Automated Docking of Highly Flexible Ligands by Genetic Algorithms: A Critical Assessment

MARCO CECCHINI, PETER KOLB, NICOLAS MAJEUX, AMEDEO CAFLISCH
Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received 30 May 2003; Accepted 6 August 2003

Abstract: An improved version of the fragment-based flexible ligand docking approach SEED–FFLD is tested on inhibitors of human immunodeficiency virus type 1 protease, human α-thrombin and the estrogen receptor β. The docking results indicate that it is possible to correctly reproduce the binding mode of inhibitors with more than ten rotatable bonds if the strain in their covalent geometry upon binding is not large. A high degree of convergence towards a unique binding mode in multiple runs of the genetic algorithm is proposed as a necessary condition for successful docking.

 J Comput Chem 25: 412–422, 2004

Key words: docking; genetic algorithm; SEED; FFLD; ligand flexibility; HIV-1 protease

Introduction

Computer-aided approaches for docking small molecules into proteins of known structure are useful tools for drug design.[1–4] The importance of automatic docking procedures keeps growing because of the ever increasing number of 3D structures of pharmacologically relevant enzymes and receptors. Further, combinatorial and parallel synthesis techniques have generated a significant number of libraries of compounds with good pharmacological properties[5] and in certain cases tailored for specific targets.[6,7] Automatic approaches are available for docking flexible ligands of up to about 10 rotatable bonds into rigid[8–10] and partially flexible targets,[11–17] and several review articles have been published.[2,18–20] Ligands with a larger amount of rotatable bonds are much more difficult to dock,[13,14] even using a rigid protein.[21,22]

In this article we present the improvements with respect to the original version of SEED–FFLD[10,23] and evaluate the new version on difficult test cases. The SEED–FFLD approach is based on the decomposition of ligands into mainly rigid fragments. First, the most favorable fragment positions and orientations in the receptor binding site are determined by the program SEED according to an accurate binding energy that includes electrostatic solvation effects.[24] The optimal binding modes of the fragments are then used in FFLD as binding site descriptors to guide the placement of ligand conformations generated by a genetic algorithm (GA). The SEED–FFLD approach was tested by docking known nanomolar inhibitors with about 10 rotatable bonds in the active site of the uncomplexed and complexed conformations of thrombin and the human immunodeficiency virus type 1 protease (HIV-1 PR).[10]

The present work was motivated by two main questions: Is it possible to dock ligands with more than 10 rotatable bonds into HIV-1 PR? Are the predicted binding modes affected by the protein conformation (choice of the crystal structure)? The docking results show that redocking is almost always successful whereas cross-docking is problematic mainly because of strain in the covalent geometry of the ligand. Large and flexible ligands might be of limited relevance in the context of drug design. Yet, we think that testing docking programs in cases where the conformational space is very large is useful to find weaknesses and suggest improvements that will be in any case beneficial also for smaller and/or more rigid ligands.

Materials and Methods

The docking approach is a three-step strategy based on the decomposition of a flexible ligand into rigid fragments. First, the program SEED is used to dock the fragments into the binding site of the receptor.[23,25] Second, the ligand is docked by a genetic algorithm (FFLD) that uses a fast scoring function.[10] The genetic algorithm perturbations affect only the conformation of the ligand; its placement in the binding site is determined by the SEED anchors and a least-square fitting method.[26] In this way the position and orientation of the ligand in the binding site are determined by the best binding modes of its fragments previously docked using an accurate energy function with electrostatic solvation.[24] The scoring

Correspondence to: A. Caflisch; e-mail: caflisch@bioc.unizh.ch

function used in FFLD is based on van der Waals and hydrogen bond terms and does not explicitly include solvation for efficiency reasons. Solvation effects are implicitly accounted for as the binding modes of the fragments are determined with continuum electrostatics.

Third, the FFLD results are postprocessed by CHARMM minimization. The ligand decomposition into fragments, the choice of functional groups (preferable large aromatic rings) as anchors, the identification of rotatable bonds, and the definition of the binding site have recently been automated (P. Kolb et al., unpublished results). The other modifications of the SEED–FFLD approach with respect to the original procedure[10] are listed in the next subsections.

SEED

Fragment Docking

The SEED input parameters used for this application are identical to those in Table I of the original SEED article[23] except for the following three: (1) The interior dielectric constant is set to 2.0 to partially account for the electronic polarizability and dipolar reorientation effects of the solute. (2) The number of apolar points on the receptor is increased from 100 to 300 because of the very large buried binding site of HIV-1 PR. For human α-thrombin and the estrogen receptor β 300 and 150 apolar points were used, respectively. (3) To discard polar and apolar receptor vectors that point outside of the binding site, a selection using an angle criterion is performed. Initially, the minimal and maximal distances between the end points of the vectors and a set of points in the binding site (e.g., the positions of the heavy atoms of the ligand) are evaluated. A vector is discarded if the angle it spans with the closest point is larger than a cutoff. This selection uses a permissive cutoff of 100° for vectors close to the binding site points and a stricter one (70°) for distant vectors. Using the atoms of a ligand from a known complex to define the binding site does not introduce a bias in cross-docking and corresponds to the situation in an advanced drug design program, where one or more crystal structures of protein–ligand complexes have already been solved.

Postprocessing of the Optimal Binding Modes of the Fragments

The fragment binding modes are sorted by binding energy and clustered in SEED according to their position and orientation in the binding site using a conservative criterion based on distances between similar atom types.[23,27,28] Cluster representatives are subsequently grouped according to the coordinates of the geometric centers using a threshold of 1 Å. The geometric centers of the first five cluster representatives are removed from the clustering procedure described in the following and directly used for docking. For each cluster an average geometric center ($\mathbf{r}_{\mathrm{AGC}}$) is calculated with the following procedure. First, all the positions with a binding energy greater than 3 kcal/mol with respect to the cluster representative are discarded. Then, the $\mathbf{r}_{\mathrm{AGC}}$ of a given cluster is evaluated as an energy-weighted mean

$$\mathbf{r}_{\mathrm{AGC}} = \sum_{i=1}^{N} \omega_i \mathbf{r}_i,$$

$$\omega_i = \begin{cases} E_i \Big/ \sum_i E_i & \text{if } E_{\max} < 0 \\ (E_i - (E_{\min} + E_{\max})) \Big/ \sum_i (E_i - (E_{\min} + E_{\max})) & \text{if } E_{\min} > 0 \end{cases} \quad (1)$$

where $E_{\min}$ and $E_{\max}$ are the minimum and maximum energy within a cluster, respectively, and the sum runs over the N geometric centers $\mathbf{r}_i$ of the fragments in the cluster. In the sporadic case of $E_{\min} > 0$, by subtracting $(E_{\min} + E_{\max})$ from the fragment binding energies E_i it is possible to give more weight to positions with small absolute energy values. For $E_{\min} \leq 0 \leq E_{\max}$, average centers for the subsets of favorable ($E_i < 0$) and unfavorable ($E_i > 0$) binding modes in a cluster ($\mathbf{r}_-$ and $\mathbf{r}_+$, respectively) are computed by eq. (1) and the $\mathbf{r}_{\mathrm{AGC}}$ is evaluated as

$$\mathbf{r}_{\mathrm{AGC}} = 0.8\mathbf{r}_- + 0.2\mathbf{r}_+ \quad (2)$$

where the multiplicative factors 0.8 and 0.2 are somewhat arbitrary and were not optimized. The first 15 $\mathbf{r}_{\mathrm{AGC}}$ are kept for FFLD. Therefore, the binding site maps used for docking are defined by 20 points, 5 corresponding to geometric centers of fragments optimally docked by SEED and 15 average geometric centers. In this way, the final list of geometric centers used for docking is a compromise between the accuracy of the SEED binding energy and the diversity derived from the clustering procedure. The postprocessing of the optimal binding modes of the fragments leads to more heterogeneous binding site maps than using only the most favorable ones. To store this information for efficient placement of the ligand in the binding site three hash tables are generated as described in ref. 10.

