ELSEVIER

Development of filter functions for protein–ligand docking

Martin Stahl and Hans-Joachim Böhm

Molecular Design and Bioinformatics, Hoffmann-La Roche, Ltd., Pharmaceuticals Division, Basel, Switzerland

Current docking methods can generate bound conformations of a ligand close to the experimentally observed structure of a protein–ligand complex. However, the scoring functions used to evaluate the potential solutions are not yet reliable enough at giving the highest ranks to the best structure predictions. One approach to this problem is the use of filter functions that are applied to all docked conformations to remove structures with certain energetically unfavorable properties. We present a computationally efficient scheme for such a postprocessing of docking results. For each of the conformations generated for a given protein–ligand complex, four properties are calculated: the fraction of the ligand volume buried inside the binding pocket, the size of lipophilic cavities along the protein–ligand interface, the solvent-accessible surface (SAS) of nonpolar parts of the ligand, and the number of close contacts between non-hydrogen-bonded polar atoms of the ligand and the protein. These four terms were used to filter out the majority of the calculated solutions and to rescore the remaining ones. On a test set of 32 protein–ligand complexes, this protocol significantly improves the accuracy of the structure predictions. © *1999 by Elsevier Science Inc.*

Keywords: drug design, molecular docking, scoring functions, structure prediction

INTRODUCTION

The prediction of the bound conformation of a ligand in the active site of a protein is a major challenge in current structure-based drug design. Docking algorithms designed to perform this task can be formally split into two parts that are computationally demanding for different reasons. One part is the search for the conformational, translational, and rotational space of a ligand in the receptor pocket, which must be performed thoroughly yet quickly. The other part is the energy function that is used to evaluate the interactions between the receptor and the ligand and to score the resulting solutions. Here the challenge is to find functions that incorporate important contributions to the binding in a consistent manner. Ideally, such prioritization methods should also be able to provide an estimate of the binding constant.

Several methods have been proposed for docking of flexible ligands into protein pockets (for reviews see Ref. 1), e.g., the programs FlexX[2,3] GOLD[4,5] and DOCK 4.0.[6] The current experience with these programs indicates that they can generate docked conformations of a ligand close to the experimentally observed structure of the protein–ligand complex. However, it has also become apparent that a major limitation of these approaches relates to problems in differentiating between the correct orientation and artificial orientations generated by the docking algorithm that are in disagreement with the experimental data. The problem is mainly due to the limited accuracy of the present scoring function used in the docking algorithms.[7] Several scoring functions have been proposed for the prioritization of docked ligands. Some docking programs use intramolecular as well as nonbonded attractive and repulsive terms from molecular mechanics force fields.[8,9] This approach has the major drawback that the resulting energies are not a measure of binding free energies in aqueous solution. Therefore, there have been efforts to include additional solvation terms.[10] The approach taken by many groups,[11,12] including ourselves,[13,14] was to use simple empirical functions that are fitted to measured binding affinities for a set of experimentally determined protein–ligand complexes and that can be evaluated rapidly. They contain additive terms for hydrogen bonds, ionic interactions, ligand flexibility, and lipophilic contacts. The score is a direct (if rather crude) estimate of the free energy of binding. While this has led to fast scoring functions, one major limitation is the total lack of penalty terms for energetically unfavorable arrangements. Because protein–ligand hydrogen bonds and salt bridges contribute much to the calculated score, this often results in large binding energies being assigned to structures with bad steric complementarity between protein and ligand, e.g., owing to holes along the interface or solvent exposition of large lipophilic parts of the ligand.

Address reprint requests to: Dr. M. Stahl, Molecular Design and Bioinformatics, Pharmaceuticals Division, Hoffmann-La Roche, Ltd., CH 4070 Basel,Switzerland.E-mail:martin.stahl@roche.com,hans-joachim.boehm@roche.com.

Journal of Molecular Graphics and Modelling 16, 121–132, 1998

655 Avenue of the Americas, New York, NY 10010

1093-3263/98/$–see front matter
PIIS1093-3263(98)00018-7

In this work, we have employed the docking program FlexX[2,3] for structure prediction. It contains a slightly modified version of an empirical scoring function published by one of the authors.[13] One major advantage of the FlexX program is its speed, which makes it feasible to handle libraries consisting of several thousand entries. In the following, we use the term "pose" to denote a conformation together with its orientation in space relative to the protein coordinates. This term is used for experimental structures as well as for structures generated by a docking program.

Apart from computational speed, the quality of docking programs can be assessed in two ways, in both of which the scoring function is the most critical issue.[15,16] One way is to measure how well it performs the task of ranking a set of molecules with respect to the same receptor, which is important in lead finding and optimization. The other is to measure its ability to rank different poses of one ligand such that the pose closest to the experimental structure is scored best. In this work, we have concentrated on the latter criterion, since we believe that successful structure prediction is a prerequisite for the more complex task of selecting putative ligands for a receptor.

We have implemented and tested three penalty functions that calculate the size of lipophilic cavities along the protein—ligand interface, the lipophilic part of the solvent-accessible surface (SAS) of the ligand, and the fraction of the ligand volume buried in the cavity. These functions, although somewhat interrelated, together reflect the fact that a tightly binding protein–ligand complex must possess a high degree of steric complementarity. Since FlexX uses a stepwise buildup procedure of the ligand inside the active site, these functions cannot be part of the scoring function and are therefore applied as filters after all poses have been generated. As an additional term we have investigated close contacts between polar atoms that are not taking part in hydrogen bonds. We present a computational protocol that combines all four functions and improves the quality of the structure predictions of 32 protein–ligand complexes, at the same time greatly reducing the number of poses generated per docking run.

COMPUTATIONAL METHODS

For each protein–ligand complex, a discrete grid-based model of the active site was generated as a first step. A reference ligand pose (X-ray structure) was used to define the active site by means of its geometric center and the radius of the smallest sphere into which the ligand fitted. All protein atoms further than 7 Å away from the surface of this sphere were removed. A rectangular grid with 1-Å spacing was placed on the remaining protein atoms. Grid points were marked as occupied by the protein if the distance to a protein atom center was less than its atomic radius plus a tolerance of 0.8 Å. To arrive at a description of the active site cavity, each of the remaining grid points was assigned an accessibility variable with theoretical values between 0 and 14 in the following way (Figures 1a and b illustrate this procedure for an analogous two-dimensional case): the lines connecting each grid point with the corners and face centers of a hypothetical cube placed around it define 14 directions. Each of these directions was followed on the grid. If a grid point encountered along one of these directions was occupied by the protein, a counter variable was incremented. It was not incremented if the grid boundaries were reached or more than 10 steps in one direction had been taken. The rationale for this 10-Å cutoff is that distant portions of the protein surface should have no influence on the accessibility of a given point and that the surface properties of large and shallow cavities do not differ much from those of convex surface patches. With the accessibility values assigned to each cavity grid point, an outer limit of the cavity was defined by means of an empirical cutoff value. For all proteins in the test set, a cutoff value of 9 (all points with accessibility values >9 form the cavity) proved sensible and yielded reasonable cavity borders approximately on a level with the surface of the protein.

As a next step, the fraction of the ligand volume located inside the cavity was determined. For each of the poses generated by FlexX, the grid points it occupied were determined with the same criteria as for the protein. The number of ligand grid points that fell inside the cavity were then divided by the total number of ligand grid points.

