Two-Stage Method for Protein−Ligand Docking

Daniel Hoffmann,*,† Bernd Kramer,† Takumi Washio,‡ Torsten Steinmetzer,§ Matthias Rarey,† and Thomas Lengauer†

German National Research Center for Information Technology, Institute for Algorithms and Scientific Computing (GMD-SCAI), Schloss Birlinghoven, D-53754 Sankt Augustin, Germany, C&C Research Laboratories NEC Europe Ltd., D-53757 Sankt Augustin, Germany, and Friedrich-Schiller-Universität Jena, Biologisch-Pharmazeutische-Fakultät, Institut für Biochemie & Biophysik, Philosophenweg 12, D-07743 Jena, Germany

Received May 31, 1999

A two-stage method for the computational prediction of the structure of protein−ligand complexes is proposed. Given an experimentally determined structure of the protein, in the first stage a large number of plausible ligand conformations is generated using the fast docking algorithm FlexX. In the second stage these conformations are minimized and reranked using a method based on a classical force field. The two-stage method is tested for 10 different protein−ligand complexes. For 9 of them experimentally determined structures are known. It turns out that the two-stage method strongly improves the predictive power as compared to that of the fast docking stage alone. The tenth case is a bona fide prediction of a complex of thrombin with a new inhibitor for which no experimentally determined structure is available so far.

Introduction

A major goal of rational drug design is the development of ligands that bind specifically and with high affinity to proteins that play key roles in diseases. Computational molecular modeling can contribute significantly to this development process by offering detailed insights into protein−ligand interactions. An important problem that is handled with molecular modeling techniques is the so-called docking problem: how can one predict the 3D-structure of a protein−ligand complex, given only the conformation of the protein and the chemical structure of the ligand?

In recent years a number of computational tools have been devised to deal with the docking problem.[1−6] Typically, docking programs generate large numbers of conformations and try to rank them according to their (free) energy. The conformation of lowest energy is then considered the best prediction for the true conformation of the protein−ligand complex. To tackle the problem of finding this conformation in a very large conformational space within a reasonable amount of time, most docking programs use elaborate algorithms for efficient conformational searches and simplified energy functions that can be evaluated quickly. Although the predicted conformations often are essentially correct, there remain many instances where this is not the case. To a first approximation the reasons for these failures fall into two categories: either the conformational sampling is insufficient—the correct conformation is not generated, or the energetic ranking is not good enough—the correct conformation is not identified.

In this article a two-stage strategy is proposed to tackle the second problem. In the first stage the conformational space is sampled using the fast FlexX algorithm,[4,7] and in the second stage low-energy conformations are reranked according to conformational energy and solvation energy calculated with a more detailed energetic model mainly based on classical force fields. In contrast to previous reranking work,[8,9] in this paper the reranking stage involves an energy minimization with a fully flexible ligand.

To assess whether the two-stage method is suitable for ranking, we investigate 10 protein−ligand systems, most of which pose hard problems if only fast docking tools such as FlexX are used and no subsequent reranking takes place. Interestingly, some of these systems are difficult challenges for other established docking methods such as GOLD as well.[5] Hence, we think that the scope of this paper is not restricted by the use of FlexX as docking tool. Also included in the test set is a bona fide prediction for a complex of thrombin with a new inhibitor for which no experimental 3D-structure is available so far.

The remainder of the paper is organized as follows. First the main features of the two-stage method are presented with an emphasis on the second stage. Then the results of the 10 test cases are discussed in detail. From these data we finally draw conclusions concerning the power and limitations of the proposed method, and we also outline possible further improvements.

Methods

The method for protein−ligand docking proposed here is a combination of two techniques. In the first stage a very fast docking with a simple molecular model is performed using FlexX, version 1.7.0.[4] This typically results in a few hundred plausible but distinct conformations of the ligand in the binding pocket of the protein. In the second stage the ligand conformations having the lowest FlexX energies are first minimized in the binding pocket using CHARMM[10] using the CHARMm22 force field (MSI, San Diego, CA) and reranked using CAMLab, version 1.0,[11] with a more detailed molecular model including also solvation effects.

Because FlexX has been described in depth elsewhere,[4] we here sketch only the main features of the algorithm. FlexX

* Corresponding author. E-mail: Daniel.Hoffmann@GMD.DE. Fax: +49 2241 14 2656.
† GMD-SCAI.
‡ C&C Research Laboratories.
§ Friedrich-Schiller-Universität.

10.1021/jm991090p CCC: $18.00
Published on Web 09/24/1999

docks flexible ligands into binding pockets of rigid proteins. The docking proceeds incrementally by adding fragment after fragment so that the energy of the complex is always locally minimal. For a better sampling of the conformational space, the build-up of the ligand is allowed to ramify into diverse but energetically favorable regions. To make the search fast, the conformational space of the ligand is discretized and a simple energetic model derived from LUDI[12] is used for scoring.

The second stage starts with the selection of those ligand conformations that will be reranked with the more detailed energetic model. Generally the N conformations having the lowest FlexX energies are considered for reranking. Since for the reranking the CPU time per conformation is about the same as for the whole FlexX run, N should be chosen not too large if time is a limitation. On the other hand if N is too small, the best conformation may be missed. A rule of thumb is that the more rotatable bonds and the less specific the binding interactions, the more conformations should be considered in the reranking stage. In this paper we choose values of N between 60 and 500. Experience shows that in many cases it is sufficient to restrict oneself to the first 100 conformations, but in other cases, like the one of complex 2mcp below, this restriction may prevent accurate results.

To prepare the calculations in the second stage, a rectangular box is determined, that contains all N selected ligand conformations plus a safety margin of 12.5 Å. Only the part of the protein within this box is taken into account for the detailed energy calculations. In this way CPU time and computer memory are saved without a significant distortion of the energy landscape.

All docking cases in this paper use protein structures that have been determined crystallographically and that are published in the Protein Data Bank (PDB).[13,14] Hydrogen atoms are added to these structures using CHARMM.[10] Each of the N selected ligand conformations is subjected to an energy minimization of a maximum of 1500 steps using the conjugated gradient method in CHARMM. The protein structure is held fixed during this minimization, whereas the ligand is allowed to change its conformation and position freely under the influence of the force field. The purpose of the minimization is to relax the ligand conformation into a neighboring local energy minimum that essentially retains the binding pattern of the original conformation generated by FlexX. Because only a local optimization is intended, we screen the long-range Coulombic interactions that otherwise could lead to larger conformational changes during minimization. This screening is achieved by using a distance-dependent dielectric ($\epsilon = r$) in the minimization.

The energy minimization consumes most of the CPU time in the reranking. Since in many applications time is a limiting factor, it is important to emphasize the necessity of the minimization step. The energy landscape modeled by the all-atom force field is very rugged: small structural changes often correspond to large changes in energy, mainly because of atomic overlaps or stretched bonds. In nature, such strained conformations will occur very rarely, and the partition function will by far be dominated by structures around minima of the force field energy. In this sense it is necessary and sufficient to compare minimized structures.

For each minimized complex the total energy in aqueous solution is then computed using CAMLab,[11] a program that can perform computations of force field and solvation energies. The total energy E_t can be written as

$$E_t = E_{ff} + E_{es} + E_{np} \quad (1)$$

where E_{ff} is the CHARMm22 force field energy[10] with constant dielectric ($\epsilon = 4$), E_{es} is the electrostatic part of the solvation energy, and E_{np} is the nonpolar part of the solvation energy. The force field comprises bonded energy terms (bonds, bond angles, torsions, improper torsions) and nonbonded terms (Lennard−Jones and Coulomb) with the usual functional forms.[10]

E_{es} is obtained from the solution of the Poisson equation[15,16] using a fast multigrid finite difference solver.[11] Atomic radii and charges are taken from the CHARMm22 force field. The charges are discretized on a relatively coarse grid with a grid spacing of 1 Å. The Poisson equation is solved twice: first, in a homogeneous dielectric with $\epsilon = 4$ everywhere and, second, with $\epsilon = 4$ inside the molecular surface of the protein−ligand complex and $\epsilon = 80$ outside. From the difference of the two solutions the electrostatic part of the solvation energy E_{es} is obtained (or more precisely the electrostatic part of the transfer energy between media with $\epsilon = 4$ and $\epsilon = 80$).[17]

The nonpolar part of the solvation energy E_{np} is approximated by the total solvent-accessible surface of the complex times a surface tension constant of 84 J/Å^2.[18−20]