FFLD

The following subsections contain the improvements with respect to the original version of FFLD.[10] The FFLD scoring function consists of an intraligand van der Waals energy, a ligand–protein soft core van der Waals, and an intermolecular polar energy term.[10] The two latter terms were modified as described below.

Intermolecular Soft Core van der Waals

A soft core van der Waals is used to generate a smooth energy landscape by reducing the frustration originating from the steep repulsive part.[29] In the present work, the linearization of the Lennard–Jones potential used to smooth the energy surface was slightly modified with respect to ref. 10. Originally, upon ligand placement, for every atom located in the binding site (including a 1-Å layer below the protein surface) a van der Waals energy with the closest protein atom was evaluated at first. The interaction energy was linearized if its value was higher than a cutoff; otherwise, a grid-based interpolated interaction energy including all

contributions of the receptor atoms was considered. In the original procedure small atomic interpenetrations with the protein surface were overpenalized without taking into account the favorable contributions of the neighboring atoms. In the version used in the present study, the van der Waals interaction with the closest protein atom is compared to the grid-based interpolated interaction energy. If the latter is more favorable, the attractive contributions of the receptor atoms dominate the interaction and the grid-based energy is more appropriate. Otherwise, the atomic interpenetration is significant and the repulsive contribution dominates the interaction. In this case, the linearized interaction with the closest protein atom is used.

Polar Interactions

The polar term approximates electrostatic interactions between ligand and protein:

$$E_{\text{polar}}^{\text{inter}} = \sum_{i=1}^{N_{\text{HB}}} E_i^{\text{HB}} + \sum_{i=1}^{N_{\text{UP}}} E_i^{\text{UP}} \tag{3}$$

where N_{HB} and N_{UP} are the number of hydrogen bonds (HBs) and unfavorable polar contacts (UPs), respectively. A significant improvement with respect to the original version[10] is the replacement of step functions, which allow different binding modes with the same energy value with smooth functions. Smooth functions allow the optimization of the hydrogen bonding pattern, avoiding discontinuities on the energy surface. A dependence on the distance and the angle in the polar term of the scoring function was introduced to reduce the noise arising from the energy degeneration and improve the convergence of the docking runs. The criteria used for the definition of unfavorable polar contacts and hydrogen bonds and the smooth functions implemented are described below.

Unfavorable Polar Contacts. An interaction between two H-bond donors or two H-bond acceptors is an unfavorable polar contact whose energy is a function of the distance between the interacting heavy atoms. A sigmoidal function is used

$$E^{\text{UP}}(r) = \begin{cases} E_{\text{bad}} & r < r_{\text{on}} \\ E_{\text{bad}}(1 + e^{\beta(r-\alpha)})^{-1} & r_{\text{on}} \le r \le r_{\text{off}} \\ 0 & r > r_{\text{off}} \end{cases} \tag{4}$$

where r is the distance between the heavy atoms, E_{bad} is the maximal penalty for an unfavorable polar contact, α is the interatomic distance corresponding to the inflection point of the function, and β is related to the steepness of the sigmoidal. The value of $E_{\text{bad}} = 3.0$ kcal/mol is the same as used previously.[10] The values of r_{on} and r_{off} depend on the choice of α and β, which were fixed at 2.8 Å and 15.5 Å^{-1}, respectively. The value of r_{on} is defined as the point where the sigmoidal reaches a value of 0.99 E_{bad}. For symmetry reasons, r_{off} is fixed by the choice of r_{on}. Hence, $r_{\text{off,on}} = \alpha \pm 4.65/\beta$.

Hydrogen Bonds. The hydrogen bonding model of the MAB force field[30]

$$E^{\text{HB}}(r, \theta) = W_{\text{H}} B_R(r) \Theta_D(\theta) \tag{5}$$

was implemented, where r is the distance between the donor (D) and the acceptor (A) atoms, θ is the angle at the H atom (D—H $\cdots$ A), and W_{H} is an atom-type dependent parameter that defines the strength of the bond. The energy is determined by the parameter W_{H}, while the distance dependence and the directionality of the hydrogen bonds follow bathtub-shaped functions:

$$B_R(r) = -\left[1 - \left(\frac{r - r_{\text{eq}}}{w_r}\right)^n\right]^m \tag{6}$$

$$\Theta_D(\theta) = \left[1 - \left(\frac{\theta - \theta_{\text{eq}}}{w_\theta}\right)^n\right]^m \tag{7}$$

Equations (6) and (7) apply whenever the expression in the outer brackets is positive; otherwise, B_R and Θ_D vanish. The exponents n and m determine the curvature of the bathtub-shaped functions and were chosen as $n = 2$ and $m = 4$. Compared to ref. 30, the implemented model, while preserving an all-atom description, does not take into account the angle at the acceptor atom in the hydrogen bond energy evaluation. Although this could lead to a loss in accuracy, it avoids the need for additional parameters for describing the valence state of the atoms involved in the hydrogen bond. The following parameters were used: the well depth W_{H} and the optimal distances r_{eq} were chosen according to the atom types of the donor and acceptor atoms involved;[23,30] θ_{eq} was set to π because of the linearity of optimal hydrogen bonds; w_r was 5.0 Å when $r \le r_{\text{eq}}$ or 1.25 Å when $r > r_{\text{eq}}$ and w_θ was 1.5 radian to mimic the original stepwise functions.

Local Search

Following a comparative study of several search engines for AutoDock,[31] a hybrid search procedure was implemented in FFLD. The hybrid search combines a global optimization procedure based on a genetic algorithm to overcome energy barriers with a local minimization algorithm to explore regions within energy basins. Local optimization has been shown to dramatically improve the success rate of the genetic algorithm search without any loss in efficiency.[31] The use of a local optimizer increases the fitness of the individuals and accelerates convergence. In the hybrid search, a loop over generations is performed until the maximum number of generations or the maximum number of energy evaluations is reached. In a generation, five different stages follow one another: evolution, mapping, fitness evaluation, local search, and similarity test. At the beginning of each cycle, the *old population* is evolved by means of the genetic operators, one-point crossing over and mutation, and a *new population* is generated. The new genetic material is decoded and the binding energy is evaluated. For the best 10% of the individuals, the local search is performed to improve the ligand fitness. Finally, the individuals of the *old population* are replaced by new chromosomes, taking into account both the energy difference and the conformational similarity.[32] The latter dramatically improves the efficiency of the hybrid search because the local search can easily cause the system to get trapped in local minima. In fact, to avoid premature con-

vergence it is important to keep structural diversity during the evolution. As a local optimizer, the Solis and Wets algorithm[33] was used with the following parameters: The maximum number of iterations per search was set to 300; the local search stepsize (ρ) was 0.1 radian; the maximum number of consecutive successes before increasing ρ was 5, while the maximum number of consecutive failures before decreasing ρ was 3; the lower bound on ρ, i.e., the termination criterion for the local search, was 0.01.

