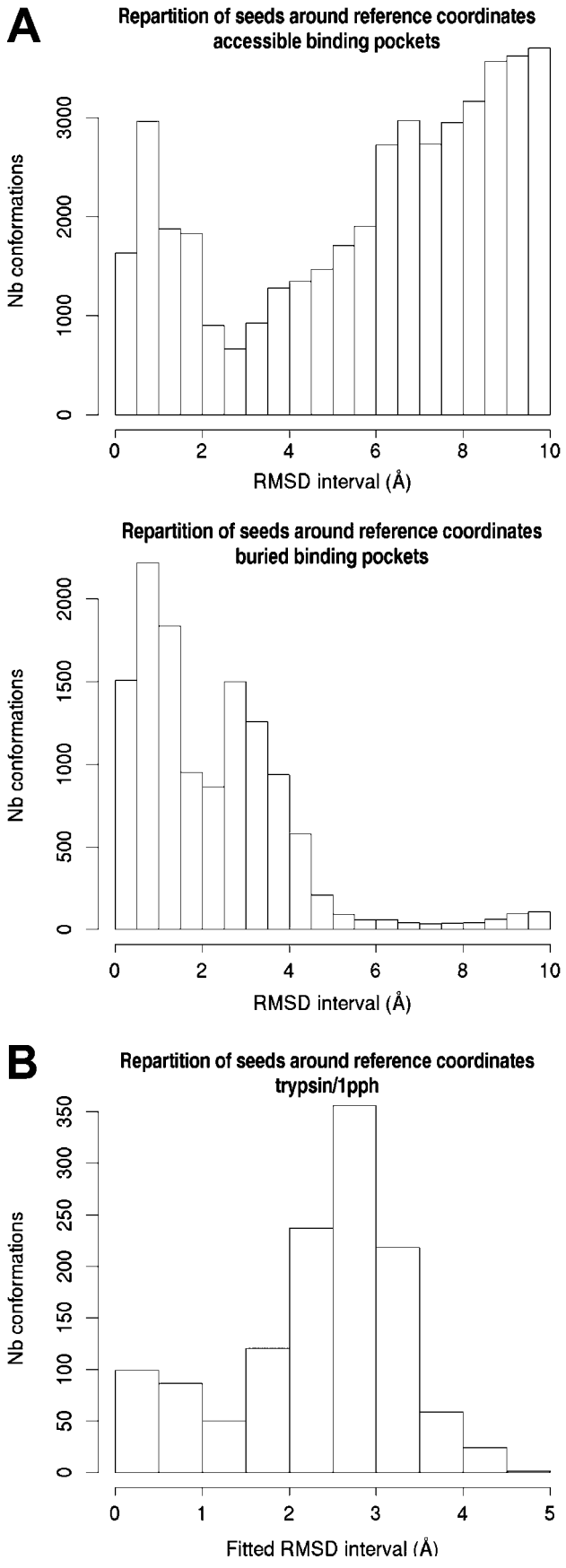


Fig. 3. Histograms of $RMSD_{BC}$ and $RMSD_{AD}$ (see Material and Method) for a representative test case (ribonuclease/1gsp). As can be seen, the distribution of $RMSD_{BC}$ (black lines) is tighter than the distribution of $RMSD_{AD}$ (red bars), and corresponds to much lower RMSD. This supports the hypothesis that the minima of the *SimpleFitness* are also minima of the *FullFitness*. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

The intrinsic ability of "smart operators" to generate low energy conformations was assessed. Because of the automatic operator scheduling policy implemented in EADock, this ability is reflected by the probability of an operator to be applied. As described in Methods, the algorithm can adjust the probability of operators from 0.2 to 0.3 depending on the fitness of the children they produced. These probabilities were averaged over the 140 dockings performed. The adaptive probabilities of the smart operators are larger than those of the stochastic operators, illustrating their competitive advantage over standard operators (see Fig. 6).

To measure the impact of smart operators on convergence, two docking assays were realized with seeds ranging from 8 to 11 Å RMSD to the crystal structure, one with classical operators, the other one with both classical and smart operators. The speed of convergence increases with the use of smart operators, especially during the early generations (see Fig. 7), as a consequence of their better adaptation to the physical nature of the system than classical operators. This allows a more thorough exploration of the search space that could not be achieved by stochastic operators.

