PROTEINS: Structure, Function, and Genetics 40:623–636 (2000)

Similarity-Driven Flexible Ligand Docking

Xavier Fradera, Ronald M.A. Knegtel, and Jordi Mestres*
Department of Molecular Design & Informatics, N.V. Organon, Oss, The Netherlands

ABSTRACT **A similarity-driven approach to flexible ligand docking is presented. Given a reference ligand or a pharmacophore positioned in the protein active site, the method allows inclusion of a similarity term during docking. Two different algorithms have been implemented, namely, a similarity-penalized docking (SP-DOCK) and a similarity-guided docking (SG-DOCK). The basic idea is to maximally exploit the structural information about the ligand binding mode present in cases where ligand-bound protein structures are available, information that is usually ignored in standard docking procedures. SP-DOCK and SG-DOCK have been derived as modified versions of the program DOCK 4.0, where the similarity program MIMIC acts as a module for the calculation of similarity indices that correct docking energy scores at certain steps of the calculation. SP-DOCK applies similarity corrections to the set of ligand orientations at the end of the ligand incremental construction process, penalizing the docking energy and, thus, having only an effect on the relative ordering of the final solutions. SG-DOCK applies similarity corrections throughout the entire ligand incremental construction process, thus affecting not only the relative ordering of solutions but also actively guiding the ligand docking. The performance of SP-DOCK and SG-DOCK for binding mode assessment and molecular database screening is discussed. When applied to a set of 32 thrombin ligands for which crystal structures are available, SG-DOCK improves the average RMSD by ca. 1 Å when compared with DOCK. When those 32 thrombin ligands are included into a set of 1,000 diverse molecules from the ACD, DIV, and WDI databases, SP-DOCK significantly improves the retrieval of thrombin ligands within the first 10% of each of the three databases with respect to DOCK, with minimal additional computational cost. In all cases, comparison of SP-DOCK and SG-DOCK results with those obtained by DOCK and MIMIC is performed. Proteins 2000;40:623–636.**

Key words: ligand-ligand similarity; protein-ligand docking; binding mode assessment; molecular database screening

INTRODUCTION

The prediction of binding modes and affinities of small ligands to biologically relevant target proteins are important aspects for discovering and optimizing new lead compounds in pharmaceutical industry. With the availability of experimentally determined protein structures, docking methods soon emerged as attractive computational techniques to tackle those aspects. Based on the lock-and-key principle, protein-ligand docking methods have the potential of identifying compounds that complement the steric and electrostatic characteristics of the target receptor.[1–5] However, docking ligands into a receptor structure is not a simple task and the final outcome depends on many factors. For instance, early methods were restricted to docking rigid ligands into rigid receptors.[6–8] This simplification reduces the complexity of the docking problem to mainly two factors, namely, sampling the orientation of the ligand in the receptor cavity and scoring the ligand-receptor interaction for each orientation. However, it also limits the scope of docking methods to fairly rigid ligands and receptors. The increase in computer power has permitted recent docking methods to account for ligand flexibility[9–16] and receptor flexibility.[17–24] Although proper coverage of the receptor and/or ligand conformational space adds yet another level of complexity to sampling and scoring,[25,26] the recent development of more efficient approaches allows for flexible ligand docking within a reasonable amount of time. This finding extends the applicability of docking methods to basically any problem for which a protein structure is available.

At the initial stages of a ligand design project, protein structures may not always be available. In these cases, protein-ligand docking methods cannot be applied. What is usually known at these early stages of the project is one or more ligands with confirmed activity for the target receptor. Based on the similarity-property principle, ligand-ligand similarity methods become then the 3-dimensional structure-based alternative to ligand design.[27,28] Similarity methods have the ability to score target ligands on the basis of their relative superposition with respect to a reference ligand.[29] As happened before with docking methods, early developed similarity methods assumed that the molecules to be superimposed are rigid.[30–32] Because most of the molecules of pharmaceutical interest have some degree of flexibility, treating molecules as rigid substan-

X. Fradera's present address is Institut de Química Computacional, Universitat de Girona, 17071 Girona, Catalonia (Spain).

R.M.A. Knegtel's present address is Vertex Pharmaceuticals (Europe) Ltd., 88 Milton Park, Milton, Abingdon, Oxon OX14 4RY, United Kingdom.