Postprocessing by CHARMM Minimization

For every docking experiment, 10 genetic algorithm runs were performed. For each run, the binding mode with the lowest FFLD energy was postprocessed by CHARMM minimization[34] using the CHARMm22 force field (Accelrys, Inc.). The structure of the protein, including the critical bridging water, was held fixed. A distance-dependent dielectric function [$\varepsilon(r) = r$] was used and the conjugate gradient minimization was stopped when the root mean square of the energy gradient reached a value of 0.01 kcal mol^{-1} $Å^{-1}$. The CHARMM-minimized ligand conformations were sorted according to their binding energy and the lowest-energy solution was compared to the reference ligand (see Preparation of the Ligands section below). The heavy atom root mean square deviation (RMSD) from the reference was determined as a quantitative measure of the docking reliability. No least-square superposition is used in calculating the RMSD because the rigid protein establishes a fixed frame of reference.

System Setup

Test Cases

To evaluate the performance of the improved version of the SEED–FFLD procedure, five HIV-1 PR protein–ligand complexes were considered. Inhibitors of HIV-1 PR are peptidomimetic molecules with a dozen or more rotatable bonds and, as such, they present a challenging target for automated docking techniques. Moreover, a large number of crystallographic structures of HIV-1 PR protein–ligand complexes are available from the Protein Data Bank (PDB) database.[35] The crystal structures used in this study were 1hvr, 1hbv, 1htg, 1hvs, and 5hvp.[36–40] The corresponding set of ligands is characterized by a wide range of torsional degrees of freedom (between 10 and 22 rotatable bonds) and contains a certain amount of diversity due to the different functional groups (Fig. 1). The ligand 1hvr has the lowest number of rotatable bonds (10) and is the only nonpeptidic inhibitor of the set. It includes aromatic fragments suitable as anchors and contains a large central scaffold, the cyclic urea, which significantly reduces the complexity of its conformational space by decoupling most of the torsional degrees of freedom. Moreover, the cyclic urea itself strongly interacts with the protein, forming on one side hydrogen bonds with the flaps and on the other side hydrogen bonds with the aspartyl dyad. Therefore, the ligand 1hvr is expected to be the simplest test case of the set. The ligands 1hbv and 1htg present a larger amount of flexibility (15 and 17 rotatable bonds, respectively) with respect to 1hvr and therefore represent a test system having an intermediate level of difficulty. Eventually, the ligands 1hvs and 5hvp (21 and 22 rotatable bonds, respectively) provide very difficult test cases because of the large conformational space.

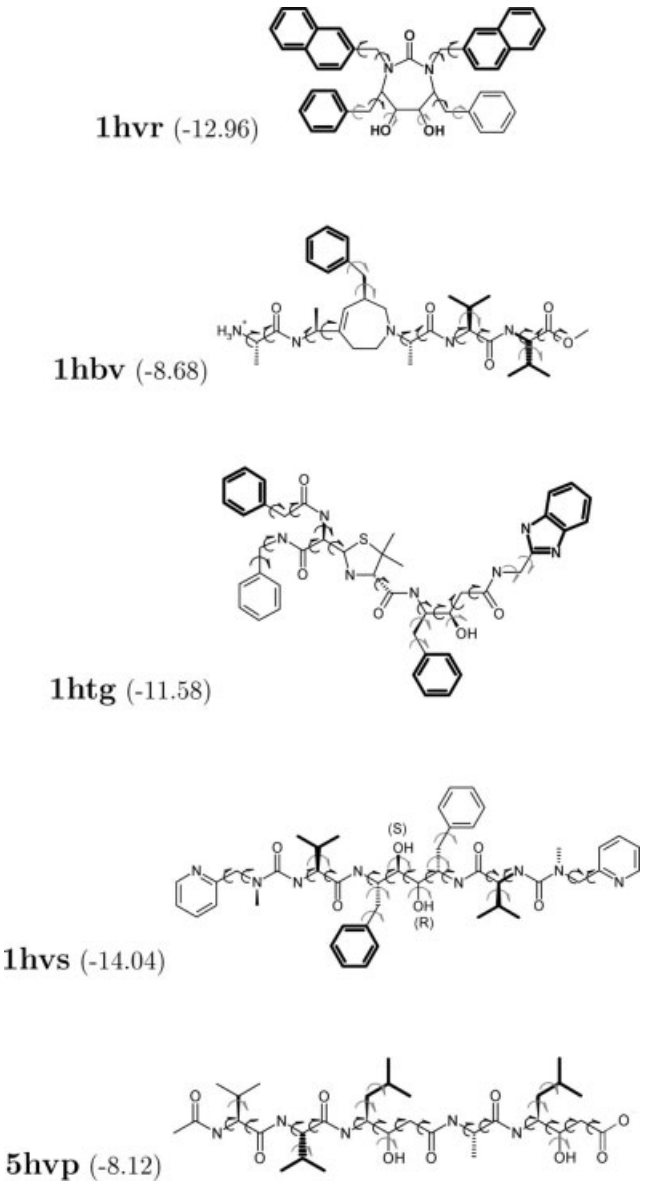

Figure 1. HIV-1 PR inhibitors used in this work. Coordinates for each complex were obtained from the PDB[35] using the accession codes given in bold. Bonds that were treated as flexible are marked by curly arrows; gray and black arrows indicate side-chains[14] and main-chain, respectively. Fragments used in SEED are in bold. Experimental free energies of binding (kcal/mol) are given in parentheses; they were obtained from the primary reference for each crystallographic structure.[36–40]

In addition, the *Nα*-((2-naphthylsulfonyl)glycyl)-DL-*p*-amidinophenylalanyl-piperidine (NAPAP)/human α-thrombin complex[41] and the complex between lig15 and the estrogen receptor β[42] were investigated to measure the robustness of the docking method. The former is a cross-docking experiment in the uncomplexed structure of the

human α-thrombin (1hgt). The latter is a redocking simulation in the "native" conformation of the estrogen receptor β (1nde). NAPAP is a peptidomimetic molecule with eight rotatable bonds. It includes two large hydrophobic fragments, naphthalene and piperidine, and the positively charged benzamidine. Lig15 is a 1, 3, 5-triazine-based molecule with 11 rotatable bonds. The ligand presents a trisubstituted planar central scaffold (the triazine) with flexible linkers to hydrophobic substituents.

Preparation of the HIV-1 PR Structures

The five crystal structures were downloaded from the PDB database.[35] The ligand and all water molecules but one were removed. The water bridging the two flaps was retained except for the docking runs with either the 1hvr protease structure or the 1hvr ligand (cyclic urea inhibitor). The side-chains of lysine and arginine residues were protonated, as well as the side-chains of histidine residues because the optimal pH of HIV-1 PR is around 5 and they are exposed to the solvent. The carboxylate groups of aspartic and glutamic acid were ionized. Particular attention was addressed to the ionization state of the cleavage site, which contains the aspartyl dyad (Asp25/Asp25′). At optimal pH for enzymatic activity (~5–6), the aspartyl dyad is most probably monoprotonated in the uncomplexed enzyme. Besides the pH, different inhibitors can have an effect on the ionization state of the active site because they can stabilize the neutral dyad or the negatively or dinegatively charged forms. According to Piana et al.,[43] the monoprotonated state is accompanied by the presence of two strong hydrogen bonds between the aspartyl dyad and H-bond donors belonging either to the inhibitor or to an ordered water molecule. Hence, a monoprotonated state was considered for the proteases 1hvr and 1hvs, while a diprotonated state was chosen for the others. Hydrogen atoms were added to the structures and minimized with the program CHARMM[34] and the CHARMm22 force field (Accelrys, Inc.). Partial charges were assigned using the MPEOE method.[44,45] Finally, the binding site was defined as the smallest zone that encompasses the HIV-1 PR residues with more than 50% of the atoms within a distance of 5 Å from any atom of the inhibitor in the X-ray structure of the complex.