For the calculation of the lipophilic cavities, a similar protocol was applied. The procedure to define the active site cavity was repeated for every ligand pose, this time regarding only those grid points that were occupied neither by the ligand nor by the protein. A two-dimensional example of a cavity defined in this way is given in Figure 1c. A flood fill algorithm was used to select all grid points that were situated at the interface between the protein and the ligand. This was necessary in order to remove small additional cavities within the protein. The algorithm starts at the grid points marking the boundary of the ligand and selects all contiguous sets of grid points with accessibility values equal to 14. This means that only those cavities with no direct access to the outside of the protein were retained. Note, however, that the resulting cavities differ from those obtained by a conventional flood fill algorithm[17] filling the active site from the outside. Such an algorithm would find only those cavities that are completely surrounded by protein or ligand atoms. In the two-dimensional illustration in Figure 1c, for example, a conventional flood fill algorithm would find only the upper cavity portion, whereas the lower part would be flooded. The final step of the cavity calculation is the removal of all those grid points that have polar protein or ligand atoms as their nearest neighbors. All atoms except for carbon and sulfur atoms (only when part of thioethers and disulfide bridges) were regarded as polar in this context.

The SAS of the ligand was calculated numerically. Around each ligand atom, 200 points were evenly distributed on the surface of a sphere with the van der Waals radius of the atom plus a solvent radius of 1.4 Å. All points lying within the SAS of other (protein or ligand) atoms were removed. The number of remaining points is a measure of the atom's SAS. The lipophilic ligand SAS was calculated by adding the SAS contributions of all lipophilic atom types. The same definition of lipophilic atom types was used as for the cavity calculations. In addition, the surface contributions of all carbon atoms that are part of carbonyl or carboxylic acid groups or chemical derivatives of thereof (esters, imines) were disregarded.

Close contacts between polar atoms (N, O) were calculated in the following way: first, all hydrogen bonds between protein and ligand were determined by means of a simple distance criterion. This was necessary because in hydrogen-bonding networks rather close distances between donor and acceptor atoms can be observed. A cutoff value of 2.6 Å was used for all hydrogen acceptor and metal acceptor distances. With this parameter, the algorithm

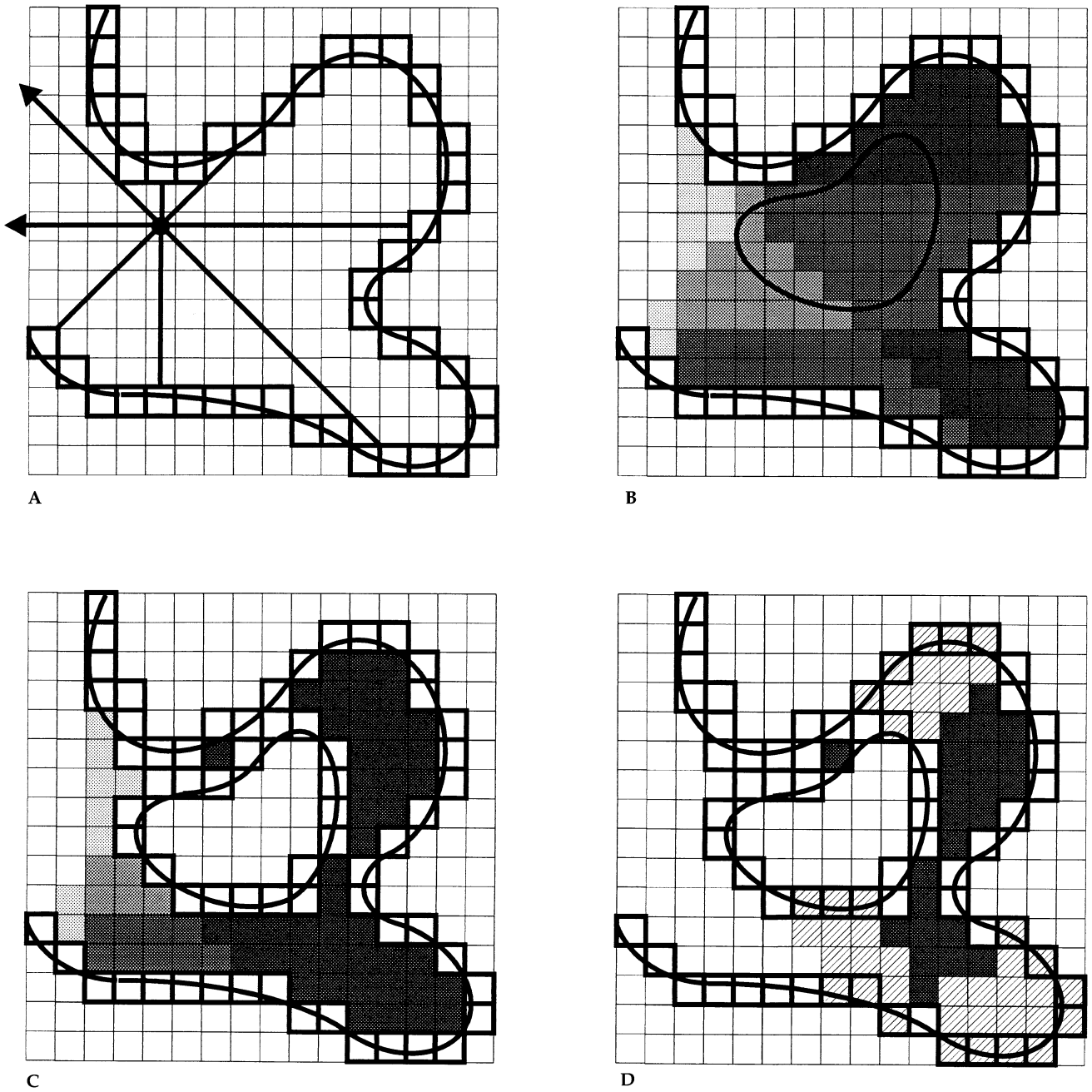

Figure 1. Two-dimensional illustrations of the grid-based filter functions. (a) Each grid point in the active site is assigned an accessibility value by counting the number of times the protein surface is hit when moving along the eight directions. This leads to a "coloring" of the cavity as shown in (b) (dark gray is buried deepest). (c) The same procedure is repeated with the ligand inside the pocket. (d) Cavities where either the ligand or the protein is polar (hatched squares) are removed.

was able to identify all H bond-type favorable polar interactions in the protein–ligand complexes of the data set. Second, all protein–ligand atom pairs that did not take part in a favorable polar interaction were investigated. If their distance was less than 2.6 Å, this arrangement was counted as an electrostatic clash.

The filter routines were implemented as a series of ANSI C modules. On an R5000 SGI O2 computer, the calculation of all four terms is fast enough to be performed interactively for all protein–ligand complexes in the test set. The initialization phase of the receptor took 5–15 s. Subsequently, depending on the size of the ligand and the active site, 3 to 10 ligand poses could be processed per second of CPU time.

A test set of 32 protein–ligand complexes (Table 1) was compiled from the PDB.[18] The test complexes were selected to cover cases found to be problematic with FlexX and representative of some of the major areas of structure-based drug design. All ligands are highly flexible molecules that contain at least three rotatable bonds.