*Correspondence to: Jordi Mestres, Department of Molecular Design & Informatics, N.V. Organon, Molenstraat 110, P.O. Box 20, 5340 BH, Oss, The Netherlands. Email: j.mestres@organon.oss.akzonobel.nl

Received 21 October 1999; Accepted 21 April 2000

tially limits the scope of similarity methods. This finding has promoted the development of methods that include conformational flexibility, at least to some extent.[33–38] The recent development of similarity methods allowing for fast flexible ligand superposition[39–41] has raised their applicability to the level where they can be considered complementary, or even competitive, to docking methods for both binding mode assessment and molecular database screening.[41]

However, docking and similarity methods focus only on one part of the structural information present in experimentally determined protein-ligand complexes. Docking methods use only the structural information from the protein perspective, with no consideration of how the co-crystallized ligand is actually binding to the receptor, whereas similarity methods use only the structural information from the ligand perspective, without taking into account any information from the receptor structure. On the other hand, the number of ligand-bound protein structures available in the Protein Data Bank (PDB)[42] has increased dramatically since the emergence of the first docking methods. Therefore, it is probably appropriate to devise docking methods that have the possibility of maximally exploiting all the structural information present in ligand-bound protein structures, both from the protein and ligand perspectives. Enforcing some degree of similarity with observed binding modes could help overcome some of the current shortcomings of conformational sampling and scoring in protein-ligand docking.

The objective of this work is to present the implementation of similarity-driven strategies into a flexible ligand docking method. To illustrate the strategy proposed, DOCK 4.0[12] and MIMIC[38] were selected as the protein-ligand docking and ligand-ligand superposition programs, respectively. The choice for DOCK and MIMIC was based on availability of the source code and familiarity with the programs. Implementation of the similarity-driven strategies proposed here in other flexible ligand docking codes should be straightforward. In particular, two similarity-driven approaches to flexible ligand docking have been evaluated, namely, similarity-penalized docking (SP-DOCK) and similarity-guided docking (SG-DOCK). The two approaches differ in the way they use the MIMIC similarity score to modify the DOCK energy score: SP-DOCK includes a similarity correction only at the end of the ligand incremental-construction process, whereas SG-DOCK considers similarity corrections throughout the entire ligand incremental-construction process. Details of the implementation and applications to binding mode assessment, for a set of 32 thrombin inhibitors for which crystal structures are available, and molecular database screening, for three databases containing a 1,000 diverse molecules of different degree of drug-likeness, are given in the next sections.

MATERIALS AND METHODS

Docking

All docking calculations were done with the program DOCK 4.0.[12] DOCK 4.0 allows for performing rigid as well as flexible docking based on an incremental construction algorithm. Such an algorithm considers each molecule as formed by a collection of rigid segments connected by rotatable bonds. The program automatically selects those segments best suited as initial base fragments (so-called anchors). Then, for each anchor, the best docking orientation into the receptor is searched and, starting from the best-scoring anchor placements, the rest of the ligand is constructed. A docking score is calculated each time a new layer (one or more segments) is added to the growing ligand, keeping always several different trees to allow for a reasonable covering of the number of tried fragment orientations (so-called seeds). For flexible docking calculations, besides the intermolecular scoring, an approximate intramolecular scoring is also taken into account. To speed up computations, grid files are used to map the electrostatic and Van der Waals potentials, as well as the regions where ligands cannot be placed because of steric clashes (so-called bumps) with the receptor. Finally, to find adequate initial placements for the entire ligand or its fragments, a simplistic receptor model is defined as a set of spheres generated to be complementary to the protein cavity surface.

In this work, the AMBER inter- and intramolecular force field energy was used. DOCK 3.5 AMBER energy[43] scoring grids were generated with CHEMGRID,[44] by using a 0.3 Å grid spacing and a 4r distance-dependent dielectric. No interpolation was used to calculate force field scores. The number of accepted bumps between receptor and ligand was set to three. A minimal anchor size was set to four atoms. A maximum of 100 cycles and a convergence threshold of 0.1 were set for energy score minimization. Sphere centers, to be used for matching with ligand atoms during docking, were generated in the ligand binding site with SPHGEN.[6] Initial sphere sets were manually reduced by using Sybyl 6.5 (Tripos, Inc., St. Louis, MO) to yield sets of 20–30 well-dispersed spheres covering all the receptor active site. The scope of the sampling is related to the number of starting orientations that can be tried for each base fragment (anchor) and the number of fragment orientations (seeds) that are kept at each stage of the ligand incremental-construction process. The number of starting orientations and the number of seeds were set to 500 and 25, respectively, for binding mode assessment applications, and to 50 and 5, respectively, for molecular database screening applications. In the reminder of the study, these options will be referred to as "high-sampling" and "low-sampling" options, respectively.