Table 1. Numerical Results for Eight Test Cases[a]

complex	N_{rot}	$rmsd_{flexx}$ (Å)	$rmsd_{min E_t}$ (Å)
1poc	26	9.49	3.11 (20)
4hmg	12	5.45	1.55 (61)
1dwd	10	1.12	1.71 (52)
2cgr	9	6.38	0.54 (10)
2mcp	7	6.86	1.30 (320)
1imb	7	5.10, 0.84	6.01 (23), 1.03 (45)
1tni	5	2.71, 3.30	2.50 (2), 1.86 (88)
1mup	4	3.82, 3.76	3.54 (57), 2.16 (156)

[a] N_{rot} is the number of rotatable bonds of the ligand. $Rmsd_{flexx}$ is the rmsd between corresponding non-hydrogen atoms of the ligand in the crystal structure and the ligand conformation at rank 1 (lowest energy) according to FlexX. $Rmsd_{min E_t}$ is the rmsd between crystal structure and the ligand conformation of lowest energy E_t; the number in parentheses is the rank of this conformation before reranking. For 1imb, 1mup, and 1tni there are two rmsd values given. The first respective value refers to calculations without consideration of X-ray waters; the second value results from calculations with explicitly modeled X-ray waters. 1icn is missing because of ambiguities in the experimentally determined structure (see main text).

Results and Discussion

We assume that the experimentally determined complex conformations correspond to structures of minimum (free) energy. If our energy function E_t is adequate, and if our conformational search covers the region of minimal energy, then the two-stage method should identify a conformation close to the experimentally determined structure as conformation of minimum energy E_t in the set of generated structures. Another formulation of this requirement is that in the list of conformations arranged in order of ascending energy, the conformation at rank 1 should be close to the experimentally determined structure. This is the criterion that we use to judge the quality of the proposed method. As a measure for the closeness of predicted structures to the experimentally determined structure of the ligand, we choose the root-mean-square-deviation (rmsd) of the Cartesian coordinates of all non-hydrogen atoms. Numerical results in terms of the rmsd of the conformation of lowest energy are summarized in Table 1. It is not expected that after reranking a positive correlation exists over the whole range of energies and rmsd's, because it is possible to generate structures that are close to the crystal structure but have, for instance, atomic overlaps and hence have much higher energies than relaxed structures far away from the crystal structure. Ideally, if the experimentally determined structure is submitted to the reranking procedure, the energy E_t of this structure should be particularly low. This is also investigated in our test cases below.

We study the performance of our two-stage method in 10 test cases. Eight of these cases are purposefully chosen such that FlexX performs badly: it fails to assign

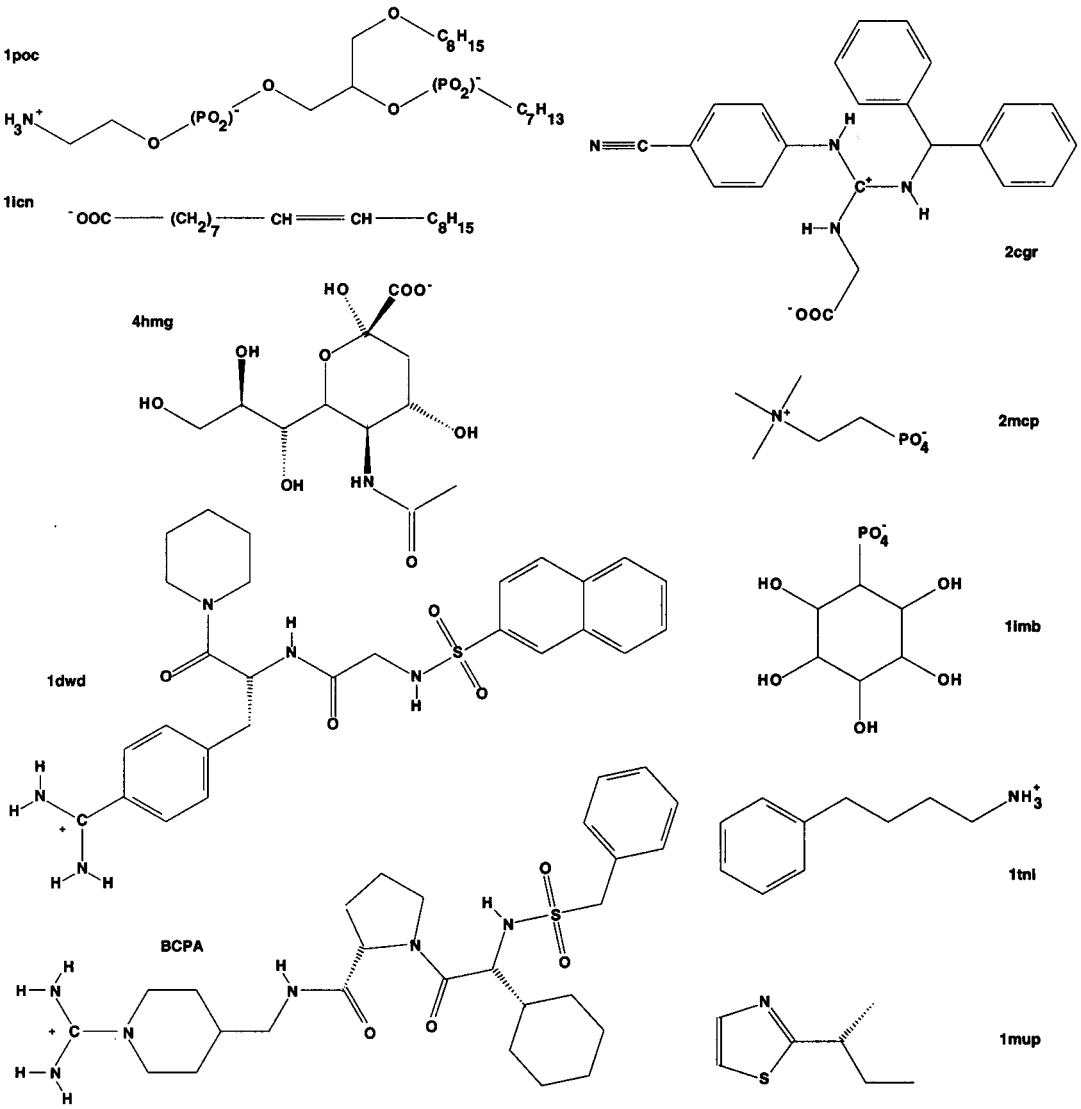

Figure 1. Chemical structures of the ligands in the 10 protein−ligand complexes investigated. Labels indicate the respective complexes (see text).

the top rank or the lowest FlexX energy to a ligand conformation that is essentially correct. However, we will show that in two of these eight cases (1icn and 1imb) the problem is probably not rooted in FlexX itself.

We also present results for complexes of thrombin with two inhibitors: first, the complex of thrombin and NAPAP is a reference case where FlexX performs well and, second, in a bona fide prediction, we apply our two-stage method on the complex of thrombin with a new ligand for which no experimentally determined structure is available. Except for the last example we label each case with the PDB code[13,14] of the respective experimentally determined structure. The discussion proceeds from larger to smaller ligands (Figure 1) and concludes with the results of the bona fide prediction.

1poc. 1poc[21] is a complex of phospholipase A_2 and a phospholipid-like transition-state analogue as a ligand. This ligand is difficult for FlexX to dock because it has many rotatable bonds and hence a large conformational space. Guided by the channels that accommodate the two alkyl chains, FlexX nevertheless manages to find a set of reasonable structures with rmsd's of about 3 Å (Figure 2). In this set only the positively charged part of the lipid headgroup is pointing in a direction that is different from that seen in the crystal structure. However, this group of structures is not top-scoring according to the energy function of FlexX.

After reranking, the energy E_t (eq 1) of the minimized structures distinguishes three separate groups (Figure 2): a group with rmsd of about 3 Å has the lowest energy, and there are two groups with rmsd values of 9−10 Å with much higher energies. The value of E_t of the crystal structure after energy minimization is still considerably lower than that of the 3 Å group, or in other words, the reranking procedure clearly identifies the experimental structure as that of lowest energy (diamond in Figure 2). The energetic components contributing most to this discrimination are Coulomb energy, Lennard−Jones energy, and bonded energy terms. In the crystal structure the charged parts of the headgroup are matched by opposite charges of the protein and the ligand fills the binding site without being strained.