The ligand coordinates were extracted from the PDB files and energy minimized with the MAB force field[19] imple-

Table 1. Test set of 32 protein–ligand complexes from the PDB

Complex No.	PDB entry	Enzyme	Ligand	Resolution (Å)	FlexX changes from default settings
1	1abf	L-Arabinose binding protein	Fucose	1.9	—
2	1cbx	Carboxypeptidase A	L-Benzylsuccinate	2.0	—
3	1die	D-Xylose isomerase	Deoxynojurimycine	2.5	—
4	1eap	Cat. antibody 17E8	Antigen	2.5	—
5	1etr	ε-Thrombin	MQPA	2.2	—
6	1ets	ε-Thrombin	NAPAP	2.3	—
7	1ett	ε-Thrombin	4-TAPAP	2.5	—
8	1glq	Glutathione-*S*-transferase	Glutathione derivative	1.8	Increased sampling, clash factor 0.8
9	1hsl	Histidine binding protein	Histidine	1.9	—
10	1hvr	HIV-1 protease	XK-263	1.8	Increased overlap
11	1lic	Adipocyte lipid binding protein	Hexadecane sulfonic acid	1.6	Increased sampling
12	1mnc	Neutrophil collagenase	PLH	2.1	—
13	1nsc	Neuraminidase	*N*-Acetyl neuraminic acid	1.7	—
14	1nsd	Neuraminidase	DANA	1.8	—
15	1poc	Phospholipase A_2	TS analog	2.0	Increased sampling
16	1ppc	Trypsin	NAPAP	1.8	—
17	1pph	Trypsin	3-TAPAP	1.9	—
18	1rnt	Ribonuclease	Guanylic acid	1.9	—
19	1stp	Streptavidin	Biotin	2.6	—
20	1tmn	Thermolysin	Leu-Trp derivative	1.9	—
21	2cgr	Immunoglobulin IgG_{2b}	Antigen	2.2	—
22	2ctc	Carboxypeptidase A	Phenyl lactate	1.4	—
23	2ifb	Fatty acid binding protein	Palmitate	2.0	Increased sampling
24	2tmn	Thermolysin	*N*-Phosphoryl leucinamide	1.6	—
25	2tsc	Thymidylate synthase	CB3717	2.0	—
26	3cpa	Carboxypeptidase A	Gly-Tyr	2.0	—
27	3tpi	Trypsinogen	Ile-Val	1.9	—
28	4dfr	Dihydrofolate reductase	Methotrexate	1.7	—
29	4phv	HIV-1 protease	L700,417 (first binding mode)	2.1	—
30	4tln	Thermolysin	Leucine hydroxylamine	2.3	—
31	4tsl	Tyrosine transferase	Tyrosine	2.5	—
32	8gch	γ-Chymotrypsin	Gly-Ala-Trp	1.6	Increased sampling, increased overlap

mented in our in-house modeling package MOLOC. All ligands were docked back into their corresponding protein environment using the default parameter settings of FlexX. In six cases, FlexX was unable to generate structures that resembled the X-ray structures closely enough to be useful for comparison. Docking calculations were rerun for these cases with slightly changed parameter settings as indicated in Table 1. The final test set consisted of between 50 and 800 docked structures per complex, among which there was at least one ligand pose with an root-mean-square (rms) deviation of less than 2.3 Å (in two cases up to 3.1 Å) from the X-ray geometry of the ligand. By subjective judgment, the best prediction assumed the experimentally observed binding mode in all cases.

RESULTS AND DISCUSSION

We have established a computational protocol that applies a series of filters to a set of poses and rescores the remaining ones. The protocol consists of the following steps:

1. From the original pose list, all structures exhibiting close contacts of nitrogen and oxygen atoms that do not take part in protein–ligand hydrogen bonds are discarded.
2. Of the remaining poses, the maximum and minimum values of the buried volume fraction of the ligand are determined. All poses with values less than the average of these two values are discarded.
3. The minimum value of lipophilic cavity volume that occurs in any of the remaining poses is determined. All poses with a volume at least 25 $Å^3$ above this value are discarded.
4. Of the remaining poses, the pose with minimum lipophilic SAS area is determined. The new score for a pose *i* is then $S_{i,new} = S_{i,old} + 0.2\ (surface_i - surface_{min})$.
5. Finally, the poses are reordered according to the new score values.

Note that apart from the coefficient of 0.2 kJ/$Å^2$ in rule 4, the filter rules were not derived by optimizing cutoff values for the test set, but originate from general considerations. In the following section we motivate each of the rules and describe the results obtained with each of them separately, but in the order in which they are applied in the preceding filtering protocol. Finally, results of the protocol as applied to the test set are analyzed and problematic cases highlighted.

Repulsive polar interactions

The function that finds close contacts between non-hydrogen-bonded nitrogen and oxygen atoms simply returns the number of times per pose that such arrangements occur. In general, for 10–50 poses per entry at least one such electrostatic clash was found. In all but two cases, by visual inspection all these poses were identified to be highly unnatural arrangements of functional groups. The exceptions were 2tmn and 1poc, where a phosphonic acid group (2tmn) or a phosphoric acid ester (1poc) binds to a metal cation. FlexX failed to reproduce the bidentate binding mode of these functional groups. As a result, the $–PO_2–$ groups are rotated by about 60° relative to the X-ray orientation and the noncomplexed oxygen atom can get close to nearby polar residues of the protein. Small changes in torsional angles would suffice to remove these clashes. Since this problem could be solved by changing the FlexX parametrization, we did not change our filter function and further on used it as a "first pass" filter for all docked structures. We believe that such a filter should be implemented directly into the scoring function, since it can be applied to every ligand fragment separately already during the buildup phase of the ligand.

Buried fraction of the ligand

The filter function that calculates the fraction of the ligand poses buried in the active site pocket was intended to remove all poses erroneously docked outside the actual active site at the surface of the protein. Poses with a small percentage of buried volume should be filtered out. In Figure 2, the range of the buried fractions found in all poses is plotted as a gray bar for each complex and the value for the X-ray structure is marked as a black diamond. It is immediately obvious that FlexX is unable to construct poses that are situated significantly deeper in the binding pocket than the X-ray structure. This result is not surprising because one of the major determinants of protein–ligand complexes is a good steric complementarity.

