pharmACOphore: Multiple Flexible Ligand Alignment Based on Ant Colony Optimization

Oliver Korb,[†,‡] Peter Monecke,[§] Gerhard Hessler,[§] Thomas Stützle,[||] and Thomas E. Exner*[,†]

Fachbereich Chemie and Zukunftskolleg, Universität Konstanz, Konstanz, Germany, R&D LGCR/Structure, Design & Informatics, Sanofi-Aventis Deutschland GmbH, Frankfurt, Germany, and IRIDIA, CoDE, Université Libre de Bruxelles (ULB), Brussels, Belgium

Received January 13, 2010

The flexible superimposition of biologically active ligands is a crucial step in ligand-based drug design. Here we present pharmACOphore, a new approach for pairwise as well as multiple flexible alignment of ligands based on *ant colony optimization* (ACO; Dorigo, M.; Stützle, T. *Ant Colony Optimization*; MIT Press: Cambridge, MA, USA, 2004). An empirical scoring function is used, which describes ligand similarity by minimizing the distance of pharmacophoric features. The scoring function was parametrized on pairwise alignments of ligand sets for four proteins from diverse protein families (*cyclooxygenase-2, cyclin-dependent kinase 2, factor Xa and peroxisome proliferator-activated receptor γ*). The derived parameters were assessed with respect to pose prediction performance on the independent *FlexS* data set (Lemmen, C.; Lengauer, T.; Klebe, G. *J. Med. Chem.* **1998**, *41*, 4502−4520) in exhausting pairwise alignments. Additionally, multiple flexible alignment experiments were carried out for the pharmacologically relevant targets *trypsin* and *poly (ADP-ribose) polymerase* (PARP). The results obtained show that the new procedure provides a robust and efficient way for the pairwise as well as multiple flexible alignment of small molecules.

INTRODUCTION

The binding of a ligand to its protein target requires complementarity of both binding partners in terms of shape and electrostatics. Optimization of such interactions to increase potency is a major goal in drug design. Although the number of three-dimensional protein−ligand structures is permanently increasing, there are still many pharmaceutically relevant proteins for which no three-dimensional structure is known. In this situation, ligand-based design techniques like 3D-QSAR (*three-dimensional quantitative structure−activity relationship*) or pharmacophore-based methods are used for the optimization of ligand potency. A prerequisite for these methods is the molecular alignment of biologically active ligands.

Different approaches to the alignment problem have been proposed in the literature. Only few of these will be highlighted, while we refer to the literature for a more detailed overview.[1,2] A flexible alignment algorithm inspired by the docking algorithm *FlexX*,[3] called *FlexS*,[4] is based on a combinatorial matching procedure. It allows for the flexible superimposition of a ligand structure onto a rigid template molecule. Like in FlexX, the ligand structure is divided into fragments, which are reassembled during the search process guided by a similarity-based scoring function. Other approaches[5−8] use *genetic algorithms* (GA).[9] The approach published by Jones et al.[5] is inspired by the genetic algorithm used in the docking approach GOLD[10] and allows for the multiple flexible alignment of ligand structures. In Handschuh et al.,[7] the GA is combined with a numerical optimization method and allows two ligand structures to adapt flexibly to each other. In FLAME[8] (*FLexibly Align MolEcules*), pairwise and multiple flexible ligand alignment can be performed. A multiple flexible alignment approach called QUASI is proposed by Todorov et al.[11] The ligands are aligned flexibly with respect to a receptor model, which is coevolved simultaneously with the flexible ligand alignment. In this way, in addition to the ligand superimposition also a receptor interaction model is retrieved, which can be used for ligand-based virtual screening. The optimization itself is carried out using a Monte Carlo stochastic tunneling procedure. ROCS[12−14] (*Rapid Overlay of Chemical Structures*) from OpenEye is a shape matching application maximizing the overlap of smooth Gaussian functions representing the molecular volumes of the ligand structures. Shin et al.[15] used a modified SEAL similarity index[16] combined with an energy penalty term to generate pairwise alignments. LigMatch[17] is based on a geometric hashing method. Other approaches generate multiple alignments by combinations of pairwise ones. Jones et al.[18] select the pairwise alignments to be combined by a genetic algorithm. Another method is to identify pharmacophore features in pairwise alignments and then to merge all alignments exhibiting significant subsets of corresponding features.[19] MOGA[20−22] uses a multiobjective genetic algorithm to generate conformations of the ligands and a mapping of corresponding pharmacophoric points, which are then matched onto each other using least-squares fitting. Noteworthy, this approach also accounts for the problem of partial overlap.[21]

* Corresponding author: phone +49 7531 882015, fax +49 7531 883587, e-mail thomas.exner@uni-konstanz.de.
† Universität Konstanz.
‡ Present address: Cambridge Crystallographic Data Centre, Cambridge, U.K.
§ Sanofi-Aventis Deutschland GmbH.
|| Université Libre de Bruxelles (ULB).

10.1021/ci1000218
Published on Web 08/30/2010

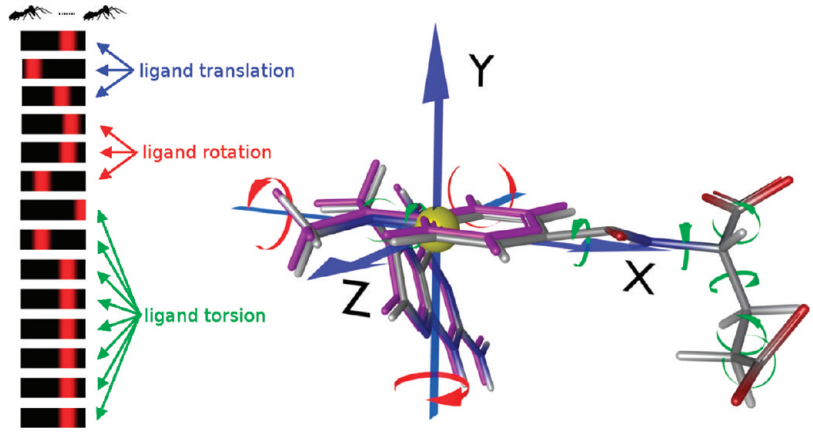

Figure 1. Illustration of the problem encoding exemplified for the pairwise alignment problem. The template structure is shown in magenta, while the flexible ligand is color-coded according to the atom type. The yellow sphere defines the origin of the local coordinate system represented by the large arrows. Translational as well as rotational degrees of freedom are defined with respect to these three principal axes. Small arrows mark torsional degrees of freedom, that is, rotatable single bonds not part of a ring system. All these degrees of freedom are discretized resulting in *pheromone vectors* as illustrated in the left part of the picture. The artificial ant colony employed in the ant colony optimization approach uses this representation to mark favorable values for each degree of freedom by depositing a pheromone trail onto the appropriate entry of the corresponding pheromone vector. The amount of pheromone deposited is proportional to the objective function value of the solution.

In recent years, numerous new methods have been described indicating that the alignment problem is not considered to be solved.[23] An active field of research is the identification of a fair balance of internal conformational energy and similarity score for 3D alignments of molecules. Also, the alignment scoring function may not be easily customizable.[8] Especially in the context of target-specific scoring functions, the ability to fine-tune the influence of specific scoring function contributions is of major importance. Here, we present a new approach called pharmACOphore for the structural alignment of multiple ligands addressing these problems. We used a novel algorithm in combination with an extensive parametrization process to obtain a similarity-based alignment scoring function of a nonphysical nature. The method is based on a hybrid *ant colony optimization* algorithm (ACO),[6] which combines a global optimization based on a MAX−MIN Ant System[24] with a local search by the Nelder Mead simplex algorithm.[25] This hybrid algorithm was adopted from our protein−ligand docking software PLANTS[26−28] (*Protein−Ligand ANT System*), since the conformational search can be described identically in both problems. While for the protein−ligand docking problem the scoring function rewards complementarity of ligand and protein, the similarity of ligands is rewarded in the case of the alignment problem. The description of molecular similarity is based on pharmacophoric features like hydrogen bond donors and acceptors as well as ring systems. The identification of corresponding pharmacophoric features in our method solely relies on the accuracy of the scoring function. Therefore we decided to carry out an extensive parametrization process. This process allows us to balance the various scoring contributions, for example, intramolecular conformational strain vs intermolecular alignment of pharmacophoric features. The new pharmACOphore approach can be applied to the multiple flexible alignment of ligands as well as the pairwise alignment of ligand structures. We demonstrate the successful application of the new algorithm to both problems and discuss some limitations.

MATERIALS AND METHODS

Ant algorithms are inspired by the pheromone trail laying and following behavior of some ant species, which are capable of finding a shortest path between their nest and a food source. These ants use indirect communication in the form of pheromone trails to mark paths between their nest and a food source. They tend to choose paths with high pheromone intensities with a higher probability, thereby reinforcing the pheromone trail. In ACO, an artificial ant colony is employed to mimic the trail-laying behavior of real ants. The ants deposit artificial pheromone trails to mark solution components of the given optimization problem. The amount of pheromone deposited is thereby usually dependent on the solution quality. This artificial pheromone trail information is modified in subsequent iterations to increase the probability of generating high-quality solutions. In the following, we will describe how the ACO metaheuristic can be used in the context of molecular superimposition. The pharmACOphore approach is based on the same algorithmic outline as our protein−ligand docking algorithm, PLANTS, which is described in detail elsewhere.[26−28] Therefore, we will only highlight the modifications necessary to apply the approach in ligand-based drug design here.