As expected the energy minimization does not lead to strong improvements in rmsd but changes structures only locally. For instance the minimization is not able to move the positively charged part of the headgroup to its correct position. One could think of using an un-

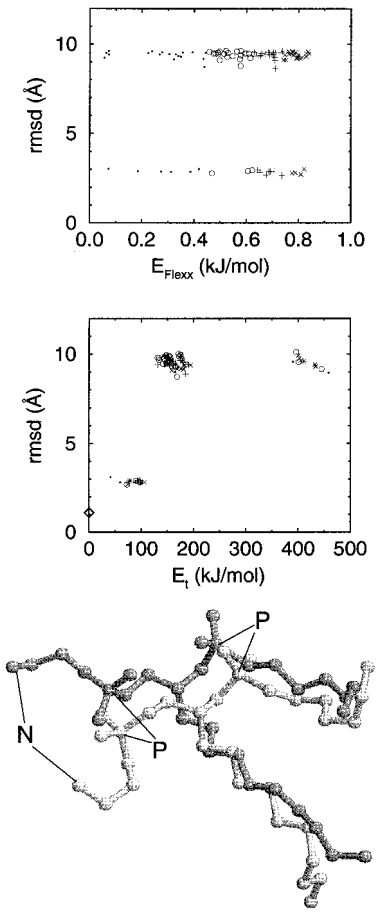

Figure 2. Energies, rmsd values, and structures of FlexX-generated predictions for 1poc. Rmsd values are given with respect to the crystal structure. In the top part the energy function is that used in FlexX, and in the center part it is E_t (eq 1), the energy used for reranking. Energy scales are shifted such that the structure of lowest energy is assigned zero energy to facilitate comparisons between figures. Four symbols are used (dot, circle, plus, and cross), one for each respective quartile of the FlexX ranking. The same symbol is used for corresponding ligand structures in both plots to make visible the sorting effected by the reranking. Note that the rmsd values can change from the top part to the center part due to the energy minimization. The minimized crystal structure is indicated by a diamond in the center part. For clarity, only the first 100 FlexX-generated structures are shown in this and the following figures if not mentioned otherwise. The bottom part shows the superposition of crystal structure (dark gray) and structure of lowest E_t (light gray) with P and N atoms indicated for orientation. Hydrogen atoms are omitted here and in the following figures.

screened Coulomb interaction during minimization to facilitate larger conformational corrections to the position of the charged headgroup. However, this does not lead to better results: the distances between the corresponding positively charged nitrogens of the crystal structure and the minimized structures are still larger than 6 Å. In terms of rmsd the prediction based on the minimum value of E_t is still 3 Å off the crystal structure (compared to more than 9 Å before reranking). Given the large size and high flexibility of the ligand with its 26 rotatable bonds, this result of the two-stage method is satisfactory.

1icn. In the complex 1icn an oleate molecule is bound to a fatty acid-binding protein.[22] According to the published crystal structure the carboxylate group of the oleate is buried at the bottom of the hydrophobic pocket and the hydrocarbon tail extends from there up to the surface of the protein. The carboxylate is disordered, and three alternate positions are reported in the PDB.

Two FlexX runs are carried out. In the first run, the carboxylate is assumed to be deprotonated and hence negatively charged. FlexX generates 147 structures, but only 9 of these with ranks of 130 and higher had rmsd's of less than 3 Å and none less than 2 Å. By far the majority of structures has rmsd values of 8−12 Å. In these structures the volume of the oleate superimposes well with the conformation deposited in the PDB. It only differs in the orientation of the molecule, in that the carboxylate group is not located at the bottom of the hydrophobic pocket but exposed to the solvent and to polar groups on the surface of the protein.

In the second FlexX run the carboxylate group is assumed to be protonated. This protonation state is probably correct for each of the three positions of the carboxylate given in the PDB. Otherwise the negative charge of the carboxylate would be buried in a hydrophobic environment with no countercharge to compensate the effect of desolvation. To estimate the pK_a of the carboxylate in the hydrophobic pocket, we use the MEAD program.[23,24] For each of the three positions of the carboxylate group, the calculated pK_a is higher than 11, meaning that under physiological conditions the carboxylate in the pocket should be protonated. In the FlexX run with protonated carboxylate some structures with small values of rmsd come closer to rank 1. However, the structures of best rank and the majority of conformations are still oriented as before with the carboxylate exposed to the solvent and hence have large values of rmsd (top part of Figure 3).

The reranking in the second stage of our method does not seem to improve the ranking (center part of Figure 3). The three alternate structures given in the PDB have energies that are significantly higher than those of the six highest ranking conformations generated by FlexX, that all have solvent-exposed carboxylates and rmsd's of 11 or more. At first sight this looks as if the two-stage method has failed completely for 1icn. However, it may be that this complex is one of the rare cases where the predicted structure is closer to the truth than the structure found in the PDB. The crucial hint comes from the crystallographers themselves, who observe "additional weak and continuous electron density that extends beyond the location of the terminal methyl of oleate in the wild-type holoprotein. This weak density is J-shaped and extends over a distance equivalent to four methylenes".[22] The results of our calculations suggest that this electron density could not be due to the hydrocarbon tail of oleate but more probably due to its carboxylate which may be exposed to the solvent. Furthermore we suspect that the disordered "carboxylates" given in the PDB are more likely alternate positions of the hydrocarbon tail. Both changes to the structure are in accord with our computational results and probably also with the crystallographic electron density. The orientation of the oleate in the PDB may be an artifact of the refinement procedure where the wild-type protein with bound myristate was used as an initial model. In contrast to the wild-type protein which has an arginine at the bottom of the binding pocket, in

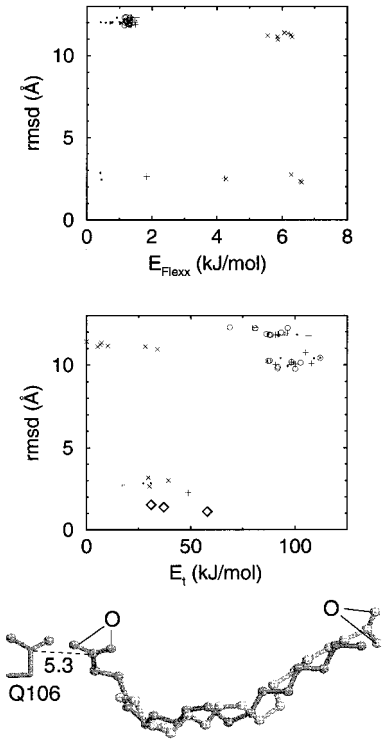

Figure 3. Rmsd and energy before and after reranking for 1icn. Note that there are three alternate positions of the carboxylate of the oleate ligand given in the PDB, corresponding to the three diamonds in the center part. Rmsd values are relative to the first position of carboxylate. The bottom part shows the superposition of crystal structure (dark gray) and structure of lowest E_t (light gray). Carboxylate oxygens, position of side chain of glutamine 106, and distance between glutamine and carboxylate are indicated for orientation.

the mutant protein in 1icn this position is replaced by a neutral glutamine, and hence the ability to form a salt bridge with the carboxylate of the fatty acid is lost. Interestingly, independent investigations using a different docking tool[5] could also not reproduce the PDB structure but rather predicted conformations similar to that preferred by our method.

4hmg. 4hmg[25] is a complex of influenza virus hemagglutinin with its ligand sialic acid. Sialic acid is smaller and more compact than the previous two ligands, but having 12 torsional degrees of freedom it is quite flexible, too. With a negatively charged carboxylate, a carbonyl group, and 6 hydrogen bond donors the ligand is very hydrophilic. The largest hydrophobic group of the ligand is a methyl group. The protein has three binding sites for sialic acid of which we consider the one located on chain A.

