Journal of Computer-Aided Molecular Design, **16:** 129–149, 2002.
KLUWER/ESCOM

Flexible docking under pharmacophore type constraints

Sally A. Hindle*, Matthias Rarey, Christian Buning & Thomas Lengauer†
Fraunhofer Institute for Algorithms and Scientific Computing (FhI-SCAI), Schloss Birlinghoven, D-53754 Sankt Augustin, Germany

Received 5 December 2001; Accepted 6 March 2002

Key words: constraint, flexible docking, molecular modeling, pharmacophore, protein-ligand interaction, virtual screening

Summary

FLEXX-PHARM, an extended version of the flexible docking tool FLEXX, allows the incorporation of information about important characteristics of protein-ligand binding modes into a docking calculation. This information is introduced as a simple set of constraints derived from receptor-based type pharmacophore features.

The constraints are determined by selected FLEXX interactions and inclusion volumes in the receptor active site. They guide the docking process to produce a set of docking solutions with particular properties. By applying a series of look-ahead checks during the flexible construction of ligand fragments within the active site, FLEXX-PHARM determines which partially built docking solutions can potentially obey the constraints. Solutions that will not obey the constraints are deleted as early as possible, often decreasing the calculation time and enabling new docking solutions to emerge. FLEXX-PHARM was evaluated on various individual protein-ligand complexes where the top docking solutions generated by FLEXX had high root mean square deviations (RMSD) from the experimentally observed binding modes. FLEXX-PHARM showed an improvement in the RMSD of the top solutions in most cases, along with a reduction in run time. We also tested FLEXX-PHARM as a database screening tool on a small dataset of molecules for three target proteins. In two cases, FLEXX-PHARM missed one or two of the active molecules due to the constraints selected. However, in general FLEXX-PHARM maintained or improved the enrichment shown with FLEXX, while completing the screen in considerably less run time.

Introduction

As the number of drug-like molecules stored in high-throughput libraries reaches into the millions, virtual screening techniques offer a faster and more cost effective alternative to conventional screening methods. Nowadays, especially in view of the expanding rate at which target protein structures are being solved, structure-based techniques such as flexible protein-ligand docking play an increasingly important role in the identification of potential lead compounds in the drug discovery process.

There are various approaches used in flexible protein-ligand docking to model how a ligand may bind in the active site of a protein. For all approaches it is necessary to have a complete description of at least the protein active site, often in the form of the 3D structure as found in the Protein Data Bank (PDB) [1]. Good [2] and Gane et al. [3] offer comprehensive reviews on the recent advances in structure-based virtual screening and structure-based rational drug design respectively, including flexible docking techniques, while a review of high-throughput docking has also recently been published by Abagyan et al. [4]. When docking algorithms are used within a screening scenario, the emphasis has to be on speed. For this, the docking program FLEXX [5] relies on an incremental construction algorithm. In a virtual screening experiment, each molecule is docked and scored and finally

*Author to whom correspondence should be addressed. Current address: BioSolveIT GmbH, An der Ziegelei 75, D-53757 Sankt Augustin, Germany. E-mail: hindle@biosolveit.de
†Current address: Max-Planck Institut für Informatik, Stuhlsatzenhausweg 85, D-66123 Saarbrücken, Germany

the molecules in the database are re-ranked according to those scores. The main aim of a virtual screening experiment is to bring molecules predicted to be active with respect to the target protein to the top of the database hence giving an enrichment of potential leads.

A pharmacophore is defined as being a set of structural features in a molecule that is recognized at a receptor site and is responsible for that molecule's biological activity [6]. It is used to represent the minimal structural requirements which are essential for receptor recognition, receptor binding and biological response. A pharmacophore is usually generated from a set of known active ligands of a target protein, which infer likely binding conformations taken up by key bioreactive groups. Typically, a pharmacophore consists of three or four sites such as hydrogen bonding groups, charged centers or hydrophobic groups, along with a set of geometrical rules that describe how these features are related in 3D space. Exclusion volumes are also often included to represent the presence of the protein. Pharmacophores are used as 3D similarity search queries in virtual screening to identify new leads [7–9]. These types of similarity searches identify compounds in a database that can adopt the correct conformation of bioreactive groups but, despite the use of exclusion volumes, they take no account of how the ligand may bind in the actual protein active site.

The advantage of flexible docking techniques is that they explore the conformational space of the ligand within the active site of a protein, leading to a highly diverse docking solutions set. This could however, in terms of screening applications, be seen as a drawback. A flexible docking algorithm does not make use of any knowledge that might be available beforehand about the binding conformation of a similar ligand or about the properties of ligands that are already known to be active in the target molecule. When such beneficial information is available, time and effort are wasted docking molecules in the wrong conformations or, of particular importance during a virtual screening experiment, docking molecules that should be rejected because they contradict the information at hand. By combining the advantages of 3D similarity searching techniques and docking, it should be possible to refine the results of virtual screening. Furthermore, it should also be possible to speed up docking-based virtual screening; although docking algorithms are becoming increasingly faster, docking-based virtual screening is still not as fast as other 3D searching techniques and timing remains critical.

Relevant information may be available to the user in the form of an experimentally resolved structure of a similar protein-ligand complex to that in which they are interested. These structures show whereabouts in the site the ligand binds and in what conformation, and may also hold interesting clues about important interactions that are formed between ligands and the protein in the bound conformation. When such direct information is unavailable information can be derived from various sources. Analysis of active site properties allows the identification of key interaction 'hotspots' [9–12] and the derivation of so-called receptor-based pharmacophores, while pharmacophore knowledge may be obtained from the study of the properties of known active ligands, using, for example, the active analogue approach, clique detection techniques and more recently genetic algorithms (for an overview of pharmacophore techniques in general see [9]). One important source of information that still plays a role in pharmacophore development comes from the understanding the user themselves may have about the biochemical system.

Previously, Fradera et al. [13] published a variation of the DOCK 4.0 [14] program that uses the experimentally resolved structure of a protein-ligand complex to guide the flexible docking of similar ligands in the active site of that protein. During the docking process, the position of a ligand is compared to that of the crystal structure and the docking score is weighted according to the similarity of the two. Meanwhile, Thomas IV et al. [15] developed pharmacophore docking also within the DOCK program, in order to dock pharmacophores rather than ligands into protein active sites. This method was developed as one way of incorporating conformational ligand information into the DOCK rigid body docking algorithm.

For our purposes, a receptor-based pharmacophore type descriptor lends itself easily to application in docking. If pharmacophore information has been derived from the side of the ligand only, the pharmacophore sites in the ligand can be translated on to the receptor site perhaps from an analysis of the results of docking calculations. With a receptor-based descriptor, interactions are defined within the active site on the side of the protein rather than on the side of the ligand, while exclusion volumes are already described by the 3D structure of the site itself. Consequently, as some description of the active site is a prerequisite for docking, a receptor-based pharmacophore provides an

ideal way of conferring information into the docking calculation.

In the following we present FLEXX-PHARM, an extended version of the flexible docking program FLEXX, in which the user can incorporate pharmacophore features as constraints into a flexible docking calculation. All docking solutions must possess the properties prescribed by the set of constraints. The FLEXX docking calculation is modified to accommodate the constraints not by introducing penalties or weighting into the scoring scheme, but rather by using filters to keep or reject solutions in one of two ways. Firstly, and most simply, the set of constraints is used as a post-docking filter applied to the final set of docking solutions generated by FLEXX. Or, secondly, the set of constraints is used as a filter applied during the reconstruction of the ligand in the active site. This is the most advantageous method because solutions are filtered out early in the course of the calculation. This leads not only to a more specific set of docking solutions but also to a speed-up in the calculation and the opportunity for novel docking solutions to appear due to a more intensive search of the conformational space defined by the constraints.

The following sections describe in detail the form of the pharmacophore type constraints used in FLEXX-PHARM, the implementation of various look-ahead filtering checks and finally the application of FLEXX-PHARM in docking problems and in virtual screening scenarios.

Methods

The full details of the models and algorithms used in FLEXX are described in detail elsewhere [5, 16, 17]. However, a brief outline of the relevant points is given below.

In FLEXX, the protein is referred to as the 'receptor' and therefore the term 'receptor' will be used in the text from here on. The receptor active site and the ligand interacting groups in FLEXX are described by the LUDI interaction model [18]. An interacting group on the surface of the site is assigned an interaction type and, accordingly, an interaction center and surface. The interaction center forms the center of a sphere on which the interaction surface lies. Interactions are formed when interaction groups on the ligand are matched with appropriate groups in the active site. If the interaction center of one group lies close to the interaction surface of the other, as in Figure 1,

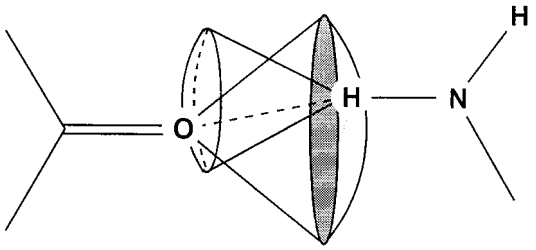

Figure 1. A receptor-ligand interaction at the ideal geometry, formed between a hydrogen bond acceptor (left) and a hydrogen bond donor (right).

then the interaction has an ideal geometry. (Deviations from the ideal geometry are tolerated in FLEXX but are penalized in the scoring function by a scaling factor.) The types of interaction currently available in FLEXX are listed in Table 1. Note that for algorithmic purposes, the interaction surfaces on the receptor are approximated as a finite set of discrete points during the calculation.

The ligand is divided into fragments at rotatable bonds. Various base fragments are selected and placed independently into the active site. The remaining fragments are then built on to the base fragments in an iterative procedure. During each fragment placement, FLEXX identifies the new interactions formed between the site and the fragment according to the interaction geometry parameters. The interactions are then used to optimize the ligand position and to score the placement by means of an empirical scoring function based on that of Böhm [19]. A greedy heuristic is used to retain the k best placements (default parameters give $k = 800$ for most FLEXX calculations) for the next iterative step.