Problem Representation and Algorithm. In order to apply the hybrid ant colony optimization algorithm as employed in PLANTS, the alignment problem must be represented appropriately (for an illustration, see Figure 1). In general, the approach takes into account translational, rotational, and torsional degrees of freedom of all ligand structures to be aligned. Ring flexibility is not accounted for in this publication. Nevertheless, the approach is capable of performing ring corner flipping. When multiple flexible ligand alignment is performed, the translational and rotational degrees of freedom for one ligand of the given ligand set, S_{lig}, can be neglected because these degrees of freedom only influence the position and orientation of the final alignment. Hence, the problem dimension becomes $n = 6(|S_{lig}| - 1) + \sum_{l \in S_{lig}} r_l$, where r_l is the number of rotatable bonds in ligand l. For the pairwise alignment problem, where all degrees of

freedom for one ligand (*template*) are kept fixed, the problem dimension equals $n = 6 + r$, where r is the number of rotatable bonds in the flexible ligand to be aligned. All continuous variables are discretized, resulting in a *pheromone vector* for each degree of freedom with as many entries as result from the discretization. In the case of a rotational or torsional degree of freedom, a discretization step-size of 1° is used resulting in 360 pheromone vector entries. Translational degrees of freedom are discretized in 0.1 Å steps, and ligand translation is performed with respect to the origin of its local frame of reference (illustrated by a sphere in Figure 1). The translational degrees of freedom of all nonfixed ligands are restricted to a spherical domain, that is, a representative point of each ligand (placed in the center of the molecule) is not allowed to leave an enclosing sphere. This sphere contains the heavy atoms of all fixed ligands representing template structures extended by 5 Å. This methodology follows binding site definitions used in protein−ligand docking calculations, where usually all protein residues up to 5−6 Å away from any ligand heavy atom as given in an experimentally determined structure are considered.

The artificial ant colony uses the pheromone vector representation to mark favorable values for each degree of freedom by depositing a pheromone trail onto the appropriate entry of the corresponding pheromone vector. The amount of pheromone deposited directly depends on the solution quality; that is, more pheromone is deposited for higher quality solutions as defined by the scoring function. In each iteration of the ACO algorithm, each ant of the artificial colony probabilistically constructs a new solution taking the already existing pheromone trails into account. Thus, vector entries with a higher pheromone intensity are chosen with a higher probability. The constructed solutions are then locally minimized with the Nelder Mead simplex algorithm, and the iteration-best ant updates the pheromone trails. While the ACO algorithm explores the search space by global optimization, the local minimization step focuses on the identification of high-quality solutions. This combination of search exploration and exploitation is especially beneficial in the case of very rugged fitness landscapes as observed for the alignment or docking problem. The use of upper and lower pheromone limits in the *MAX−MIN* Ant System[24] prevents the algorithm from prematurely converging to suboptimal solutions. Since each possible value for a degree of freedom has a nonzero probability of being selected at any stage of the optimization process, the independent selection of solution components in the probability-based solution construction step allows it to escape from local minima. Therefore diverse solutions can still be constructed in subsequent iterations of the algorithm. In pharmACOphore, the number of iterations carried out by the ACO algorithm is set to $\sigma \cdot 100$, where the scaling factor σ is set to 1 in standard settings. For a detailed description of the discretization procedure, as well as the combination of a *MAX−MIN* Ant System with the Nelder Mead simplex algorithm, we refer to our previous work on PLANTS.[26−28]

Scoring Function. The scoring function employed in pharmACOphore consists of a user-configurable part for describing the ligand similarity, as well as a clash and a torsional potential for intraligand interactions:

$$f_{\text{align}} = f_{\text{sim}} + f_{\text{clash}} + f_{\text{tors}} \tag{1}$$

Atom types and the intraligand potentials, that is, the torsional potential, f_{tors}, and the clash potential, f_{clash}, considered for each ligand structure separately, are the same as in the empirical scoring functions used in PLANTS.[26] The similarity part, f_{sim}, is dependent on the distance between identical pharmacophoric features in the different structures to be aligned. Internally, pharmACOphore automatically recognizes several pharmacophoric feature classes, which are distinguished in distance-based and directional classes. Since the distance-dependent potential is evaluated for each pair of features of a specific kind, there is no explicit one-to-one correspondence of features in multiple ligands. All combinations of desired features are considered and partial alignments may be generated if a specific feature is not present in all ligands. While no explicit volume overlap is calculated between ligands, this is implicitly accounted for by the distance-dependent features, which contribute favorably to the total score if many feature points overlap.

Distance-Based Pharmacophoric Features. Purely distance-based pharmacophoric features are used for the classes *donor* (hydrogen bond donor), *acceptor* (hydrogen bond acceptor), *donor_acceptor* (atom can act both as donor and acceptor), *nonpolar* (nonpolar atom), and *nonpolar_no_ring* (nonpolar atom not part of a ring system). In these cases, each pharmacophoric feature is represented by an atom. A limitation of this approach is that aligned hydrogen bond features for which the counter groups in the protein point in opposite directions will also contribute favorably to the total score. This particular problem could be resolved by using directional hydrogen bonding features. But for doing so, the position of an optimal hydrogen partner has to be defined, which is not possible unambiguously. Thus, multiple directional features would be needed, highly increasing the complexity of the problem. Additionally, when we look at crystal structure superimpositions, the directions of overlaid hydrogen bonding groups in different ligands of the same target do not necessarily coincide. Thus, also the introduction of directional features can prevent the identification of the correct overlay.

The similarity for these distance-based features is given by the following potential, taking into account the actual distance r between two pharmacophoric feature points:

$$f_{\text{dist}}(r, w, r_{\text{opt}}) = w \cdot \begin{cases} 0 & \text{if } r > r_{\text{cut}} \\ 1 & \text{if } r < r_{\text{opt}} \\ \left(1 + s\dfrac{r^2 - r_{\text{opt}}^2}{r_{\text{cut}}^2 - r_{\text{opt}}^2}\right)^{-1} & \text{otherwise} \end{cases} \tag{2}$$

A weight w is assigned for each pharmacophoric feature class. For example, aligning a donor−donor pair may be assigned an attractive weight of −6, while for mismatching pairs, for example, a donor−acceptor pair, a weight of +6 may be assigned acting as a penalty and resulting in a repulsive interaction. Hence, the scoring function not only allows rewarding of desired matches of corresponding features but also enables disfavoring matches of improper features. It is assumed that the distance up to which the optimal score is assigned, r_{opt}, is lower than the total cutoff-

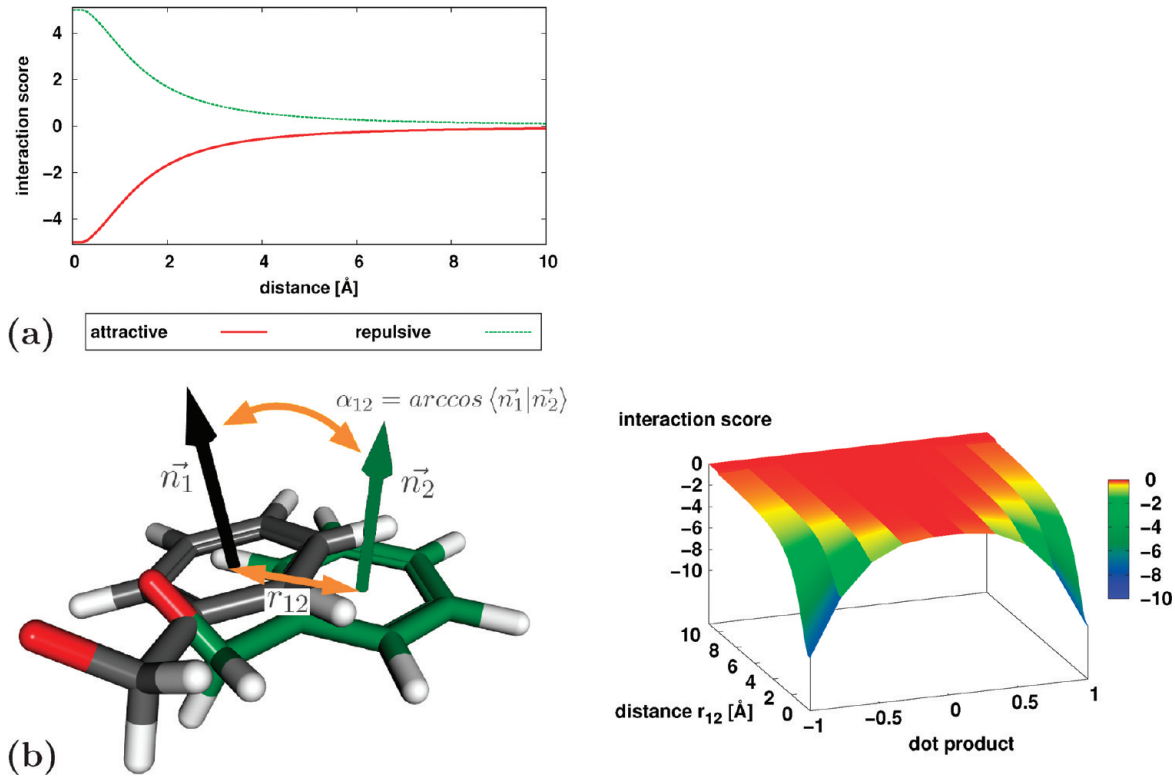

Figure 2. (a) Pairwise similarity function for attractive and repulsive interactions. (b) Vector-similarity function illustrated using a small molecular fragment. The distance r_{12} is defined between the two ring centers and the dot product $\langle \vec{n}_1 | \vec{n}_2 \rangle$ is a measure for the angular deviation α_{12} of the two normal vectors $\vec{n}_1$ and $\vec{n}_2$.

radius used for all pairwise interactions, r_{cut}, which is set to 10 Å. Parameter s defines the sharpness of the potential and is set to 50 in all calculations. Figure 2a shows both potentials given in the example above for $r_{opt} = 0.25$. For all selected classes, all possible pairs of features in different ligands are created and stored in a set called D. This set is subsequently used to calculate the distance-based part of the alignment score (see eq 4). As already mentioned above, the all-to-all comparison allows for the matching of multiple features of one ligand to one or more features of the other one. One example is hydrophobic regions, for which the definition of one single representative pharmacophoric point is difficult. In the approach presented here, a hydrophobic part is represented by one pharmacophoric point per heavy atom. A hydrophobic region in a ligand is therefore modeled by a cloud of hydrophobic points. Superimposing two or more of these clouds using the all-to-all comparison inherently accounts for different sizes of hydrophobic regions in different ligands and thus also uncertainties in the alignment.