To assess the performance of the sampling, the cluster from the optimized population (combined last genera-



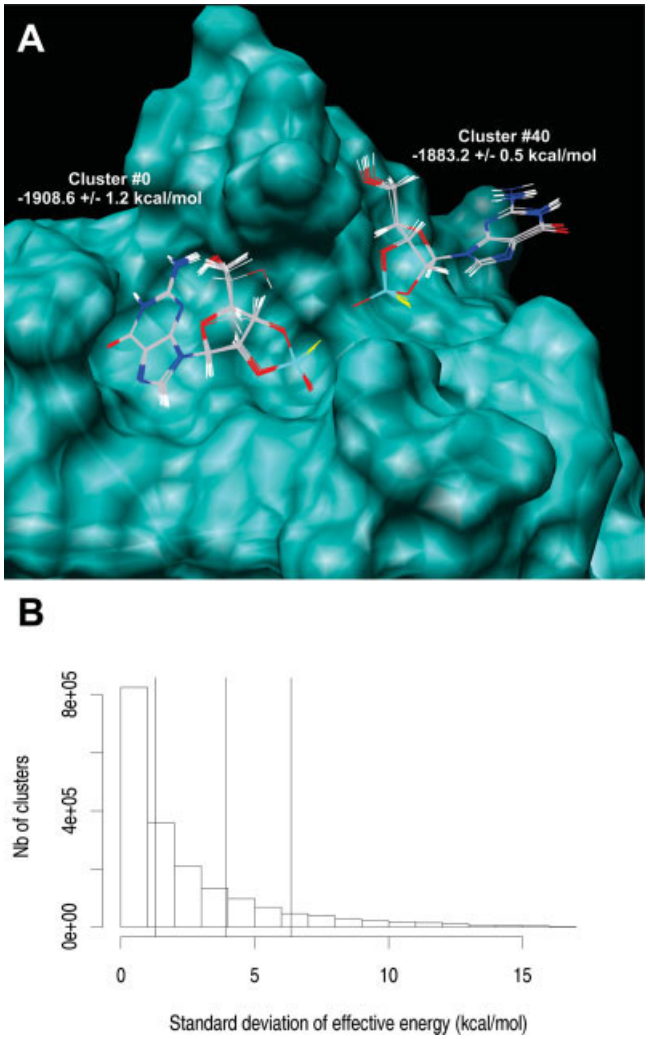


Fig. 5. **A**: variability of coordinates and energies within two top ranked clusters for ribonuclease/1gsp. Molecular graphics images were produced using the UCSF Chimera package.[54] **B**: Distribution of the standard deviation of the effective energy inside the 1.8 million clusters encountered during 140 docking runs. The vertical lines at $\sigma_{eff} = 1.29$ kcal/mol, $\sigma_{eff} = 3.93$ kcal/mol, and $\sigma_{eff} = 6.36$ kcal/mol corresponds to the percentiles 50, 80, and 90, respectively (the 1% highest standard deviations are not represented). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

tions of the five independent runs, see Methods) with the lowest average RMSD to the crystal structure was retained, whatever its rank and *FullFitness*. The success rate remained high regardless of the seed distribution (see Fig. 8). At least one acceptable binding mode was found within the five top ranked clusters for 89% of the test cases, except when seeds are generated between 8 and 11 Å RMSD to the crystal structure (79%). If all

Fig. 4. **A**: Assessment of the distribution of seeding conformations around the crystal structure: histogram of the RMSD between all seeds and the crystal structure for the 28 accessible and the 9 buried test cases. **B**: Fitted RMSD between 1250 seeds and the crystal structure for trypsin/1pph. The reference conformation (RMSD = 0) corresponds to that found in the crystal structure, which is poorly represented.

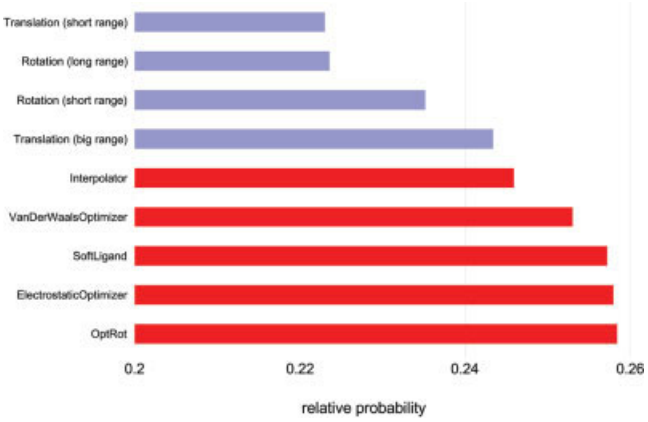


Fig. 6. Relative probabilities of operators averaged over the 140 runs. Smart operators (red) significantly outperform classical operators (blue). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

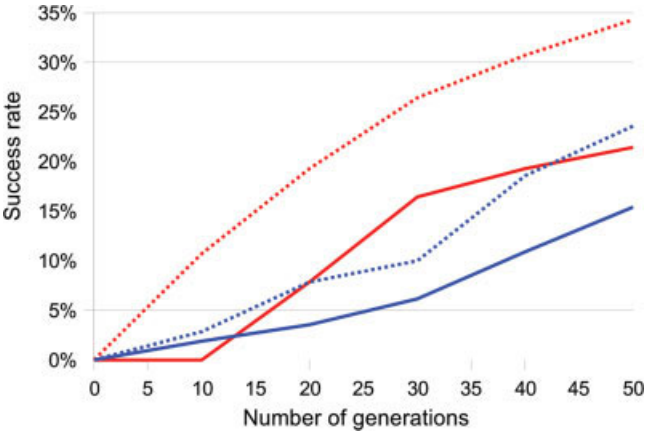


Fig. 7. Success rate considering either the top-ranked cluster (full line) or the five top-ranked clusters (dashed line), with and without smart operators (red and blue curves, respectively), as a function of the number of generations. During the early generations, the success rate is higher with smart operators. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