FlexX correctly orients the carboxylate group of the ligand toward a cluster of hydrogen bond donors around asparagine 137 and glutamine 226. However, the proposed positions of the remainder of the ligand are spreading over a large area. FlexX finds plenty of possibilities to form hydrogen bonds between the many hydrophilic residues lining the vicinity of the binding pocket and parts of the ligand. Conversely, the methyl group of the ligand does not form hydrophobic contacts for most of the proposed positions and in some cases is even fully exposed to the solvent. The resulting conformations typically have rmsd values of 5−6 Å to the crystal structure (Figure 4). A smaller distinct group of

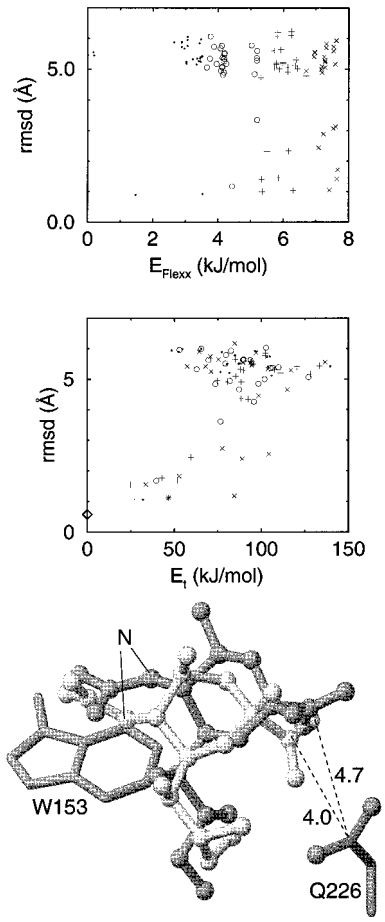

Figure 4. Rmsd and energy before and after reranking for 4hmg. The bottom part shows the crystal structure (dark gray) and the structure of lowest E_t (light gray). Also shown for orientation are the side chains of tryptophan 153 and glutamine 226, the positions of the ligand N atoms, and the distances between the ligand carboxylates and glutamine.

conformations with rmsd's of 1−3 Å correctly places the methyl group in the hydrophobic pocket between tryptophan 153 and glycine 134, but this group does not rank at the top according to FlexX.

The reranking procedure leads to a relative shift in energy of the two groups of conformations with high and low values of rmsd. Ligand conformations with methyl groups buried in the hydrophobic pocket and rmsd's of 1−3 Å now rank at the top (center part of Figure 4). Their energies are significantly lower than those of the group with rmsd's of 5−6 Å. The minimized crystal structure has an energy that is even lower, because the sum of Lennard−Jones and torsion energy is particularly low for this conformation. This means that for 4hmg the reranking works well.

1dwd. The complex[26] of human α-thrombin with *N*-α-(2-naphthylsulfonylglycyl)-D-*p*-amidinophenylalanylpiperidine (NAPAP)[27] is a system for which FlexX performs well, in that it ranks at the top ligand conformations with rmsd's of 1 Å to the crystal structure. The results on this complex demonstrate that the reranking procedure does not spoil good rankings of a successful docking run.

In the crystal structure NAPAP occupies three pockets of thrombin: in the S_1 pocket at the bottom of the binding site the amidinophenyl group forms a salt

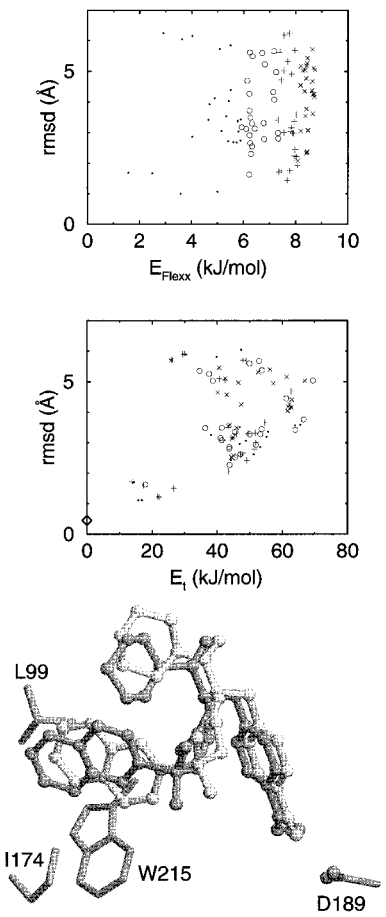

Figure 5. Rmsd and energy before (top) and after (center) reranking for 1dwd. The bottom part shows the superposition of crystal structure (dark gray) and structure of lowest E_t (light gray). Side chains of three amino acids forming pocket S_4 and aspartate 189 are also shown.

bridge with aspartate 189, the piperidine ring resides in the S_2 pocket, and the naphthyl group lies in the S_4 pocket formed by leucine 99, isoleucine 174, and tryptophan 215 (Figure 12). FlexX places amidinophenylalanyl and piperidine groups of all ligand conformations in their proper pockets. Only the naphthyl group shows a broad distribution of positions, with some instances in the correct pocket, some extending into the solvent, and others in a more hydrophilic pocket. Ligand conformations on the first three ranks are essentially correct and differ only by the orientation of naphthyl and piperidine rings within their respective pockets. The rmsd's to the crystal structure in this leading group are between 1 and 2 Å (Figure 5).

The reranking procedure sharpens the contours of the distribution of rmsd versus energy (Figure 5) by clustering structures of similar rmsd in the same energy minimum (center part of Figure 5), in particular the 11 structures with rmsd below 2 Å. According to FlexX the ranks in this group are spread quite uniformly between 1 and 75 (of 100). After reranking this group has ranks 1−10 and rank 14. Consequently, structures with the hydrophobic naphthyl ring in more hydrophilic environments are pushed to ranks worse than 10. The minimized crystal structure has a value of E_t that is still significantly lower than that of the leading group. An analysis of the energetic components (data not shown) reveals that leading group and crystal structure are favored mainly by Lennard−Jones interaction and also—to a lesser extent—by the nonpolar solvation energy E_{np}, indicating that these conformations are well-adapted to the shape of the binding site.

It could be argued that the dominance of Lennard−Jones interactions may be an artifact due to the use of a rigid protein: The rigid protein "lock" probably overemphasizes the necessity for the ligand "key" to match the geometric shape of the binding pocket. An optimal matching implies that the Lennard−Jones energy in the complex is minimized. To test whether the dominance of the Lennard−Jones interactions in this complex is lost if the conformation of the protein is changed, we docked NAPAP in thrombin taken from the crystal structure of the complex 1dwc of argatroban and thrombin.[26] The rmsd of the two proteins after optimal superposition is about 0.2 Å. The picture after reranking the conformations generated by FlexX closely resembles the center part of Figure 5 with the Lennard−Jones interactions still dominating. Qualitatively similar results were obtained for the docking of argatroban into the thrombin structures taken from 1dwd and 1dwc (not shown). Hence, the dominance of the Lennard−Jones interactions in these thrombin complexes is not a mere artifact of the rigid protein structure used in the calculations. This point is discussed further in the conclusions.

2cgr. 2cgr[28] is the complex of an immunoglobulin Fab fragment with the ligand *N*-(*p*-cyanophenyl)-*N*′-(diphenylmethyl)guanidineacetic acid (GAS). In the crystal structure the ligand entertains several specific interactions with the protein that we use to classify the conformations generated by FlexX: a charged hydrogen bond between the carboxylate and arginine 57, a hydrogen bond of the cyano group with buried water 102, and various interactions of the three phenyl rings of GAS with surrounding aromatic groups of the protein.

Most of the conformations generated by FlexX differ significantly from the crystal structure (top part of Figure 6). This is also true for the six top-ranking conformations which only have in common the charged hydrogen bond between the carboxylate and arginine 57. In this group of conformations the cyanophenyl group points into the solvent, whereas another phenyl ring sits in the pocket which in the crystal structure is occupied by the cyanophenyl group. The third phenyl ring also does not fill the pocket which it fills in the crystal structure. Consequently, the rmsd to the crystal structure lies at high values of about 6 Å and more. The first conformation with a binding mode that is essentially correct is ranked eighth with a rmsd of 1 Å.

Reranking shifts the conformations with wrong binding modes up in energy E_t relative to those conformations that are close to the crystal structure. After energy minimization, a group of four conformations with rmsd of 0.6 Å has the lowest energies. These conformations are practically identical with the minimized crystal structure, which means that the reranking was successful for 2cgr. The Lennard−Jones energy again is the most important energy term that discriminates between the leading group and the rest of the conformations. Coulombic energy also contributes to this discrimination but does so to a lesser extent.