In FLEXX-PHARM, the set of pharmacophore features in the active site constrains the docking calculation so that only solutions are produced that match the specified set of features. Thus, the term pharmacophore constraint is here used to describe a feature in the site which will impose a restriction on the docking calculation. The first modification of FLEXX is the inclusion of the pharmacophore constraints, the nature of which are described in the next section below. If the pharmacophore constraints are to be applied during the docking calculation then some pre-docking preparation steps are initiated. Firstly, all *candidate countergroups* for each constraint are located in the ligand. A candidate countergroup set contains one candidate countergroup for each constraint. Combinatorially, there is usually a list of possible candidate countergroup sets. A maximum distance look-up table

Table 1. Interaction types in FLEXX. An interaction group on the receptor can be matched by any of the countergroups in the ligand and *vice versa*

Name	Interaction Group	Countergroups
Directed interactions (level 3 interactions - strong)		
h_don	hydrogen bond donor	h_acc
h_acc	hydrogen bond acceptor	h_don
metal	metal atom	metal_acc
metal_acc	metal ligand	metal
Hydrophobic directed interactions (level 2 interactions)		
phenyl_center	phenyl ring (centroid)	phenyl_ring, ch3_phe, amide
phenyl_ring	atom in phenyl ring	phenyl_center
ch3_phe	methyl group	phenyl_center
amide	amide bond	phenyl_center
Hydrophobic undirected interactions (level 1 interactions - weak)		
ch	CH group	ch, ch2, ch3, sulfur, aro
ch2	CH_2 group	ch, ch2, ch3, sulfur, aro
ch3	methyl group	ch, ch2, ch3, sulfur, aro
sulfur	sulfur atom	ch, ch2, ch3, sulfur, aro
aro	aromatic C atom	ch, ch2, ch3, sulfur, aro

(MAXLT) is calculated for all the candidate countergroups in the ligand, while a minimum distance look-up table (MINLT) is calculated for all the constraints on the receptor. These are used to reduce the combinatorial list of candidate countergroup sets by eliminating impossible combinations of countergroups. The resulting reduced list is referred to in the text as the *master list*.

During the fragment placement, FLEXX-PHARM carries out an additional search for newly formed interactions with extended geometry parameters (larger distance and angle tolerances in deviations from the ideal geometries). These are stored separately and are not used in the optimization or scoring of the placement. The identification of these interactions is important during the look-ahead filtering.

The look-ahead filtering takes place after the base placement and each fragment building step. The filtering consists of a series of three checks designed to assess whether each of the k retained solutions will be able to match the constraints. The filters; logical checks, distance checks and directed tweak checks, increase in complexity and are explained in more detail below. If a partial docking solution satisfies one check, it passes on to the next and if it satisfies all checks then it is retained for the next fragment building iteration.

Definition of the Pharmacophore Constraints in FLEXX-PHARM

In FLEXX-PHARM two different types of constraint can be defined in the active site, *interaction constraints* and *spatial constraints*. For the first type, the user can specify a FLEXX interaction surface in the active site that must take part in an interaction with the ligand. For the second type, the user can specify inclusion volumes. In addition the constraints can be designated *essential* or *optional*. The following description makes reference to an example set of pharmacophore constraints built in the active site of carbonic anhydrase (PDB code 1azm) which are listed in Table 2 and shown in Figure 2.

Interaction constraints. An interacting group and interaction type in the active site must be specified (along with an interaction surface if more than one surface exists for that interaction). FLEXX-PHARM ensures that an interaction is formed between the specified interacting group in the active site and the ligand in a valid docking solution. In Figure 2, constraint 1 is a metal interaction on the zinc ion. A metal acceptor in the ligand must form an interaction with the zinc in a valid docking solution. A hydrogen donor (constraint

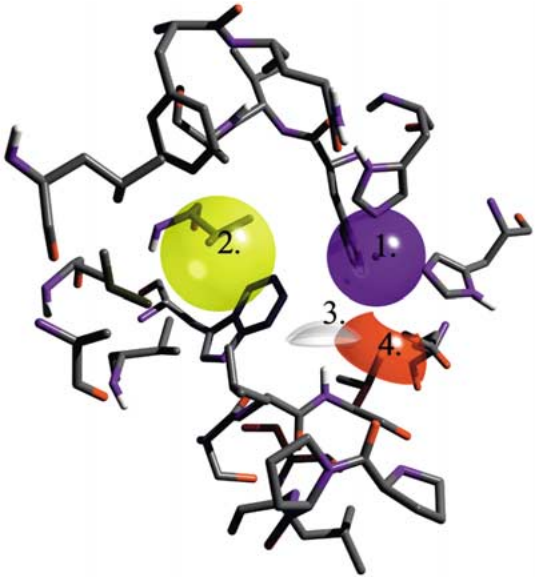

Figure 2. A set of four pharmacophore constraints in the active site of carbonic anhydrase (1azm). Constraint 1. essential metal interaction at the zinc ion, constraint 2. essential spatial constraint for a carbon atom, constraint 3. optional h_don interaction at the backbone nitrogen of THR 199, constraint 4. optional h_acc interaction at the gamma oxygen of THR 199. $P_{\min} = 1$ and $P_{\max} = 2$. See text for more details.

3) and acceptor (constraint 4) have also been selected on the amide nitrogen and the hydroxy oxygen of THR 199 respectively.

Spatial constraints. This constraint can be used to constrict ligand position in the active site and consists of a sphere plus an associated element type. For example, a carbon atom of the ligand must lie in the defined sphere of constraint 2 in Figure 2.

Essential constraints. A valid docking solution must obey all essential constraints. Constraints 1 and 2 in Figure 2 are assigned essential priority.

Optional constraints. The number of optional constraints obeyed in a valid docking solution must lie within a given interval $[P_{\min}, P_{\max}]$. Optional constraints allow for a partial pharmacophore match and hence provide a little more flexibility in the overall pharmacophore definition. $P_{\min}$ is useful, for example, when at least one interaction is required with the guanidinium group of Arginine. $P_{\max}$ can be used to limit the number of optional constraints found, specially in the case of screening experiments where there may be different classes of active molecules. Constraints 3 and 4 in Figure 2 are optional. The associated partial definition prescribes that at least one of the two constraints is fulfilled in a valid docking solution.

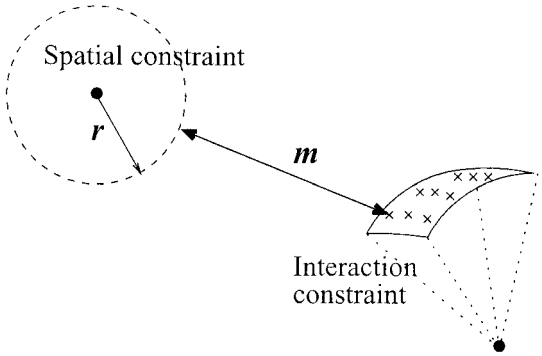

Figure 3. The minimum distance m between two pharmacophore constraints. For the spatial constraint the minimum distance is measured to the sphere (to the central point less the radius r) and for the interaction constraint the minimum distance is measured to the nearest approximation point on the interaction surface.

Pre-docking checks in FLEXX-PHARM

In order to gain the maximum potential from FLEXX-PHARM in terms of speed and optimal results, ligands that are bound to fail the pharmacophore constraints should be eliminated before the docking calculation. This can be achieved with some pre-docking checks in FLEXX-PHARM. If the ligand could potentially fit the constraints then the information calculated during the pre-docking checks is used later in the docking calculation to delete incompatible docking solutions as early as possible.

Pre-docking preparation on the receptor

The MINLT (minimum distance look-up table for the constraints) is created by calculating the minimum distance between all pairs of constraints. The minimum distance (**m** in Figure 3) is calculated to a spatial constraint by taking the nearest point on the sphere (i.e. the distance to the center less the radius **r**), while the minimum distance to an interaction constraint is calculated to the nearest approximation point on the interaction surface.

Pre-docking preparation on the ligand

The MAXLT (maximum distance look-up table for the candidate countergroups) is calculated for the maximum distances between all pairs of atoms and candidate countergroups in the ligand. A candidate countergroup for a spatial constraint is an atom of the

correct type, while a candidate countergroup for an interaction constraint is an interaction center of an appropriate countergroup for that interaction. Note that interaction centers are not atoms but defined points in space. The MAXLT is filled as follows. Firstly, all distances between each interaction center and an atom of the interacting group to which it belongs are calculated. A tolerance is added on to this distance to allow a little extra leeway in the MAXLT to accommodate the formation of interactions that deviate from the ideal interaction geometry. Then the distances between pairs of atoms are added to the table. Because all bond angles in FLEXX remain fixed, the distances between atom pairs separated by one and two bonds are taken directly from the original ligand conformation. For any rotatable bonds, the maximum possible distance between atom pairs separated by three bonds is calculated by setting the torsion to 180°. For ring systems, distances are taken from the set of valid ring conformations stored by FLEXX (as calculated by the program CORINA [20]). The resulting table usually has gaps in it which are completed using the Floyd–Warshall all-pairs shortest path algorithm [21, 22] (here applied in its variant for maximization of path length). This algorithm fills a gap with the shortest distance between two points that is consistent with all the pairwise distance relationships already known.

Finally, a list of candidate countergroup sets is compiled from all combinations of candidate countergroups in the ligand. The candidate countergroups (plus the interaction surfaces) found in the ligand from the 1azm complex are shown in Figure 4. There are several candidates for each of the constraints (given in Figure 2 and Table 2).

FLEXX-PHARM determines whether the ligand can potentially meet the constraints by comparing the MAXLT and MINLT. A distance check is carried out for each candidate countergroup set. If the MAXLT distance between any two essential candidate points is shorter than the corresponding MINLT distance, that candidate countergroup set is immediately deleted. Otherwise, the optional constraints are brought into play; if every MAXLT distance between the optional candidate countergroup and each essential candidate countergroup is greater than the corresponding MINLT distance, that optional candidate point satisfies the distance check. If the number of optional candidate countergroups which satisfy the check is greater than $P_{\min}$ then this candidate countergroup set is retained. The distances between pairs of optional candidate countergroups are not considered

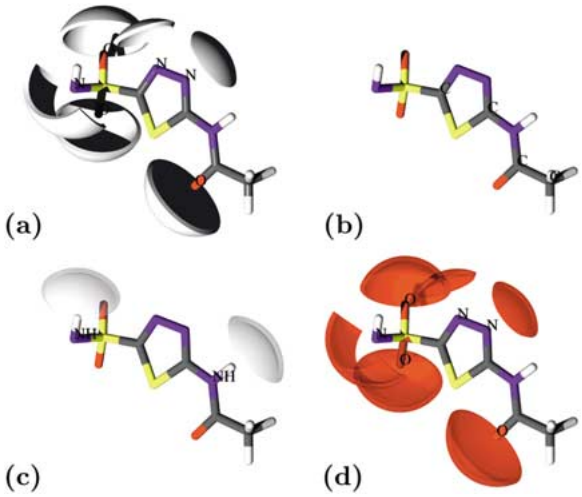

Figure 4. Candidate countergroups in the 1azm complex ligand. Shown are the: (a) interaction surfaces of the six metal_acc countergroups for constraint 1, (b) four C atoms for constraint 2, (c) interaction surfaces of the two h_don countergroups for constraint 3, (d) interaction surfaces of the six h_acc countergroups for constraint 4.

at this stage due to the very large number of comparisons that would have to be made. If, finally, after the pre-docking distance checks the list of candidate countergroup sets is empty, it is impossible for any combination of candidate countergroups in the ligand to fulfill the constraints and the ligand is rejected. If not, the remaining list of candidate countergroups sets is stored as the master list.