In a few cases, FlexX generates structures with a larger percentage of the ligand volume buried than in the crystal structure (1eap, 1poc, 1pph, 4dfr). In both 1eap and 1poc, small rotations of flexible alkyl side chains orient them further into a

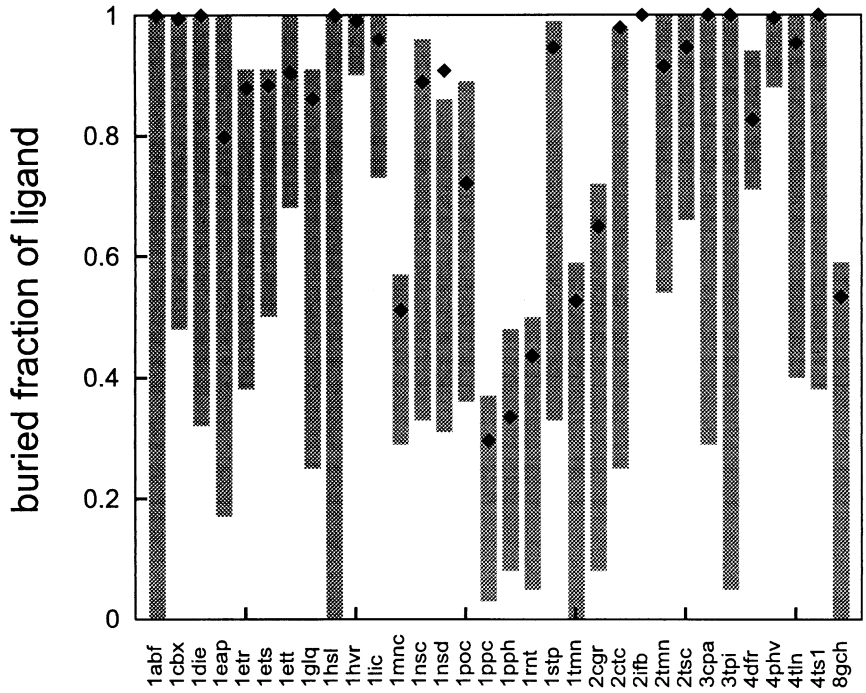

Figure 2. Ranges of buried fractions of ligand volume (gray bars) calculated for all poses generated by FlexX for each complex in the test set. Values calculated for the X-ray pose are marked as black diamonds.

buried portion of the active site than is the case in the X-ray structure. In the complex 4dfr, the ligand methotrexate can move further into the large cavity, thereby losing most of the favorable hydrogen-bonding interactions. In the complex 1pph, a large part of the ligand 3-TAPAP reaches out of the cavity and is exposed to the solvent, partly owing to conformational constraints that are not recognized by FlexX (see below).

Apart from these exceptions, the buried fraction of the ligand volume is a good filter criterion. For its application, a different cutoff criterion must be found for each protein structure, since the active sites of proteins differ vastly in size, form, and accessibility. We have therefore decided to use this filter in the following way: of all poses generated by FlexX for a particular protein–ligand complex, the maximum and minimum fractions of buried volume are determined and all poses with values less than the average of these two are removed from the pose list. We did not use the average value of all poses, since this average may be heavily biased by clusters of conformers originating from a single placement either inside or outside the pocket, the number of cluster members depending strongly on spatial restrictions in the different regions of the pocket. In all 32 complexes in the test set, the poses filtered out in this way included none of the structures close to the X-ray geometry and all of the poses one would regard as being bound outside the active site on visual inspection.

Lipophilic cavities

We now discuss the filter function that calculates the size of lipophilic cavities along the protein–ligand interface. The first question to be addressed is the accuracy of the volume calculation. Owing to the coarse grid resolution and the simple nearest-neighbor criterion used for the definition of a lipophilic volume element, changes in ligand orientation can change the size of the calculated volume to some extent. Using different values for the distance cutoff in the occupancy criterion of the ligand on the grid and different orientations of the grid on the complex coordinates, we found that volume differences between poses are accurate to ± 10 Å^3 for all complexes in the test set. This is sufficiently accurate, since only differences of volumes calculated with the same protein orientation are needed as filter criteria. Figure 3 shows that for most complexes in the test set the lipophilic cavity volume is almost zero. Similar conclusions had been drawn from the analysis of a different data set in an earlier investigation, where it was found that the fit of experimental data to an empirical scoring function could not be improved through the inclusion of a cavity term.[14] While this is true when only the crystal structures themselves are regarded, the lipophilic cavity size can obviously be used as a physically meaningful filter for poses generated by a docking program.

In Figure 3, two complexes (1lic and 2ifb) strike as outliers. In these two cases, a minimum of about 80 Å^3 of lipophilic volume is calculated for all poses. Both proteins bind long chain fatty acids and have a large active site pocket. In the crystal structure, the alkyl chains lie close to one wall of this pocket, thereby minimizing the free lipophilic volume and thus the number of water molecules captured inside the active site.

In two complexes (1glq and 1poc) the free lipophilic volume calculated for the X-ray structure is at the upper end of the range observed in the FlexX-generated poses. For 1poc it has been mentioned that small changes in torsional angle can move a side chain deeper into the cavity. This is accompanied by a reduction in lipophilic cavity volume. In the X-ray structure of complex 1glq, a methyl group is closest to a small lipophilic

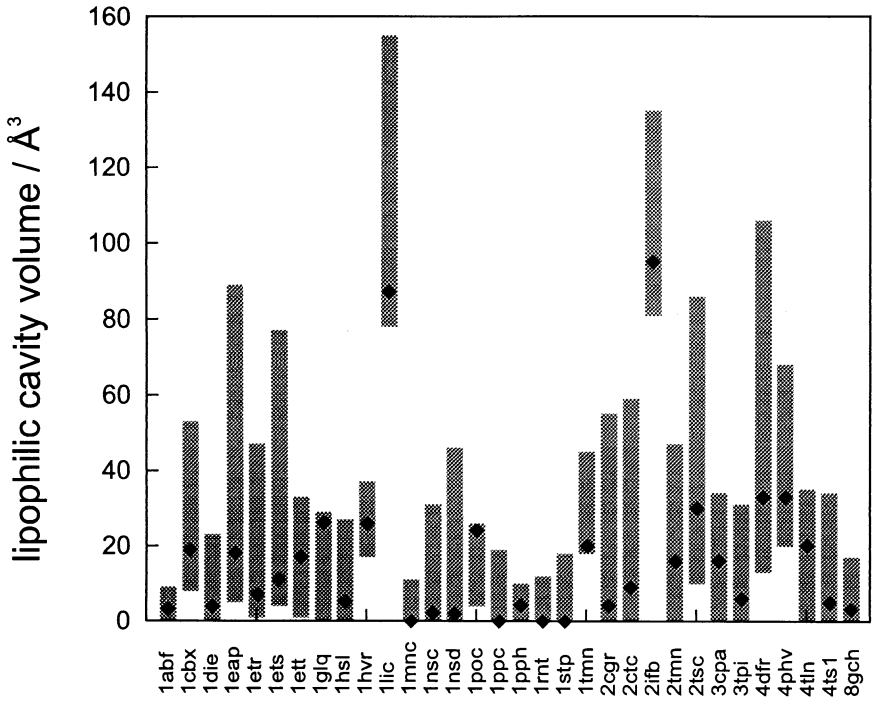

Figure 3. Ranges of lipophilic cavity volume (gray bars) calculated for all poses generated by FlexX for each complex in the test set. Values calculated for the X-ray pose are marked as black diamonds.

hole in the active site, while in all docked conformations this side chain is rotated such that a polar atom is closest to this cavity. As a consequence, the cavity is counted to be lipophilic only in the X-ray structure.

The cavity filter was applied by means of the following cutoff criterion: for every complex regarded, the minimum size of lipophilic cavity volume found in any of the generated poses was determined. All poses with a free lipophilic volume more than 25 $Å^3$ above this minimum value were removed from the pose list, since single cavities larger than this volume could accommodate one water molecule without forming any hydrogen bonds to it. This would clearly be a high-energy state of a complex that is unlikely to occur.