clusters present in the last generation are considered (between 30 and 60 depending on the test case), this increases up to 100, 96, 93, 93, and 86% when seeds are within 0–3, 2–5, 4–7, 6–9, and 8–11 Å RMSD to the crystal structure, respectively. Noteworthy, RMSD less than 1 Å are reported for 70% of the cases, and up to 90% when the RMSD between seeds and the crystal structure ranges from 0 to 3 Å (data not shown).

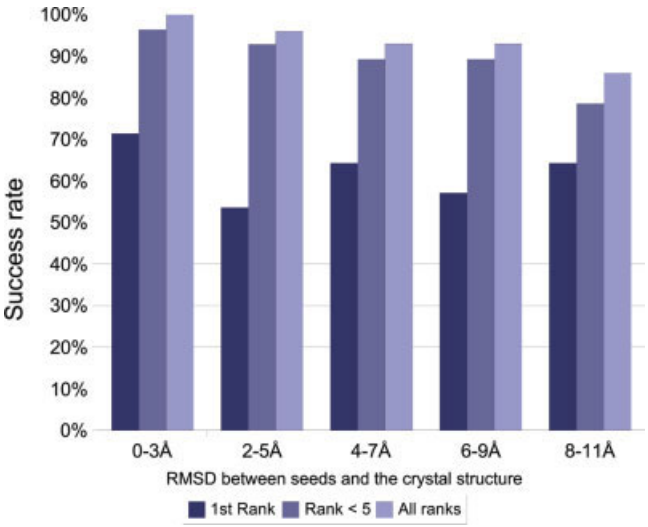


Fig. 8. Impact of the RMSD between the crystal structure and seeds on the docking outcome, considering either the top ranked, the five top ranked, or all clusters in the optimized population (deep, medium, and light blue bars, respectively). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

The algorithm was able to generate at least one conformation within a 2 Å RMSD threshold to the ligand in the crystal structure for 97% of the 140 docking assays. Only four sampling failures were reported for penicillopepsin/1apt and 1apu, thermolysin/6tmn and ε-thrombin/1etr, when starting from the most remote seeds. These four sampling failures reflect a difficulty for the sampling strategy to generate a reasonable conformation inside a poorly accessible binding pocket (see Fig. 9). The remaining failures can be attributed to our scoring, although for most of them, a significant interaction was found between the ligand and a neighboring complex in the crystal (see later).

Benchmarks

Aside from the algorithm assessment presented above, the predictive performance of EADock was benchmarked

Fig. 9. **A**: View of the single sampling failure reported, corresponding to a poorly accessible binding site (penicillopepsin/1apt). **B**: Example of a typical contact between the ligand and a neighboring complex in the crystal (thermolysin/5tln, 5.2 Å), misleading our scoring functions. **C**: Example of the limits of the *FullFitness*: bond between the ligand and the heme of the receptor (cytochrome/1phf), misleading the *FullFitness*. **D**: Example of the limits of the *SimpleFitness*, with protein/ligand interactions mediated by crystal water molecules (thermolysin/3tmn).

Fig. 10. Docking of a RGD cyclic pentapeptide on the $\alpha V\beta 3$ integrin. Seeds (pink sticks) were generated between 15 and 25 Å RMSD to the crystal structure (red and green circles, respectively). The ROI was defined by a 25 Å radius sphere encompassing 65,449 $Å^3$. The binding mode of the ligand found in the crystal is shown in ball and stick (inset). A standard docking was performed, and the top ranked cluster was found to have a binding mode identical to that of the crystal structure, as shown by the three top scoring conformations (green sticks). The average RMSD to the crystal structure for these three conformations is 1.17 Å.