Preparation of Human α-Thrombin and the Estrogen Receptor β

After downloading the crystal structures from the PDB database,[35] human α-thrombin (1hgt) and the estrogen receptor β (1nde) were prepared following a procedure similar to the one described above for HIV-1 PR.

Human α-thrombin is a trypsin-like serine protease that fulfills a central role in both hemostasis and thrombosis.[46] Several inhibitors are known to bind to the nonprime region of the active site, i.e., pockets S3–S1. The S3 and S2 pockets are hydrophobic. S3 is occupied by an aromatic ring in most of the known inhibitors and by the Phe side-chain in the natural substrate. The S2 pocket is usually filled by aromatic or aliphatic side-chains. The S1 pocket is a cylindrical cavity with an Asp at the bottom. It is usually filled by a positively charged side-chain (Arg, Lys, benzamidine, etc.) involved in a salt bridge with the Asp. During the preparation of the protein, particular attention was addressed to the protonation state of the active site residues (His, Asp, Ser). In particular, the catalytic His was protonated at the δ nitrogen. Its monoprotonated state is fully compatible with its partially buried position in the binding site and allows the formation of two intramolecular hydrogen bonds between the residues of the catalytic tryad.

The estrogen receptor β is a ligand-activated transcription factor that plays a key role in the modulation of gene expression. It binds a wide range of steroidal and nonsteroidal ligands with moderate to high affinity, with the minimal requirement of at least one paramonosubstituted phenol as the basic pharmacophore.[42] The estrogen receptor β binding site is almost buried and predominantly hydrophobic. Nevertheless, the paramonosubstituted phenol increases the ligand binding affinity because its hydroxyl group fits optimally into the gap between Arg394 and Glu353, accepting a hydrogen bond from the guanidinium and donating a hydrogen bond to the carboxy group.[42] In the structure downloaded from the PDB database,[35] the coordinates of 24 residues were missing due to poor electron density. These incomplete regions were far away from the binding site and were neglected in the docking experiments.

Preparation of the Ligands

The initial coordinates were downloaded from the PDB database.[35] Formal charges were assigned by ionizing carboxylic-acid groups and protonating amino groups. Hydrogen atoms were added and partial charges were assigned (see previous subsection). The molecular structure was minimized in two different ways with the program CHARMM[34] and the CHARMm22 force field (Accelrys, Inc.).

The first minimization was carried out in the binding site with the protein fixed. For redocking, the ligand structure optimized within its own protein conformation is used as reference for the calculation of the RMSD values. For cross-docking, protein coordinates were first superimposed by fitting the C_α atoms. The ligand conformation then optimized in the "non native" protein is used as the RMSD reference structure. In both cases, the reference structures used for evaluating docking results differ from the experimental conformation of the bound ligand. Nevertheless, the comparison between reference and predicted conformations is more consistent in this way because both structures correspond to minima of the same energy function (CHARMm22, Accelrys, Inc.). For redocking, the RMSD between X-ray and reference structures was 0.4, 0.6, 0.7, 1.0, 1.1, and 1.0 Å for 1hvs, 1htg, 1hvr, 1hbv, 5hvp, and 1nde, respectively. The average RMSD between X-ray and reference structures for cross-docking was 1.3 ± 0.4 Å, with a maximum value of 2.4 Å.

The second ligand minimization was performed outside of the receptor to remove any bias originating from the PDB structure of the protein–ligand complex [see type (3) docking experiments below]. Both minimizations were carried out using the same protocol as described before for the postprocessing.

Results and Discussion

The results on HIV-1 PR are discussed first while the human α-thrombin and estrogen receptor β results are at the end of this

section. Each of the five HIV-1 PR ligands was docked to each of the 5 protein structures, yielding a matrix of 25 docking experiments. The complexes along the matrix diagonal correspond to redocking experiments and are expected to more easily match the crystallographic structures because any "induced fit" due to the inhibitor is already present in the protein conformation. To study the effect of both ligand flexibility and geometry (covalent bond distances and bond angles in the ligand input structure) on the docking results three kinds of docking experiments were performed:

1. ***Biased geometry and partial flexibility.*** The covalent geometry of the ligand was minimized with CHARMM in the binding site of the receptor before running FFLD. The docking experiments were performed with flexible ligand side-chains and rigid main-chain (same rotatable bonds as in ref. 14; gray arrows in Fig. 1).
2. ***Biased geometry and full flexibility.*** The covalent geometry of the ligand was the same as in (1) but the experiments were carried out allowing flexibility to all rotatable bonds (gray and black arrows in Fig. 1).
3. ***Unbiased geometry and full flexibility.*** The ligand structure was minimized with CHARMM outside of the receptor to remove any geometric bias. The ligand flexibility was the same as in (2).

Conformations docked within 2.4 Å heavy-atom RMSD from the reference structure are considered successes.

Biased Geometry and Partial Flexibility

In docking experiments (1), the SEED–FFLD procedure was able to correctly dock the five ligands in each of the five protein structures (Fig. 2, top). Upon CHARMM minimization, the lowest-energy conformation for each experiment reproduced the crystallographic binding mode very well with a maximal RMSD of 1.1 Å with respect to the reference structure. Moreover, all of the 10 docking runs gave the same binding mode in 23 of 25 cases. Lack of full convergence for inhibitors 1hbv and 1htg in the 1hvr HIV-1 PR conformation was probably due to the lack of the water bridging the flaps in the 1hvr structure of the protease. This structural water plays an important role in the molecular recognition process, acting as anchor in the active site. Note that the CHARMm energy of the docking solution can be up to 16 kcal/mol (1htg in 5hvp) more favorable than the energy of the reference structure despite RMSD values smaller than 1.2 Å (Fig. 2, top). This is mainly due to electrostatic contributions coming from a better placement of the hydroxyl hydrogen interacting with the catalytic dyad (hydrogens are not considered in the RMSD evaluation). Albeit successful, the docking experiments of type (1) assume the knowledge of the backbone conformation of the bound ligand and are therefore only marginally useful (e.g., to dock a library of compounds with the same backbone).

Biased Geometry and Full Flexibility

In docking simulations of fully flexible ligands, such as experiments (2) and (3), the conformational space of the ligand is much larger than in (1). Nevertheless, in experiments (2) redocking and cross-docking of inhibitors 1hvr and 1hvs gave good results while cross-docking of 1hbv, 1htg, and 5hvp was only partially successful.

Cross-docking with the ligand 5hvp and with the protease 1hvr proved difficult for the SEED–FFLD procedure. In the first case, the large conformational space of the ligand (22 rotatable bonds) and the lack of fragments suitable as anchors were crucial for the poor performance of the simulations. Moreover, the wrong binding modes have a CHARMm energy more favorable than the reference structure (Fig. 2, right) and this points to limitations of the energy model. As an example, the misdocked conformation of the ligand 5hvp in the protease 1hvr (RMSD of 3.8 Å and $E_{\text{pred}}^{\text{CHARMm}} - E_{\text{ref}}^{\text{CHARMm}} = -18.9$ kcal/mol) is completely bent in the middle, folds back onto itself, and occupies only half of the binding site "cylinder." Although both inter- and intramolecular van der Waals interactions are optimized by this binding mode, the negatively charged terminal carboxy group is buried. This is a clear indication that solvation is needed for the final ranking of the solutions, especially for very flexible ligands without anchors.