Directional Pharmacophoric Features. The second type of pharmacophoric features considered is of directional nature. Here, especially reduced representations of ring systems are used. To define these features, the center of mass of all non-hydrogen atoms constituting a ring is calculated in a first step. Then a set of vectors connecting the center of mass with each heavy atom of the ring is generated. From this set, all pairs of vectors enclosing an angle of up to 120° are used to calculate a corresponding perpendicular vector. Finally, the direction of all these perpendicular vectors is averaged, resulting in a normal vector defining a plane equation. This ring plane is utilized to recognize either planarity or nonplanarity of the ring system, thereby gaining aromaticity information at the same time. If any ring heavy atom has a distance deviation of more than 0.1 Å from the ring plane, the ring is classified as *nonplanar*, otherwise it is classified as *planar*. These two classes are further refined with respect to the number of nonpolar atoms constituting the ring system ($R_{nonpolar}$). This number directly influences the ring score, which compares the topological similarity of two ring systems without taking their relative orientations into account and is calculated by eq 3. The weighting factor w_{ring} defines the maximum contribution to the total scoring function value. This weight is scaled by indicator function I_{planar}, which returns 1 if both ring systems R_1 and R_2 are either planar or nonplanar; otherwise it returns 0.5. The last part of this equation accounts for the atomic constitution and the size of the ring systems.

$$f_{ring}(R_1, R_2) = w_{ring} \cdot I_{planar}(R_1, R_2) \cdot \left(1 - \frac{|R_{nonpolar,1} - R_{nonpolar,2}|}{\max(R_{size,1}, R_{size,2})}\right) \quad (3)$$

If both rings are of the same size (R_{size} in eq 3) and have the same number of nonpolar atoms, $R_{nonpolar}$, the maximum score is rewarded. Otherwise, the score is scaled by a factor accounting for the difference in nonpolar atoms divided by the maximum of both ring sizes. All possible pairs of rings, apart from pairs of rings in the same ligand structure, are finally stored in the vector-based set V. This set is subsequently used to calculate the directional part of the alignment score (see eq 4).

Similarity Score. Given the set of distance-based pharmacophoric features, D, and the set of directional pharmacophoric features, V, a combined similarity scoring function is defined:

$$f_{\mathrm{sim}} = \sum_{d \in D} f_{\mathrm{dist}}(d_{r_{12}}, d_{w_{12}}, d_{r_{12,\mathrm{opt}}}) + \sum_{v \in V} f_{\mathrm{dist}}(v_{r_{12}}, f_{\mathrm{ring}}(v_{R_1}, v_{R_2}), v_{r_{12,\mathrm{opt}}}) \cdot \langle v_{\vec{n}_1} | v_{\vec{n}_2} \rangle^3 \quad (4)$$

In this equation, $d_{r_{12}}$ and $v_{r_{12}}$ represent the actual distances between two pharmacophoric points (1 and 2) or between two ring centers, respectively. Similarly, $d_{r_{12,\mathrm{opt}}}$ and $v_{r_{12,\mathrm{opt}}}$ define the upper distances, up to which the maximum weights $d_{w_{12}}$ and $f_{\mathrm{ring}}(v_{R_1}, v_{R_2})$ contribute to the total score. Both parameters, $d_{r_{12,\mathrm{opt}}}$ and $v_{r_{12,\mathrm{opt}}}$, are set to 0.25 Å here. In the case of directional features v, the distance-based similarity score as returned by f_{dist} is further scaled with the third power of the dot product between the two normal vectors $v_{\vec{n}_1}$ and $v_{\vec{n}_2}$. An illustration of scoring two ring systems as well as the effect of distance and angular deviation on the actual scoring function value can be found in Figure 2b.

Parametrization of the Alignment Scoring Function. Due to the nonphysical nature of the similarity-based scoring function, a parametrization was derived on the basis of alignment experiments using experimentally determined protein−ligand complexes. An exhaustive search in a discrete parameter space was performed to identify reasonable weights for the different pharmacophoric features. The weights for the pharmacophoric feature classes *donor*, *acceptor*, *donor_acceptor*, and *nonpolar_no_ring*, the ring weight w_{ring}, and the torsional potential w_{tors} were optimized with respect to the values given in Table 1. A weight $w_{\mathrm{hb\text{-}ideal}}$ is introduced for pharmacophoric classes accounting for hydrogen bonding. This parameter is used if pharmacophoric features of class *donor*, *acceptor*, or *donor_acceptor* are aligned onto their own class, for example, *donor* onto *donor*. If a *donor* or *acceptor* feature is aligned onto a *donor_acceptor* feature, a second weight w_{hb} is applied. This parameter is set to have half the contribution of $w_{\mathrm{hb\text{-}ideal}}$, favoring the superimposition of identical pharmacophoric features. Finally, parameter w_{nonpolar} is used if atoms part of class *nonpolar_no_ring* are superimposed. The training set consists of four protein targets from different protein families with up to eight ligands, which were extracted from the Protein Data Bank (PDB).[29] The four test sets used are *cyclin-dependent kinase 2* (CDK2), *cyclooxygenase-2* (COX-2), *factor Xa* (fXa), and *peroxisome proliferator-activated receptor γ* (PPARγ). All PDB codes used for the individual targets can be found in Table 2. A reference alignment of the ligands was obtained with Relibase[30] from the structural superimposition of the corresponding protein. The resulting ligand coordinates were used as the reference for the rmsd calculations. For the alignment, all ligand structures have been recreated with the 3D structure generator Corina[31] to obtain an unbiased ligand structure. Care has been taken for correct chirality and protonation states.

Pairwise alignment experiments were carried out for the four test sets using all 108 parameter settings resulting from all possible parameter value combinations as given in Table 1 in a grid search. The crystal structure conformation was used as the fixed template while all other Corina-generated ligand structures were subsequently aligned flexibly onto the template. This was carried out for all ligands of the corresponding data set. The quality of the resulting alignments was assessed by a rmsd (*root-mean-square deviation*) measure similar to the one used in protein−ligand docking calculations. The rmsd was calculated between the heavy atom coordinates of the predicted and the experimentally determined ligand structure. We used a rmsd bound of 2.5 Å for the assessment of a successful prediction to take account for differences in the protein structures and for the resulting shifting of the experimental ligands produced by the overlay of active-site residues of the crystal structures. This success criterion is comparable to rmsd bounds used in cross-docking studies,[32] where ligand conformations are predicted in non-native protein structures. There is some criticism on using rmsd values to judge ligand poses since it was shown in docking studies[33] that even poses with a wrong binding motif can fulfill this criterion. Manual comparison of many generated poses fulfilling the 2.5 Å rmsd criterion indicated that such wrong binding motifs are not observable in the complexes investigated in this publication. We will show some examples later in which, although the criterion of 2.5 Å is not fulfilled, the binding motif is actually identified correctly in the alignment (see Results and Discussion). Additionally, results are much easier to compare between different studies when using rmsd values compared to other proposed criteria.[4,34]

Table 1. Parameter Values Used for the Scoring Function Optimization Process[a]

parameter	values	pharmacophoric pair
$w_{\mathrm{hb\text{-}ideal}}$	{−2, −4, −6}	*donor*−*donor* *acceptor*−*acceptor* *donor_acceptor*−*donor_acceptor*
w_{hb}	$0.5 w_{\mathrm{hb\text{-}ideal}}$	*donor_acceptor*−*donor* *donor_acceptor*−*acceptor*
w_{nonpolar}	{0, −0.1, −0.25, −0.5}	*nonpolar_no_ring*−*nonpolar_no_ring*
w_{ring}	{−5, −10, −15}	
w_{tors}	{1, 2, 3}	

[a] For parameters representing the optimum interaction weight of pharmacophoric classes, *X*−*Y* denotes the interaction of classes *X* and *Y*.