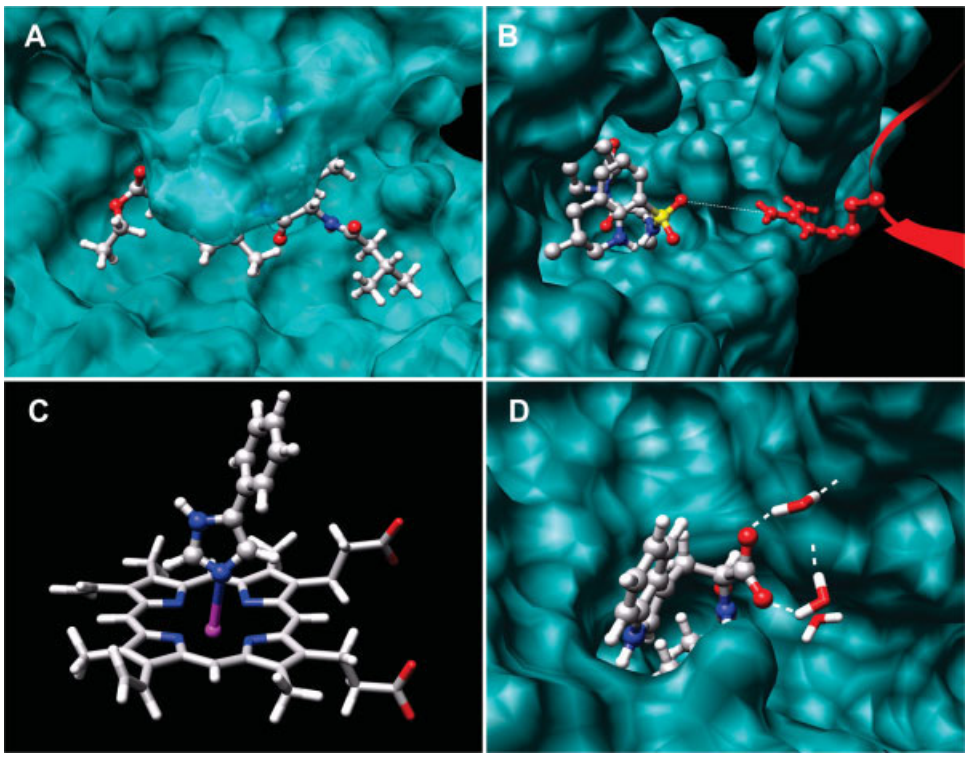


Figure 9.

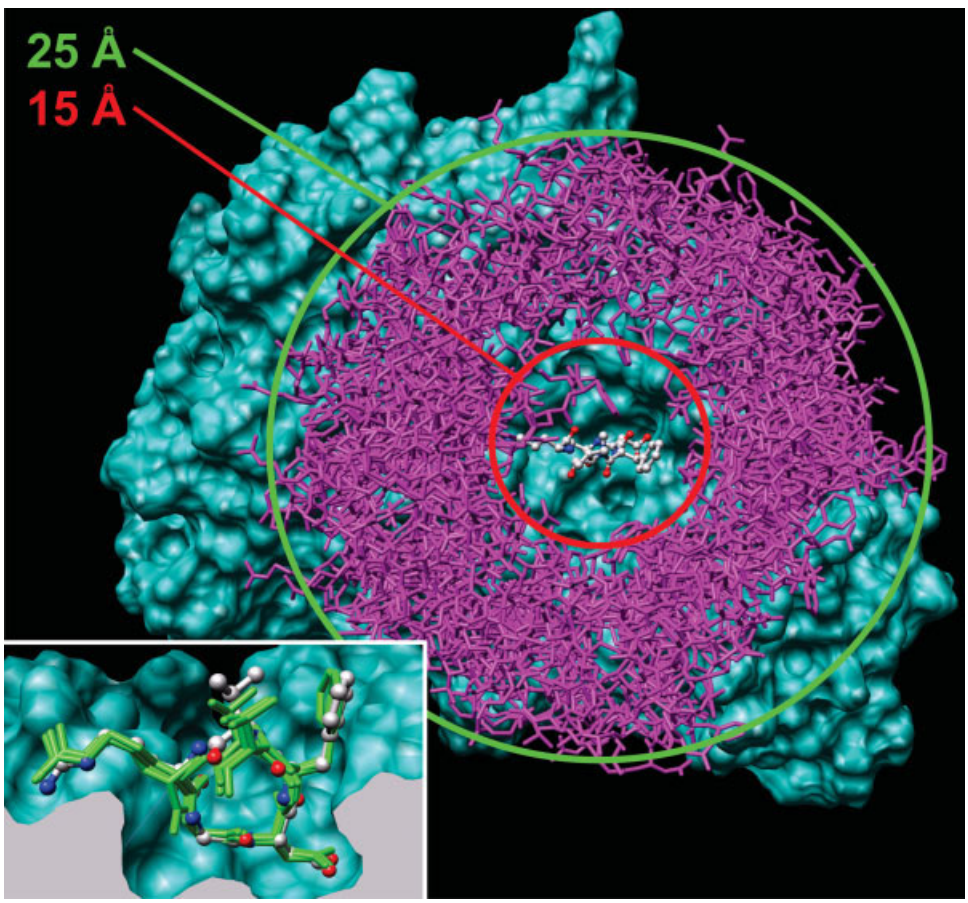


Figure 10.

for the 37 test cases using a realistic seeding (see Material and Methods) ranging from 3 to 10 Å RMSD to the crystal structure. Again, the crystal conformation was excluded from the seeds. All predictions are shown in Table II. A cluster with an average RMSD to the crystal structure lower than 2 Å was ranked first for 68% of the test cases. When considering the five top-ranked solutions or all solutions surviving through the evolution, the success rate increased up to 78 and 92%, respectively. To compare the accuracy of the predicted poses to other programs benchmarked in Ref. 38, we focused on the 11 test cases corresponding to native docking experiments in both studies. Despite the exclusion of the crystal structure from the seeds, the average RMSD between the best clusters predicted by EADock and crystal structures is 0.75 Å. This is significantly better than what was reported for ICM (1.04 Å), AutoDock (2.46 Å), GOLD (3.31 Å), FlexX (3.85 Å), and DOCK (3.87 Å).