Similarity

All similarity calculations were done with the program MIMIC.[38] Two types of Gaussian-based molecular fields were used to evaluate molecular similarities. An atom-centered steric-volume field (MSV)[45] and an united-atom point-charge electrostatic potential (MEP)[32] are used to represent the steric and electrostatic features of a molecule, respectively. Although these are rather crude approximations, they are generally adequate to reproduce qualita-

tively most of the important steric and electrostatic features of a molecule.[46–48]

A field-based similarity measure (Z_{AB}) between two molecules is computed as

$$Z_{AB} = \int F_A(\mathbf{r})F_B(\mathbf{r})d\mathbf{r} \quad (1)$$

where $F_A(\mathbf{r})$ and $F_B(\mathbf{r})$ are the molecular fields associated to the reference molecule, A, and the target molecule, B, respectively. Then, to compare similarity values from different pairs of molecules, similarity measures are normalized by using a cosine-like similarity index, S_{AB}, defined as

$$S_{AB}(\mathbf{t}, \mathbf{\Theta}, \boldsymbol{\tau}_B) = \frac{Z_{AB}(\mathbf{t}, \mathbf{\Theta}, \boldsymbol{\tau}_B)}{(Z_{AA}Z_{BB}(\boldsymbol{\tau}_B))^{1/2}}. \quad (2)$$

S_{AB} depends on the mutual superposition of molecules, as well as on their conformations. Therefore, assuming a rigid reference molecule A, optimization of S_{AB} depends on three translational ($\mathbf{t}$) and three rotational (Θ) degrees of freedom, for rigid superposition of B on A,[46,47] as well as on several conformational degrees of freedom for the target molecule (τ_B, set of rotatable bonds), for flexible superposition of B on A.[48] Depending on the molecular field used to calculate the similarity, S_{AB} can take values between 0 and 1, for positive definite molecular fields, such as MSV, and between -1 and +1, for nonpositive definite fields, such as MEP. Following previous validation studies,[46–48] throughout the work a similarity index defined by a weighted combination of MSV and MEP similarity indices in a 2:1 ratio ($S_{AB} = 0.67 \cdot S_{AB}^{MSV} + 0.33 \cdot S_{AB}^{MEP}$) was used.

A MSV asymmetric similarity index, defined as $S_{AB}^{(A)} = Z_{AB}/Z_{AA}$, was used to evaluate the similarity between a reference pharmacophore structure and the target molecule.[49] $S_{AB}^{(A)}$ is the asymmetric similarity index between two molecules with respect to the reference molecule, A, and gives a measure of how well the target molecule is overlapping the reference molecule. Note that $S_{AB} = (S_{AB}^{(A)} \cdot S_{AB}^{(B)})^{1/2}$. The use of asymmetric similarity indices has been found more adequate for identifying molecules sharing a similar substructure characteristics with respect to the reference molecule.

Finally, to find the optimal superposition between the reference and the target molecules, the orientational space is explored by using a spherical systematic search.[38] Basically, the reference molecule is kept fixed and the adapting molecule is systematically placed in several orientations around the reference molecule, from which optimization of the molecular similarity as defined by eq. (2) will proceed following normal gradient-seeking techniques. The scope of the sampling is related to the number of starting orientations generated, which depends on the rotational step used to define the spherical search. Also, the gradient convergence criteria can be modified to achieve premature convergence of the similarity in cases where extreme accuracy is not needed. Therefore, within MIMIC no implicit conformational sampling is performed. Instead, as similarity is optimized from each starting orientation in all translational, rotational and conformational degrees of freedom, each starting orientation leads to the conformation of the target molecule that optimally mimics the conformation presented by the reference molecule.[48] The number of starting orientations and gradient convergence criteria were set to 104 (45-degree search) and 0.001, respectively, for binding mode assessment applications, and to 12 (90-degree search) and 0.01, respectively, for molecular database screening applications. As for docking, in the remainder of the article, these options will be referred as "high-sampling" and "low-sampling" options, respectively.

Similarity-Driven Docking

The similarity-driven docking approach proposed in this work is a docking procedure in which, at certain steps of the calculation, the similarity of a target ligand to a reference ligand or pharmacophore structure is used as a weighting factor correcting the docking energy score. If a ligand-bound protein structure is available, the same co-crystallized ligand can be used as reference ligand. Alternatively, a pharmacophore structure can be placed in the active site of the protein and used as the reference structure.