The lack of the structural water in the protease 1hvr precluded the reproduction of the experimental results. Although successful, most of the redocking and cross-docking experiments did not converge in all cases (Fig. 2, middle). This is due to the large space accessible to ligands and in particular to the presence of flexibility in the main-chain. In fact, only the convergence for the ligand 1hvr, which has a rigid central ring rather than a flexible main-chain, was not affected by the increased number of rotatable bonds.

Unbiased Geometry and Full Flexibility

As for the docking experiments of type (1), experiments (2) require the knowledge of the bound complex, at least for determining the geometry of the input ligand structure. Therefore, to reproduce the typical high-throughput docking conditions completely unbiased docking simulations, experiments (3), were performed. Here, redocking was successful in three of five cases (1hvr, 1htg, and 1hvs), while cross-docking gave good results only for the highly flexible ligand 1hvs (Fig. 2, bottom). Experiments of type (3) are much more prone to fail with respect to experiments of type (2), even though both deal with the same amount of ligand dihedral flexibility. With respect to this point, the docking simulations of the ligands 1hvr and 1hbv are in particular significant. For the 1hvr inhibitor, while experiments (2) always converged to the reference structure experiments (3) completely misdocked the ligand in the majority of the cross-docking simulations. In the case of ligand 1hbv, while partially successful in experiments (2) and (3) did not reproduce the experimentally determined binding mode.

The explanation for this clear worsening in the performance of docking calculations is to be searched for in the covalent geometry of the ligand structures used as input in experiments (2) and (3). Therefore, for each docking simulation the biased and unbiased input structures were superimposed and compared. In most cases, the biased structure, i.e., the geometry of the bound ligand, was far away from the energy minimum and the CHARMM minimization performed outside of the receptor yielded geometries not compatible with the binding site (Fig. 3). In the ligand 1hvr, the covalent angle at the methylene carbons linking the naphthalenes to the

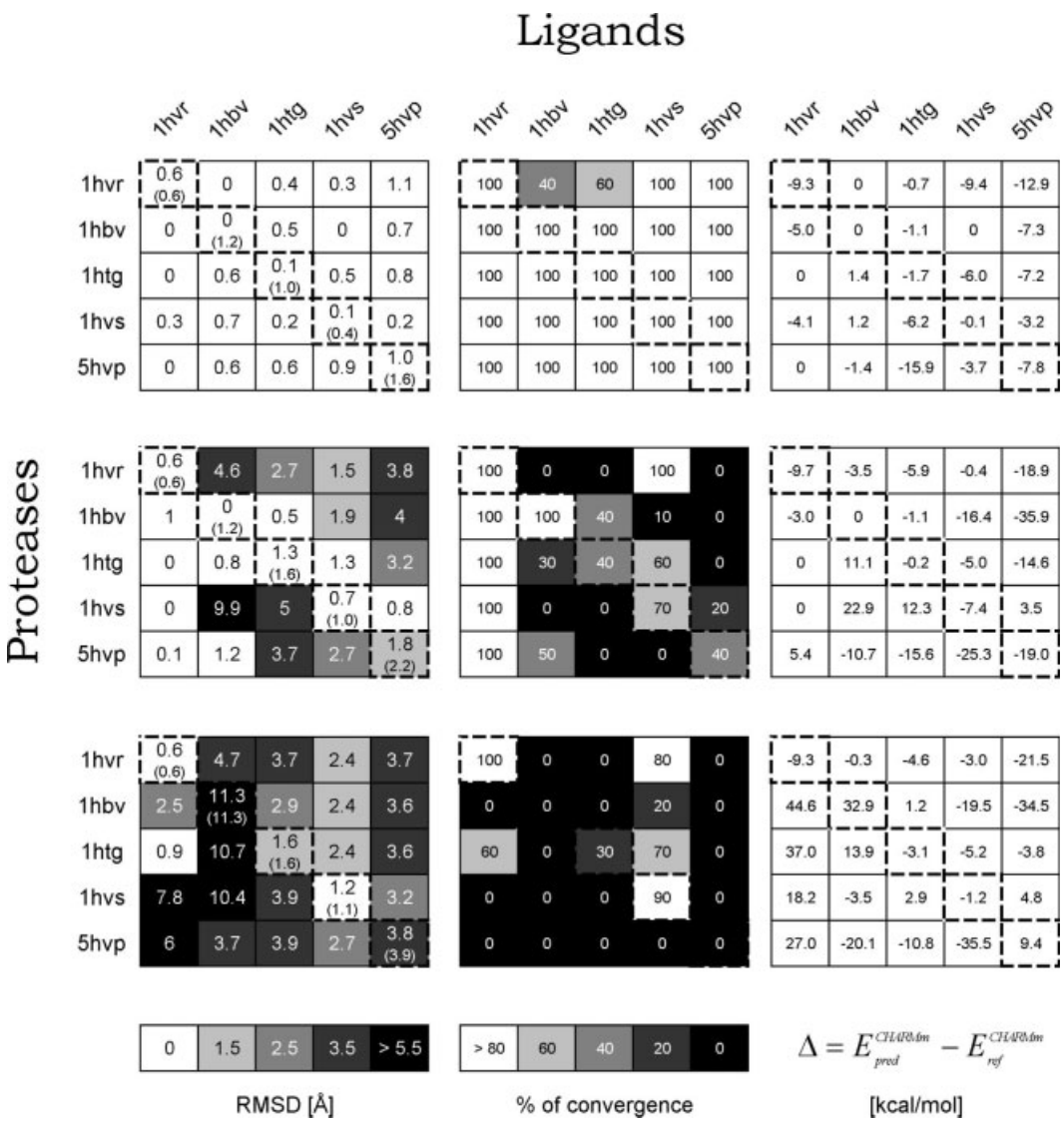

Figure 2. Results of the docking simulations. Redocking (boxes along the diagonals) and cross-docking experiments of types (1), (2), and (3) are presented from top to bottom, respectively. In the left column, heavy-atom RMSD values between the docked conformation of most favorable energy and the reference structure are given (Å). For redocking experiments, RMSD values to the unminimized experimentally determined ligand positions are reported in parentheses. In the middle column, the convergence of docking experiments are given in terms of percent ratio of successes. Docking predictions within 2.4 Å RMSD of the reference structure were considered successes. In the right column, the CHARMm energy difference between the docking solution and the reference structure is reported. Large negative values indicate limitations in the scoring function whereas positive values may point to uncomplete sampling.

cyclic urea is stretched to about 120° in the bound conformation to optimize the van der Waals interactions between the naphthalenes and the protein. In the minimization outside of the receptor, the covalent angle at the methylene carbons relaxes to a value of about 113°.

Analogous considerations can be made for the ligand 1hbv (Fig. 3). Here, the geometry of the nitrogen atom belonging to the cyclic scaffold is stretched to about 120° in the bound conformation. Upon minimization outside of the receptor, the geometry at the nitrogen atom relaxes and moves to a more pronounced tetrahedral geometry. In both cases, the CHARMM minimization outside of the receptor significantly modifies the covalent geometry of the ligands so that the SEED–FFLD approach cannot reproduce the experimental binding mode only by dihedral modifications of the input structure.