A single sampling failure was observed for penicillopepsin/1apt that has a poorly accessible binding site [Fig. 9(A)]. However, when the complex corresponding to the crystal structure is generated, it is successfully identified by the two fitness functions (data not shown).

For trypsin/1tni, the cluster with the most favorable energy is rejected as the RMSD to the crystal structure is 2.1 Å, hardly above the 2 Å threshold defining a successful prediction. This is due to the different conformations of the phenyl ring. Interestingly, the *B*-factors reported for the corresponding atoms in the original PDB file are high (47.1 $Å^2$ on average, to be compared with an average of 17.6 $Å^2$ over all other atoms), suggesting that this part of the ligand is more flexible. In addition, a contact was observed between this phenyl ring and a neighboring complex of the crystal unit cell.

Similar crystal contacts were identified for ε-thrombin/1etr, carbocypeptidase/6cpa and thermolysin/5tln. For the latter, a ionic interaction was found between the ligand and a neighboring complex in the unit cell [Fig. 9(B)].

The ligands of cytochrome/1phf and cytochrome/1phg, are linked to the heme of the receptor, while our molecular mechanics force field-based scoring function does not account for bonds between interacting partners [Fig. 9(C)]. These bonds were not reported by Bursulaya et al.[38] although discussed in the original article.[55]

Other benchmarked programs reported in Ref. 38, except ICM, systematically failed at identifying the correct binding modes for all test cases belonging to the carbonic anhydrase family. This could be due to a limitation of the force fields in describing the interaction between the ligand and the zinc ion in the binding pocket.[29] It is worth mentioning that the ROI used with ICM was not described, and it is not possible to exclude that it encompasses only a restricted number of putative binding modes.

Our validation test set was too small to significantly quantify the improvement provided by using GB-MV2 (*FullFitness*) over $\varepsilon = 1$ (*SimpleFitness*). However, using this sophisticated implicit solvation model led to successes for 3tmn, 1ets, 1etr, and 3cpa. For example, three water molecules mediate interactions between the ligand and the protein in the 3tmn complex (see Fig. 9). According to the *SimpleFitness* (i.e. *in vacuo*), at least two of these water molecules are needed in order to rank the crystal structure first [Fig. 2(B)]. Without these structural water molecules, the crystal structure corresponds only to a local minimum of the *SimpleFitness*. However, this local minimum is deep enough to be represented by a cluster, which is ranked first according to the *FullFitness* [Fig. 2(B)]. This is a strong argument in favor of realistic solvent models such as GB-MV2. Noteworthy, the docking of the 3tmn test case was not successful when the tabu list (see Material and Methods) was not used. This highlights the need to limit the sampling of uninteresting regions of the search space, according to the *FullFitness*, when discrepancies are found between the two fitness functions. In a previous study, only ICM was able to identify this binding mode correctly but, again, its ROI was not described and its sampling might have been limited inside the binding pocket.[38]

DISCUSSION

Benchmarking Docking Algorithms

Two questions must be addressed to benchmark a docking algorithm: (1) its ability to generate the good solution through sampling, and (2) its ability to recognize this solution as the correct one by its scoring function. Unfortunately, evaluations of sampling and scoring are tightly coupled, as scoring failures cannot be identified if the good solutions are not even generated. Also, the crystal structure is believed to be at least a local minimum of the scoring function. Therefore, if the seeds are too close to the crystal structure or if the ROI is too small, an algorithm with a poor sampling is likely to succeed even if its scoring is deficient, since it will be unable to sample a remote (physically irrelevant) global minimum, and will converge toward the closest local one (not far from the crystal structure). The quality of the algorithm is thus not assessed and the benchmark is not relevant, because successful predictions can either be attributed to a good scoring function or to a lack of sampling.[45] In addition, using seeds too close to the crystal structure also implies that the latter is *a priori* identified, which does not correspond to a real prediction (see list in Ref. 45).

Our Approach

To avoid such problems during the benchmark of EADock, a minimal distance of 3 Å RMSD between the seeds and the crystal structure was enforced after random rotations and translations of the ligand, as well as modifications of its internal degrees of freedom. As a result, to be successful, the sampling must be able to identify minima deeper than those provided by this high quality seeding.