Given a receptor structure, R, a reference ligand structure, A, and a target ligand structure, B, one can obtain a docking score for the interaction between the receptor and the target ligand, D_{RB}, and a similarity score between the reference and target ligands, S_{AB}. Then, a similarity-weighted docking score, T_{RB}, can be defined as

$$T_{RB} = D_{RB} \cdot S_{AB}. \quad (3)$$

Because both D_{RB} and S_{AB} can have positive and negative values, only those docking solutions having negative D_{RB} values (favorable receptor-ligand interaction) and nonnegative S_{AB} values (some degree of similarity between reference and target ligands) are considered. By using eq. (3), T_{RB} will be equal to D_{RB} only if $S_{AB} = 1$, which is only possible if the reference and target ligands are perfectly superimposed. In cases with $0 \leq S_{AB} < 1$, D_{RB} will be penalized proportionally to the degree of similarity between the reference and target ligands. Alternatively, cases with $S_{AB} < 0$ correspond to orientations where electrostatic complementarity between the ligands (overlap between regions of positive and negative electrostatic potential) prevails over steric similarity between them and this will allow for eliminating noncomplementary orientations at early stages of the docking procedure. Therefore, the use of a similarity-weighted docking score as defined in eq. (3) is expected to promote target ligand orientations after the binding mode observed for the reference ligand and penalize those diverging from the observed binding mode. The final best docking solution according to the T_{RB} score should be able to produce docking orientations of the ligand that provide a balance between maintaining favorable interactions with the receptor without deviating too much from the observed binding mode of the reference ligand or pharmacophore structure used.

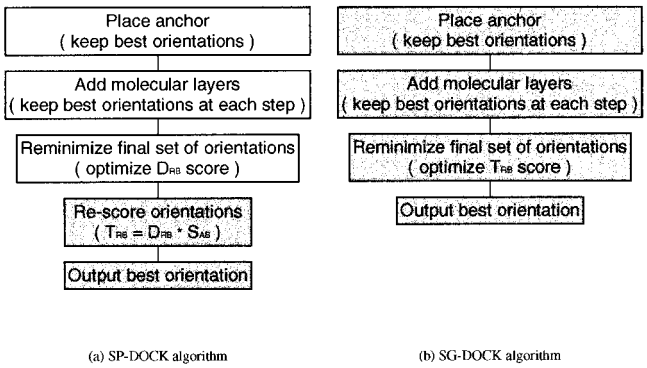

Fig. 1. Schematic representation of the two similarity-driven docking approaches implemented. Boxes in white and grey background indicate the use of the original docking score (D_{RB}) and the similarity-corrected docking score (T_{RB}), respectively.

Figure 1 delineates the two different computational schemes of similarity-driven docking that have been implemented, namely, a similarity-penalized docking (SP-DOCK) and a similarity-guided docking (SG-DOCK). Details of the two implementations are given in the next sections.

Similarity-Penalized Docking (SP-DOCK)

Within this approach, a standard docking calculation is performed and the docking scores are corrected according to the similarity to a reference structure only at the end of the docking procedure, as schematically illustrated in Figure 1a. This means that, during flexible ligand docking, different starting anchors and different starting orientations are tried, and after ordering, clustering and minimization, several top-scoring (according to D_{RB}) orientations are kept. Then, S_{AB} values for all the orientations are calculated, the D_{RB} scores corrected following eq. (3), and the orientations reordered according to the new T_{RB} scores. Note that, within this scheme, the T_{RB} similarity-corrected scores are never used for deriving ligand orientations and conformations or minimizing them. Thus, the role of similarity is not producing ligand orientations different from the ones obtained within a pure docking approach but to modify the scores of the orientations found. With respect to the original docking procedure, this method can occasionally lead to a reordering of the top-scoring orientations of a ligand, producing a different final best solution. Even when the best orientation is the same as found by the original docking approach, the final score will be different. Therefore, the ordering of a set of ligands docked to a certain receptor according to their similarity-penalized docking scores can be different than the ordering obtained from the original docking scores. Because the number of similarity calculations is relatively small and restricted only to a subset of all the orientations sampled during the docking process, the computational cost associated with a SP-DOCK calculation is approximately the same as for the original DOCK calculation.

Similarity-Guided Docking (SG-DOCK)

Within this approach, the similarity between the reference and target ligands is calculated every time a DOCK energy calculation is performed, and the similarity index obtained is used to correct the energy score, as schematically illustrated in Figure 1b. Thus, in a flexible ligand docking the similarity-weighted docking score T_{RB} is used throughout all the incremental ligand construction process and also during the final minimization step. This gives similarity in the SG-DOCK approach a more important role than in the SP-DOCK approach, because it can effectively guide the incremental construction procedure in such a way that the target ligand grows approximately following the reference ligand. However, this approach is computationally more expensive, because the number of similarity calculations to perform is relatively large.