For redocking 1hvr and cross-docking of 1hvr into 1htg, this problem was overcome by CHARMM postprocessing. The lowest-energy conformation found by the SEED–FFLD procedure was

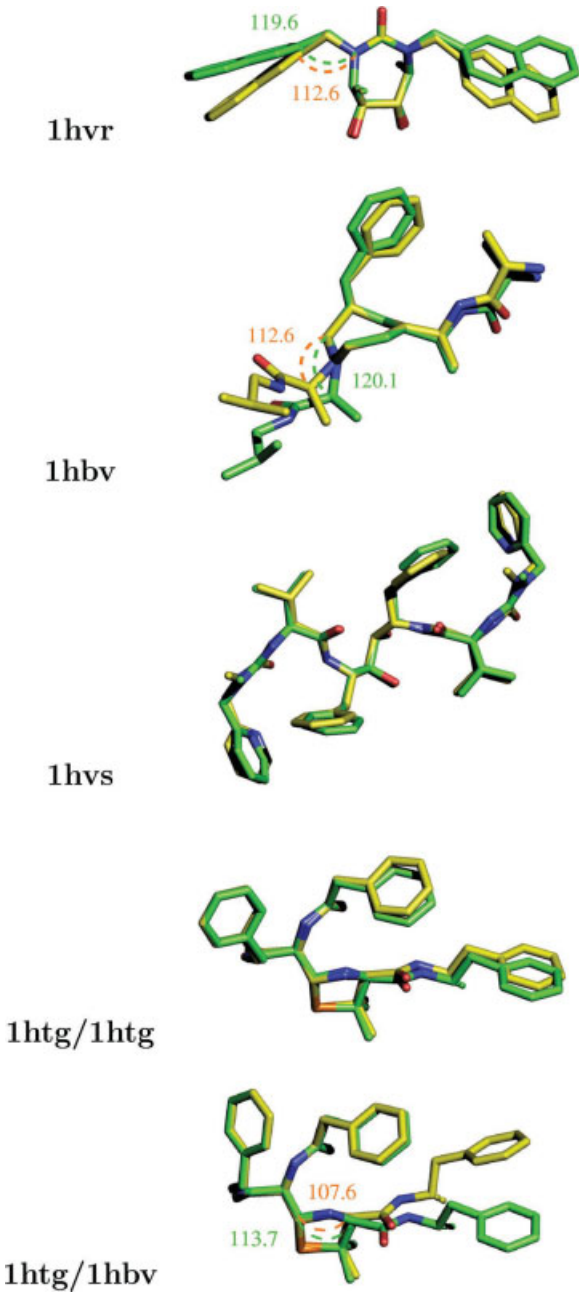

Figure 3. Comparison between ligand structures with biased geometry (green carbons) and unbiased geometry (yellow carbons) used as input for redocking 1hvr, 1hbv, 1hvs, and 1htg and for cross-docking 1htg with the protease 1hbv (from top to bottom, respectively). Significant deviations in the covalent angles are marked by broken arcs. (The pictures of the ligands were drawn using the program PyMOL.[48])

close to the X-ray structure so that during CHARMM postprocessing the experimental binding mode was reproduced. However, it has to be mentioned that this was a fortuitous event. In fact, cross-docking of the 1hvr inhibitor failed in three of four experiments (Fig. 2, bottom). It is important to note that the CHARMm energy of the docking solution is much poorer than the one of the reference structure not because the sampling in dihedral space was incomplete. Rather, the reason is that the optimal covalent geometry (obtained by minimization outside of the binding site) prevents the FFLD docking from reaching the basin of the CHARMm energy that contains the X-ray structure. Hence, the docking procedure has no chance to succeed, even after CHARMM postprocessing. This happened also with the ligand 1hbv, where even redocking was not successful. Here, the apparently small deformation of the central ring (Fig. 3) played a crucial role in docking and was the main reason for the observed failures.

The ligand 1hvs has similar biased and unbiased input structures (Fig. 3), i.e., there is not any significant deformation in the bound conformation. Hence, the unbiased geometry cannot prevent SEED–FFLD from finding a solution close to the correct one and the performance of experiments (2) and (3) are comparable. However, the biased and unbiased geometry structures are not identical and the ligand poses predicted by experiments (3) are slightly worse than those in (2) (Fig. 2).

Another interesting case is the redocking of the ligand 1htg and its cross-docking with the protease 1hbv. As shown in Figure 2, using a biased input structure (middle) both docking simulations were successful while using an unbiased input structure (bottom) only redocking succeeded. To understand how the covalent geometry influenced the performance of the simulations, the biased and unbiased input structures are analyzed. Figure 3 shows that for redocking the molecular conformations can be well superimposed and that no important deformation of the covalent geometry occurs upon binding. On the contrary, for cross-docking the geometry of the nitrogen in the central ring of the biased conformation is stretched and the unbiased structure significantly differs from the one minimized in 1hbv. These deformations can preclude the finding of the solution and were responsible for the cross-docking failure in experiments (3).

A careful analysis of the results shown in Figure 2 suggests that there is a correlation between the quality of the docking prediction and the convergence of multiple runs of the genetic algorithm. This is partially due to the fact that it is more difficult to dock highly flexible ligands for which the lack of convergence is a consequence of the large conformational space. In principle, the lack of convergence could then be used as a criterion for judging the quality of a docking prediction and as a valuable indicator in virtual high-throughput screening projects. To evaluate the reliability of a convergence-based criterion, only the cross-docking experiments of type (c) were considered, yielding a test sample of 20 docking simulations. The redocking simulations were discarded because they implicitly contain information of the bound complex. For each of the 20 docking simulations, convergence toward the lowest-energy conformation (not necessarily the experimental structure) in 10 FFLD runs with different random generator seeds was first determined (Fig. 4, top). The convergence values were then used to build the density plot shown in Figure 4 (bottom). The density plot is darker in the top right region, which indicates that convergence is a necessary condition for reliable predictions. Undoubtedly, the sample used for testing the convergence-based criterion is rather small (only 20 docking simulations) and therefore it is difficult to draw general conclusions. Nevertheless, the density plot shows a clear trend and implies that docking experi-

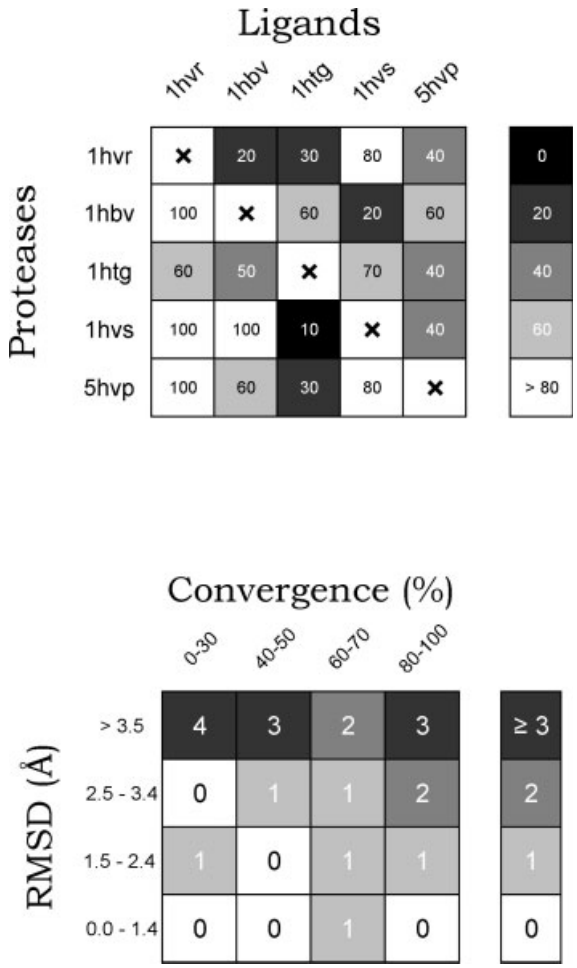

Figure 4. Evaluation of the convergence-based criterion proposed for judging the quality of the docking predictions. Top, convergence toward the lowest energy solution (not necessarily the experimental structure) in 10 docking runs is shown for the cross-docking experiments of type (3). Bottom, the density plot reports the frequency of finding a certain RMSD between the lowest-energy ligand pose and the experimental structure for a given amount of convergence; darker colors represent higher-frequency values.

ments with convergence lower than 60% may have not found the global minimum of the CHARMm22 energy surface and should be discarded during the analysis of a library screening project.