Lipophilic solvent-accessible surface

Just as for the cavity filter discussed above, we found that the size of the lipophilic SAS is a valuable measure for structure prediction (Figure 4). It is obvious that for most complexes, FlexX cannot locate poses that have a significantly smaller exposed surface area than the X-ray structure of the ligand. Again, this is a physically meaningful result, since the burial of lipophilic surface is one of the major driving forces of protein–ligand binding in aqueous solution.

In 21 of the 32 complexes, a surprisingly good structure prediction would be possible on the basis of the lipophilic SAS alone as a scoring function. In these cases, among the five poses with the smallest exposed surface there was at least one structure with an rms deviation (rmsd) not more than 0.4 Å away from the best overall prediction. However, in the remaining 11 cases (1abf, 1die, 1lic, 1poc, 1ppc, 1pph, 1rnt, 2ifb, 4phv, 4tln, and 8gch), the poses with the smallest surface areas are either very different from the X-ray structure or the ligand is highly polar, such that the lipophilic SAS is not a meaningful property. For this reason, we could not find a way to use the lipophilic solvent-exposed surface as an additional filter that further reduced the number of poses per complex without significantly deteriorating the structure prediction for some complexes. Instead, it seemed reasonable to use the lipophilic surface area as a penalty term in combination with the score calculated by FlexX for each pose. This was done in the following way: for each complex, the pose with the minimum surface area was determined and its surface area subtracted from the areas calculated for all other poses. The remaining area was multiplied by a coefficient and added to the score. Values between 0.05 and 0.3 kJ mol^{-1} $Å^{-2}$ were used as coefficients. Optimum results were obtained with a value of 0.2 kJ mol^{-1} $Å^{-2}$, which is close to the value of 0.17 kJ mol^{-1} $Å^{-2}$ used for the lipophilic contact surface in the FlexX scoring function. This combination of the score with the solvent-accessible surface increased the quality of the FlexX structure predictions substantially. An example is given in Figure 5. For both diagrams, only those poses that pass filters 1–3 have been taken into account. Since the poses clustering at about 4.5 Å rms deviation have a greater lipophilic surface area than the other poses with lower rmsd, they are shifted toward lower binding energy. Consequently, the highest ranking pose now has an rmsd of about 2 Å instead of 4 Å and there are two poses with an rmsd of ca. 1.5 Å among the 10 highest ranking poses.

Filters 1–3 reduce the size of the pose list considerably (Table 2). For more than half of the complexes in the test set, the number of poses left after filtering is reduced to less than 80. For seven complexes there were than 200 poses left after filtering.

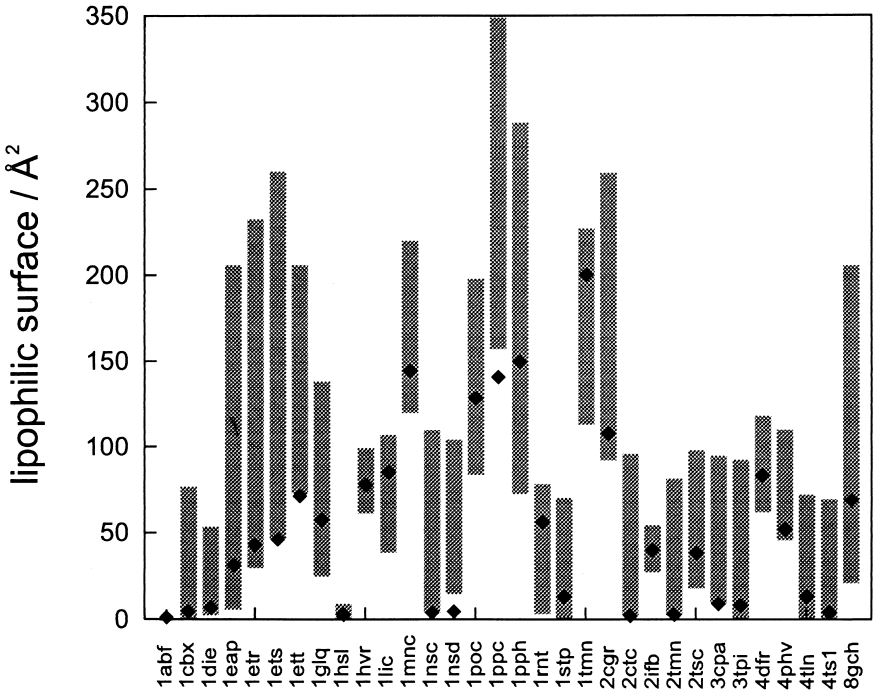

Figure 4. Ranges of lipophilic solvent-accessible surface (gray bars) calculated for all poses generated by FlexX for each complex in the test set. Values calculated for the X-ray pose are marked as black diamonds.