Table 2. PDB Codes for the Targets Used in the Scoring Function Parameterization Process

target	PDB codes
CDK2	1dm2, 1e9h, 1fvt, 1fvv, 1g5s, 1h1q, 1h1s
COX-2	1cx2, 3pgh, 4cox
fXa	1ezq, 1f0r, 1f0s, 1ksn, 1nfu, 1nfw, 1nfx, 1nfy
PPARγ	1fm6, 1i7i, 1k74, 1knu, 1nyx, 1rdt, 2ath, 2gtk

Search Algorithm Parameter Optimization. Alignment sampling efficiency is influenced by parameters like σ, which is scaling the number of iterations of the ACO algorithm, the evaporation factor ρ, the number of nonimproving iterations until an update with solution s^{db} is forced, as well as the simplex tolerance values for the local search ($\mathrm{nms_{tol}}$) and the refinement local search (ref-$\mathrm{nms_{tol}}$) (see ref 26 for details). The search algorithm parameters were optimized for the problem of pairwise alignment based on protein-based reference ligand superimpositions of the four targets shown in Table 2. The goal was to identify parameter settings capable of yielding correct alignments of test sets from different targets. Besides the following changes, the same experimental setup as described for the search parameter optimization of PLANTS[26] was used. Only 10 instead of 25 validation runs have been performed due to the extended number of test set entries used for the training set. It consisted of 29 pairs constituted of selected entries from the training

set used for the scoring function parametrization. For CDK2, one template structure, the ligand of Protein Data Bank (PDB) code 1fvv, was used, while ligands of PDB codes 1fvt, 1fvv (self-alignment), 1h1q, 1h1s, 1e9h, and 1jsv were aligned onto it. In the case of fXa, 16 alignment pairs were considered consisting of all possible combinations of ligands of PDB codes 1nfx, 1ezq, 1nfw, and 1nfy. Finally, for PPARγ, template structure 1rdt was used to align the seven ligands of PDB codes 2gtk, 2ath, 1k74, 1rdt (self-alignment), 1i7i, 1fm6, and 1knu. Three parameter settings corresponding to different average search times were identified to scale between speed and accuracy. The standard search setting *speed 1* was selected to perform approximately 500 000 scoring function evaluations on average. For *speed 2* and *speed 4*, approximately 250 000 and 125 000 scoring function evaluations are carried out, respectively. The optimized search algorithm parameters for these settings can be found in the Supporting Information (Table S1).

Pairwise Alignment Test Set. The comprehensive *FlexS* data set[4] was used to assess the alignment performance of pharmACOphore. This test set consists of 14 different targets, for which 2 to 12 ligands have been superimposed. The superimposed crystal structures were used as reference for the rmsd calculations and also as the template structure. The minimized ligand structures, also available in this test set, were then aligned flexibly onto one of these template structures. Due to the stochastic nature of pharmACOphore, all pairwise alignment experiments were carried out 25 times with standard alignment settings (*speed 1*). A prediction was assessed as correct if the rmsd between the predicted and the experimentally observed ligand conformation was lower than 2.5 Å. The reported success rate for each experiment is the average over the 25 runs. Thus, it evaluates not only the predictive power of the scoring function but also the sampling reliability of the optimization procedure and is thus a performance measure for the overall method. Alignment timings were measured on an Intel Xeon X5365 CPU processor with 3 GHz.

Multiple Flexible Alignment Test Sets. For multiple flexible alignment experiments, the target *trypsin* from the *FlexS* data set was used. The *trypsin* set consists of PDB codes 1pph, 1tnh, 1tni, 1tnj, 1tnk, 1tnl, and 3ptb. Additionally, five complexes of *poly (ADP-ribose) polymerase* (PARP) were selected from PDB (codes 1efy, 1pax, 2pax, 3pax, and 4pax). The reference alignment of the ligands was generated by superimposition of the ligand binding pockets of the protein structures using Relibase.[30] Prior to the alignment with pharmACOphore, unbiased ligand conformations were generated with Corina[31] using the standard settings. Alignment results were assessed by visual inspection as well as rmsd calculations between the predicted and the experimentally observed conformations.

RESULTS AND DISCUSSION

As described above, the pharmACOphore approach is capable of producing pairwise, as well as multiple, flexible ligand alignments. While the pairwise mode is quite fast and thus suited for ligand-based virtual screening, the more time-consuming multiple flexible alignment mode is aimed at the generation of consistent superimpositions, to be subsequently used for example in the field of QSAR applications. In the following sections, results for both alignment modes will be discussed.

Table 3. Average Success Rates, Alignment Times, and Number of Scoring Function Evaluations Obtained for the Pairwise Alignments of the *FlexS* Set[a]

	success rates [%]					
target	nat only	best	worst	avg	time [s]	eval [10^6]
carboxypept. A	100	99	50	69	6.9	0.79
concanavalin	100	100	100	100	1.0	0.37
DHFR	100	64	12	38	7.7	0.86
elastase	98	33	0	14	11.0	1.07
endothiapepsin	77	42	0	14	106.4	3.22
fructose	100	100	100	100	2.5	0.47
glyc. phosphorylase	100	67	0	50	1.8	0.44
HIV-protease	72	12	0	4	71.8	2.79
immunoglobulin	100	50	25	40	1.3	0.31
rhinovirus	100	43	43	43	4.3	0.69
streptavidin	100	100	97	99	1.4	0.39
thermolysin	99	28	0	13	6.5	0.81
thrombin	100	90	64	75	6.9	0.81
trypsin	100	79	0	54	0.8	0.24
average	96	65	35	51	16.5	0.95
average (reduced set)	100	71	41	58	4.3	0.60

[a] For each target, average values over 25 independent experiments are presented. A success was obtained if the top-ranked prediction of the flexible ligand had an rmsd lower than 2.5 Å. Column *nat only* refers to the success rates for the self-alignments only, that is, superimposing the in vacuo minimized ligand onto its native crystallographic template structure. Average success rates for the best, the worst and over all template structures are reported excluding the self-alignments. The last rows report average values over the whole test set, as well as for a reduced set excluding targets *endothiapepsin* and *HIV-protease*. Average alignment times were measured on an Intel Xeon X5365 CPU processor with 3 GHz.

Parameter Optimization. The parameter optimization process revealed two parameter settings (out of the 108 possible combinations of parameters) capable of reproducing 100 of 201 alignment pairs within 2.5 Å, which corresponds to a success rate of 50%. While this pose prediction result is clearly suboptimal, especially when compared to average success rates obtained in the case of protein−ligand docking, it should be noted again that in the ligand-based case no information about the protein environment is available and only one ligand structure was used as the template for the prediction of all other ligand poses belonging to the corresponding protein. From the two best-performing scoring models, the one with the better average pose prediction rmsd of 3.59 Å, that is, the average rmsd over all performed pairwise alignments, was selected as the standard setting for pharmACOphore. The weights used in this model are $w_{\text{hb-ideal}} = -6$, $w_{\text{hb}} = -3$, $w_{\text{nonpolar}} = -0.25$, $w_{\text{ring}} = -10$, and $w_{\text{tors}} = 2$.

All cross-alignment results obtained for this standard setting on the training set can be found in Tables S2 and S3 of the Supporting Information. From these tables two general conclusions can be drawn. First, as expected, all ligands can be aligned onto themselves with rmsd values lower than 1 Å. Note that all rmsd values are calculated between the crystal structure conformation and the predicted conformation derived from a Corina-generated input structure. Thus, the observation of rmsd values around zero is unlikely because of potentially subtle differences in bond lengths and angles

Table 4. Pairwise Alignment Results for the Three Targets *Immunoglobulin*, *Streptavidin*, and *Carboxypeptidase A*[a]

immunoglobulin

	template				
flex.	1dbb	1dbj	1dbk	1dbm	2dbl
1dbb	**100**	0	0	100	100
1dbj	0	**100**	100	0	0
1dbk	0	100	**100**	0	0
1dbm	100	0	0	**100**	100
2dbl	100	0	0	100	**100**
avg. [%]	60	40	40	60	60

streptavidin

	template				
flex.	1srf	1srg	1srh	1sri	1srj
1srf	**100**	100	88	100	100
1srg	100	**100**	100	100	100
1srh	100	100	**100**	100	100
1sri	100	100	100	**100**	100
1srj	100	100	100	100	**100**
avg. [%]	100	100	98	100	100

carboxypeptidase A

	template				
flex.	1cbx	2ctc	3cpa	6cpa	7cpa
1cbx	**100**	100	100	100	100
2ctc	100	**100**	100	100	100
3cpa	100	100	**100**	100	100
6cpa	0	0	0	**100**	96
7cpa	0	0	0	88	**100**
avg. [%]	60	60	60	98	99

color code: < 30 | $[30, 70)$ | ≥ 70

[a] A superimposition was assessed as correct if the rmsd of the top-ranked solution was lower than 2.5 Å. All success rates are averaged over 25 independent experiments. Additionally, for each *template* structure the average success rate is presented.