RESULTS AND DISCUSSION

The two similarity-driven flexible ligand docking implementations, SP-DOCK and SG-DOCK, have been applied to the prediction of binding modes for a set of 32 thrombin ligands, for which ligand-bound protein crystal structures are available, and to the screening of three different molecular databases of 1,000 diverse compounds, containing also the set of 32 thrombin ligands. The thrombin case was selected mainly for two reasons. On one hand, it has been used repeatedly as a validation test for other computational methods; thus, it is relevant for comparative purposes. On the other hand, it has been shown recently for the same set of 32 ligands that, despite having a well-defined active site, thrombin can still be a challenge for docking methods aiming at obtaining small root-mean square deviations (RMSD) in binding mode assessment applications[50] and good active-molecule enrichments in molecular database screening applications.[51] The structures of the 32 thrombin ligands used in this work have been presented in detail in previous studies.[50,51] The coordinates of their ligand-bound conformations were obtained either from the PDB[42] or an unpublished internal database.[52,53] All calculations were performed on a SGI/R10000 machine.

One of the objectives of this study was to examine the dependence of the similarity-driven docking methods on the ligand structure taken as reference. Therefore, a similarity analysis of the 32 thrombin ligands was performed to assess the structural diversity of the ligands in our set. For this purpose, the pairwise field-based similarity indices for all possible molecular pairs were computed with MIMIC by using the bound conformations in their observed relative orientations. Then, a principal component analysis of the 32 $\times$ 32 similarity matrix was done. The first three principal components of the analysis accounted for 79% of the total variance. Figure 2 shows the 32 ligands projected in the plane formed by the first and third principal components, which were chosen to provide a good visual perspective. Several clusters of compounds can be visually identified in Figure 2. Cluster I collects the

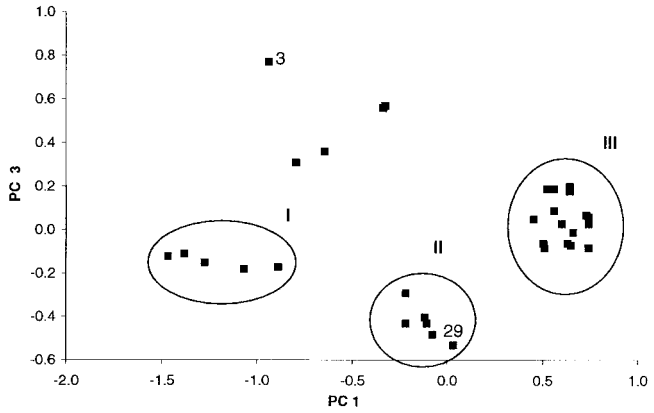

Fig. 2. Principal component analysis of the similarity matrix derived from the set of 32 thrombin inhibitors. For the sake of clarity, only the first and third principal components are shown. The Argatroban (3) and NAPAP (29) thrombin ligands are labeled according to Knegtel et al.[50, 51]

five small ligands present in our set,[†] cluster II contains six ligands,[‡] and cluster III agglomerates sixteen ligands.[¶] Five additional ligands appear more dispersed in the principal component space.[#] From this analysis, two reference structures can be objectively selected, namely, Argatroban and NAPAP. On the one hand, Argatroban does not belong to any of the three major clusters identified in Figure 2, which will ensure limited bias toward a particular structural class of ligands in our set showing a similar binding mode. On the other hand, NAPAP belongs to cluster II, which contains only six ligands (18.75% of our ligand set), and it is located at an intermediate position between clusters I and III. It is obvious that the choice of any ligand from cluster III as the reference ligand in SP-DOCK and SG-DOCK calculations would potentially deliver good results for all the ligands belonging to cluster III (containing 50% of our ligand set); therefore, they were excluded as reference structures.

The structures of human α-thrombin with Argatroban (1DWC, 3.0 Å resolution) and bovine ϵ-trhombin with NAPAP (1ETS, 2.3 Å resolution) were extracted from the PDB. Throughout the article, the nomenclature 1DWC and 1ETS will indifferently refer to the protein structure containing the bound ligand (for SP-DOCK and SG-DOCK), the protein structure only (for DOCK), or the bound ligand structure only (for MIMIC), depending on the method being used. In addition, a "pharmacophore" reference structure derived from a native-substrate analogue bound to thrombin (PDB code: 1UCY) was also considered as an alternative to using the complete structures of Argatroban and NAPAP. The "pharmacophore" used includes the guanidino group, the leucine side-chain, and the phenyl group of the substrate present in the P1, P2, and P3 pockets of the thrombin active site, respectively. The carbonyl group of the substrate making a polar interaction with the backbone of Gly-216 was also considered. It is expected that the use of a "pharmacophore" reference structure for penalizing the final docking orientations of the target ligands will be less biased toward the structure of a particular reference ligand. Throughout the work, SP-DOCK calculations using this "pharmacophore" structure in combination with the 1DWC and 1ETS protein structures will be labeled as SP-DOCK(1DWC/phar) and SP-DOCK(1ETS/phar). For both SP-DOCK(phar) calculations, the steric asymmetric similarity index was used (vide supra).[49] Applications of all five methods (DOCK, SP-DOCK(phar), SP-DOCK, SG-DOCK, and MIMIC) to binding mode assessment and molecular database screening are presented in the next sections.