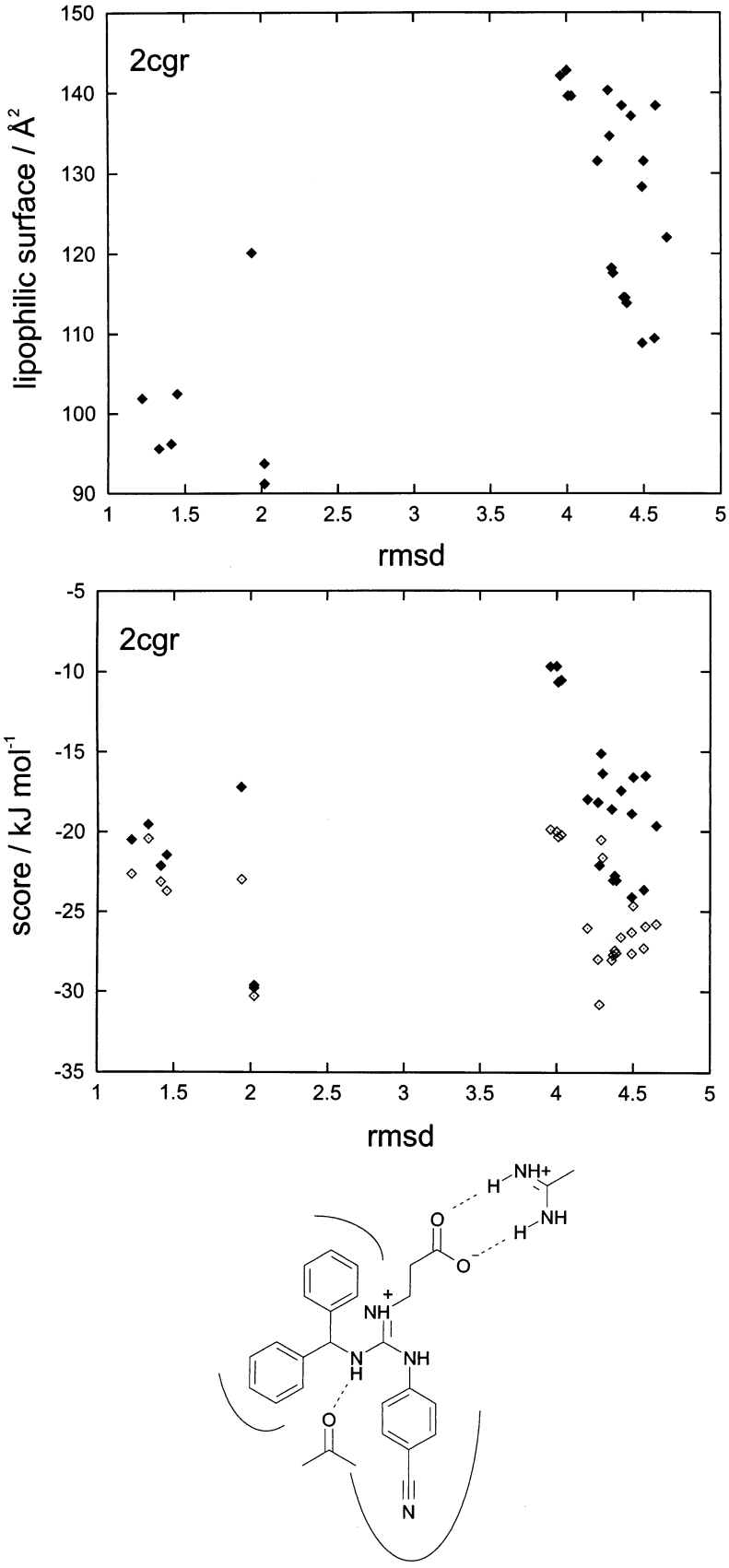

Figure 5. Top and middle: Plots of lipophilic SAS and score against rmsd for all poses of the immunoglobulin–antigen complex 2cgr that are left after treatment with the first three filters. Middle: Open diamonds represent the FlexX score, and filled diamonds represent the score after adding the surface term. Bottom: Structure of the antigen and main interactions with the protein.

Table 2. Reduction in the number of poses by filters 1–3 in the filtering protocol[a]

Complex No.	PDB	Poses before filtering	Poses removed by:			Poses after filtering
			Filter 1	Filter 2	Filter 3	
1	1abf	129	0	2	0	127
2	1cbx	324	41	108	75	100
3	1die	206	10	40	0	156
4	1eap	844	15	347	246	236
5	1etr	310	38	202	8	62
6	1ets	411	6	237	125	43
7	1ett	226	14	104	31	77
8	1glq	388	47	56	4	281
9	1hsl	104	18	3	1	82
10	1hvr	67	0	1	0	66
11	1lic	461	0	442	12	7
12	1mnc	338	7	98	0	233
13	1nsc	288	17	142	101	28
14	1nsd	205	27	50	55	73
15	1poc	746	669	0	0	77
16	1ppc	90	0	31	0	59
17	1pph	337	20	90	0	227
18	1rnt	225	39	58	0	128
19	1stp	262	9	73	0	180
20	1tmn	90	0	44	23	23
21	2cgr	271	54	81	109	27
22	2ctc	226	11	155	9	51
23	2ifb	244	0	0	101	143
24	2tmn	621	34	513	34	40
25	2tsc	293	142	64	49	38
26	3cpa	154	15	71	8	60
27	3tpi	517	1	98	60	358
28	4dfr	233	49	62	48	74
29	4phv	148	30	18	42	58
30	4tln	393	9	131	43	210
31	4tsl	253	42	20	2	189
32	8gch	448	2	119	100	209

[a] Filter 1, polar clashes; filter 2, buried volume; filter 3, lipophilic cavities. See text for details.

Relationships between the filters

The second and fourth steps in the filter protocol are of course closely interrelated. Poses with a large fraction of volume outside the cavity necessarily have a large SAS. However, the buried volume calculation is the fastest filter and usually removes many structures. The more time-consuming surface calculation is then needed only to "fine-tune" the score for the remaining poses. On the other hand, there is usually no correlation between the size of lipophilic cavities and the lipophilic SAS area (exceptions are proteins with large cavities that can accommodate water). The cavity volume is a local property that can change with little effect on the surface area. The only relation that exists between these two properties is the fact that poses with large cavity volumes usually cannot have a small exposed surface.

Improvement of structure predictions

Table 3 summarizes the results obtained with this protocol for all 32 complexes in the test set. Three criteria were used to assess the quality of the structure predictions: the rmsd of the highest ranking solution, the smallest rmsd of the 20 highest ranking predictions, and the rank of the best prediction. In the following discussion only differences between rms deviations of ± 0.1 Å are regarded as relevant. It is obvious that the filtering and reranking protocol improves the structure predictions significantly: the solution with the lowest rmsd is never filtered out and its rank improves in all cases but two, 1rnt and 2tsc. In 11 cases, the rmsd of rank 1 improves; in 14 cases there is a pose with a smaller rmsd between ranks 1 and 20 after the filter application. This is illustrated in Figure 6. In both diagrams, the uppermost curve describes the theoretical case that the best prediction for each complex is on rank 1, which can be regarded as the limit of the data set. It is apparent that with the FlexX + filter protocol, for most complexes the best predictions are now found among the 20 highest ranking poses.

The most notable exception to this general improvement of structure prediction is the complex 1pph. This is the only complex where both the average rmsd of poses 1–10 and the rmsd of rank 1 increase on filtering. During the discussion of the surface area filter it was mentioned that the ligand 3-TAPAP has a large SAS compared with other poses generated by FlexX. Therefore, this result is not surprising. However, a closer look at the poses ranking higher than the best structure prediction shows that in all of them the ligand adopts highly strained conformations with *syn*-pentane interactions between the piperidine α-carbon atoms and the carbon atom at position 5 relative to them (Figure 7). Such an interaction is unlikely to occur in a protein–ligand complex, since it adds about 7–10 kJ mol^{-1} to its energy.[20] Obviously, the decrease in SAS cannot make up for this strain. When all poses with *syn*-pentane arrangements are removed from the pose list, a pose with an rmsd of 1.9 Å becomes the highest ranking solution. 1pph is not the only complex for which FlexX predicts poses with *syn*-pentane arrangements on high ranks. Poses with such interactions are also generated for other complexes with related ligands: 1ppc, 1ets, 1ett, and 1etr. Fortuitously, in these cases the strained poses also have a higher SAS than the low-energy conformations. It becomes clear, however, that the internal energy of the ligand conformations should not be disregarded.

Up to this point, the discussion has concentrated on structure prediction. Another major area for the application of docking programs is in the comparison of binding energies of various ligands in different receptors. Since our filtering protocol changes the score of all poses, we must make sure that the score remains a comparable entity for different protein–ligand complexes. Figure 8 shows a plot of the score as obtained after filtering versus the FlexX score. The score of the highest ranking solution has been used in both cases, so the poses are not necessarily the same on both axes. Both scores correlate well, the main difference being the offset of 3.3 kJ mol^{-1}. This offset originates from the fact that in most of the complexes, the highest ranking FlexX solution is different from the one with the lowest lipophilic SAS and therefore a solution with a lower *a priori* calculated binding energy assumes rank 1. We can conclude that rescoring by means of the SAS does not alter the general applicability of the empirical scoring scheme. It is

Table 3. FlexX results on the test set of 32 protein–ligand complexes, with and without subsequent application of the filtering and rescoring protocol[a]