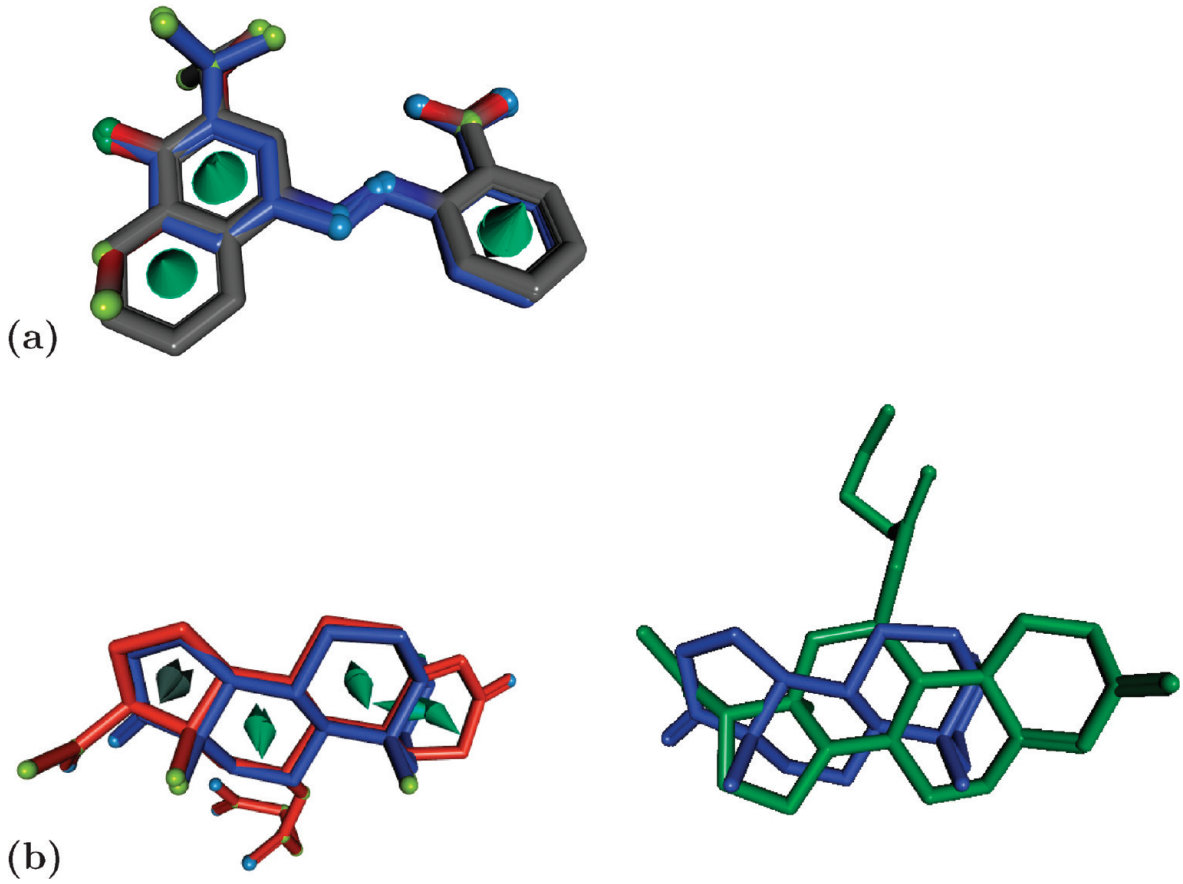

Figure 3. Pairwise alignment results for *immunoglobulin* and *streptavidin*. Cones represent pharmacophoric ring features, yellow spheres represent nonpolar atoms, which are not part of a ring system (class *nonpolar_no_ring*), blue spheres represent acceptors (class *acceptor*), magenta spheres represent donors (class *donor*,) and green spheres represent donor/acceptor atoms like OH-groups (class *donor_acceptor*). (a) *Streptavidin*. Superimposition of all pairwise alignment results generated for template structure 1srf (blue). Note that all ligand structures could be superimposed correctly. (b) *Immunoglobulin*. Incorrect alignment of ligand 1dbm (red) onto template 1dbj (blue). The experimentally observed conformation of 1dbm is shown in green. The failure can be attributed to the alignment of the ring systems which are not perfectly aligned in the experimentally observed structure.

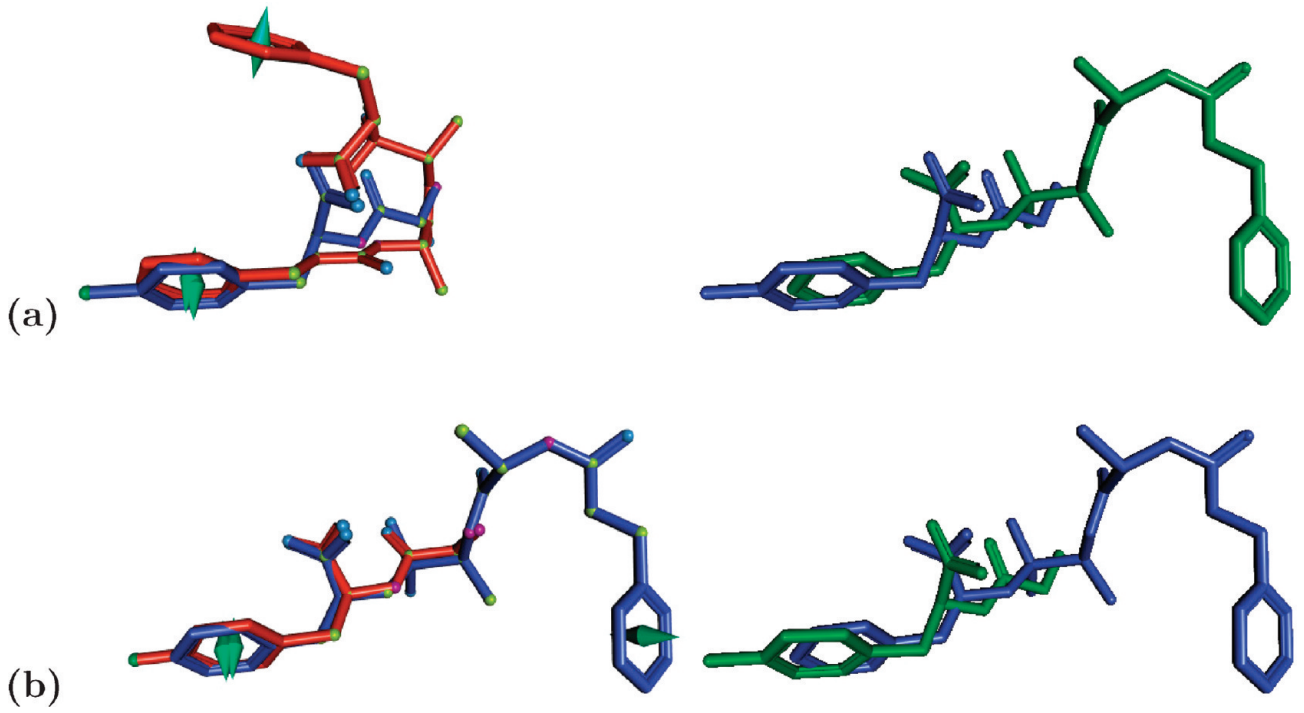

Figure 4. Pairwise alignment results for *carboxypeptidase A*. Cones represent pharmacophoric ring features, yellow spheres represent nonpolar atoms, which are not part of a ring system (class *nonpolar_no_ring*), blue spheres represent acceptors (class *acceptor*), magenta spheres represent donors (class *donor*), and green spheres represent donor/acceptor atoms like OH-groups (class *donor_acceptor*). (a) Incorrect alignment of ligand structure 6cpa (red) onto template structure 3cpa (blue). The experimentally observed conformation of 6cpa is shown in green. (b) Correct alignment of ligand structure 3cpa (red) onto template structure 6cpa (blue). The experimentally observed conformation of 3cpa is shown in green.

in the two conformations. Second, some ligands are better suited as template than others, which can be explained not only by different interaction patterns but also by different sizes of the ligands (see below for a more in-depth discussion).

Pairwise Alignment. We use pairwise alignments as a first step in validating the pharmACOphore parameter setting where one ligand is kept rigid and fixed to keep a defined reference frame and the other ligand is aligned flexibly onto the reference ligand. Such an alignment mode is also often used in ligand-based virtual screening applications. The *FlexS* data set was used because it consists of different protein targets and covers a broad range of chemically diverse ligands. In all cases, the protein-derived superimposition formed the reference state. An overview of the results can be found in Table 3. For each target, the success rates for the self-alignment (aligning the Corina-generated conformation of a ligand onto its experimentally determined reference structure), for the best and the worst template structure as well as the average success rate over all template structures is reported (for the latter three, the self-alignment results are not considered). Additionally the average alignment times and the average number of scoring function evaluations needed are presented. Looking at the self-alignment performance only, an average success rate of 96% is observed. Most of the ligand poses are perfectly reproduced but some targets, like *HIV-protease* or *endothiapepsin*, only reach average success rates around 70−80%, which can be explained by the large size of the ligands and sampling problems arising from this size even if the search time is already increased by an order of magnitude compared to the other targets. As these two cases represent exceptions compared to the rest of the targets, they will be excluded from the following discussion.