Binding Mode Assessment

The first objective is to assess the binding mode of the selected set of 32 ligands into the thrombin active site and analyze the dependence of the quality of the orientations produced on the method and reference structure used. Rule-based conformations were generated with the program CORINA.[54] For both bound and rule-based conformations, hydrogen atoms were automatically added and Gasteiger-Marsili atomic charges[55] generated by using Sybyl 6.5. In all cases, both low- and high-sampling parameters were used (vide supra).

Table I summarizes the results obtained for the binding-mode-assessment application. For all flexible ligand docking and similarity calculations, the quality of the results is expressed in terms of the average RMSD obtained for both the highest-ranking solution (best-scoring) and the closest solution to the bound conformation (best-RMSD). To determine the inherent limitations of the different methodologies used, optimization calculations were also performed on the rigid bound conformations. Starting from the experimentally determined binding mode, all docking-based methods find, irrespective of the reference structure, minimum-energy solutions close to the observed bound orientations, with average RMSD values between 0.53 and 0.79 Å. Because no orientational sampling is carried out, DOCK, SP-DOCK(phar), and SP-DOCK produce exactly the same solutions, although their energy scores are different. Results from SG-DOCK do not differ much from those obtained by DOCK. On average, they are slightly better than DOCK when the NAPAP ligand is used as a reference during docking (1ETS reference structure) but slightly worse when the Argatroban ligand is used during docking (1DWC reference structure). Optimization of the similarity between a reference ligand and each target ligand starting from the relative orientation observed in the

[†]Ligands in cluster I: AMH(2), Benzamidine(4), FBA(28), PPA(30), PTA(31). In parenthesis, the labeling according to Knegtel et al.[50]

[‡]Ligands in cluster II: BM51.0986(14), BM51.1011(15), BM51.1023(17), BMS186282(25), BMS186090(26), NAPAP(29). In parenthesis, the labeling according to Knegtel et al.[50]

[¶]Ligands in cluster III: BM14.1224(6), BM14.1238(7), BM14.1241(8), BM14.1243(9), BM14.1244(10), BM14.1248(11), BM18.0537(12), BM18.0540(13), BM51.1022(16), BM51.1031(18), BM51.1037(19), BM51.1045(20), BM51.1047(21), BM51.1059(22), BM51.1081(23), BM51.1110(24). In parenthesis, the labeling according to Knegtel et al.[50]

[#]Ligands not assigned to any major cluster: 3-TAPAP(1), Argatroban(3), BM14.1203(5), DAPA(27), 4-TAPAP(32). In parenthesis, the labeling according to Knegtel et al.[50]

TABLE I. Average RMSD Values for the 31 Thrombin Ligands (the Native or Reference Ligand Is Always Excluded) Obtained From Rigid-Body Optimization of Bound Conformations (X-Ray/Rigid) and Flexible Sampling of Corina-Generated Structures (Corina/Flexible) for the Five Different Methods, the Two Reference Structures, and the Two Sets of Sampling Parameters Used[†]

Reference protein and/or ligand	Method	X-ray/rigid optimization	Corina/flexible			
			Low-sampling best-scoring	Low-sampling best-RMSD	High-sampling best-scoring	High-sampling best-RMSD
1DWC	DOCK	0.67	4.10	3.13 (2)	4.16	2.46 (8)
	SP-DOCK (phar)	0.67	3.81	3.11 (2)	3.21	2.41 (9)
	SP-DOCK	0.67	4.13	3.19 (4)	4.06	2.48 (13)
	SG-DOCK	0.79	3.67	2.82 (4)	3.15	2.05 (13)
	MIMIC	1.50	3.64	3.03 (2)	3.50	2.66 (3)
1ETS	DOCK	0.60	5.57	4.36 (4)	3.88	2.35 (14)
	SP-DOCK (phar)	0.60	5.26	4.26 (4)	3.60	2.35 (10)
	SP-DOCK	0.60	5.20	4.17 (4)	3.77	2.34 (14)
	SG-DOCK	0.53	4.06	3.43 (3)	2.66	1.90 (8)
	MIMIC	0.99	2.51	2.17 (2)	2.39	1.89 (3)

[†]RMSD, root mean square deviation; SP, similarity penalized; SG, similarity guided. The average rank of the best-RMSD solution is given in parentheses.

ligand-bound protein structures finds similarity maxima which are on average much farther away from the observed relative orientation when compared with docking-based methods. The average RMSD values obtained with MIMIC are 1.50 and 0.99 Å when Argatroban (1DWC) and NAPAP (1ETS) are taken as ligand reference structures, respectively. Considering these results as the best attainable from each method, MIMIC seems in principle more handicapped than all docking-based methods when attempting to assess the binding mode of this set of 32 thrombin ligands from flexible calculations.