		FlexX			FlexX + filter		
Complex No.	PDB entry	rmsd of rank 1	best rmsd between ranks 1–20	Lowest rmsd (rank)	rmsd of rank 1	best rmsd between ranks 1–20	Lowest rmsd (rank)
1	1abf	0.8	0.6	0.6 (9)	1.4	0.6	0.6 (9)
2	1cbx	1.2	0.6	0.6 (9)	1.1	0.6	0.6 (4)
3	1die	2.9	2.4	1.7 (151)	2.9	2.7	1.7 (81)
4	1eap	5.3	5.3	2.3 (839)	5.3	5.0	2.3 (169)
5	1etr	7.5	3.8	1.5 (77)	1.5	1.5	1.5 (1)
6	1ets	2.3	1.4	1.4 (12)	2.3	1.4	1.4 (10)
7	1ett	4.4	3.9	1.4 (130)	1.4	1.4	1.4 (1)
8	1glq	6.3	6.0	2.3 (142)	10.6	2.4	2.3 (31)
9	1hsl	1.0	0.6	0.6 (2)	1.0	0.6	0.6 (2)
10	1hvr	1.0	0.9	0.8 (29)	0.9	0.9	0.8 (26)
11	1ic	5.9	4.9	2.1 (88)	7.6	2.1	2.1 (5)
12	1mnc	2.5	1.9	1.1 (264)	2.5	2.4	1.1 (89)
13	1nsc	1.3	1.0	1.0 (18)	1.3	1.0	1.0 (5)
14	1nsd	5.8	3.2	2.8 (26)	3.2	2.8	2.8 (2)
15	1poc	5.3	5.3	3.1 (505)	5.3	3.2	3.1 (28)
16	1ppc	3.7	2.0	1.6 (39)	3.9	1.6	1.6 (20)
17	1pph	3.8	3.8	1.9 (129)	5.3	3.8	1.9 (31)
18	1rnt	3.2	1.5	1.1 (95)	3.6	1.6	1.2 (102)
19	1stp	0.5	0.5	0.5 (163)	0.5	0.5	0.5 (132)
20	1tmn	1.7	0.6	0.6 (7)	0.6	0.6	0.6 (1)
21	2cgr	4.3	1.5	1.2 (26)	2.0	1.2	1.2 (11)
22	2ctc	0.6	0.4	0.4 (2)	0.6	0.4	0.4 (2)
23	2ifb	2.8	1.8	1.6 (34)	2.0	1.7	1.6 (28)
24	2tmn	6.3	6.0	1.8 (610)	2.5	1.8	1.8 (11)
25	2tsc	2.6	2.3	2.3 (27)	2.6	2.3	2.3 (29)
26	3cpa	3.3	1.0	0.7 (36)	1.9	0.7	0.7 (21)
27	3tpi	0.6	0.6	0.6 (1)	0.6	0.6	0.6 (1)
28	4dfr	0.9	0.6	0.6 (50)	1.2	0.6	0.6 (8)
29	4phv	1.2	1.2	1.2 (1)	1.2	1.2	1.2 (1)
30	4tln	3.7	3.1	1.8 (169)	3.7	2.3	1.9 (74)
31	4tsl	1.5	0.7	0.7 (14)	1.2	0.7	0.7 (10)
32	8gch	7.5	5.7	2.9 (162)	5.7	3.2	2.9 (32)

[a] Note that the total number of poses is smaller in the FlexX + filter column (see Table 2).

interesting that the complexes 1ppc and 1etr appear as outliers on the plot in Figure 8. This is no longer the case when the strained conformations are removed as discussed above. The r^2 value of the correlation then increases to 0.97.

A few final remarks are in place about the quality of our structure predictions in relation to the results of other docking programs. Unfortunately, there is little room for meaningful comparisons of performance, since for some docking programs the published test sets are small (e.g., Welch et al.[21]). FlexX performed well in the docking section of the CASP2 contest.[16] The program GOLD has been tested extensively on a set of 100 protein–ligand complexes.[5] Among these, there are 17 of the complexes of our present test set. In the GOLD evaluation paper, nine of these were classified as "good" predictions (1cbx, 1poc, 1stp, 2cgr, 2ctc, 3tpi, 4dfr, 4phv, and 8gch), five as "close" (1die, 1dwd, 1glq, 1tmn, and 3cpa), and three as "partially correct" (1eap, 1etr, and 1lic). The FlexX + filter predictions showed significant deviations from the crystal structure in four cases only, namely 1eap, 1glq, 1poc, and 8gch. The level of accuracy is therefore similar for both programs. In general, GOLD seems to perform better for large ligands containing long, flexible chains without well-defined base fragments, because the ligand is regarded as a whole during the docking procedure. However, GOLD calculations yield only one pose per run and are far slower than FlexX calculations. Considering the tradeoff between computer time and accuracy, the FlexX + filter protocol appears therefore superior. This is also true in comparison with flexible docking with DOCK 4.0,[22] for which published results are not yet available.

SUMMARY AND CONCLUSION

We have presented a computationally efficient scheme for the post-processing of docking results. The docking program

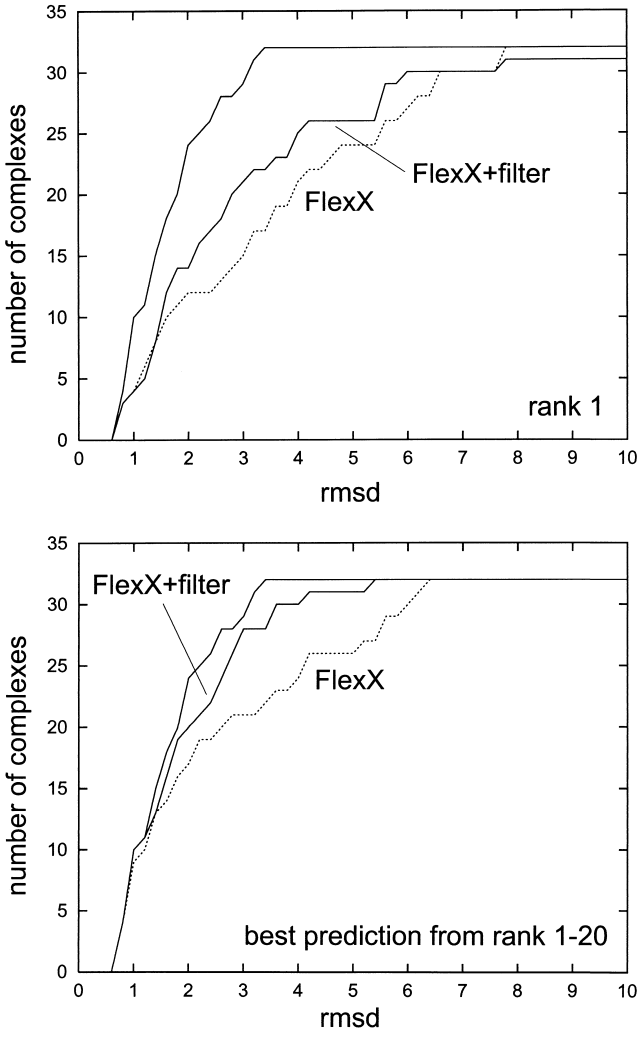

Figure 6. Total number of complexes predicted better than a threshold rmsd value plotted versus rmsd (see text for details). The uppermost solid curve in both plots was generated using the best prediction of each FlexX run regardless of its rank.