Although the success rates are lower, we have excluded the self-alignment experiments from the analysis of the best, worst, and average success rates, since they are not really relevant for alignment studies aiming at identifying new actives (the corresponding results including self-alignment are given in the Supporting Information). On average, an encouraging success rate of 71% can be obtained if the best template structure is used. However, for the individual complexes, a wide range of success rates ranging from 28% (*thermolysin*) to 100% (*concanavalin, fructose, and streptavidin*) can be observed. Additionally, a considerable drop in the performance is visible when looking at the average or worst success rates for each target over the template structures. For example, for *trypsin*, the success rate decreases by almost 80% (79% and 0% for the best and the worst template structure, respectively). The latter result emphasizes that the choice of the template structure is extremely important. A general trend is that larger ligands are better suited as a template than smaller ones. If a large ligand is aligned onto a small ligand, those parts of a large ligand that do not have corresponding features in the small template are orientated arbitrarily and are not forced into the correct orientation. In contrast, if large templates are used, the partial alignment is already sufficient to get the correct pose of the small ligands. However, if substructures of the ligands occupy different cavities of the active site, not all necessary feature information may be available in a single ligand that may be solved by using multiple template structures. Overall, these observations are in line with alignment results published in other studies.[34]

Table 5. Pairwise Alignment Results for *Elastase*[a]

elastase							
	template						
flex.	1ela	1elb	1elc	1eld	1ele	t003	t003
1ela	**100**	0	0	100	100	0	0
1elb	0	**88**	0	0	0	0	0
1elc	0	0	**100**	0	0	0	0
1eld	100	0	0	**100**	100	0	0
1ele	100	0	0	100	**100**	0	0
t003	0	0	0	0	0	**100**	0
t003	0	0	0	0	0	0	**100**
avg. [%]	43	13	14	43	43	14	14

color code: < 30 | $[30, 70)$ | ≥ 70

[a] A superimposition was assessed as correct if the rmsd of the top-ranked solution was lower than 2.5 Å. All success rates are averaged over 25 independent experiments. Additionally, for each *template* structure the average success rate per template structure is presented.

The results of all individual pairwise alignments are presented in the Supporting Information. Here, we will highlight only four examples representing different categories of alignment results: (i) alignments in agreement with experimental superimposition, (ii) alignments with separation into two different superimposition clusters, (iii) failing alignments due to size effects, where the reference ligand is significantly smaller than the ligand to be aligned, and (iv) failing alignments because of unexpected changes in the experimentally observed binding modes. It should be noted that the latter two cases are general limitations of purely ligand-based methods.

Case i: A quite striking performance across all template structures can be observed for target *streptavidin* (see Table 4). All template structures are able to reproduce all ligand structures correctly at excellent average success rates between 88% and 100%. The visualization of the alignment for template structure 1srf is presented in Figure 3a, which shows a perfectly overlaid common scaffold.

Case ii: For *immunoglobulin* (see Table 4), two independent subsets can be identified, for which all entries can be aligned correctly onto each other. The first set is constituted of ligands of PDB codes 1dbb, 1dbm, and 2dbl, while the second set consists of ligands of PDB codes 1dbj and 1dbk. Thus, the best-performing template structures for this target are all ligands part of set 1 capable of reproducing the structures of 3 ligands correctly at success rates of 100%. Figure 3b illustrates the reason of failure for the alignment of structure 1dbm onto 1dbj. The ring systems of the template structure 1dbj (shown in blue) and the experimentally observed structure of 1dbm (shown in green) do not coincide, while pharmACOphore generates a perfect match of the ring systems (shown in red).

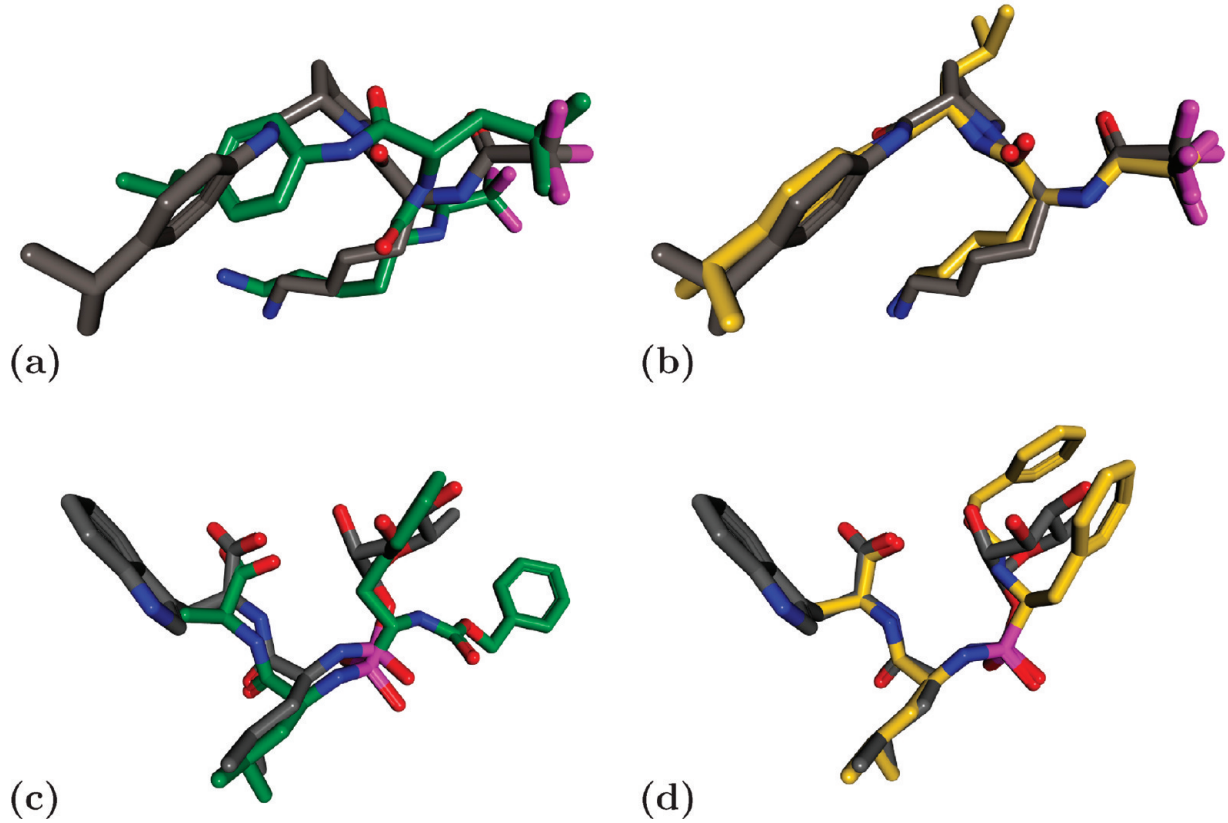

Figure 5. Two examples depicting reasons for the failure of pharmACOphore. (a) Relative orientation of *elastase* ligands in 1ela (gray) and 1elb (green) resulting from the superimposition of the active site residues of the crystal structures. The two ligands, although highly similar, show quite different binding modes as discussed in detail in the original publication of PDB code 1ela.[35] (b) pharmACOphore alignment of ligand of 1elb (yellow) onto the ligand template of 1ela (gray). The binding mode of the ligand of 1elb cannot be reproduced by the ligand-based approach. (c) *Thermolysin* ligands 1tlp (gray) and 4tmn (green) in their crystal structure conformation, as well as (d) the pairwise alignment result of 4tmn (yellow) with 1tlp (gray) as template. According to the success criterion, this superimposition is a failure because of a heavy atom rmsd of 3.73 Å. Nevertheless, the main interactions are predicted correctly, and only the large substituents cannot be placed properly. This can be attributed to missing corresponding features in the template.

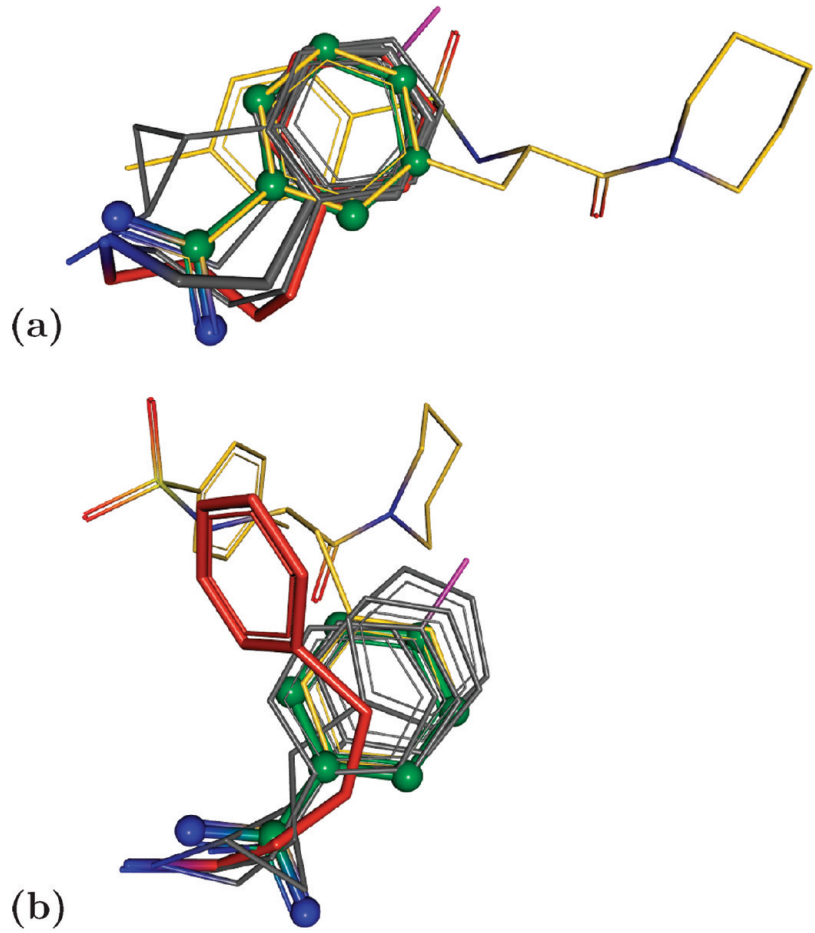

Figure 6. Superimposition of 7 *trypsin* structures. The ligand structures of PDB code 1tni and 1pph are highlighted in red and yellow, respectively. Ligand 3ptb defining the reference frame is shown in a green ball-and-stick representation in both figures. (a) Predicted superimposition. The alignment with the best overall score over the 25 runs is visualized. All ligands except 1tni (red) and 1pph (yellow) are reproduced within a rmsd of 2.5 Å. (b) Crystal structure superimposition.