Such challenging conditions make a direct comparison between EADock and the programs previously bench-

TABLE II. Summary of Predictions for Each Test Case, Presenting the Distance Between the Seeds and the Crystal Structure, the Fraction of the Receptor Surface Included in the ROI, the Rank of the First Acceptable Cluster and the Average RMSD Between its Members and the Crystal Structure

				First acceptable cluster			
Seeding	Binding site accessibility	PDB codes of complexes	% Rec. Acc. Surf in the ROI	Rank	Average RMSD (Å)	Δ FullFitness (kcal/mol)	Possible explanation
3-10 Å RMSD between seeds and the native binding mode	Accessible	*Carbonic anhydrase*					
		1cil	10.0%	49	1.08	26.77	Poor description of the interaction between the ligand and a Zn atom
		1okl	9.7%	58	1.57	34.53	
		1cnx	12.3%	46	1.65	55.28	
		Neuraminidase					
		1nsc	8.1%	1	0.52	—	—
		1nsd	8.1%	1	0.69	—	—
		1nnb	8.8%	1	0.74	—	—
		Ribonuclease					
		1gsp	31.1%	1	0.85	—	—
		1rhl	32.0%	1	1.12	—	—
		1rls	32.4%	1	0.98	—	—
		Trypsin					
		3ptb	15.9%	1	0.49	—	—
		1tng	16.1%	1	0.21	—	—
		1tnj	17.5%	1	0.86	—	—
		1tnk	17.2%	1	1.19	—	—
		1tni	17.1%	2	1.98	2.96	Crystal contact
		1tnl	17.0%	1	0.99	—	—
		1tpp	15.0%	1	0.35	—	—
		1pph	15.4%	1	0.49	—	—
	Poorly accessible	*Carbocypeptidase*					
		1cbx	9.7%	1	0.58	—	—
		3cpa	9.6%	1	0.85	—	—
		6cpa	10.8%	2	0.98	0.88	Crystal contact
		Penicillopepsin					
		1apt	13.3%	Sampling failure			—
		1apu	12.9%	6	0.68	12.03	—
		Thermolysin					
		3tmn	11.5%	1	0.57	—	—
		5tln	11.1%	7	1.86	28.98	Crystal contact
		6tmn	10.1%	Fitness failure			Water mediated interactions
		ε-Thrombin					
		1etr	12.6%	Fitness failure			Crystal contact
		1ets	12.0%	1	1.16	—	—
		1ett	12.4%	1	0.79	—	—
3–5 Å RMSD between seeds and the native binding mode	Buried	*Cytochrome P-450cam*					
		1phf	0.3%	2	1.83	1.62	Bond between the ligand and the receptor
		1phg	0.2%	3	1.65	0.97	
		2cpp	0.2%	1	0.19	—	—
		Intestinal FABP					
		1icm	1.4%	1	0.66	—	—
		1icn	1.0%	1	1.87	—	—
		2ifb	0.8%	1	0.76	—	—
		L-Arabinose					
		1abe	0.3%	1	0.18	—	—
		1abf	0.3%	1	0.68	—	—
		5abp	0.3%	1	0.64	—	—

If the best ranked acceptable cluster is not ranked first, the *FullFitness* difference with the top ranked cluster is shown. If the first acceptable cluster is not ranked first, a possible explanation is given.

marked in Ref. 38 on the same validation set difficult for two reasons. First, in the previous study, the crystal structure was kept in the seeds. Second, the ROI used with EADock in our study encompass 14% of the receptor surface, whereas the previous study defined ROI encompassing only 10% (AutoDock), 4.6% (Gold), and 1.7% (DOCK and FlexX). While such limited ROI are not suitable for the benchmark of docking programs (see above), they are relevant if the binding pocket is known prior to the docking study. Despite unfavorable conditions, the search realized by EADock converges within only 50,000 *SimpleFitness* and 15,000 *FullFitness* evaluations on average, depending on the evolutionary path explored. This highlights the efficiency of our sampling strategy, as it can be compared to the 2,500,000 poses evaluated by AutoDock in Ref. 38. Another recent study based on a different test set, reported between 200,000 and 400,000 evaluations to converge,[34] depending on the software used. This also corresponds to the number of fitness evaluations usually observed.[56] This good performance of the sampling of EADock is also supported by the successful docking of all ligands if the RMSD is used as a fitness, no matter the distance between the seeds and the crystal structure (data not shown). This points out that the success rate of EADock is limited by our scoring functions. Interestingly, most scoring failures may be explained by crystal contacts or bonds between the ligand and the receptor.

The estimation of the binding free energy requires that the predicted and the native binding modes are as close as possible. When considering successful predictions, the average RMSD between crystal structures and binding modes proposed by EADock is only 0.75 Å. This gain over other approaches is expected to play a key role for rational lead optimization.