FLEXX-PHARM *docking filters*

Application of pharmacophore constraints during docking

The actual placement of fragments by the FLEXX base placement and fragment building algorithms is not altered by FLEXX-PHARM. One incremental construction step begins with the set of partial docking solutions remaining from the previous step. A new fragment is added to the placed fragments in all possible conformations providing there is no significant overlap with the receptor. The FLEXX placement algorithm identifies the new interactions between the fragment and the receptor which are used for optimization and scoring. Meanwhile, FLEXX-PHARM identifies extra interactions with the extended angle and distance tolerance parameters. The use of these extended parameters is necessary because the position of the partial solutions may change during the rest of the incremental construction process due to lig-

and placement optimizations. FLEXX-PHARM could otherwise exclude a solution wrongly because the necessary interaction may later come to exist when groups on the ligand and receptor move closer together.

The fragment building step is completed when a fragment has been added to all partial solutions, optimized, scored and the k best solutions have been selected. The k best solutions are then filtered against the constraints by FLEXX-PHARM before the next fragment building step begins. At this stage a valid solution is one in which the constraints are either fulfilled or can still be fulfilled during the placement of the remaining unplaced ligand.

Logical checks. The logical checks ascertain whether the correct combinations of candidate countergroups exist in the placed and unplaced parts of the ligand. For each pharmacophore constraint, FLEXX-PHARM looks to see whether the constraint is already fulfilled in the placed part of the ligand or, if not, whether candidate countergroups still exist in the unplaced part of the ligand (i.e. whether it is still possible to fulfill the constraint).

An interaction constraint is fulfilled when an interaction containing the constraint interaction surface exists in amongst the interactions identified by FLEXX or FLEXX-PHARM (with extended parameters) during the fragment placement. A spatial constraint is fulfilled when a ligand atom of the correct type lies within the radius plus a tolerance of the constraint sphere. A tolerance is required on the sphere again to account for any movement of placed atoms during subsequent optimizations. Note that more than one candidate countergroup can fulfill one constraint. If a constraint is not fulfilled then it can be designated possible to fulfill if countergroups remain available in the unplaced part of the ligand.

If the fragment added was the final fragment of the ligand then all constraints must be fulfilled (in accordance with the partial constraints definition). The extra interactions identified by FLEXX-PHARM and tolerances on spatial constraint spheres are *not* included as there can be no further movement of the ligand atoms.

Simple logical conditions are used to decide whether the partial docking solution can be retained. The solution is deleted if one of the following conditions is true:

- any essential interaction constraint is not fulfilled or not possible;
- fewer than $P_{\min}$ optional constraints are fulfilled or possible;
- greater than $P_{\max}$ optional constraints are fulfilled (this condition can only be applied if the fragment added was the final fragment because the number of fulfilled optional constraints may otherwise change due to optimizations).

The logical checks are illustrated by means of an imaginary docking situation with the carbonic anhydrase active site, ligand and constraints in Figure 5.

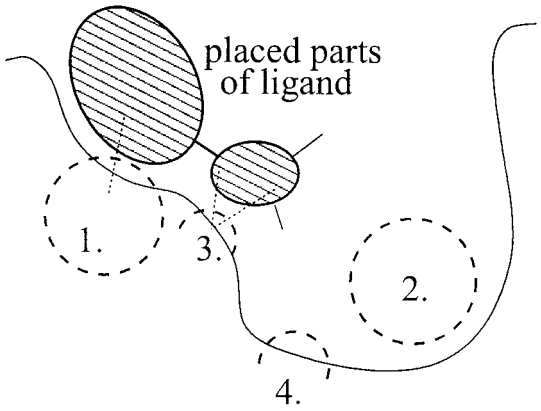

Figure 5. Schematic of the 1azm active site with the pharmacophore constraints. After placement of two fragments of the ligand, constraint 1 (essential) is fulfilled by one candidate countergroup in the ligand and constraint 3 (optional) is fulfilled by two. Constraints 2 (essential) and 4 (optional) have not yet been fulfilled. It must be possible to fulfill constraint 2 as this is an essential constraint. It does not matter at this stage whether constraint 4 is still possible because the optional requirements are already satisfied by constraint 3.

If a partial (i.e. not a final) docking solution is retained then a list of candidate countergroup sets is created from the ligand placement. Several combinatorial possibilities may exist due to the fact that constraints can be fulfilled by more than one countergroup and it may still be possible to fulfill constraints with more than one countergroup. All combination sets formed are checked against the master list; if that set is not present in the master list then it has already been eliminated because the maximum possible distances between the candidate countergroups are already too short for the constraints. This newly created list is local to that partial docking solution and the *local lists* vary from solution to solution. A solution with an empty local list is immediately deleted.

Distance checks. The remaining filter checks focus only on the constraints that remain possible and not those that are already fulfilled. This also means there is no further checking if the fragment added was the final fragment.

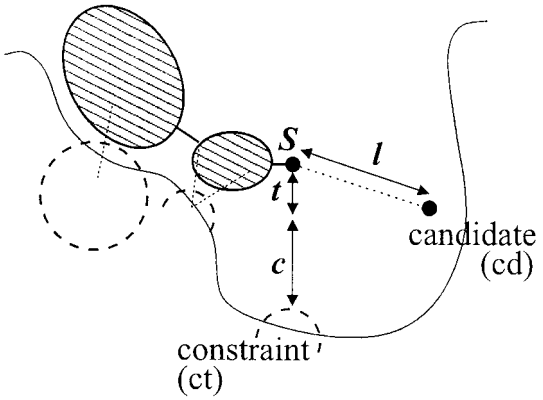

Figure 6. During distance checking, the distance c from atom S to the constraint must be shorter than distance l (in the MAXLT) from atom S to the candidate countergroup. t is a tolerance allowing for movement of S in future ligand optimizations.

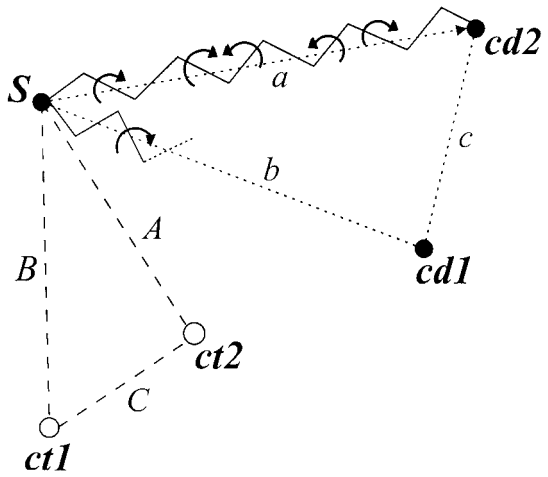

Figure 7. The application of directed tweak in FLEXX-PHARM. The set of distances (small letters) between candidate points and S in the ligand are altered to match the distances (capital letters) between the constraints and S in the active site via a torsion angle optimization process.

As soon as part of the ligand is placed, it lies in a conformation where distances between placed candidate countergroups and candidate countergroups in the unplaced part of the ligand are likely to be shorter than the distances contained in MAXLT. For every partial solution there is a leading atom among the placed fragments to which the next fragment will be added. This is always attached to the placed part of the ligand with one rotatable bond and has no other placed atoms attached to it. For this discussion this atom is named S. The distances from S to the candidate countergroups in the unplaced part of the ligand can now be compared to the distances from S to the respective constraints.

As earlier (Figure 3), the distance to an interaction constraint is calculated to the nearest approximation point, while the distance to a spatial constraint is to surface of the sphere. This time a tolerance t (see Figure 6) allows for movement in the position of S during subsequent ligand optimizations. The resulting distance c is compared to the distance l between S and the respective candidate countergroup taken from the MAXLT.

For each candidate countergroup set, the distance l from atom S to the candidate points must be greater than distance c for all essential constraints and for more than $P_{\min}$ optional constraints. If not, the candidate countergroup set is deleted from the local list and consequently any partial docking solution with an empty local list is removed from the docking calculation.

Directed tweak checks. Directed tweak [23] is a technique developed for flexible 3D searching. It adjusts the rotatable bonds of a query molecule to try and produce the conformation prescribed by the target. The conformational search is carried out via a minimization in torsion space. The variables are distance-based and the derivatives are analytical which means that the technique is relatively fast. It has been evaluated and applied for pharmacophore pattern matching in the UNITY tool [24, 8] and also in 3D flexible ligand design [25] and superposition [26].

The cost function that is minimized in a directed tweak calculation takes the form

$$\sum_{j<i}(d_{ij} - d_{ij}^0)^2, \tag{1}$$

where d_{ij} represents a distance between two points in the query molecule and d_{ij}^0 represents the distance between corresponding points in the target. The two points usually belong to a set of points and hence the sum is taken over all pairwise distances between the points in the set.

In FLEXX-PHARM the set of points in the query is taken to be S plus the candidate countergroups (labeled cd1,2 in Figure 7) and where d_{ij} is the set of pairwise distances a,b,c. The set of target points consists of S plus the constraints (labeled ct1,2) on the receptor and d_{ij}^0 is the set of pairwise distances A,B,C.

The path between a pair of candidate countergroups is defined by the shortest path of bonds through the ligand. In the case that an interaction center does not lie over an atom, the interaction center is joined to the atom by a temporary bond which is then included

in the path. The target points (other than S) are points that are chosen to represent the constraints in the receptor. The representative point for a spatial constraint or an interaction constraint with a spherical interaction surface is taken to be the center of the sphere. Otherwise, it is taken to be the closest approximation point on the interaction surface.