2mcp. The complex 2mcp of an immunoglobulin Fab

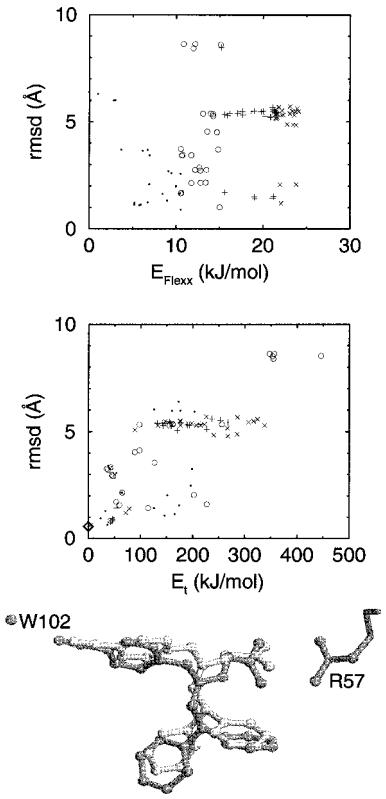

Figure 6. Rmsd and energy before (top) and after (center) reranking for 2cgr. In the bottom part the crystal structure (dark gray) and the structure of lowest E_t (light gray) are superposed. Water 102 (W102) and arginine 57 mentioned in the text are also shown.

with phosphocholine[29] is an example that clearly demonstrates the effectiveness of the two-stage method. In the crystal structure the ligand lies in a pocket in which the two charged groups—the positive choline and the negative phosphate—interact with groups of opposite charge or polarity, for instance, glutamate 35 and aspartate 97 and arginine 52. The more hydrophobic choline group is mainly buried in a pocket, whereas the more hydrophilic phosphate is partly exposed to the solvent.

FlexX basically generates two groups of conformations (top part of Figure 7): a larger group with rmsd values of around 7 Å with ranks 1−320 and a smaller group with rmsd values between 1 and 4 Å with ranks 320−400. The first group does not occupy the phosphocholine binding pocket but rather sits on the surface of the protein. The only feature in common with the crystal structure is the binding of the phosphate group to arginine 52, but with the phosphate approaching the guanidino group from a different side. The choline groups are spread over a wide area forming contacts with various polar groups. The second group of conformations inserts the choline groups into the phosphocholine binding pocket but has the phosphate positions spread over a space spanning several angstroms. Some conformations of the second group are close to the conformation of the crystal ligand.

Reranking reverses the relative rankings of the two groups of ligand conformations. Conformations of the group with rmsd values of 7 Å now have higher energies than the group close to the crystal structure. The ligand structure with the lowest value of E_t previously had a

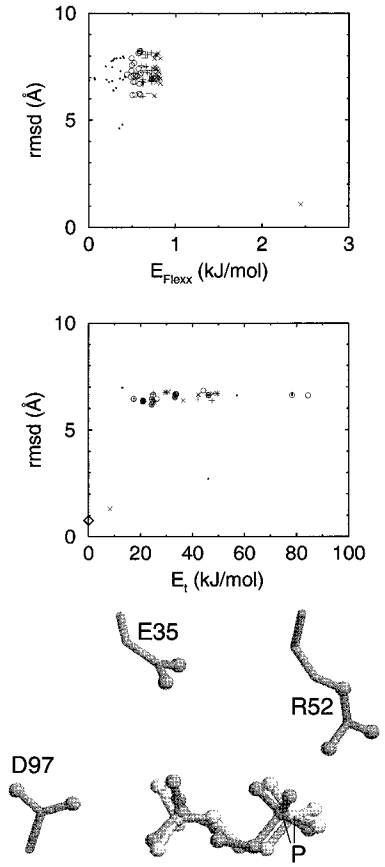

Figure 7. Rmsd and energy before (top) and after (center) reranking for 2mcp. The top part only shows the first 100 conformations generated by FlexX and the conformation on rank 320 (single cross in lower right-half of top part). The center part shows these conformations and in addition the minimized crystal structure (diamond) after reranking. Note that FlexX generates more than one structure with rmsd lower than 2 Å, but these have ranks higher than 320 and are not shown here. Bottom part: superposition of crystal structure (dark gray) and structure of lowest E_t (light gray). Also shown are three amino acids mentioned in the text and positions of ligand P atoms.

FlexX rank of 320. After minimization the rmsd of this conformations is 1.3 Å. The rmsd of the minimized crystal structure to the original crystal structure is 0.8 Å, and its E_t is still lower than that of the best FlexX structure. Hence, E_t after energy minimization successfully recognizes conformations close to the experimental structure. The energetic components contributing most to this result are the Lennard−Jones term, the Coulomb term, and the nonpolar solvation term.

1imb. The complex 1imb of inositol monophosphatase with its ligand L-*myo*-inositol-1-phosphate[30] probably is a difficult task for any docking algorithm, if the positions of crystal waters are not known in advance. This is because half of the binding pocket is filled with more than 10 water molecules that are in contact with the ligand. An oxygen atom of the phosphate group of the ligand is part of an octahedral coordination shell of a Gd^{3+} ion. The sugar moiety forms hydrogen bonds to crystal waters and to acceptors of the protein.

In the first docking attempt we ignore the presence of the crystal waters. Whereas many phosphate positions of the structures generated by FlexX are near the gadolinium ion, the sugar moiety is scattered over

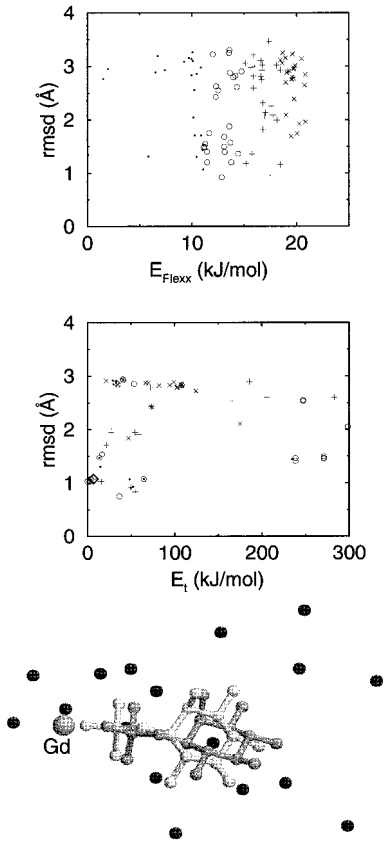

Figure 8. Rmsd and energy before and after reranking for 1imb. Seventeen crystal water molecules in the binding pocket have been taken into account to generate these data. These molecules are shown in the bottom part (black spheres) together with the superposition of crystal structure (dark gray) and ligand structure with lowest E_t. Also indicated is the position of the Gd^{3+} ion.

large parts of the binding pocket. Conformations on ranks 1−75 have rmsd's in the range of 4−7 Å, and the first ligand conformation with an acceptable rmsd of 1.8 Å is on rank 76 (see Table 1). Reranking of these conformations after minimization does not improve the situation. Minimization often leads to large movements of the ligand within the pocket, partly due to the strong Coulombic interactions between the charged groups lining the surface of the pocket, namely the gadolinium ion and six carboxylates, and the charged or polar parts of the ligand. Many ligand positions after minimization overlap with the water positions in the crystal structure. Even for the crystal structure the ligand moves by more than 3 Å during energy minimization.

These results suggest to explicitly consider crystal waters in the calculations. Therefore, 17 crystal waters with oxygens closer to the ligand than 6 Å are reintroduced into the binding pocket. Hydrogen positions are generated and energy-minimized with CHARMM. During this minimization the water oxygens are fixed and the ligand is not present. The ligand is then docked using FlexX. The new results are strikingly different from the former ones. Now, we find a structure on rank 1 with a rmsd of 0.8 Å. Moreover, the range of rmsd's is restricted to values below 4 Å (Figure 8). Since the result of the docking stage is already quite accurate, the reranking cannot improve much. The minimized ligand of the crystal structure is almost identical to the minimized best FlexX structures with rmsd's from the original crystal structure of about 1 Å. Interestingly, there are some ligand conformations with rmsd's below 1 Å but higher values of E_t, mainly due to strained bond angles and Lennard−Jones interactions. In conclusion we find that for 1imb the crystal waters are an integral structural part of the binding pocket that has to be taken into account explicitly for meaningful results. These water molecules provide specific interactions with the ligand as well as excluded volume that constrains the ligand to a pocket that is effectively smaller. We suspect that the problem of predicting the correct structure of this particular protein−ligand complex in absence of the crystal waters is intractable for any docking algorithm available today.

1tni. 1tni is a complex of trypsin with the small inhibitor 4-phenylbutylamine.[31] In the crystal structure the amine of the ligand binds to aspartate 189 and neighboring carbonyl groups in the specificity pocket of the protein. The rest of the ligand mainly forms hydrophobic contacts to nonpolar parts of the pocket, and there is also a weak CH···O hydrogen bond between the oxygen of crystal water 269 and the phenyl ring. The temperature factors of the ligand are high with values in the range of 40−50, indicating that the ligand position is not as well-defined as in the other cases investigated, where *B*-values typically are between 10 and 30. This is not surprising since the ligand is small in comparison to the volume of the binding pocket and does not interact very specifically with the protein.