To demonstrate the efficiency of our approach under extreme conditions on a real application, the RGD cyclic pentapeptide was docked on the αVβ3 integrin. The same docking parameters were used, except that the ROI was set to a sphere with a radius of 25 Å centered on the binding site, encompassing 65,000 $Å^3$. Seeding conformations were generated far away from the binding pocket, between 15 and 25 Å RMSD to the crystal structure. Despite these difficult conditions EADock is able to identify the crystal structure, with a RMSD of only 1.17 Å (see Fig. 10). This points out the efficiency of our sampling and scoring in a real application, and opens the field of a rational design of active compounds derived from cilengitide, which is of major interest given its clinical impact.[57]

All our results are obtained by the combination of a multi-objective optimization taking into account the solvation free energy of the complex, and an efficient sampling able to converge with a very limited number of docking modes evaluation. Taking into account the solvation free energy with a force-field based scoring has three drawbacks: it is too sensitive to small variation in the atomic coordinates, it does not provide a clear driving force, and it is computationally demanding. Our docking algorithm proposes convenient solutions for each of these three limitations. First, variations of the effective energy according to the coordinates are smoothed by an averaging over several docking modes within the same cluster. Second, a driving force is introduced into the evolutionary process: the *SimpleFitness* is used as a filter for the identification of a reasonable docking pose. Third, as the *SimpleFitness* is fast to compute, it is a first approximation to the expensive *FullFitness*. These three improvements allow EADock to use a fitness function based on the solvation free energy calculated with the CHARMM package. This provides several advantages over other approaches such as universality (i.e. it is not limited to docking on proteins) and its average description of the solvent effect. The use of this universal scoring is believed to increase the transferability of our results to other receptor/ligand families (e.g. ligand/DNA) as well as other docking strategies like, for example, molecular fragments. The latter opens the door to a unified approach toward fragment-based drug design.

ACKNOWLEDGMENTS

We thank the VITAL-IT project of the Swiss Institute of Bioinformatics, and the Cluster versus Cancer Project and its Foundation (www.clustervscancer.org) for providing the computational resources. We thank Michel Cuendet, Ute Röhrig, Simon Bernèche and Theres Fagerberg for careful reading of the manuscript.

REFERENCES

1. 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;3:935–949.
2. 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. I. Method and assessment of docking accuracy. J Med Chem 2004;47:1739–1749.
3. 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;15:411–428.
4. Claussen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 2001;308:377–395.
5. Miller MD, Kearsley SK, Underwood DJ, Sheridan RP. FLOG: a system to select quasi-flexible ligands complementary to a receptor of known three-dimensional structure. J Comput-Aided Mol Des 1994;8:153–174.
6. Welch W, Ruppert J, Jain Hammerhead AN. Fast, fully automated docking of flexible ligands to protein binding sites. Chem Biol 1996;3:449–462.
7. Cho AE, Wendel JA, Vaidehi N, Kekenes-Huskey PM, Floriano WB, Maiti PK, Goddard WA. The MPSim-Dock hierarchical docking algorithm: application to the eight trypsin inhibitor cocrystals. J Comput Chem 2005;26:48–71.
8. Budin N, Majeux N, Caflisch A. Fragment-based flexible ligand docking by evolutionary optimization. Biol Chem 2001;382:1365–1372.
9. Wu G, Robertson DH, Brooks CL, Vieth M. Detailed analysis of grid-based molecular docking: a case study of CDOCKER—a CHARMm-based MD docking algorithm. J Comput Chem 2003;24:1549–1562.
10. Vieth M, Hirst JD, Kolinski A, Brooks CL. Assessing energy functions for flexible docking. J Comput Chem 1998;19:1612–1622.
11. Goldberg DE. Genetic Algorithms in Search Optimization and Machine Learning. Boston: Addison-Wesley Professional; 1989.