The directed tweak method attempts to match the distances between the target points exactly. In our case, this is unnecessary because a spatial constraint is satisfied as soon as the candidate countergroup lies within a sphere or within a certain distance of the interaction surface. Equation 1 becomes

$$\sum_{j<i} D_{ij} \quad \text{where}$$

$$D_{ij} = \begin{cases} 0: & d_{ij} - d_{ij}^0 \;\leq\; T_{ij}, \\ (d_{ij} - d_{ij}^0)^2: & d_{ij} - d_{ij}^0 \;>\; T_{ij}, \end{cases} \quad (2)$$

where T_{ij} is the sum of allowed deviations from the target points plus the tolerance in the position of S.

The minimization is carried out using a quasi-Newton approach. Only essential constraints are entered into the directed tweak calculation as consideration of the various combinations of optional constraints is too time consuming for our current purpose. The calculations are carried out up to a specified maximum number of rotatable bonds per calculation. The final value of the cost function, zero or non-zero, is all that is required. Directed tweak is used here as a method to judge whether the remaining unplaced ligand can achieve a conformation to satisfy the remaining constraints or not. It provides a more informative alternative than simply using a minimum distance look-up table for the ligand. If a minimized cost function results from an unlikely ligand conformation and FLEXX-PHARM retains the partial docking solution, then this is of little consequence because FLEXX will not build such an invalid ligand conformation and the partially built ligand will proceed towards an alternative conformation.

The directed tweak test is called for each candidate countergroup set remaining in the local list for each partial docking solution. A candidate countergroup set satisfies the directed tweak test if the cost function was minimized to zero. As soon this occurs then the partial docking solution satisfies all the filtering checks and is retained for the next fragment building step. If none of the candidate sets pass the directed tweak checking the docking solution is rejected.

FLEXX-PHARM *post-docking filter*

In the post-docking filter mode, the constraints are used as a filter for the final docking solutions produced by FLEXX. Each solution is retained or rejected depending purely on the numbers of essential and optional constraints fulfilled. The main disadvantage with using this mode lies in the fact that there is no chance for novel solutions to appear in the final set. The post-docking filter mode can be selected by the user or is selected automatically according to certain criteria.

As discussed below in the Results and discussion section for thermolysin, long lists of candidate countergroup sets result in a long run time that is undesirable in a screening scenario. FLEXX-PHARM was modified so that the post-docking filter mode was used for those specific cases rather than the look-ahead filtering mode. The automatic selection of the post-docking filter mode must be made before the docking of a ligand actually begins, i.e., during the pre-docking preparation. The filtering mode is chosen depending on the length of the master list. The default maximum length for the master list is set to 20 000; if the length is shorter then FLEXX-PHARM uses look-ahead filtering for that ligand, otherwise it selects the post-docking filter.

Results and discussion

To assess the performance of FLEXX-PHARM we carried out three sets of comparisons against unconstrained docking with FLEXX, named here classes one, two and three. The first two classes were designed to show how the docking performance of FLEXX was improved by including pharmacophore constraints, while the third class shows how FLEXX-PHARM was used to enhance the performance of FLEXX in screening experiments. The sets of constraints used in the class one and two comparisons are listed in Table 2.

Class one: improving ranking of correct binding modes

Complexes were selected from the standard FLEXX test set consisting of 200 complexes taken from the PDB (FLEXX200 set [27]). The strategy employed in selection was to pick out cases where it was clear that FLEXX (with default parameters) could generate good docking solutions with respect to the crystal structure position of the ligand, but where these docking

Table 2. Details of pharmacophore constraints for class one and two comparisons of FLEXX and FLEXX-PHARM performance

Class One					Class Two				
	Constraint type	Detail[a]	e/o[b]	P_{min}, P_{max}[c]		Constraint type	Detail[a]	e/o[b]	P_{min}, P_{max}[c]
1azm	spatial	C	e		1bma	ch3_phe	CG2 VAL 224(A)	e	
	metal	ZN	e			h_don	N VAL 224(A)	e	
	h_don	N THR 199	o			h_acc	O VAL 224(A)	o	
	h_acc	OG1 THR 199	o	1,2		h_acc	O SER 222(A)	o	1,2
1etr	h_acc	O GLY 216(H)	e		1eap	ch3_phe	CG2 VAL 37(B)	e	
	h_acc	OD1 ASP 189(H)	o			h_don	N GLY 102(B)	e	
	h_acc	OD2 ASP 189(H)	o	1,2					
1ivf	phenyl_center	CE3 TRP 178	e		1ele	h_acc	O VAL 224(E)	e	
						h_acc	O SER 222(E)	e	
1thy	ch3_phe	CB ALA 220	e		1hdc	h_don	OH TYR 152(A)	e	
						h_don	N THR 185(A)	e	
2tmn	spatial	C	e		1rob	h_don	OG1 THR 45	e	
	spatial	O	e						
	spatial	O	e						
4tln	spatial	C	e		8gch	amide	C TRP 215	e	
	spatial	O	e			h_don	N GLY 216	e	
	spatial	O	e			h_acc	O GLY 216	e	

[a]If interaction constraint, then the name of the receptor atom (PDB nomenclature: atom name, amino acid code, amino acid number and optionally chain identifier). If a spatial constraint, then the element type associated with that constraint.
[b]e denotes essential constraint, o optional constraint.
[c] P_{min} is the minimum number of optional constraints allowed, P_{max} is the maximum number of optional constraints allowed.

solutions lay on low ranks*. The pharmacophore constraints were built individually for each complex with the goal of improving the ranking of solutions with low root mean square deviation (RMSD) from the ligand position in the crystal structure. Comparisons of FLEXX and FLEXX-PHARM docking results for PDB complexes 1azm, 1etr, 1ivf, 1thy, 2tmn and 4tln are shown in Table 3. Three examples are discussed below (see also Figure 8 for the ligand structures and Figure 9 for the docking results).

1azm

1azm contains the complex of carbonic anhydrase with acetazolamide (AZM.1azm) shown in Figure 8a. Carbonic anhydrase II is a metalloprotease containing a zinc ion in its active site. The ligand is mainly anchored by interactions of the sulfonamide group to the zinc ion and hydrogen bonds to THR 199. The N-acetyl group is involved in interactions to solvating water molecules; these were not present in the docking calculation. Meanwhile, the terminal methyl group points into a hydrophobic pocket formed by LEU 131, LEU 141 and PHE 91.

The constraints were previously described for 1azm (see Figure 2 and the related discussion). They are designed to ensure that the ligand in the docking solutions forms an interaction with the zinc ion and one or two hydrogen bonding interactions with THR 199, and that a carbon atom is found in the spatial volume, thus directing the ligand into the hydrophobic area in the binding site.

The top scoring FLEXX solution deviated from the crystal structure ligand position by 2.61 Å RMSD, while the best solution with an RMSD of less than 1.5 Å lay low down at rank 176. When the ligand was docked with FLEXX-PHARM, the RMSD of the top ranking solution decreased to 1.84 Å and the best solution with RMSD less that 1.5 Å moved up to rank 12. The positions of the ligand in the top ranking docking solutions from FLEXX and FLEXX-PHARM can

*In these discussions the highest ranking solution is that on rank one, while a high score is actually one with a more negative score (a high absolute value); high scoring solutions lie on higher ranks.

Table 3. Class one results for comparison of FLEXX and FLEXX-PHARM

	FLEXX				FLEXX-PHARM			
	Score rank 1	RMSD rank 1 (Å)	Best RMSD < 1.5 Å[a]	Rank best RMSD	Score rank 1	RMSD rank 1 (Å)	Best RMSD < 1.5 Å[a]	Rank best RMSD
1azm	−24.17	2.61	1.41	176	−15.42	1.84	1.41	12
1etr	−39.55	7.24	1.44	11	−39.11	1.53	1.28	2
1ivf	−34.50	6.97	1.49	5	−41.63	1.22	1.22	1
1thy	−24.98	2.67	1.44	76	−20.62	1.96	1.45	5
2tmn	−29.47	5.16	0.73	187	−22.18	1.07	1.07	1
4tln	−28.37	3.68	1.26	253	−13.26	1.19	1.19	1

[a]Best scoring solution with RMSD < 1.5 Å, or if no such solution exists, solution with best RMSD.

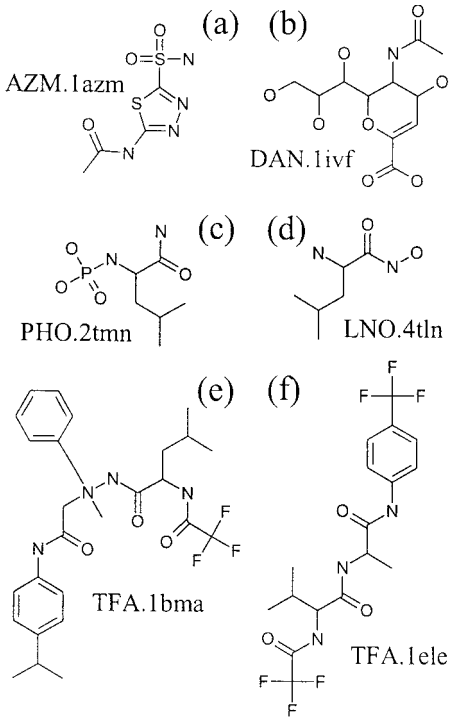

Figure 8. The structures of the ligands from the complexes used in the class one comparisons; (a) 1azm, (b) 1ivf, (c) 2tmn, (d) 4tln and in the class two comparisons; (e) 1bma, (f) 1ele.

be seen compared to the crystal structure position in Figure 9a.

The top ranking solutions now have a much lower score as most of the solutions with a high score are rejected. However, when these scores are compared to the score allocated by FLEXX to the crystal structure (−15.15), they are similar. The problem with the highest ranking solutions was that, although the two ends of the ligand are found to lie in positions close to that of the crystal structure, the position of the ring deviated considerably. 1azm proves to be a tricky case for FLEXX and FLEXX-PHARM because there are many possibilities for placing the ligand so that it makes more directed interactions with the receptor than it does in the crystal structure position.

1ivf

1ivf contains influenza neuraminidase complexed with the substrate analogue ligand DAN.1ivf shown in Figure 8b. The active site of the neuraminidase is polar containing a large number of ionized side chains, especially arginines. In the crystal structure, the ligand forms hydrogen bonding interactions with an arginine residue near the entrance to the active site (ARG 371). There are also hydrogen bonding interactions deeper in the pocket allowing the methyl group to sit in a hydrophobic region in the interior of the site, while water mediates other hydrogen bonding interactions to side chains (again water molecules were not included in the docking calculations).