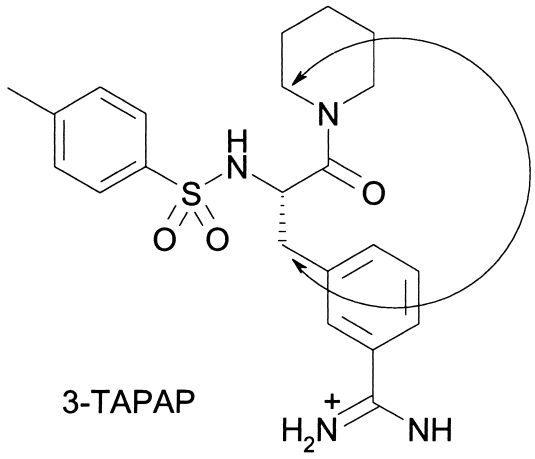

Figure 7. Structure of the thrombin inhibitor 3-TAPAP. The arrow indicates the positions that have close contact in a syn-pentane arrangement.

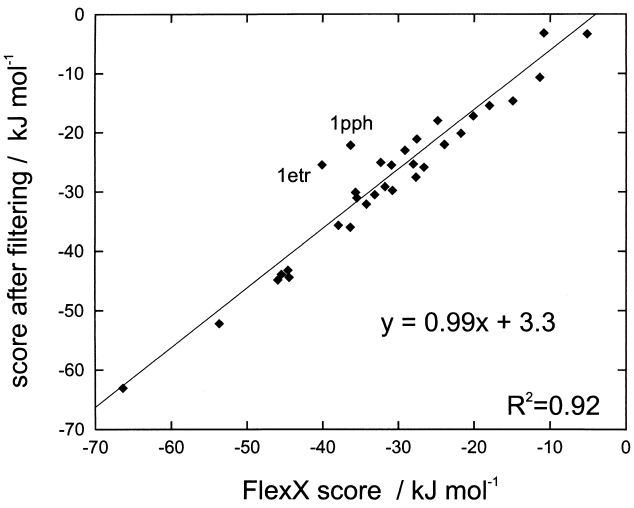

Figure 8. Plot of the score after filtering against the original FlexX score (rank 1 poses).

FlexX has been used for structure generation, but we believe that a similar scheme will be useful in conjunction with other docking tools as well. For each of the poses generated for a given protein-ligand complex, four properties are calculated, the fraction of the ligand volume buried inside the binding pocket, the size of lipophilic cavities along the protein-ligand interface, the area of the SAS of unpolar parts of the ligand and the number of close contacts between non-hydrogen-bonded polar atoms of the ligand and the protein. These four terms were used to filter out the majority of the poses and to rescore the remaining ones. On a test set of 32 protein-ligand complexes, this protocol significantly improves the accuracy of the structure predictions. Larger test sets will be necessary when an optimization of the individual cutoff values is desired. We have, however, demonstrated that filter functions based on simple geometric criteria provide a feasible way of improving structure prediction. The filters could also be tailored in such a way that they meet the requirements for *in silico* screening against a specific target. During the development of the filter protocol it has become apparent that some of the filter functions might also be implemented into a scoring function, which would allow to weed out undesirable ligand poses while they are not yet completely built. This would require a closer communication between the scoring and pose building algorithms. In particular, the buried volume of a partially assembled ligand pose (or a fragment thereof) might be used to assess if subsequent building steps might lead to promising results. Again, such a method might be useful for virtual screening applications when the active site pocket is the same for all ligands.

ACKNOWLEDGMENTS

The authors thank Dr. Matthias Rarey for a copy of the FlexX program and Dr. Gareth Jones for a discussion on the definition of enzyme binding pockets. Furthermore, we are grateful to Dr.

Daniel Bur, Dr. Stefan Fischer, Dr. Paul Gerber, and Dr. Gisbert Schneider for fruitful discussions.

REFERENCES

1 Jones, G., and Willet, P. Docking small-molecule ligands into active sites. *Curr. Opin. Biotechnol.* 1995, **6,** 652–656; Rarey, M., and Lengauer, T. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* 1996, **6,** 402–406; Gschwend, D.A., Good, A.C., and Kuntz, I.D. Molecular docking towards drug discovery. *J. Mol. Recognit.* 1996, **9,** 175–186

2 Rarey, M., Kramer, B., Lengauer, T., and Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* 1996, **261,** 470–489

3 Rarey, M., Kramer, B., and Lengauer, T. Multiple automatic base selection: protein–ligand docking based on incremental construction without manual intervention. *J. Comput.-Aided Mol. Design* 1997, **11,** 369–384

4 Jones, G., Willet, P., and Glen, R. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* 1995, **245,** 43–53

5 Jones, G., Willet, P., Glen, R.C., Leach, A.R., and Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 1997, **267,** 727–748

6 Ewing, T.J.A., and Kuntz, I.D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* 1997, **18,** 1175–1189

7 Dixon, S. Evaluation of the CASP2 docking section. *Proteins Struct. Funct. Genet.* 1997, Suppl. 1, 198–204

8 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R.L., and Ferrin, T.E. A geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.* 1982, **161,** 269–688

9 Goodsell, D.S., Morris, G.M., and Olson, A.J. Docking of flexible ligands: Applications of AutoDock. *J. Mol. Recognit.* 1996, **9,** 1–5

10 Trotov, M., and Abagyan, R. Flexible protein–ligand docking by global energy optimization in internal coordinates. *Proteins Struct. Funct. Genet.* 1997, Suppl. 1, 215–220

11 Elridge, M., Murray, C.W., Auton, T.R., Paolini, G.V., and Mee, R.P. Empirical scoring functions. I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Design* 1997, **11,** 425–445

12 Jain, A.N. Scoring noncovalent protein–ligand interactions: A continuous differentiable function tuned to compute binding affinities. *J. Comput.-Aided Mol. Design* 1996, **10,** 427–440

13 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. Design* 1994, **8,** 243–256

14 Böhm, H.-J. Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de-novo design or 3D-database search programs. *J. Comput.-Aided Mol. Design* 1998, **12,** 309–323

15 Ajay, and Murcko, M. Computational methods to predict binding free energy in ligand–receptor complexes. *J. Med. Chem.* 1995, **38,** 4953–4967; Clark, D.E., Murray, C.W. and Li, J. Current issues in *de novo* molecular design. In: *Reviews in Computational Chemistry* (Lipkowitz, K.B., and Boyd, D.B., eds.), Vol. 11. Wiley-VCH, New York, 1997, pp. 67–125

16 Kramer, B., Rarey, B., and Lengauer, T. CASP2 experiences with docking flexible ligands using FlexX. *Proteins Struct. Funct. Genet.* 1997, Suppl. 1, 221–225

17 Ho, C.M.W., and Marshall, G.M. Cavity search: An algorithm for the isolation and display of cavity-like binding regions. *J. Comput.-Aided Mol. Design* 1990, **4,** 337–354

18 Bernstein, F.C., Koetzle, T.E., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977, **112,** 535–537

19 Gerber, P.R., and Müller, K. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput.-Aided Mol. Design* 1995, **9,** 251–268

20 Goto, H., Osawa, E., and Yamato, M. How many conformers are there for small *n*-alkanes? Consequences of asymmetric deformation in GG′ segment. *Tetrahedron* 1993, **49,** 387–396

21 Welch, W., Ruppert, J., and Jain, A.J. Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* 1996, **3,** 449–462

22 Todd Ewing, personal communication.