As for 1imb we run two series of calculations: first without crystal waters and second with seven crystal waters in the binding pocket closer than 6 Å to the crystal ligand. FlexX generates ligand structures that occupy the correct binding pocket in the right orientation but differ from the crystal structure by up to 3.3 Å in terms of rmsd (Figure 9). This is because the ligand is small in comparison with the volume of the pocket and many interactions stabilizing different conformations are available, none of which is particularly strong. The effect of the crystal waters in the docking step differs from that seen for 1imb in that the structure on rank 1 shifts from 2.7 to 3.3 Å rmsd when the crystal waters are introduced.

After reranking the ligand structure of lowest E_t is still 2.5 or 1.9 Å away from the crystal structure for the calculation without or with crystal waters, respectively. These rmsd values may seem relatively large, but we think that the result is acceptable because the overall binding mode and orientation seen in the crystal structure are retained: the amino group of the ligand is interacting with the aspartate 189 and the phenyl group occupies the upper part of the pocket. Moreover, the high temperature factors mentioned above suggest that the ligand position is only loosely defined. In this respect it is not astonishing that the crystal structure moves by 1.5 Å during the minimization step and is overtaken in rank by other structures. Energy terms with the largest impact on the ranking are the Lennard−Jones term, followed by bond angle and torsion energies.

1mup. The complex 1mup of a pheromone-binding protein with its ligand 2-(*sec*-butyl)thiazoline[32] is difficult to treat because the ligand is highly hydrophobic and does not form hydrogen bonds or other directional

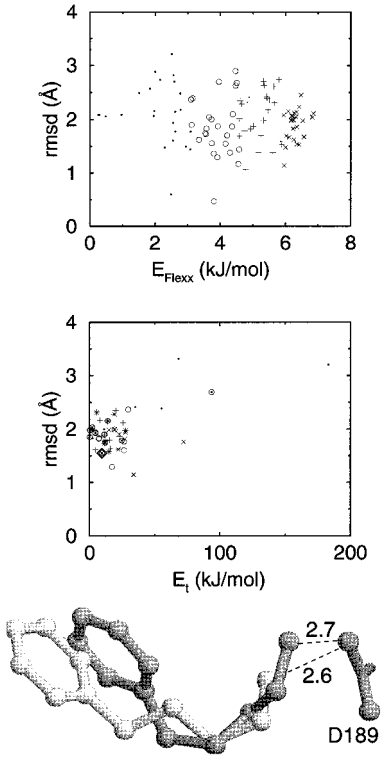

Figure 9. Rmsd and energy before and after reranking for 1tni. Seven crystal water molecules have been considered in the calculations. Bottom part: superposition of crystal structure (dark gray) and structure with lowest E_t (light gray). Also shown are the side chain of aspartate 189 and distances to ligand N atoms.

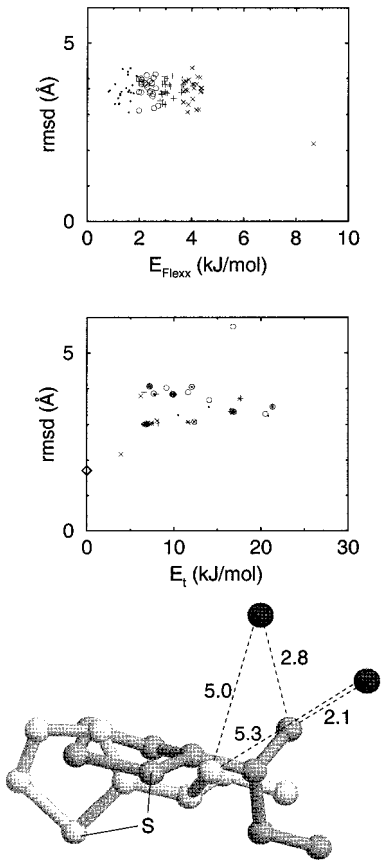

Figure 10. Rmsd and energy before and after reranking for 1mup. Data refer to calculation in which the two crystal water molecules contained in the binding pocket have been considered. Shown are data of the first 100 structures of highest rank according to FlexX, and the structure on rank 156 (single cross in right-half of top part) which has the lowest rmsd to the crystal structure. The bottom part shows the superposition of crystal structure (dark gray) and structure with lowest E_t (light gray). Also indicated are the position of the ligand S atoms and the distances of methyl groups to two water molecules (black spheres) mentioned in the text.

interactions with the protein. The ligand is completely enclosed by the protein in a hydrophobic binding pocket. Two crystal waters cover the only hydrophilic patch of the pocket and also have van der Waals contacts with the ligand.

As for the two previous complexes we carry out two series of calculations: one without and one with the two crystal waters. In both runs the majority of structures have rmsd's of between 3 and 5 Å (Figure 10), and on rank 1 we find a structure with a rmsd of 3.8 Å. Only few structures have rmsd values below 3 Å, and none falls below 2 Å. FlexX mainly generates structures with the thiazoline ring occupying one of three niches of the binding site, whereas in the crystal structure this ring sits in the center of the pocket fixed by hydrophobic groups protruding from the walls of the pocket. The introduction of the crystal waters does not change this situation but rather shifts some of the thiazoline positions overlapping with these waters to another niche.

In the reranking stage the introduction of the two crystal waters leads to a significant improvement. For the complex without waters the structure with lowest E_t is still 3.5 Å off the crystal structure. During minimization the ligand molecule is in many cases shifted to positions that overlap with the crystal waters. This is true in particular for the crystal ligand itself. The reranking does not significantly improve rmsd's. This changes with the introduction of the crystal waters. Now reranking brings the structure ranked 156th by FlexX to rank 1 according to E_t. This structure has an rmsd value of 2.2 Å, which is the minimum value of all structures generated by FlexX. In this sense the reranking is also successful for 1mup. The energetic component mainly responsible for the ranking is the Lennard–Jones term of E_t.

Thrombin and Inhibitor BCPA. One of the best means to test a method that claims to be predictive is bona fide predictions, because the application of the method and the assessment of the results are carried out independently. Here, we predict the structure of the complex of thrombin with the new inhibitor BCPA or benzylsulfonyl-D-cyclohexylglycyl-prolyl-4-(amidomethyl)amidinopiperidine. The chemical structure of BCPA is closely related to that of the members of a group of new thrombin inhibitors.[33,34] Benzylsulfonyl-D-diphenylalanyl-prolyl-*trans*-4-(amidomethyl)cyclohexylamine[33] can be considered the template inhibitor. For the complex of this template inhibitor with thrombin a X-ray structure exists,[33] which is not available to us. Based on this template we have synthesized the new inhibitor I11 or benzylsulfonyl-D-cyclohexylalanyl-prolyl-4-(amidomethyl)amidinopiperidine that binds to thrombin with a K_i of 0.27 nM.[34] There exists no experimentally determined structure of the complex of I11 and thrombin. Recently, it has been found for a related inhibitor that the substitution of the D-cyclohexylalanine by a D-cyclohexylglycine residue was fol-

lowed by a reduced affinity to the fibrinolytic enzymes plasmin and tissue-type plasminogen activator, while sufficient antithrombin potency was maintained.[35] As a consequence this inhibitor showed no inhibitory effects on endogeneous fibrinolysis even at very high doses, which is important for clinical development. This suggests to replace the D-cyclohexylalanine of I11 by D-cyclohexylglycine, which leads to BCPA. As for I11, there exists no experimentally determined structure of the BCPA−thrombin complex.

The ligand BCPA is docked into the thrombin structure taken from complex 1dwd.[26] Two FlexX runs with different starting conditions (detailed below) are performed to generate a greater conformational diversity. The conformations of the two runs are then collectively submitted to the reranking procedure.

The first run is carried out in analogy to the docking for complex 1dwd above. FlexX automatically chooses and places the base fragment and generates 171 ligand structures. The structures ranked best by FlexX have the amidinopiperidine of BCPA positioned in the S_1 site similar to the benzamidine group of NAPAP. The positions of the N-terminal benzyl ring are scattered over a wide range between phenylalanine 60 and isoleucine 174. There is a tendency toward elongated ligand conformations.