One simple interaction constraint with the ring of TRP 178 was used to ensure that in all solutions some part of the ligand must form an interaction in the hydrophobic region of the site.

FLEXX found a solution with low RMSD for 1ivf (RMSD 1.49 Å on rank 5) but this solution lay alone among many solutions that took up alternate binding modes, including the solution on rank one (RMSD 6.97 Å). Docking with FLEXX-PHARM resulted in the best solution with an RMSD less than 1.5 Å lying on rank one and all the top scoring solutions showing the correct binding mode (all top 20 solutions had an RMSD less than 2.0 Å). The top solution had a better score (−41.63) than the top FLEXX solution (−34.50).

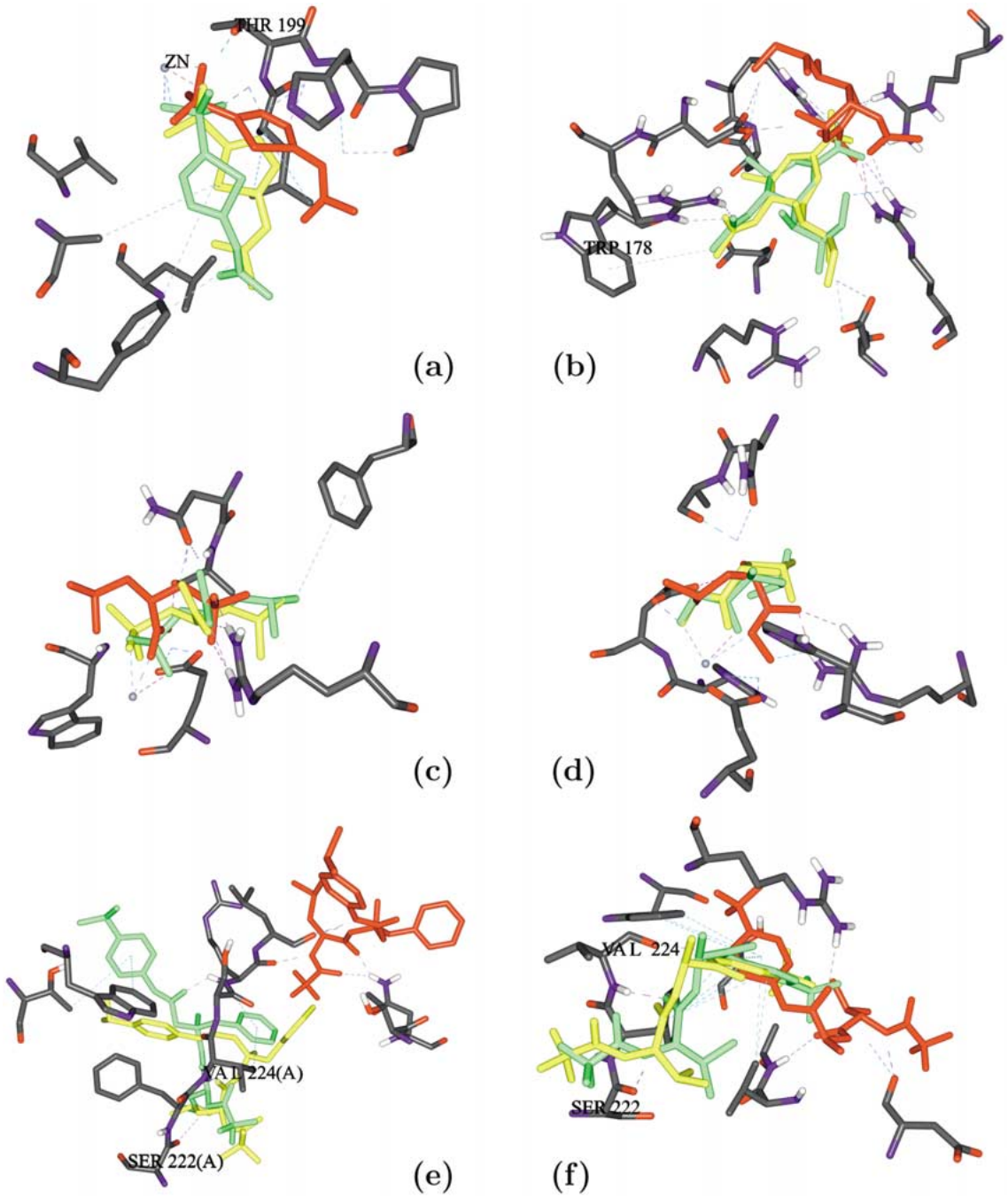

Figure 9. The ligand positions in the rank one docking solutions of FLEXX (red) and FLEXX-PHARM (green) and the crystal structure (yellow) for the class one comparisons: (a) 1azm, (b) 1ivf, (c) 2tmn, (d) 4tln and the class two comparisons: (e) 1bma, (f) 1ele. Only the interacting amino acids are shown for clarity while interaction constraint amino acids are labeled.

The solutions found by FLEXX with a high RMSD are generated because FLEXX finds a binding mode far from that of the crystal structure with very favorable interactions. The ligand is often placed in the region of ARG 371 where there are also many other hydrogen bonding residues and in a conformation where it does not extend deeper into the site (see Figure 9b). These interactions are rewarded with a high score which means these solutions are ranked highly along with correct binding mode. With the constraint, the methyl group was always placed in the correct region deeper in the active site.

The fact that the top scoring FLEXX-PHARM solution had a better score than that of FLEXX is an example of how FLEXX-PHARM allows for the appearance of novel solutions. This is a consequence of the reduced search space imposed by the constraints. The constraints restrict the conformational space of the ligand meaning more partial solutions found in this space with low score at the start of the incremental construction remain in the calculation and eventually achieve a high score at the end. In FLEXX these solutions drop out as there are many more partial solutions with a better score at each step. (This phenomena is explored further in the Class Two results section.) Of course simply enlarging the parameters that control how many partial solutions proceed to the next incremental construction step (increasing k) in FLEXX allows the appearance of these higher scoring solutions (and in this case the solution with score -41.63 found with FLEXX-PHARM). However, the runtime of FLEXX is highly dependent on k and FLEXX-PHARM can be used to find the same result much faster, which is especially important in screening applications.

2tmn (and 4tln)

2tmn contains the complex of thermolysin with the ligand PHO.2tmn shown in Figure 8c. Thermolysin is a metalloprotease with a zinc ion in its active site. The site itself is quite small and narrow and the zinc is not deeply buried. The ligand is positioned in the site so that the phosphate group coordinates to the zinc and also forms hydrogen bonds with water molecules. The carbamoyl group forms an hydrogen bonding interaction with ARG 203 and the isopropyl group lies deeply buried in a hydrophobic pocket, interacting with LEU 133, VAL 139, VAL 192 and LEU 202.

Many inhibitors of thermolysin such as PHO.2tmn are transition state analogues. In PHO.2tmn the phosphate group mimics the tetrahedral transition state at the peptide carbonyl carbon after addition of water, where two of the phosphate oxygens complete coordinations to the zinc. For this reason, two spatial constraints were chosen to contain oxygen atoms in the regions where the zinc ion coordinations are completed. These spheres were made as large as possible without overlapping (radius 1.2 Å). An additional spatial constraint to contain a carbon atom was also built in the hydrophobic region.

FLEXX found binding solutions with a low RMSD for the thermolysin complex 2tmn but on a very low rank (RMSD 0.73 Å on rank 187). With the pharmacophore constraints, FLEXX-PHARM produced the best solution with an RMSD less than 1.5 Å on rank one. The top FLEXX docking solutions revealed a binding mode which is more or less reversed to that in the crystal structure (see Figure 9c). The carbamoyl oxygen atom formed a single coordination to the zinc ion, leaving two of the phosphate oxygens free to interact with the ARG 203 and hence maximizing directed interactions. With these constraints, FLEXX-PHARM threw away the solutions with the reversed binding mode. Solutions with a lower score but the correct binding mode with respect to the crystal structure moved up to higher ranking positions.

The 4tln complex is another thermolysin complex and is brought in at this point because the same constraints were used for this complex as for 2tmn. The ligand, LNO.4tln (Figure 8d), contains a hydroxomate group that coordinates to the zinc ion, has deeply buried methyl groups and binds in a very similar mode to the ligand in 2tmn. In this case, FLEXX also found a reversed binding mode (see Figure 9d). With the constraints, FLEXX-PHARM again placed the solutions with the correct binding modes on the top ranks.

As both complexes contain thermolysin and the same constraints were used for both proteins, simple cross-docking tests (i.e. docking the ligand from one complex into the protein from the other) were carried out. The results are shown in Table 4. Again, FLEXX-PHARM was able to improve on the performance of FLEXX by bringing the solutions with better RMSD up to the higher ranks.

General observations

One point that must be mentioned first is the constraints themselves. The derivation of sets of pharmacophore constraints is not a straightforward task and has become an area of research within itself. We try to demonstrate here how FLEXX-PHARM functions and how it can be applied to certain problems but not how to derive pharmacophore models.

Table 4. Cross-docking results for comparison of FLEXX and FLEXX-PHARM with the 2tmn and 4tln PDB complexes

Receptor	Ligand	FLEXX				FLEXX-PHARM			
		Score rank 1	RMSD rank 1 (Å)	Best RMSD < 1.5 Å[a]	Rank best RMSD	Score rank 1	RMSD rank 1 (Å)	Best RMSD < 1.5 Å[a]	Rank best RMSD
2tmn	4tln	-24.14	3.22	1.41	173	−14.69	1.55	1.41	2
4tln	2tmn	-34.60	5.30	1.43	343	−20.04	1.77	1.45	2

[a]Best scoring solution with RMSD < 1.5 Å, or if no such solution exists, solution with best RMSD.

As expected, FLEXX-PHARM improved the rank of docking solutions containing the ligand in the correct binding mode, i.e. with low RMSD, in all cases. This was done by rejecting docking solutions with higher RMSD via the constraints. FLEXX-PHARM also ran faster than FLEXX in these six examples; the running times for FLEXX-PHARM ranged from 67.6% (2tmn) to 96.0% (1etr) of the corresponding FLEXX run time (the average FLEXX-PHARM run time was 26.3 s on a 750 MHz Sun ULTRASparc III processor).