The results obtained from orientational and conformational sampling of CORINA-generated structures provide a more realistic test of the actual limitations of the different methodologies. As can be observed in Table I, with respect to the RMSD values obtained from optimization calculations, average RMSD values obtained from flexible calculations are much higher for all methods. In general, increasing the orientational and conformational sampling improves the average RMSD values, especially for the 1ETS system. However, the overall trend is maintained, with SG-DOCK, MIMIC, and SP-DOCK(phar) methods producing the best results and SP-DOCK, being on average comparable to DOCK. At this point, it should be stressed that, although lower sampling yields less accurate binding orientations, this method does not necessarily have to compromise the retrieval of active molecules in database searches.[51] The performance of the 1DWC and 1ETS systems in database searches, as will be discussed later on, will illustrate this issue. Even though the 1ETS reference system produces worse average RMSD with low sampling than 1DWC, its ability to retrieve actives from databases can still be comparable to that of the 1DWC system (vide infra).

The differences between the average best-attainable RMSD values from optimization and the average RMSD values obtained for the best-scoring solutions from flexible ligand docking and superposition reveal limitations in sampling and scoring in all methods. Inspection of the scores obtained with docking protocols reveals that, for some ligands, they are on average less favorable than those obtained by rigid minimization, whereas this is not the case for MIMIC. This indicates that the high sampling parameters used in the docking calculations is still unable to locate the closest solution to the observed binding mode. The use of even a more refined sampling (25 seeds and 1500 total orientations) in all docking-based methods did not improve significantly the results obtained with the parameters referred to as "high-sampling" in this work. Therefore, to our view, this finding may be caused by the use of discrete torsion angles in ligand reconstruction and discrete grids for scoring and optimizing. Interestingly, although increased sampling produces better dockings in terms of the average best-RMSD values obtained, solutions docked closest to the observed bound conformations obtain lower ranks (higher values of ranking order) with higher sampling. MIMIC yields similar rankings independent of sampling. This result suggests some degeneracy of the scoring function as more sampling identifies additional conformations with favorable scores that are farther away from the experimental result. In conclusion, the observation that some of the best scorings still have less favorable scores than the minimized structures suggests that sampling is still incomplete, whereas the lack of correlation between RMSD and rank indicates imperfections in scoring.

However, a general trend can be also observed from all average RMSD values reported in Table I, both from optimization of rigid structures and sampling of flexible structures. The more important the similarity term is to a certain method, the more dependent are the results on the reference ligand structure. For optimization calculations of rigid structures, the absolute differences between the average RMSD values obtained when using 1DWC or 1ETS as reference structures are 0.07, 0.07, 0.07, 0.26, and 0.51 for DOCK, SP-DOCK(phar), SP-DOCK, SG-DOCK,

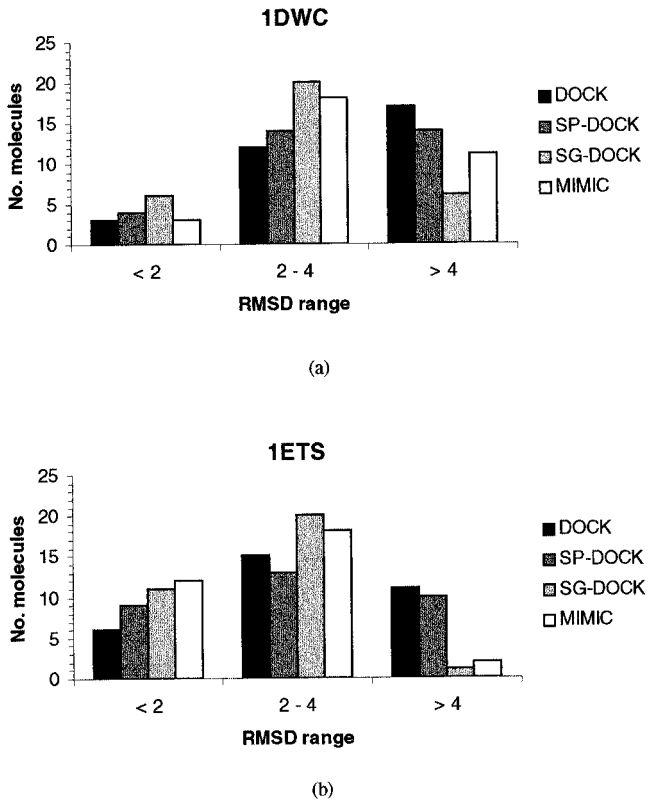

Fig. 3. Frequency distributions of the RMSD values of the best-scoring solutions with respect to the bound conformations for the 32 thrombin inhibitors when using CORINA-generated conformations in flexible ligand calculations. High-sampling parameters were used for all methods.