Convergence toward a unique binding mode in multiple runs of the genetic algorithm is a necessary but not sufficient condition for judging the quality of a prediction. It is not sufficient because an oversimplified energy function with a funnel-like profile together with a protein conformation that does not allow the reproduction of the crystal structure will yield convergence on a wrong binding mode.

Finally, to completely remove the geometric bias of the crystal structure a conformational search of the ligand 1hvs was performed by high-temperature molecular dynamics in the absence of the protein. The simulation was run for 2 ns at 400 K using the Berendsen thermostat and a distance-dependent dielectric function $[\varepsilon(r) = 4r]$. The final snapshot was minimized and its RMSD from the X-ray conformation after optimal overlap is 5.8 Å, which indicates that all of the information was lost and the initial conformation for docking was fully unbiased. Redocking was successful with a RMSD of 1.1 Å and a convergence of 80%. This result can not be generalized. On the contrary, it is likely that the majority of experiments (3) would deteriorate by using ligand covalent geometries without any memory of the crystal structure.

Results on Human α-Thrombin and the Estrogen Receptor β

To test the approach on binding sites with different physicochemical properties the same docking procedure was applied to human α-thrombin and the estrogen receptor β. Human α-thrombin presents an asymmetrical binding site with two hydrophobic pockets and a hydrophilic cavity specific for positively charged amino acids (Lys or Arg). Estrogen receptor β has a predominantly hydrophobic and almost completely buried binding site (see above). In both cases, starting from an unbiased and fully flexible conformation of the ligand [see type (3) docking experiments above] the SEED–FFLD approach was able to correctly reproduce the experimentally determined binding modes with RMSD smaller than 1 Å and convergences larger than 90%.

Judging Search Methods

To evaluate the performance of the hybrid search procedure it was compared with the GA of the original version of FFLD.[10]

For this purpose, unbiased redocking experiments were repeated without using the local optimizer during the evolution and with the same amount of energy evaluations. The simulations clearly show that a hybrid search is more efficient than GA because it always reaches a lower-energy conformation. The results of two redocking experiments carried out with both search methods are presented in Figure 5. For redocking 1hvr, a hybrid search is more efficient than the GA, especially at the beginning of the simulation, where the energy gap is large. At about 60% of the evolution the gap decreases and the performance of the two methods is comparable. For redocking 1hvs, the hybrid search performs better during the entire simulation and the energy gap increases until the end. Moreover, the standard deviation of the hybrid search evolutions (shown as error bars in Fig. 5) is larger, indicating that it is less prone to premature convergence than the GA.

The comparison shows that the local search accelerates convergence of the simulations and dramatically improves the quality of the docking predictions in case the conformational space of the ligand is large and its torsional degrees of freedom are strongly coupled (main-chain flexibility). This is mainly due to the fact that random perturbations of binary strings performed by the GA during the evolution correspond to radical jumps in the energy landscape and may be too large. On the contrary, the local optimizer is able to refine large perturbations due to crossover and mutations and leads to a better investigation of the energy landscape. The results of the present docking study suggest that hybrid search methods should be preferred to canonical genetic algorithms.

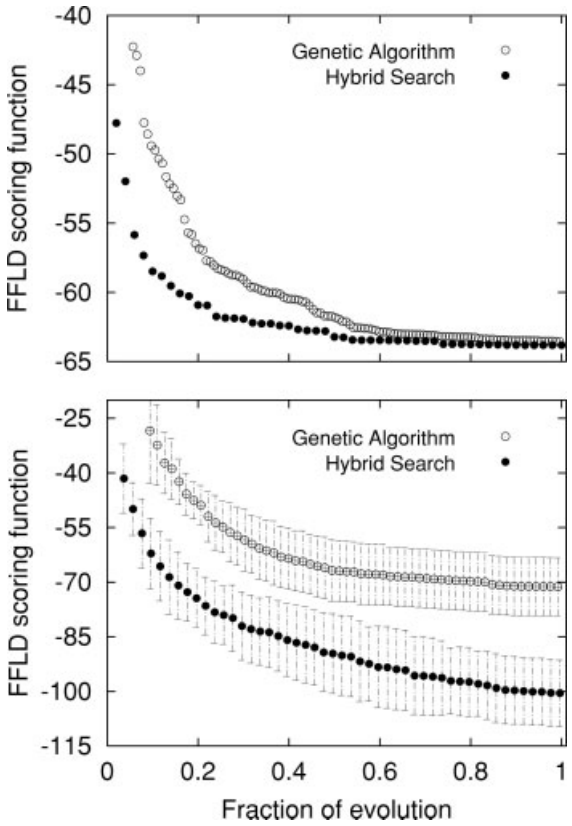

Figure 5. Evolution of the best individual of the population averaged over 10 docking runs for 2 experiments of type (3) using the same number of energy evaluations. Open and solid bullets indicate evolutions performed by genetic algorithm and hybrid search procedure, respectively. Redocking of the ligands 1hvr ($3 \cdot 10^5$ energy evaluations) and 1hvs ($1 \cdot 10^6$ energy evaluations) are shown from top to bottom. In the bottom plot, the vertical bars show the standard deviation computed over 10 docking runs.

Computational Requirement

All docking simulations were carried out on 1.6-GHz Athlon processors. For experiments of type (1), because of the limited amount of ligand flexibility fast docking calculations were performed, having a maximum number of 50 hybrid search cycles per run. The computational time required for a single docking varied from 5 to 19 min, yielding an average time of 12 min/run. For experiments of types (2) and (3), both carried out with full ligand flexibility, more extensive calculations were performed, using approximately 1 million energy evaluations per run. The computational time required for a single docking varied from 123–214 min, yielding an average time of 168 min/run. In these experiments, the number of energy evaluations per run was intentionally overestimated to make sure that the stochastic algorithm used for docking reached convergence in all cases. The computational requirements given above do not include docking of molecular fragments by SEED,[23] which was performed only once for each protein conformation.

Conclusions

Four main conclusions can be drawn from the docking results. First, a hybrid approach consisting of a local search and genetic algorithm significantly improves the quality of the SEED–FFLD docking predictions at a moderate additional computational cost. This is not a new finding[31] but simply provides further evidence with more stringent test cases, i.e., cross-docking experiments with highly flexible ligands.

Second, the quality of docking predictions depends on the degree of ligand flexibility and we suggest that validation of docking approaches should be always done with full dihedral flexibility of the ligands. In this respect, it would be interesting if the study of Österberg et al.[14] could be repeated without holding the main-chain of the peptidic inhibitors rigid.