Class two: improving binding mode prediction

Complexes were chosen from the FLEXX200 set on the basis that FLEXX (with default parameters) produced no good solutions with respect to RMSD. The pharmacophore constraints were built to guide FLEXX-PHARM to identify the correct binding mode. Comparisons of the FLEXX and FLEXX-PHARM docking results for the PDB complexes 1bma, 1eap, 1ele, 1hdc, 1rob and 8gch are shown in Table 5. Two examples are discussed in more detail below (see also Figure 8 for the ligand structures and Figure 9 for the docking results).

1ele

1ele shows elastase bound with the ligand TFA.1ele (Figure 8f). This complex is characterized by a rather small number of directed interactions for the ligand size. Two of these interactions were selected to provide the constraints; hydrogen bonding interactions with VAL 224 and SER 222.

For 1ele, FLEXX places the ligand in the defined active site but in binding modes with high RMSDs, particularly on the top rank (10.73 Å). With the two constraints in FLEXX-PHARM the best solution with respect to RMSD (1.32 Å) was placed on rank one. The ligand positions are shown in Figure 9f.

1bma

1bma again contains elastase but complexed with the ligand TFA.1bma (Figure 8e). This larger ligand binds in a similar mode to that in 1ele but with only the same number of similar directed interactions. This means that 1bma is a difficult example for the FLEXX fragment placement algorithm. The constraints finally chosen consisted of three main directed interactions with the protein (hydrogen bonding interactions with VAL 224 and SER 222) and a hydrophobic interaction with VAL 224 for placement of the the phenyl ring.

For this complex, FLEXX missed the experimentally observed binding mode completely, with the best RMSD (12.54 Å) lying on rank 143. Using constraints in FLEXX-PHARM resulted in docking solutions with an improved RMSD of approximately 4 Å on the top rank and an overall best RMSD of 3.87 Å. The solutions produced by FLEXX did not lie in the defined active site but in a region that could be described as being on the 'outside' of the protein where it could form more directed interactions. FLEXX-PHARM produced much improved docking solutions where the various branches of the ligand lie in the correct regions of the active site. However, the overall ligand position still deviated considerably from the crystal structure, as can be seen in Figure 9e.

The 1bma constraints used here were constructed individually for the 1bma complex, which contains a phenyl ring not present in 1ele. The 1bma constraints set contained the constraints from 1ele (optionally with at least one from the two required) plus extra directed and hydrophobic interactions. These two examples were used to test constraint transferability, i.e. how does a set of constraints derived for one complex perform in another? The 1ele constraints applied to 1bma gave surprisingly successful results (best RMSD of 2.24 Å on rank seven) even though the constraints were much less restrictive. As expected, however, applying the 1bma constraints to 1ele resulted in no

Table 5. Class two results for comparison of FLEXX and FLEXX-PHARM

	FLEXX				FLEXX-PHARM			
	Score rank 1	RMSD rank 1 (Å)	Best RMSD < 1.5 Å*	Rank best RMSD	Score rank 1	RMSD rank 1 (Å)	Best RMSD < 1.5 Å*	Rank best RMSD
1bma	−10.59	13.41	12.54	143	−8.41	3.94	3.87	33
1eap	−29.12	3.72	3.48	25	−19.06	1.94	1.39	3
1ele	−18.22	10.73	4.32	18	−15.76	1.32	1.32	1
1hdc	−9.61	11.74	11.50	5	−7.02	1.16	1.16	1
1rob	−19.92	7.70	4.89	114	−15.97	1.42	1.42	1
8gch	−20.92	8.91	6.06	106	−24.44	3.56	3.32	3

[a]Best scoring solution with RMSD < 1.5 Å, or if no such solution exists, solution with best RMSD.

docking solutions because the constraints were more specific to the 1bma ligand. This example again illustrates the challenges of constraint selection and underlines the fact that more specific constraints are not necessarily more favorable.

General observations

FLEXX-PHARM was used to find the correct binding mode (RMSD less than 2.0 Å) on the top rank in four from the six test cases; 1eap, 1ele, 1hdc, 1rob. For the remaining two cases, 1bma and 8gch, the RMSDs could be greatly improved compared to the FLEXX results.

FLEXX-PHARM can find the correct docking solution when FLEXX fails to do so. This is because constraints can be used to throw away solutions that, for example, do not form key interactions found in the crystal structure. At each fragment building step, FLEXX retains only the k best solutions in terms of score for the next step (the default values are 400 plus 100 per alternative base fragment). This means a partial ligand placement that matches the crystal structure may not be retained for the next step because it had a low score. In FLEXX-PHARM, many partial ligand placements will be rejected because they cannot match the constraints, in which case the correct partial ligand placement will be accepted into the k best solutions. In this way, FLEXX-PHARM allows the conformational space around the constraints to be explored more rigorously. New solutions can therefore appear in the final solutions set that are not seen with FLEXX. The fact that FLEXX-PHARM guides the docking process by rejecting solutions may also mean that, in cases where FLEXX finds no good docking solutions at all, no set of pharmacophore constraints can improve the docking results.

For five of the six class two examples, the FLEXX-PHARM run time was shorter than that of FLEXX. However, for 8gch, FLEXX-PHARM ran in 104% of the time taken by FLEXX.

The numbers of solutions thrown away at each fragment building step varies considerably between ligands. The filter test that rejects these solutions is most often the distance checks. On average for the twelve examples given in the class one and class two comparisons, 25% of all solutions rejected were rejected by the logical checks, 65% by the distance checks and 10% by the directed tweak checks. By omitting the directed tweak checks it was possible to speed up the calculations further in half of the twelve examples (the calculations taking longer or approximately the same time for the other half). But this had other consequences for the docking results – less solutions are rejected at each step meaning there was less chance for lower scoring placements to be accepted into the k best solutions. For several of the examples, particularly for class two where improvement of the docking results relies on the appearance of novel solutions in the final set, inclusion of the directed tweak checks was necessary to achieve the results shown in Tables 3 and 5.

Class three: application of FLEXX-PHARM *to virtual screening*

Whilst the objectives in the class one and two comparison tests were to improve a single run of FLEXX, the objectives in a virtual screening test are a little different. This time the relative scores of the top docking solution for different ligands is of utmost importance (in terms of re-ranking the database). As shown above, the correct binding modes of ligands often have a

lower score than the top FLEXX solution which could be problematic. However, in virtual screening experiment, ligands that bind well to the target protein must only have a *better* score than ligands that do not. Therefore, it is more important that a set of pharmacophore constraints rejects ligands that fail to meet the criteria of the screen (or FLEXX-PHARM allows only unlikely low scoring solutions that fit the constraints).

Another major aspect of virtual screening is speed. One of the main ideas of incorporating pharmacophore constraints into docking for virtual screening is that totally unsuitable molecules can be detected without having to carry out a docking calculation at all, thus speeding up the calculation. A further way of speeding up docking calculations with FLEXX is to reduce the numbers of partial solutions taken through from one fragment building step to the next (i.e. reduce the parameters for k). FLEXX-PHARM might maintain the quality of the docking results with smaller k values simply because FLEXX-PHARM often retains fewer solutions at each step. This hypothesis is tested in this section by using parameter values for k of 200 plus 50 per alternative base fragment (half the default values).

The targets in this section were chosen mainly because there were sets of active molecules available to us. The negative dataset consisted of a subset of 356 drug-like molecules from the CSD dataset [27] plus all ligands from the FLEXX200 dataset. The sets of actives for each example ranged in number from 10 to 69 and are described in more detail for each target below. The dummy molecules were scanned using the program FTREES [28] to ensure that none of the active molecules were duplicated in the negative set. The pharmacophores for these tests were selected using two methods; overlaying of protein-ligand complexes to allow the selection of constraints and taking pharmacophore constraints from the literature. The pharmacophore constraints are shown in Table 6 and in Figure 10. The experiments were all carried out on a 750 Mhz Sun ULTRASparc III processor. FLEXX and FLEXX-PHARM timing results are compared in Table 7 while the enrichment achieved for each dataset is displayed in Figure 11.

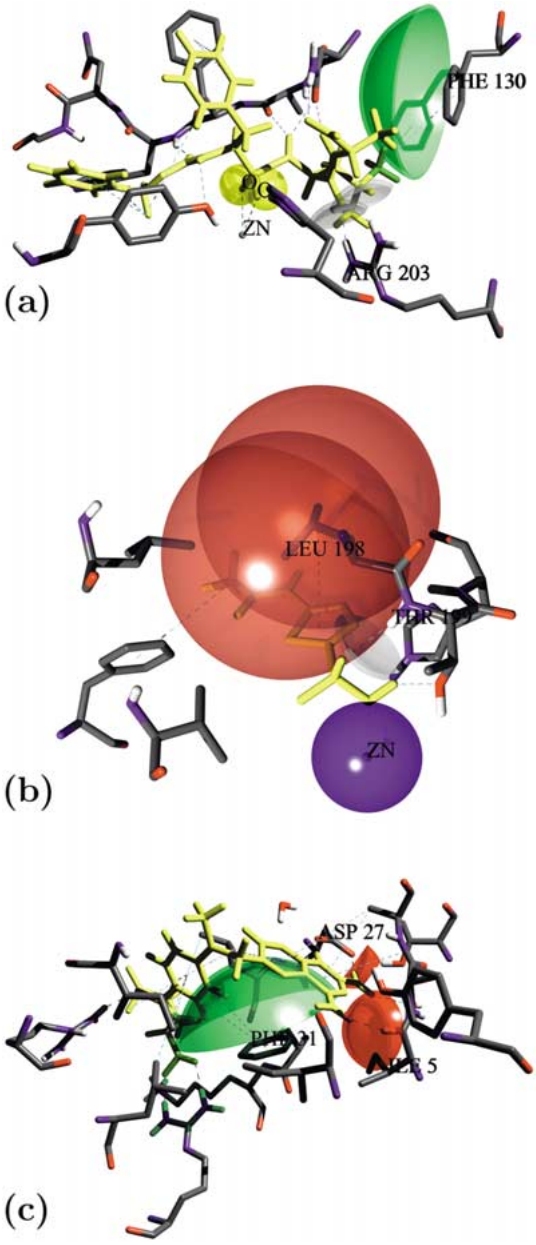

Figure 10. The pharmacophore constraints in the active sites of (a) thermolysin (4tmn), (b) carbonic anhydrase (1azm) and (c) DHFR (4dfr). Only the interacting amino acids are shown for clarity.