Case iii: In the case of *carboxypeptidase A* (see Table 4), at least two template molecules (PDB codes 6cpa and 7cpa) are able to reproduce all ligands correctly. Figure 4a and b shows the alignment results obtained with pharmACOphore. While the larger ligand structure of PDB code 6cpa cannot be aligned correctly onto the smaller template ligand structure of PDB code 3cpa (Figure 4a), a correct alignment of 3cpa onto 6cpa (Figure 4b) can be obtained. The dependence of the alignment performance on the template size is a commonly observed problem. If a large molecule is aligned onto a smaller one, all parts with no corresponding features in the small molecule will be placed arbitrarily, very likely resulting in a conformation exhibiting a large rmsd compared to the crystal structure.

Case iv: For *elastase* and *thermolysin*, pharmACOphore showed only a poor prediction performance. In most cases, each ligand could only be aligned correctly onto itself (see Table 5 for elastase results; *thermolysin* results can be found in the Supporting Information). Some reasons for failure are shown in Figure 5. Figure 5a shows the superimposition of two structurally similar *elastase* ligands in their quite dissimilar experimentally observed conformations. This remarkable observation is discussed in great detail in the original publication.[35] In this case, ligand-based approaches in general are expected to predict the wrong conformation for any of both ligands taking the other one as the template. A second reason for failure is presented in Figure 5c and d (target *thermolysin*). Here, only parts of the ligands are aligned correctly due to missing corresponding interactions resulting in an overall large rmsd value.

Multiple Flexible Alignment. *Trypsin.* The pharmACOphore approach has been tested additionally for the problem of multiple flexible alignment using all seven *trypsin* ligand structures of the *FlexS* data set (PDB codes 1pph, 1tnh, 1tni, 1tnj, 1tnk, 1tnl, and 3ptb). As the multiple flexible alignment mode is usually not time-critical, parameter σ scaling the number of iterations carried out by the ACO algorithm was significantly increased compared to the default setting. It was set to $\sigma = 7$ corresponding to the number of ligands superimposed in parallel. To allow for the assessment of the *pose prediction* accuracy, the protein structure reference frame was kept by fixing the translational and rotational degrees of freedom for one of the smallest ligands (benzamidine, PDB code 3ptb) to its crystallographic conformation. Only one torsional degree of freedom in this ligand (rotation of amidino group) was freely optimized during the search. All degrees of freedom of the other ligands were optimized so that a fully flexible alignment of the ligands in the reference frame of the protein structure was obtained. The experiment was repeated 25 times. The alignment time for this data set was around 11 min on an Intel Xeon E5420 CPU processor with 2.5 GHz.

Five of seven ligand structures could be reproduced within a rmsd of 2.5 Å in all 25 runs. Only the ligands of PDB

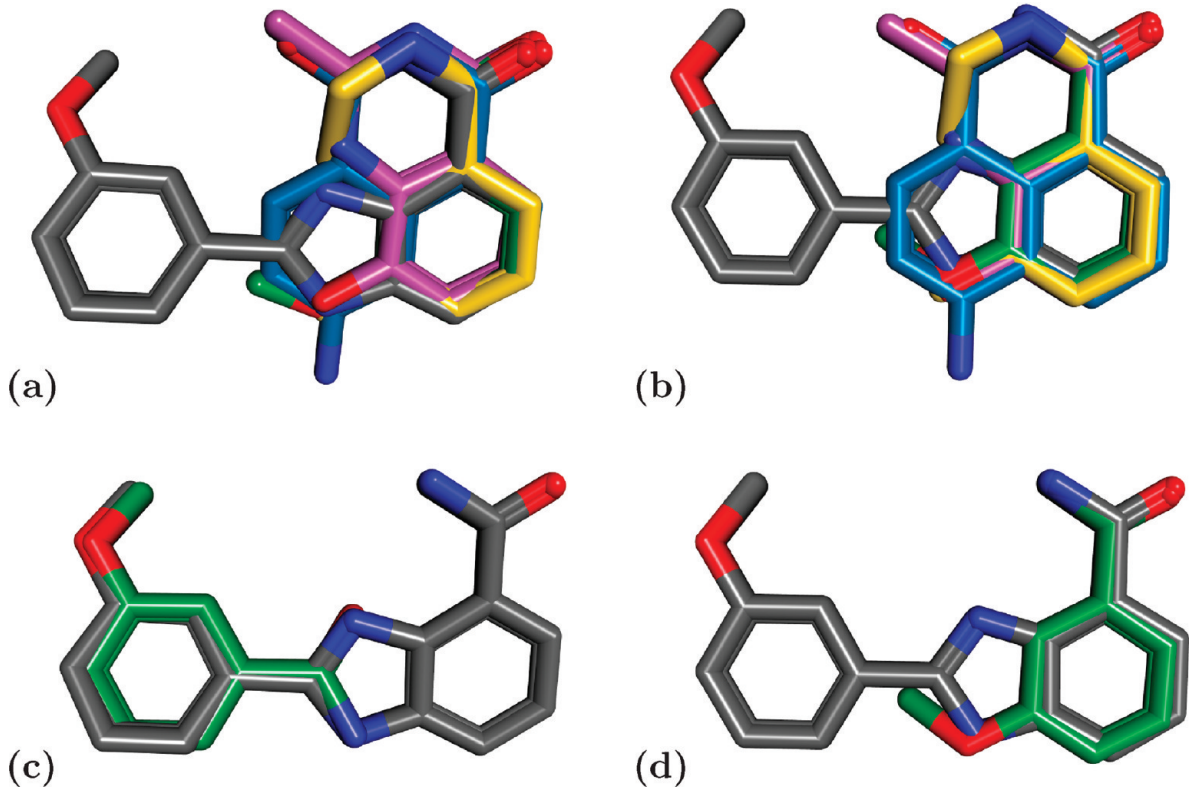

Figure 7. Alignment of 5 PARP ligands (1efy, gray; 1pax, yellow; 2pax, light blue; 3pax, green; 4pax, magenta). (a) Superimposition of the crystallographic complex structures. (b) Multiple flexible alignment produced by pharmACOphore. (c) Incorrect pairwise alignment of 3pax onto template structure 1efy. (d) Correct superimposition of 3pax and 1efy taken from the multiple flexible alignment.

code 1tni and 1pph deviate significantly from the experimental alignment and do not pass the rmsd criterion. The predicted and experimentally observed alignment can be found in Figure 6a and b, respectively. For most ligands, the predicted overlay of the ring systems and the charged donor groups is consistent with the experiment. From the crystal structure superimposition, it is obvious that 1tni has a slightly different binding mode compared to the other ligands. Therefore, the failure to reproduce the correct pose has to be attributed to the ligand-based technique and not to the scoring function or the optimization procedure. The large ligand of PDB code 1pph exhibits the problem already described in the pairwise alignment section. Since there are no corresponding features for the large flexible substituents, they are placed arbitrarily resulting in a large rmsd value compared to the experimental structure. This becomes evident if one compares the rmsd values in the individual alignment runs. 1pph shows rmsd values between 4 and 8 Å. But at the same time, all these structures have almost the same score (standard deviation of only 1.8 scoring function units). Thus, these large differences in the position of the substituent are not a sampling problem but result from the fact, that the location of this group has no influence on the score. Nevertheless, in all runs the main binding motif, the benzamidine substructure, is aligned correctly. Concluding this part, multiple alignments are useful to identify common binding motifs but they do not necessarily represent the protein-bound conformations of the ligands especially for parts, which are only present in a small number of ligands.

Poly (ADP-Ribose) Polymerase. Our second test case is the multiple flexible alignment of five *poly (ADP-ribose) polymerase* (PARP) ligands from the PDB. PARP is involved in the repair of DNA strand breaks and, in this way, in the resistance of cancer cells to certain DNA-damaging agents. For this target, a scaling factor of $\sigma = 5$ proportional to the number of ligands to align was used. All five ligands show almost exactly the same conformation and relative orientation as in the superimposition of the crystal structures (see Figures 7a and 7b). This can also be quantified by the low rmsd values of all predicted ligand conformations compared to their experimentally observed conformations, which are all below 1 Å (1efy, 0.44 Å; 1pax, 0.16 Å; 2pax, 0.19 Å; 3pax, 0.27 Å; 4pax, 0.34 Å; for the PARP experiments no fixed template was used and, thus, because of the missing reference frame in the multiple flexible alignment, rmsds were calculated after rigid superposition of the predicted and experimental ligand alignments taking the atoms of all ligands into account). In contrast, if the ligands are aligned pairwise on a reference structure, alignments are found, that differ from the protein-based alignment. For example, if ligand 3-methoxybenzamide (ligand of PDB code 3pax) is matched onto ligand 2-(3′-methoxyphenyl) benzimidazole-4-carboxamide (ligand of PDB code 1efy), the methoxyphenyl substructures of the two molecules are overlaid, resulting in a plausible but nevertheless wrong alignment (rmsd = 5.75 Å, see Figure 7c).