Third, automatic approaches that sample only in dihedral space can give misdocked predictions if the covalent geometry of the ligand (i.e., its bond angles and lengths) is strained upon binding to its target. Therefore, a reliable validation of a docking approach should be performed without using any information on the conformation of the bound ligand, i.e., after a conformational search outside of the receptor. This was not done in previous works by us and others.[10,14,22,29,31,47] For docking a limited set of compounds, approaches that allow for full flexibility (including bond angles and lengths) of the ligands, albeit computationally expensive, should be preferred.[11,12]

Fourth, the docking results indicate that convergence toward the same docking solution in multiple runs of the genetic algorithm is a necessary (but not sufficient) condition for reliable predictions.

Acknowledgments

The authors are grateful to N. Budin, F. Dey and D. Huang for helpful discussions and to the referees for interesting comments and useful suggestions. The simulations were performed on a Beowulf cluster running Linux and the authors thank U. Haberthür and F. Rao for their help in setting up and maintaining the cluster. A. Widmer (Novartis Pharma, Basel) is thanked for providing the molecular modeling program Wit!P, which was used for visual analysis of the docking results. This work was supported by the Swiss National Competence Center in Structural Biology (NCCR).

References

1. Kuntz, I. D. Science 1992, 257, 1078.
2. Apostolakis, J.; Caflisch, A. Comb Chem High Throughput Screen 1999, 2, 91.
3. Glen, R. C.; Allen, S. C. Curr Med Chem 2003, 10, 763.
4. Walters, W. P.; Namchuk, M. Nature Rev Drug Discov 2003, 2, 259.
5. Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. J Med Chem 1994, 37, 1385.
6. Zuckermann, R. N.; Martin, E. J.; Spellmeyer, D. C.; Stauber, G. B.; Shoemaker, K. R.; Kerr, J. M.; Figliozzi, G. M.; Goff, D. A.; Siani, M. A.; Simon, R. J.; Banville, S. C.; Brown, E. G.; Wang, L.; Richter, L. S.; Moos, W. H. J Med Chem 1994, 37, 2678.
7. Chang, Y.; Gray, N. S.; Rosania, G. R.; Sutherlin, D. P.; Kwon, S.; C.;

N. T.; Sarohia, R.; Leos, M.; Meijer, L.; Schultz, P. G. Chem Biol 1999, 6, 361.
8. Wang, J.; Kollman, P. A.; Kuntz, I. D. Proteins 1999, 36, 1.
9. Jones, G.; Willett, P.; Glen, R. C. J Mol Biol 1995, 245, 43.
10. Budin, N.; Majeux, N.; Caflisch, A. Biol Chem 2001, 382, 1365.
11. Given, J. A.; Gilson, M. K. Proteins 1998, 33, 475.
12. Apostolakis, J.; Plückthun, A.; Caflisch, A. J Comput Chem 1998, 19, 21.
13. Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. J Mol Biol 2001, 308, 377.
14. Österberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Proteins 2002, 46, 34.
15. Verkhivker, G. M.; Bouzida, D.; Gelhaar, D. K.; Rejto, P. A.; Freer, S. T.; Rose, P. W. Curr Opin Struct Biol 2002, 12, 197.
16. Wong, C. F.; McCammon, J. A. Annu Rev Pharmacol Toxicol 2003, 43, 31.
17. Kallblad, P.; Dean, P. M. J Mol Biol 2003, 326, 1651.
18. Halperin, I.; Ma, B. Y.; Wolfson, H.; Nussinov, R. Proteins 2002, 47, 409.
19. Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. J Comp-Aided Mol Design 2002, 16, 151.
20. Lyne, P. D. Drug Discov Today 2002, 7, 1047.
21. Verkhivker, G. M.; Bouzida, D.; Gelhaar, D. K.; Rejto, P. A.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. J Comput-Aided Mol Design 2000, 14, 731.
22. Sotriffer, C. A.; Gohlke, H.; Klebe, G. J Med Chem 2002, 45, 1967.
23. Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Proteins 1999, 37, 88.
24. Scarsi, M.; Apostolakis, J.; Caflisch, A. J Phys Chem A 1997, 101, 8098.
25. Majeux, N.; Scarsi, M.; Caflisch, A. Proteins 2001, 42, 256.
26. Kabsch, W. Acta Crystallogr 1976, A32, 922.
27. Kearsley, S. K.; Smith, G. M. Tetrahedron Comp Meth 1990, 3, 615.
28. Klebe, G.; Mietzner, T.; Weber, F. J Comp-Aided Mol Design 1994, 8, 751.
29. Verkhivker, G. M.; Rejto, P. A.; Gelhaar, D. K.; Freer, S. T. Proteins 1996, 25, 342.
30. Gerber, P. R.; Müller, K. J Comp-Aided Mol Design 1995, 9, 251.
31. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. J Comput Chem 1998, 19, 1639.
32. Budin, N.; Majeux, N.; Tenette-Souaille, C.; Caflisch, A. J Comput Chem 2001, 22, 1956.
33. Solis, F. J.; Wets, R. J. B. Math Oper Res 1981, 1, 19.
34. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
35. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. Nucl Acids Res 2000, 28, 235.
36. Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C. H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Erickson-Viitanen, S. K. Science 1994, 263, 380.
37. Hoog, S. S.; Zhao, B.; Winborne, E.; Fischer, S.; Green, D. W.; DesJarlais, R. L.; Newlander, K. A.; Callahan, J. F.; Moore, M. L.; Huffman, W. F.; Abdel-Meguid, S. J Med Chem 1995, 38, 3246.
38. Jhoti, H.; Singh, O. M.; Weir, M. P.; Cooke, R.; Murray-Rust, P.; Wonacott, A. Biochemistry 1994, 33, 8417.
39. Baldwin, E. T.; Bhat, T. N.; Liu, B.; Pattabiraman, N.; Erickson, J. W. Nature Struct Biol 1995, 2, 244.
40. Fitzgerald, P. M. D.; Mckeever, B. M.; Vanmiddlesworth, J. F.; Springer, J. P.; Heimbach, J. C.; Leu, C. T.; Werber, W. K.; Dixon, R. A. F.; Darke, P. L. J Biol Chem 1990, 265, 14209.
41. Skrzypczak-Jankun, E.; Carperos, V. E.; Ravichandran, K. G.; Tulinsky, A.; Westbrook, M.; Maraganore, J. M. J Mol Biol 1991, 221, 1379.
42. Henke, B. R.; Consler, T. J.; Go, N.; Hale, R. L.; Hohman, D. R.; Jones, S. A.; Lu, A. T.; Moore, L. B.; Moore, J. T.; Orband-Miller, L. A.; Robinett, R. G.; Shearin, J.; Spearing, P. K.; Stewart, E. L.; Turnbull, P. S.; Weaver, S. L.; Williams, S. P.; Wisely, G. B.; Lambert, M. H. J Med Chem 2002, 45, 5492.
43. Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. J Am Chem Soc 2001, 123, 8730.
44. No, K.; Grant, J.; Scheraga, H. J Phys Chem 1990, 94, 4732.
45. No, K.; Grant, J.; Jhon, M.; Scheraga, H. J Phys Chem 1990, 94, 4740.
46. Tapparelli, C.; Metternich, R.; Ehrhardt, C.; Cook, N. S. Trends Pharmacol Sci 1993, 14, 366.
47. Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Chem Biol 1995, 2, 317.
48. DeLano, W. The PyMOL Molecular Graphics System; DeLano Scientific: San Carlos, CA, 2002.