Thermolysin 4tmn

The target protein for the thermolysin screening experiment was taken from the 4tmn complex. The active molecules were taken from the thermolysin complexes 8tln, 7tln, 6tmn, 5tmn, 5tln, 4tmn, 3tmn, 2tmn, 1tmn, 1tlp, 1lne and 1lna7. The crystal structures of all sites were overlayed in order to examine the binding properties of the complexes. Most importantly, all the ligands form a double coordination with the zinc ion mimicking the transition state found in the natural substrate (see thermolysin 2tmn in the class one discussions above). Many other ligands bind to thermolysin without forming this double coordination. Therefore, two spatial constraints to contain oxygen

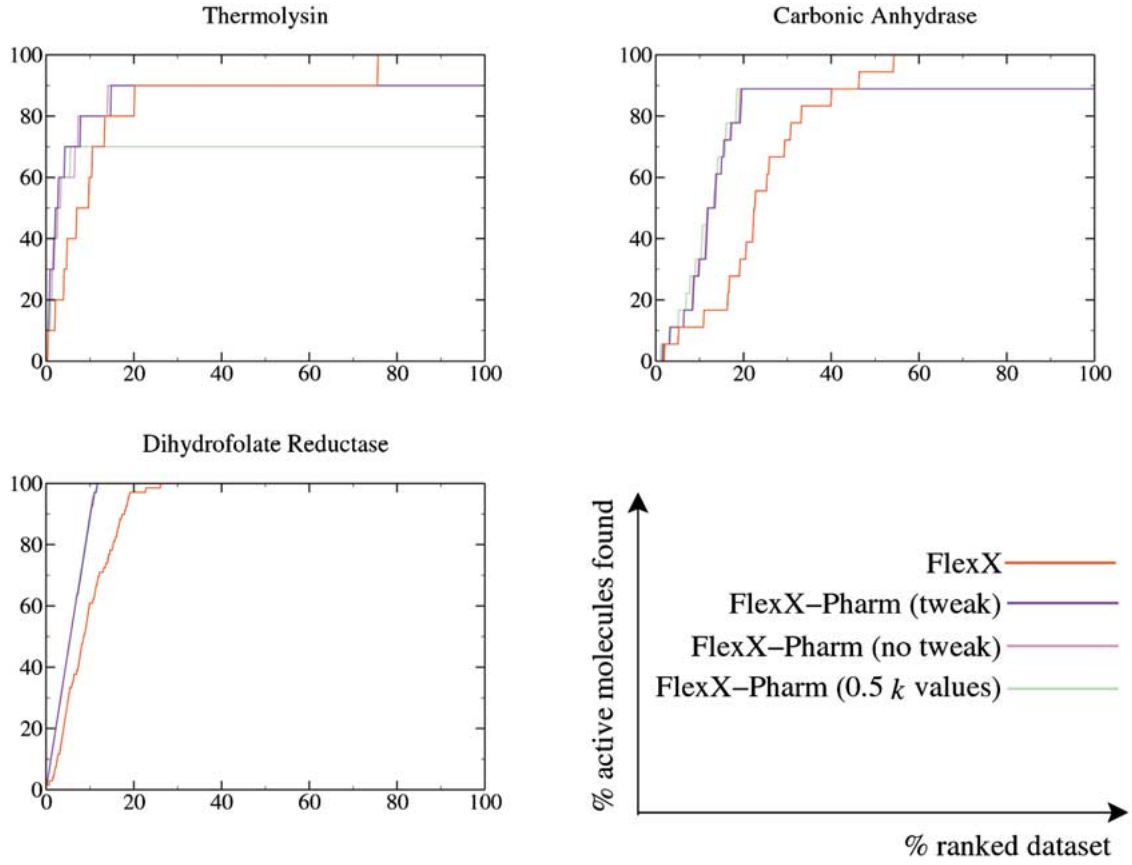

Figure 11. Enrichment curves for FLEXX-PHARM compared to FLEXX for thermolysin (4tmn), carbonic anhydrase (1azm) and DHFR (4dfr). Refer to the texts for more details about the various FLEXX-PHARM plots labeled in the legend.

atoms were built in the site as with thermolysin in class one. The other constraints chosen were a hydrophobic interaction with PHE 130 and optional hydrogen bonding interactions with ARG 203 (see Figure 10a).

The timings for the screening runs are shown in Table 7 and the enrichment curves can be seen in Figure 11. FLEXX-PHARM with directed tweak checking completed the screening run in approximately the same time as FLEXX, requiring about 62 s per dataset molecule. However, when directed tweak checking was omitted FLEXX-PHARM took significantly less time; about 13 s less per molecule. FLEXX-PHARM also improved the enrichment. The enrichment curves are very similar with and without directed tweak checking. At about 15% of the re-ranked dataset, FLEXX-PHARM and FLEXX both identified 90% of the active molecules. FLEXX-PHARM failed to identify the remaining active molecule which FLEXX ranked at approximately 75% of the dataset. FLEXX-PHARM rejected about 85% of molecules in the dataset because no docking solution could be found that fitted the constraints.

Reducing the k parameters to half (with no directed tweak checking) reduced the screening time further from about 50 s per molecule to about 41 s per molecule. However, this also caused a deterioration in the screening results. The enrichment curve follows a similar path to the curves from FLEXX-PHARM with default parameters for k but one can see that FLEXX-PHARM failed to identify a further two active molecules. When FLEXX-PHARM used the reduced parameter values no docking solutions fitting the constraints were found for about 88% of the dataset molecules.

It was clear from these screening runs that omitting the directed tweak checking had the benefit of considerably speeding up the calculations without having a detrimental effect on the results. More importantly, the screening runs with thermolysin revealed that FLEXX-PHARM does not necessarily run in around the same

Table 6. Details of the pharmacophore constraints used for the class three screening runs

	Class three			
	Constraint type	Detail[a]	e/o[b]	P_{min}, P_{max}[c]
4tmn	spatial	O	e	
	spatial	O	e	
	phenyl_center	CG PHE 130(E)	e	
	h_don	NH1 ARG 203(E)	o	
	h_don	NH2 ARG 203(E)	o	1,2
1azm	metal	ZN	e	
	h_don	N THR 199	e	
	ch3_phe	CD1 LEU 198	o	
	ch3_phe	CD2 LEU 198	o	1,2
4dfr	h_acc	OD1 ASP 27(A)	e	
	h_acc	O ILE 5(A)	e	
	phenyl_center	CG PHE 31(A)	e	

[a]If interaction constraint, then the name of the receptor atom (PDB nomenclature: atom name, amino acid code, amino acid number and optionally chain identifier). If a spatial constraint, then the element type associated with that constraint.
[b]e denotes essential constraint, o optional constraint.
[c] P_{min} is the minimum number of optional constraints allowed, P_{max} is the maximum number of optional constraints allowed.

Table 7. Timings for the database screening runs for FLEXX and FLEXX-PHARM

	4tmn		1azm		4dfr	
	Average per mol. (s)	Total time (h)	Average per mol. (s)	Total time (h)	Average per mol. (s)	Total time (h)
FLEXX	62.61	9.88	30.25	4.83	42.94	7.44
FLEXX-PHARM[1]	62.31	9.83	23.76	3.79	33.74	5.84
FLEXX-PHARM[2]	49.87	7.87	21.30	3.40	20.51	3.56
FLEXX-PHARM[3]	40.77	6.43	18.07	2.89	16.78	2.91

FLEXX-PHARM[1] FLEXX-PHARM with directed tweak checking.
FLEXX-PHARM[2] FLEXX-PHARM without directed tweak checking.
FLEXX-PHARM[3] FLEXX-PHARM without directed tweak checking and with half the default k values.
Note: Average time taken per molecule is entered in seconds, while total time taken for the screen is entered in hours.

time as FLEXX or less for all molecules. It took FLEXX-PHARM more than double the corresponding FLEXX time in certain cases. This was unacceptable, especially in a screening scenario. The long run times were due to very long candidate countergroup set lists for these molecules. For this reason, the post-docking filter mode was enabled. That meant that when the lists were very long, the post-docking filter was automatically used for that molecule and the run time for the molecule was approximately the same time as it would be in FLEXX. All results shown in class three section have the post-docking filter mode enabled with a default maximum list length of 20 000.

The post-docking filter mode was invoked for 8% of the molecules in the dataset for the thermolysin screening run. Unfortunately, this had implications because use of the post-docking filter for one of the active molecules changed the results considerably. With the look-ahead filter there were docking solutions found for the ligand from the 6tmn complex. These results would have placed this ligand at about 9% of the re-ranked dataset. However, when the post-docking filter was applied, no solutions were found that fitted the constraints. For thermolysin with these constraints, the advantage of speed in a screening experiment has to be weighed against the possibility of missing active molecules.

Carbonic anhydrase 1azm

The 1azm complex discussed previously is again used here. The active molecules consisted of six ligands taken from the PDB complexes 1azm, 1cil, 1cnw, 1cnx, 1caz and 1zsb and a set of 13 lead structures selected by Grüneberg et al. in a pharmacophore screening study [29]. The constraints were chosen by overlaying the PDB complexes and examining the binding modes. The selected constraints differed slightly from those used earlier (which are listed in Table 2) and consisted of a metal interaction with the zinc ion, a hydrogen bonding interaction with the neighboring THR 199 (as previously) and optional hydrophobic interactions with the methyl groups on LEU 198, as shown in Table 6 and in Figure 10b.

The timings for the screening runs are shown in Table 7 and the enrichment curves can be seen in Figure 11. FLEXX-PHARM with directed tweak checking completed the screens in 78% of the time of FLEXX, and FLEXX-PHARM without directed tweak checking and with halved k parameter values (no directed tweak) in 70% and 60% of the time respectively. All three FLEXX-PHARM enrichments were better than the enrichment with FLEXX; FLEXX-PHARM finds 90% of the active molecules in approximately 20% of the dataset, whereas FLEXX finds 90% of the actives in about 40% of the re-ranked dataset. FLEXX-PHARM did not find any docking solutions that fitted the constraints for 70% of the molecules in the dataset.

For two of the active molecules, FLEXX-PHARM did not find docking solutions to fit the chosen set of constraints. It transpired that one of these molecules was acetate, which is very small molecule and can not possibly match these constraints. The second molecule was lead molecule 13 from Table 1 by Grüneberg et al. [29]. While the hydroxamic acid part is able to fulfill the zinc and hydrogen bonding constraints, the ligand was rejected because there are no candidates present to fulfill the interactions with LEU 198. The ring in the dihydroimidazol part is not able to form an aromatic interaction type to LEU 198 in FLEXX, whereas the ring types in the other actives are.