12. Lawrence D. Handbook of Genetic Algorithms. New York, NY: Van Nostrand Reinhold Company; 1991.
13. 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. J Comput Chem 1998;19:1639–1662.
14. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727–748.
15. Taylor JS, Burnett RM. DARWIN: a program for docking flexible molecules. Proteins 2000;41:173–191.
16. Glaser F, Morris RJ, Najmanovich RJ, Laskowski RA, Thornton JM. A method for localizing ligand binding pockets in protein structures. Proteins 2006;62:479–488.
17. Nayal M, Honig B. On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. Proteins 2006;63:892–906.
18. An J, Totrov M, Abagyan RA. Pocketome via comprehensive identification and classification of ligand binding envelopes. Mol Cell Proteomics 2005;4:752–761.
19. Chen H, Lyne PD, Giordanetto F, Lovell T, Li J. On evaluating molecular-docking methods for pose prediction and enrichment factors. J Chem Inf Model 2006;46:401–415.
20. Simonson T, Archontis G, Karplus M. Free energy simulations come of age: protein-ligand recognition. Acc Chem Res 2002;35: 430–437.
21. Kollman P. Free energy calculations: applications to chemical and biochemical phenomena. Chem Rev 1993;93:2395–2417.
22. Holloway MK. A priori prediction of activity for HIV-1 protease inhibitors employing energy minimization in the active site. J Med Chem 1995;38:305–317.
23. Perez C, Pastor M, Ortiz AR, Gago F. Comparative binding energy analysis of HIV-1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design. J Med Chem 1998;41:836–852.
24. Böhm HJ. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput-Aided Mol Des 1994;8:243–256.
25. Böhm HJ. 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;12:309–323.
26. Head RD, Smythe ML, Oprea TI, Waller CL, Green SM, Marshall GR. VALIDATE: a new method for the receptor-based prediction of binding affinities of novel ligands. J Am Chem Soc 1996;118:3959–3969.
27. Muegge I, Martin Y. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. J Med Chem 1999;42:791–804.
28. Muegge I. A knowledge-based scoring function for protein-ligand interactions: Probing the reference state. Perspect Drug Discov Des 2000;20:99–114.
29. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL. Assessing scoring functions for protein-ligand interactions. J Med Chem 2004;47:3032–3047.
30. Camacho CJ, Vajda S. Protein docking along smooth association pathways. Proc Natl Acad Sci USA 2001;98:10636–10641.
31. Brooks BR, Bruccoleri RE, Olafson BD, States, DJ, Swaminathan S, Karplus M. CHARMm: a program for macromolecular energy minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.
32. Huang D, Lüthi U, Kolb P, Edler K, Cecchini M, Audetat S, Barberis A, Caflisch A. Discovery of cell-permeable non-peptide inhibitors of β-secretase by high-throughput docking and continuum electrostatics calculations. J Med Chem 2005;48:5108–5111.
33. Baxter CA, Murray CW, Clark DE, Westhead DR, Eldridge MD. Flexible docking using Tabu search and an empirical estimate of binding affinity. Proteins 1998;33:367–382.
34. Pei J, Wang Q, Liu Z, Li Q, Yang KL, Lai L. PSI-DOCK: towards highly efficient and accurate flexible ligand docking. Proteins 2006;62:934–946.
35. Sirockin F, Sich C, Improta S, Schaefer M, Saudek V, Froloff N, Karplus M, Dejaegere A. Structure activity relationship by NMR and by computer: a comparative study. J Am Chem Soc 2002;124:11073–11084.
36. Lee MS, Feig M, Salsbury FR, Jr, Brooks CL. New analytic approximation to the standard molecular volume definition and its application to generalized born calculations. J Comput Chem 2003;24:1348–1356.
37. Lee MS, Feig M, Brooks CL. Novel generalized born methods. J Chem Phys 2002;116:10606–10614.
38. Bursulaya BD, Totrov M, Abagyan RA, Brooks CL. Comparative study of several algorithms for flexible ligand docking. J Comput-Aided Mol Des 2003;17:755–763.
39. Hermann RB. Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area. J Phys Chem 1972;76:2754–2759.
40. Amidon GL, Yalkowsky SH, Anik ST, Valvani SC. Solubility of nonelectrolytes in polar solvents. V. Estimation of the solubility of aliphatic monofunctional compounds in water using a molecular surface area approach. J Phys Chem 1975;79:2239–2246.
41. Gohlke H, Kiel C, Case DA. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. J. Mol Biol 2003;330:891–913.
42. Hasel W, Hendrikson TF, Still WC. A rapid approximation to the solvent accessible surface areas of atoms. Tetrahedron Comput Methodol 1988;1:103–116.
43. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127–6129.
44. Zoete V, Meuwly M, Karplus M. Study of the insulin dimerization from binding free energy calculations and per-residue free energy decomposition. Proteins 2005;61:79–93.
45. Cole JC, Murray CW, Willem J, Nissink M, Taylor RD, Taylor R. Comparing protein-ligand docking programs is difficult. Proteins 2005;60:325–332.
46. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FT, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
47. Brunger AT, Karplus M. Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. Proteins 1988;4:148–156.
48. Halgren TH. Merck molecular force field. I. basis, form, scope, parameterization, and performance of MMFF94. J Comput Chem 1996;17:490–519.
49. Halgren TH. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. J Comput Chem 1996;17:520–552.
50. Halgren TH. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. J Comput Chem 1996;17:553–586.
51. Halgren TH. Merck molecular force field. IV. Conformational energies and geometries for MMFF94. J Comput Chem 1996; 17:587–615.
52. Halgren TH. Merck molecular force field. V. Extension of MMFF94 using experimental data, additonal computational data, and empirical rules. J Comput Chem 1996;17:616–641.
53. Stockwell GR, Thornton JM. Conformational diversity of ligands bound to proteins. J Mol Biol 2006;356:928–944.
54. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–1612.
55. Poulos TL, Howard AJ. Crystal structures of metyrapone- and phenylimidazole-inhibited complexes of cytochrome P-450cam. Biochemistry 1987;26:8165–8174.
56. Perola E, Walters WP, Charifson PS. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins 2004;56:235–249.
57. Burke PA, DeNardo SJ, Miers LA, Lamborn KR, Matzku S, DeNardo GL. Cilengitide targeting of the avb3 integrin receptor synergizes with radioimmunotherapy to increase efficacy and apoptosis in breast cancer xenographts. Cancer Res 2002;62: 4263–4272.