In the second FlexX run we start with the proline residue of BCPA as base fragment placing it manually in the S_2 site such that its position is exactly that of the proline residue in the complex 1hbt of thrombin with a bivalent peptidic inhibitor derived from the C-terminal part of hirudin.[36] The best structures out of the 191 generated in this run again have the amidinopiperidine positioned in the S_1 pocket. In contrast to the previous FlexX run, the new conformations have a higher content of turn-like structures with the N-terminal benzyl group closer to the amidinopiperidine moiety.

Conformations of both runs are submitted to the same reranking protocol consisting of minimization and computation of E_t. Our bona fide prediction, that is the structure with the lowest value of E_t, is shown in Figure 12 superimposed with NAPAP in the binding pocket of 1dwd. This structure is generated in the second FlexX run where it ranks 67th. Figure 11 shows rmsd's of all conformations with respect to the bona fide prediction as a function of FlexX energies and values of E_t. Generally, the structures of the second run are less than 3 Å away from the conformation with minimum value of E_t, whereas those of the first run have rmsd's of more than 3 Å to this conformation. While the FlexX energies of the two runs are spread over about the same range, the values of E_t are falling into two almost nonoverlapping intervals. The conformations of the first run after minimization have on the average significantly higher values of E_t than the structures of the second FlexX run. This means that there is a clear energetic preference for more turn-like structures with the amidinopiperidine inserted into S_1 and the phenyl ring at the other end of the molecule close to S_1. This is also true for the structure with the lowest E_t where the smallest distance between two carbons in the amidinopiperidine ring and in the phenyl ring is 3.8 Å (Figure 12). The Pro of this structure occupies the S_2 pocket formed by tyrosine 60A and tryptophan 60D of the insertion loop of thrombin

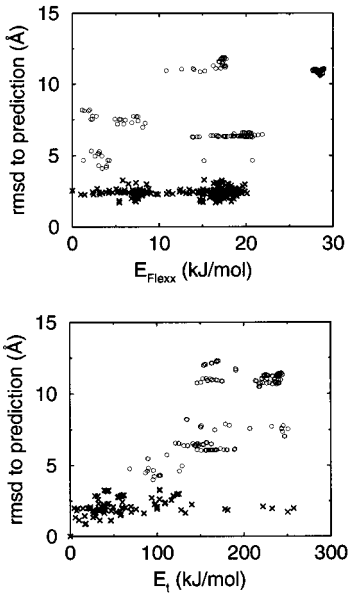

Figure 11. Bona fide prediction of the complex of thrombin and inhibitor BCPA. The data of the first FlexX run are shown as circles, the data of the second run as crosses. Rmsd values are given with respect to the predicted structure, which is the one with lowest E_t. Top half: rmsd as function of FlexX energy. Bottom half: rmsd as function of E_t after energy minimization of the FlexX-generated structures.

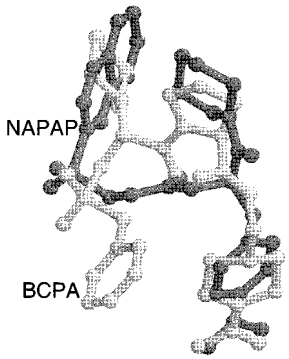

Figure 12. Bona fide predicted conformation of inhibitor BCPA (light gray) superimposed on NAPAP (dark gray) in the binding pocket of thrombin. Prediction based on protein structure taken from complex 1dwd.[26] Three of the four ring systems of BCPA occupy positions of ring systems of NAPAP with similar physicochemical features. Positions of the four ring systems of BCPA counterclockwise starting on the lower right: the amidinopiperidine sits in S_1, Pro is in S_2, the cyclohexyl ring occupies the so-called "aryl binding" site,[39] and the benzyl ring is bent back toward amidinopiperidine. Coordinates are available as Supporting Information.

and by histidine 57 and leucine 99. The cyclohexyl ring is in S_3 position between leucine 99, isoleucine 174, and tryptophan 215. Interestingly, for one of the related inhibitors a similar binding mode has been described based on a X-ray structure.[33]

Cartesian coordinates of the predicted BCPA conformation are available as Supporting Information in the World Wide Web edition of this Journal.

Conclusions and Outlook

The two-stage method, consisting of a fast docking step generating a large number of plausible conformations and a reranking step with a more detailed energy function, has been successfully applied in a number of

cases that pose difficult problems if the docking algorithm is used alone. This probably applies not only to FlexX, the docking tool used in this study, but to other docking tools as well. This is supported by the fact that three of the cases investigated here (2mcp, 1mup, 1icn) also were found intractable by the developers of the GOLD program.[5]

In most of the cases investigated, FlexX is able to generate ligand conformations close to the crystal structure, although not on rank 1. The only clear exception from this rule is the complex 1poc where FlexX is not able to sample all relevant parts of the conformational space. However, there the ligand with its 26 torsional degrees of freedom is so large and flexible that the rmsd of 3.1 Å for the best prediction after reranking is still acceptable. A signature of this sampling problem is the gap between the energy E_t of the best prediction and the much lower energy of the minimized crystal structure: the sampling did not hit the exact position of the native energetic minimum. A smaller but still notable energetic gap is also present for two other complexes with larger ligands (4hmg and 1dwd), signifying the same sampling problem. Despite this problem, in the latter two cases the best predictions using the two-stage method are satisfying in terms of rmsd to the crystal structure with values of less than 2 Å. Overall, with the exception of 1poc, the rmsd between best prediction and crystal structure lies between 0.5 and 2.2 Å (Table 1).

In general, no single component of the energy function E_t is sufficient to ensure that a ligand structure close to the experimentally determined structure is identified as the best prediction. However, in the majority of cases the Lennard−Jones term of the force field contributes most to the successful reranking. This is not surprising, since FlexX penalizes atomic overlaps only in a crude way and deliberately uses an atom model that is too soft in order to counteract the rigidity of the receptor model and other simplifications of the molecular model. Thus it avoids premature rejection of structures that are essentially correct but only slightly overlap with the receptor. Hence, reranking tends to pick among the many plausible structures generated by FlexX those ligand structures that after minimization sterically fit best into the binding pocket, that is, those which have low Lennard−Jones energies. Solvation energy turns out to be not discriminative for the complexes treated here. This may be attributed to the fact that the energy function used in FlexX implicitly considers solvation effects.[12]

There are a number of limitations of the two-stage method in its current form of which we mention two. The first one is certainly the rigid receptor model, which prevents an adequate treatment of induced fit phenomena. Fortunately, for many enzymes the rigid model is a reasonable approximation that allows to obtain meaningful results, as demonstrated in this work. The second limitation is related to the role of the crystal waters. In three of the complexes considered here, crystal waters at well-defined positions are an integral structural part of the respective binding pocket. In principle, water molecules can be treated as additional ligands that are docked as well. For instance, FlexX offers the option to place water-like particles in the binding site while the actual ligand is docked.[37] For the three complexes investigated here in which water has to be considered explicitly, this approach does not lead to improvements compared to docking results without water. The pragmatic alternative we have used in this work is to treat waters as part of the receptor if their positions are known. However, one should also consider docking calculations using various subsets of water molecules. Otherwise, ligands that replace certain structural waters could be missed, as for instance urea-based inhibitors that replace a structural water in HIV protease.[38]

The aim of this work was to study the principal viability of the proposed method to predict structures of protein−ligand complexes, rather than the performance of the method in terms of CPU time. Currently, the CPU time is dominated by far by the reranking step, which for 100 ligand conformations typically takes a few hours on a single processor of a workstation. However, the efficiency of the reranking step can be improved greatly by a number of measures. For instance the time-critical second stage can be easily accelerated by clustering similar conformations generated by FlexX and subsequent parallel processing of representatives of these clusters. Another way to increase throughput in the second stage is to carry out several cycles of minimization. In the first cycle all complex structures are submitted to a smaller number of minimization steps. Structures at the upper end of the energy scale are dropped. This is repeated through several cycles and thus can reduce the number of candidate structures and hence the amount of CPU time from hours to minutes.

Acknowledgment. We wish to thank Kees Oosterlee, Joannis Apostolakis, and Astrid Maass for helpful discussions and also the anonymous referees for constructive remarks. B.K. and M.R. acknowledge funding by the RELIMO project under Grant 0311620 of the German Federal Ministry for Education, Science, Research, and Technology (BMBF).