The better performance of the multiple flexible alignment can be explained by the fact that due to the inclusion of other ligand structures the second phenyl system (benzamide) in the ligand of PDB code 1efy is correctly identified as an important pharmacophoric feature. The three remaining ligands (PDB codes 1pax, 2pax, and 4pax) get more favorable scores according to the alignment scoring function if their aromatic system is paired with the benzoidic ring of

benzimidazol in the ligand of PDB code 1efy. Thus, because of the additional interactions with these three ligands when performing multiple alignment, the correct binding mode for the benzamide substructure of PDB code 3pax can be identified (see Figure 7d). This multiple alignment generated by pharmACOphore could be used to create a pharmacophore model using other programs to allow for ligand-based virtual screening to identify new active ligands.

CONCLUSIONS

In this paper, a new approach for the flexible alignment of two or more small molecules is introduced. The hybrid ant colony optimization algorithm previously applied to the protein−ligand docking problem[27] is combined with a new similarity-based scoring function tuned for ligand-based pose prediction. The pairwise alignment results obtained for the comprehensive *FlexS* data set[2] show that alignments can be found that are in agreement with the protein-structure-based alignment. Nevertheless, since details of the protein structure are not known to the approach, unexpected dissimilar binding modes for similar ligands can not be predicted. Failures can be attributed not only to the scoring function but also to the ligand-based approach *per se* and the rmsd-based criterion used for the assessment of a correct prediction. For some of the results assessed as incorrect according to the rmsd-based criterion actually the correct pharmacophoric pattern could be identified. In the pharmacologically relevant study of five *poly (ADP-ribose) polymerase* (PARP) ligands, a good agreement between the superimposition of the crystal structures and the one generated by pharmACOphore could be obtained.

These results show that the problems mentioned in the introduction have been, at least in parts, addressed by our new method. Our scoring function uses distance-dependent potentials for matching the pharmacophoric features in combination with an intraligand potential. It was parametrized to reproduce experimentally observed complex geometries, which requires a balance between the influence of ligand conformational energies and the similarity-based alignment score. As shown, the derived values are probably not optimal for all applications. Therefore, user-defined pharmacophoric features can be defined and the weights of the scoring function can be adapted to obtain target-specific scoring functions. Additionally, application of penalty terms for pharmacophoric mismatches like placement of a *donor* onto an *acceptor* or placement of a polar feature into a hydrophobic region could be an option to enhance alignment results. To further improve the approach, methods will be developed to generate pharmacophore models out of the produced alignments, which could be used in ligand-based virtual screening campaigns.

ACKNOWLEDGMENT

The CUSS cluster and the bwGRiD in Ulm and the HPC cluster of the University of Konstanz is acknowledged for allocating computational resources. This work was supported by a scholarship of the Landesgraduiertenförderung Baden-Württemberg awarded to Oliver Korb. Thomas Stützle acknowledges support of the Belgian F.R.S.-FNRS, of which he is a research associate.

Supporting Information Available: Optimized search algorithm parameters for the three different speed settings, the average success rates for the pairwise alignments, and the individual success rates for each pair of the *FlexS* set. This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Poptodorov, K.; Luu, T.; Hoffmann, R. D. Pharmacophore Model Generation Software Tools. In *Pharmacophores and Pharmacophore Searches*; Langer, T., Hoffmann, R., Eds.; Wiley-VCH: Weinheim, Germany, 2006.
(2) Lemmen, C.; Lengauer, T. Computational Methods for the Structural Alignment of Molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 215–232.
(3) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
(4) Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: A Method for Fast Flexible Ligand Superposition. *J. Med. Chem.* **1998**, *41*, 4502–4520.
(5) Jones, G.; Willett, P.; Glen, R. C. A Genetic Algorithm for Flexible Molecular Overlay and Pharmacophore Elucidation. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 532–549.
(6) Dorigo, M.; Stützle, T. *Ant Colony Optimization*; MIT Press: Cambridge, MA, 2004.
(7) Handschuh, S.; Wagener, M.; Gasteiger, J. Superposition of Three-Dimensional Chemical Structures Allowing for Conformational Flexibility by a Hybrid Method. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 220–232.
(8) Cho, S. J.; Sun, Y. FLAME: A Program to Flexibly Align Molecules. *J. Chem. Inf. Model.* **2006**, *46*, 298–306.
(9) Holland, J. *Adaptation in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor, 1975.
(10) Jones, G.; Willett, P.; Glen, R. C. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *J. Mol. Biol.* **1995**, *245*, 43–53.
(11) Todorov, N. P.; Alberts, I. L.; de Esch, I. J. P.; Dean, P. M. QUASI: A Novel Method for Simultaneous Superposition of Multiple Flexible Ligands and Virtual Screening Using Partial Similarity. *J. Chem. Inf. Model.* **2007**, *47*, 1007–1020.
(12) *ROCS*; OpenEye Scientific Software: Santa Fe, NM (accessed 2010).
(13) Grant, J. A.; Gallardo, M. A.; Pickup, B. A Fast Method of Molecular Shape Comparison: A Simple Application of a Gaussian Description of Molecular Shape. *J. Comput. Chem.* **1998**, *17*, 1653–1666.
(14) Rush III, T. S.; Grant, J. A.; Mosyak, L.; Nicholls, A. A Shape-Based 3-D Scaffold Hopping Method and Its Application to a Bacterial Protein−Protein Interaction. *J. Med. Chem.* **2005**, *48*, 1489–1495.
(15) Shin, W.; Hyun, S. A.; Chae, C. H.; Chon, J. K. Flexible Alignment of Small Molecules Using the Penalty Method. *J. Chem. Inf. Model.* **2009**, *49*, 1879–1888.
(16) Kearsley, S. K.; Smith, G. M. An Alternative Method for the Alignment of Molecular Structures: Maximizing Electrostatic and Steric Overlap. *Tetrahedron Comput. Methodol.* **1990**, *3*, 615–633.
(17) Kinnings, S. L.; Jackson, R. M. LigMatch: A Multiple Structure-Based Ligand Matching Method for 3D Virtual Screening. *J. Chem. Inf. Model.* **2009**, *49*, 2056–2066.
(18) Jones, G.; Gao, Y.; Sage, C. R. Elucidating Molecular Overlays from Pairwise Alignments Using a Genetic Algorithm. *J. Chem. Inf. Model.* **2009**, *49*, 1847–1855.
(19) Dror, O.; Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. Novel Approach for Efficient Pharmacophore-Based Virtual Screening: Method and Applications. *J. Chem. Inf. Model.* **2009**, *49*, 2333–2343.
(20) Cottrell, S. J.; Gillet, V. J.; Taylor, R.; Wilton, D. J. Generation of Multiple Pharmacophore Hypotheses Using Multiobjective Optimisation Techniques. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 665–682.
(21) Cottrell, S. J.; Gillet, V. J.; Taylor, R. Incorporating Partial Matches within Multiobjective Pharmacophore Identification. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 735–749.
(22) Gardiner, E. J.; Cosgrove, D. A.; Taylor, R.; Gillet, V. J. Multiobjective Optimization of Pharmacophore Hypotheses: Bias Toward Low-Energy Conformations. *J. Chem. Inf. Model.* **2009**, *49*, 2761–2773.
(23) Leach, A. R.; Gillet, V. J.; Lewis, R. A.; Taylor, R. Three-Dimensional Pharmacophore Methods in Drug Discovery. *J. Med. Chem.* **2010**, *53*, 539–558.
(24) Stützle, T.; Hoos, H. *MAX - MIN* Ant System. *Future Gener. Comput. Syst.* **2000**, *16*, 889–914.
(25) Nelder, J. A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308–313.

(26) Korb, O.; Stützle, T.; Exner, T. E. Empirical Scoring Functions for Advanced Protein−Ligand Docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49*, 84–96.
(27) Korb, O.; Stützle, T.; Exner, T. E. An Ant Colony Optimization Approach to Flexible Protein−Ligand Docking. *Swarm Intell.* **2007**, *1*, 115–134.
(28) Korb, O.; Stützle, T.; Exner, T. E., PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. In *Ant Colony Optimization and Swarm Intelligence*; 5th International Workshop, ANTS 2006 Lecture Notes in Computer Science, Vol. 4150; Dorigo, M., Gambardella, L. M., Birattari, M., Martinoli, A., Poli, R., Stützle, T., Eds.; Springer: Berlin, 2006.
(29) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(30) Hendlich, M.; Bergner, A.; Günther, J.; Klebe, G. Relibase: Design and Development of a Database for Comprehensive Analysis of Protein−Ligand Interactions. *J. Mol. Biol.* **2003**, *326*, 607–620.
(31) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic Generation of 3D Atomic Coordinates for Organic Molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537–547.
(32) Cavasotto, C. N.; Abagyan, R. A. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J. Mol. Biol.* **2004**, *337*, 209–225.
(33) Yusuf, D.; Davis, A. M.; Kleywegt, G. J.; Schmitt, S. An Alternative Method for the Evaluation of Docking Performance: RSR vs RMSD. *J. Chem. Inf. Model.* **2008**, *48*, 1411–1422.
(34) Chen, Q.; Higgs, R. E.; Vieth, M. Geometric Accuracy of Three-Dimensional Molecular Overlays. *J. Chem. Inf. Model.* **2006**, *46*, 1996–2002.
(35) Mattos, C.; Rasmussen, B.; Ding, X.; Petsko, G.; Ringe, D. Analogous Inhibitors of Elastase Do Not Always Bind Analogously. *Nat. Struct. Biol.* **1994**, *1*, 55–58.

CI1000218