Dihydrofolate reductase 4dfr

Dihydrofolate reductase (DHFR) is an example where FLEXX already performs very well in screening experiments. Despite this, we have chosen DHFR here because of the availability of well evaluated pharmacophore constraints from the literature and a large inhibitor set. The active ligands were a set of 68 known DHFR inhibitors taken from Selassie et al. [30] plus methotrexate. The constraints used were selected and adapted from a set of pharmacophore constraints used by Hoffman et al. [31]. They derived constraints from the DHFR-methotrexate PDB complex 1rh3. Different combinations of these constraints were evaluated by searching the Derwent WDI dataset for known DHFR inhibitors with success. The constraints selected for use in FLEXX-PHARM were two hydrogen bonding interactions with ASP 27 and ILE 5, plus a hydrophobic interaction with PHE 31. These are shown in Figure 10c.

The timings for the screening runs are shown in Table 7 and the enrichment curves can be seen in Figure 11. The FLEXX-PHARM screens ran faster than FLEXX, and considerably so without directed tweak checking and with the k parameters at half the default values (in 48% and 39% of the time required by FLEXX, respectively). In addition, FLEXX-PHARM improved on the enrichment shown by FLEXX and all 69 active molecules were identified. FLEXX-PHARM did not find docking solutions to fit the constraints for approximately 85% of molecules in the dataset.

General observations

The screening experiments for thermolysin demonstrated to good effect a problem often encountered in 3D pharmacophore searching; the combinatorial explosion in the number of candidate countergroup sets potentially matching the pharmacophore. In UNITY, for example, the problem is addressed by imposing a 'timeout' where the search is halted when it runs longer than a specified time. In FLEXX-PHARM this problem was overcome by enabling the post-docking filter mode. Unfortunately, the modification meant that FLEXX-PHARM failed to identify all active molecules for thermolysin. However, there was no effect on the carbonic anhydrase and DHFR screening results. Post-docking filtering was not required for any molecules in the DHFR screening run and was required for only one molecule for the carbonic anhydrase screening run. The post-docking filter mode provides a method of dealing with problematic cases in terms of computing time but does not explore the full potential of FLEXX-PHARM.

Deactivating the directed tweak checking decreased FLEXX-PHARM run times on average by 23%. The directed tweak checking does not seem to bring any apparent advantage to the screening results, only the disadvantage of longer run times. The benefits of including the tweak checking seen for the class one and two comparison tests (the appearance of solutions

with lower RMSD) is not so apparent for screening where the best score is of utmost importance. It may be more advantageous to restrict the use of the directed tweak checks to specific molecules containing, for example, longer chains. Further investigation is required to properly evaluate the influences of the directed tweak checks.

Meanwhile, reducing the default k parameters to half the default values decreases the run time further. In the case of carbonic anhydrase and DHFR the quality of the screening results was unaffected, while for thermolysin, FLEXX-PHARM failed to identify more active molecules. Both reducing the k parameters and the use of post-docking filtering illustrate the fact that in screening the advantages of speed must be weighed against the quality of the screening results.

Conclusion

We have described the program FLEXX-PHARM; an extension to the docking package FLEXX in which the docking calculation can be guided by the inclusion of pharmacophore type constraints in the active site. This is particularly useful if information about how a ligand binds to a protein is already available and in other special cases; for example, where FLEXX does not perform well or where docking should be biased around an important area of the active site.

One challenge when using FLEXX-PHARM is that the constraints must first be obtained and then entered into FLEXX-PHARM. This requires extra preparation work on the part of the user. If the resolved structure of the protein with a similar ligand is available then this can aid the selection of constraints. Otherwise, active site hotspot identification or the mirroring of established pharmacophores onto the receptor via docking studies provide alternative methods to aid constraint selection. The constraints used here were derived from families of resolved structures contained in the PDB, or taken from pharmacophore constraints in the literature derived from families of known ligands.

Examples are given where FLEXX-PHARM significantly improved the results of docking in several PDB complexes where FLEXX did not perform particularly well. The results were improved in terms of RMSD of the ligand in the highest ranking docking solutions from the ligand in the crystal structure. The structures with improved RMSD often showed a related drop in score because the top scoring FLEXX solutions display a different binding mode. The cause lies with the standard scoring function which highly favors ligand placements that maximize directed interactions with the protein binding site. There is a need for a scoring function that can overcome these problems associated with ranking of docking solutions. Although improvements have been made, with functions such as DrugScore [32] for example, a reliable, generic scoring function is not yet in sight. FLEXX-PHARM provides a pragmatic approach for dealing with the ranking of docking solutions.

FLEXX-PHARM can also be used to improve the results of virtual screening runs. Potential leads in a database often possess similar properties – a phenomenon that other 3D screening techniques exploit but virtual screening docking tools do not. FLEXX-PHARM provides a means of capitalizing on this information during docking for screening. FLEXX-PHARM succeeded in improving dataset enrichments for the targets carbonic anhydrase, dihydrofolate reductase and thermolysin. More importantly, it was able to do this in less time than required by FLEXX.

One future area of work will be to apply FLEXX-PHARM in 'real-life' virtual screening scenarios involving much larger databases of compounds and more diverse positive datasets. Another interesting possibility would be to take cross-docking investigations further using FLEXX-PHARM combined with FLEXE [33] rather than FLEXX. A rigid receptor method does not adequately deal with the variations in the active site between complexes containing the same protein and different ligands. FLEXE is a software tool for docking flexible ligands while simultaneously taking into account side chain flexibility.

The main disadvantage of including constraints in this manner is that the generality in the flexible search of the active site is lost. This is one of the strengths of using docking above other methods in drug design, in the sense that it creates the opportunity of finding new classes of leads. This point aside, FLEXX-PHARM can certainly be used to enhance the performance of FLEXX in docking and virtual screening.

Acknowledgements

The authors thank Martin Stahl (F. Hoffman-La Roche Ltd., Basel, Switzerland) for his help and suggestions – in particular relating to the application of FLEXX-PHARM as a virtual screening tool, and Martin Bohl (Tripos Inc.) for his helpful comments on the manu-

script. This work was funded by Tripos Inc., St. Louis, MO.

References

1. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., Nucl. Acids Res., 28 (2000) 235.
2. Good, A., Curr. Opin. Drug. Discov. Dev., 4 (2001) 301.
3. Gane, P.J. and Dean, P.M., Curr. Opin. Struct. Biol., 10 (2000) 401.
4. Abagyan, R. and Totrov, M., Curr. Opin. Chem. Biol., 5 (2001) 375.
5. Rarey, M., Kramer, B., Lengauer, T. and Klebe, G., J. Mol. Biol., 261 (1996) 470.
6. Gund, P., Prog. Mol. Subcell. Biol., 5 (1977) 117.
7. Martin, Y.C., Bures, M.G., Danaher, E.A., DeLazzer, J., Lico, I. and Pavlik, P.A., J. Comput. Aid. Mol. Des., 7 (1993) 83.
8. Clark, D.E., Jones, G. and Willett, P., J. Chem. Inf. Comput. Sci., 34 (1994) 197.
9. Güner, O.F. (ed.), Pharmacophore Perception, Development, and Use in Drug Design. International University Line, La Jolla, CA, USA, 2000.
10. Goodford, P.J., J. Med. Chem., 28 (1985) 849.
11. Lawkowski, R.A., Thornton, J.M., Humblet, C. and Singh, J., J. Mol. Biol., 259 (1996) 175.
12. Verdonk, M.L., Cole, J.C. and Taylor, R., J. Mol. Biol., 289 (1999) 1093.
13. Fradera, X, Knegtel, R.M.A. and Mestres, J., Prot. Struct. Funct. Genet., 40 (2000) 623.
14. Ewing, T. and Kuntz, I.D., J. Comput. Chem., 18 (1997) 1175.
15. Thomas IV, B.E., Joseph-McCarthy, D. and Alvarez, J.C., In Güner, O.F. (ed.), Pharmacophore Perception, Development and Use in Drug Design. International University Line, La Jolla, CA, USA, 2000, pp. 351–367.
16. Rarey, M., Wefing, S. and Lengauer, T., J. Comput. Aid. Mol. Des., 10 (1996) 41.
17. Rarey, M., Kramer, B. and Lengauer, T., J. Comput. Aid. Mol. Des., 11 (1997) 369.
18. Böhm, H.-J., J. Comput. Aid. Mol. Des., 6 (1992) 61.
19. Böhm, H.-J., J. Comput. Aid. Mol. Des., 8 (1994) 243.
20. Gasteiger, J., Rudolph, C. and Sadowski, J., Tetrahedron Comput. Method., 3 (1990) 537.
21. Floyd, R.W., Commun. Assoc. Comput. Machinery, 5 (1962) 345.
22. Warshall, S., J. Assoc. Comput. Machinery, 9 (1962) 11.
23. Hurst, T., J. Chem. Inf. Comput. Sci., 34 (1994) 190.
24. Tripos Associates, Inc., St. Louis, Missouri, USA. UNITY 4.2 Rapid Database Searching for Structures Matching Pharmacophore or Receptor-site Constraints, 2000.
25. Murray, C.W., Clark, D.E. and Byrne, D.G., J. Comput. Aid. Mol. Des., 9 (1995) 381.
26. Handschuh, S., Wagener, M. and Gasteiger, J., J. Chem. Inf. Comput. Sci., 38 (1998) 220.
27. Kramer, B., Rarey, M. and Lengauer, T., Protein Struct. Funct. Genet., 37 (1999) 228.
28. Rarey, M. and Dixon, J.S., J. Comput. Aid. Mol. Des., 12 (1998) 471.
29. Grüneberg, S., Wendt, B. and Klebe, G., Angew. Chem. Int. Ed., 40 (2001) 389.
30. Selassie, C.D., Fang, Z.X., Li, R.L., Hansch, C., Debnath, G., Klein, T.E., Langridge, R. and Kaufman, B.T., J. Med. Chem., 32 (1989) 1895.
31. Hoffman, R., Li, H. and Langer, T., In Güner, O.F. (ed.), Pharmacophore Perception, Development and Use in Drug Design. International University Line, La Jolla, California, USA, 2000, pp. 353–367.
32. Gohlke, H., Hendlich, M. and Klebe, G., J. Mol. Biol., 295 (2000) 337.
33. Claußen, H., Buning, C. and Rarey, M., J. Mol. Biol., 308 (2001) 377.