Supporting Information Available: Cartesian coordinates of the predicted BCPA conformation. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kuntz, I.; Blaney, J.; Oatley, S.; Langridge, R.; Ferrin, T. A Geometric Approach to Macromolecule−Ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.
(2) Goodsell, D.; Olson, A. Automated Docking of Substrates to Proteins by Simulated Annealing. *PROTEINS: Struct. Funct. Genet.* **1990**, *8*, 195−202.
(3) Welch, W.; Ruppert, J.; Jain, A. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* **1996**, *3*, 449−462.
(4) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. Predicting Receptor−Ligand Interactions by an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.
(5) Jones, G.; Willett, P.; Glen, R.; Leach, A.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(6) Baxter, C.; Murray, C.; Clark, D.; Westhead, D.; Eldridge, M. Flexible Docking Using Tabu Search and an Empirical Estimate of Binding Affinity. *PROTEINS: Struct. Funct. Genet.* **1998**, *33*, 367−382.
(7) Rarey, M.; Kramer, B.; Lengauer, T. Multiple automatic base selection: Protein−ligand docking based on incremental construction without manual intervention. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 369−384.
(8) Jackson, R. M.; Sternberg, M. J. E. A Continuum Model for Protein-Protein Interactions: Application to the Docking Problem. *J. Mol. Biol.* **1995**, *250*, 258−275.

(9) Gschwend, D. A.; Kuntz, I. D. Orientational sampling and rigid-body minimization in molecular docking revisited: On-the-fly optimization and degeneracy removal. *J. Comput. Aid. Mol. Design* **1996**, *10*, 123−132.
(10) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculation. *J. Comput. Chem.* **1983**, *4*, 187−217.
(11) Hoffmann, D.; Washio, T.; Gessler, K.; Jacob, J. Tackling Concrete Problems in Molecular Biophysics Using Monte Carlo and Related Methods: Glycosylation, Folding, Solvation. In *Proceedings of the workshop on: Monte Carlo approach to Biopolymers and Protein Folding*; Grassberger, P., Barkema, G., Nadler, W., Eds.; World Scientific: Singapore, 1998; pp 153−170; ISBN 981-02-3658-1.
(12) Böhm, H.-J. 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.
(13) Abola, E. E.; Sussman, J. L.; Prilusky, J.; Manning, N. O. Protein Data Bank Archives of Three-Dimensional Macromolecular Structures. In *Methods in Enzymology*; Sweet, R. M., Ed.; Academic Press: San Diego, 1997; pp 556−571.
(14) Sussman, J. L.; Lin, D.; Jiang, J.; Manning, N. O.; Prilusky, J.; Ritter, O.; Abola, E. E. Protein Data Bank (PDB): Database of Three-Dimensional Structural Information of Biological Macromolecules. *Acta Crystallogr.* **1998**, *D54*, 1078−1084.
(15) Warwicker, J.; Watson, H. C. Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. *J. Mol. Biol.* **1982**, *157*, 671−679.
(16) Klapper, I.; Hagstrom, R.; Fine, R.; Honig, B. Focusing of Electric Fields in the Active Site of Cu−Zn Superoxide Dismutase: Effects of Ionic Strength and Amino-Acid Modification. *Proteins: Struct. Funct. Genet.* **1986**, *1*, 47−59.
(17) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268*, 1144−1149.
(18) Hermann, R. B. Theory of Hydrophobic Bonding. II. The Correlation of Hydrocarbon Solubility in Water with Solvent Cavity Surface Area. *J. Phys. Chem.* **1972**, *76*, 2754.
(19) Eisenberg, D.; McLachlan, A. D. Solvation energy in protein folding and binding. *Nature* **1986**, *319*, 199−203.
(20) Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086−3090.
(21) Scott, D. L.; Otwinowski, Z.; Gelb, M. H.; Sigler, P. B. Crystal structure of bee-venom phospholipase A2 in a complex with a transition-state analogue. *Science* **1990**, *250*, 1563−1566.
(22) Eads, J.; Sacchettini, J. C.; Kromminga, A.; Gordon, J. I. *Escherichia coli*-derived Rat Intestinal Fatty Acid Binding Protein with Bound Myristate at 1.5 Å Resolution and I−FABP with Bound Oleate at 1.74 Å Resolution. *J. Biol. Chem.* **1993**, *268*, 26375−26385.
(23) Bashford, D.; Karplus, M. pK_a's of Ionizable Groups in Proteins: Atomic Detail from a Continuum Electrostatic Model. *Biochemistry* **1990**, *29*, 10219−10225.
(24) Bashford, D.; Case, D. A.; Dalvitt, C.; Tennant, L.; Wright, P. E. Electrostatic Calculations of Side-Chain pK_a Values in Myoglobin and Comparison with NMR Data for Histidines. *Biochemistry* **1993**, *32*, 8045−8056.
(25) Weis, W. I.; Brünger, A. T.; Skehel, J. J.; Wiley: D. C. Refinement of the influenza virus hemagglutinin by simulated annealing. *J. Mol. Biol.* **1990**, *212*, 737−761.
(26) Banner, D. W.; Hadvary, P. Crystallographic analysis at 3.0-Å resolution of the binding to human thrombin of four active site-directed inhibitors. *J. Biol. Chem.* **1991**, *266*, 20085−20093.
(27) Stürzebecher, J.; Markwardt, F.; Voigt, B.; Wagner, G.; Walsmann, P. Cyclic amides of N alpha-arylsulfonylaminoacylated 4-amidinophenylalanine−tight binding inhibitors of thrombin. *Thromb. Res.* **1983**, *29*, 635−642.
(28) Guddat, L. W.; Shan, L.; Anchin, J. M.; Linthicum, D. S.; Edmundson, A. B. Local and transmitted conformational changes on complexation of an anti-sweetener Fab. *J. Mol. Biol.* **1994**, *236*, 247−274.
(29) Padlan, E. A.; Cohen, G. H.; Davies, D. R. On the specificity of antibody/antigen interactions: phosphocholine binding to McPC603 and the correlation of three-dimensional structure and sequence data. *Ann. Inst. Pasteur. Immunol.* **1985**, *136C*, 271−276.
(30) Bone, R.; Frank, L.; Springer, J. P.; Pollack, S. J.; Osborne, S. A.; Atack, J. R.; Knowles, M. R.; Mcallister, G.; Ragan, C. I.; Broughton, H. B.; Baker, R.; Fletcher, S. R. Structural analysis of inositol monophosphatase complexes with substrates. *Biochemistry* **1994**, *33*, 9460.
(31) Kurinov, I. V.; Harrison, R. W. Prediction of new serine proteinase inhibitors. *Nat. Struct. Biol.* **1994**, *1*, 735−743.
(32) Bocskei, Z.; Flower, D. R.; Groom, C. R.; Phillips, S. E. V.; North, A. C. T. Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. *Nature* **1992**, *360*, 186−188.
(33) Tucker, T. J.; Lumma, W. C.; Mulichak, A. M.; Chen, Z.; Naylor-Olson, A. M.; Lewis, S. D.; Lucas, R.; Freidinger, R. M.; Kuo, L. C. Design of Highly Potent Noncovalent Thrombin Inhibitors That Utilize a Novel Lipophilic Binding Pocket in the Thrombin Active Site. *J. Med. Chem.* **1997**, *40*, 830−832.
(34) Steinmetzer, T.; Batdorsdhjin, M.; Kleinwächter, P.; Seyfarth, L.; Greiner, G.; Reissmann, S.; Stürzebecher, J. New thrombin inhibitors based on D-Cha-Pro-derivatives. *J. Enzyme Inhib.* **1999**, *14*, 203−216.
(35) Gustafsson, D.; Antonsson, T.; Bylund, R.; Eriksson, U.; Gyzander, E.; Nilsson, I.; Eig, M.; Mattsson, C.; Deinum, J.; Pehrsson, S.; Karlsson, O.; Nilsson, A.; Sörensen, H. Effects of Melagatran, a New Low-molecular-weight Thrombin Inhibitor, on Thrombin and Fibrinolytic Enzymes. *Thromb. Haemost.* **1998**, *78*, 110−118.
(36) Rehse, P. H.; Steinmetzer, T.; Li, Y.; Konishi, Y.; Cygler, M. Crystal structure of a peptidyl pyridinium methyl ketone inhibitor with thrombin. *Biochemistry* **1995**, *34*, 11537−11544.
(37) Rarey, M.; Kramer, B.; Lengauer, T. The Particle Concept: Placing discrete water molecules during protein−ligand docking predictions. *Proteins: Struct. Funct. Genet.* **1999**, *34*, 17−28.
(38) Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 Protease: A major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249−284.
(39) Stone, S. R. Thrombin Inhibitors. *Trends Cardiovasc. Med.* **1995**, *5*, 134−140.

JM991090P