and MIMIC, respectively. For flexible calculations with the high-sampling parameters, those values are 0.28, 0.39, 0.29, 0.49, and 1.11 for DOCK, SP-DOCK(phar), SP-DOCK, SG-DOCK, and MIMIC, respectively. The limited flexibility of the thrombin active site explains the fact that DOCK and SP-DOCK calculations are quite independent of the reference protein structure used. In comparison, the performance of SG-DOCK and MIMIC are more dependent on how representative the ligand used as reference structure is with respect to the entire set of target ligands. In general, average RMSD values obtained with the finer sampling parameters are better when using 1ETS as reference than when using 1DWC. The reason for this can be found in the fact that, as indicated by the similarity analysis of the 32 thrombin ligands used in this study (Fig. 2), the NAPAP structure (1ETS) on average represents the binding modes adopted by other ligands in the set better than the Argatroban structure (1DWC).

Although the average RMSD values gathered in Table I give an impression of the general performance of the methods under comparison, a more detailed graphical comparison is presented in Figure 3. Here, the number of ligands that are docked or aligned with an RMSD of less than 2 Å, between 2 and 4 Å, or larger than 4 Å, with respect to the observed conformation, are shown. In general, an orientation with a RMSD value of less than 2 Å is overall in good agreement with the observed orientation. When the RMSD is between 2 and 4 Å, there is normally a part of the ligand that is incorrectly oriented, but the overall orientation is still in qualitative agreement with the observed orientation. Finally, predicted orientations with a RMSD value higher than 4 Å can usually be considered to be essentially wrong. As can be derived from Figure 3, the quality of the results obtained with docking-based methods increases in the order DOCK, SP-DOCK, SG-DOCK, when using either 1DWC or 1ETS as reference. In this sense, it is remarkable that SG-DOCK yields RMSD values larger than 4 Å for very few ligands, especially when using 1ETS as reference. Figure 3 also visually illustrates the situation commented above that, despite the quality of the orientations obtained by SG-DOCK and MIMIC, results are more dependent on the choice of the reference structure than docking-based methods.

Another interesting aspect worth emphasizing from results in Table I is the reasonably good performance of the SP-DOCK(phar) calculations. When using the high-sampling parameters, SP-DOCK(phar) is able to improve the average RMSD for the best-scoring solutions with respect to DOCK by 0.9 Å when 1DWC is taken as reference, performing almost as good as SG-DOCK, whereas a similar performance compared with DOCK is obtained when using 1ETS as reference. In addition, with respect to the binned RMSD distribution obtained by DOCK, also for the best-scoring solutions (Fig. 3), SP-DOCK(phar) reduces the number of ligands docked with a RMSD>4Å from 17 to 10 and from 11 to 7 when 1DWC and 1ETS are used as references, respectively. Therefore, the important characteristic of the SP-DOCK(phar) calculations is that they are able to improve the binding mode assessment of ligands with respect to DOCK, with minimal dependence on the structure used as reference for the similarity term. Inspection of the ligand orientations obtained by DOCK and SP-DOCK(phar) provides a clue on the source of the improvements mentioned above. As reported recently,[50] sulfonamide-containing ligands are found particularly difficult to dock with accuracy in the active site of thrombin. The reason is that incorrect docking solutions with reasonable interactions with the enzyme, including a hydrogen bond between Lys60F and the sulfonamide moiety of the inhibitor, seem to be systematically preferred for the scoring function in DOCK, thus causing the ligand to adopt an orientation directed away from the P3 pocket. This situation is illustrated in Figure 4 for the BM14.1224 ligand. As stated above, the docking solution identified by DOCK (in yellow) prefers a hydrogen bonding interaction between the sulfonamide moiety and Lys60F to interactions in the P3 pocket, resulting in a RMSD of 5.95 Å. Instead, the use of a pharmacophore reference structure (in red) to penalize the different docking solutions in SP-DOCK(phar) is able to effectively guide the ligand toward the pockets where it is expected to bind, yielding a RMSD of 1.87 Å.

Finally, to illustrate the performance of SG-DOCK, an example of the evolution of the similarity-guided incremen-

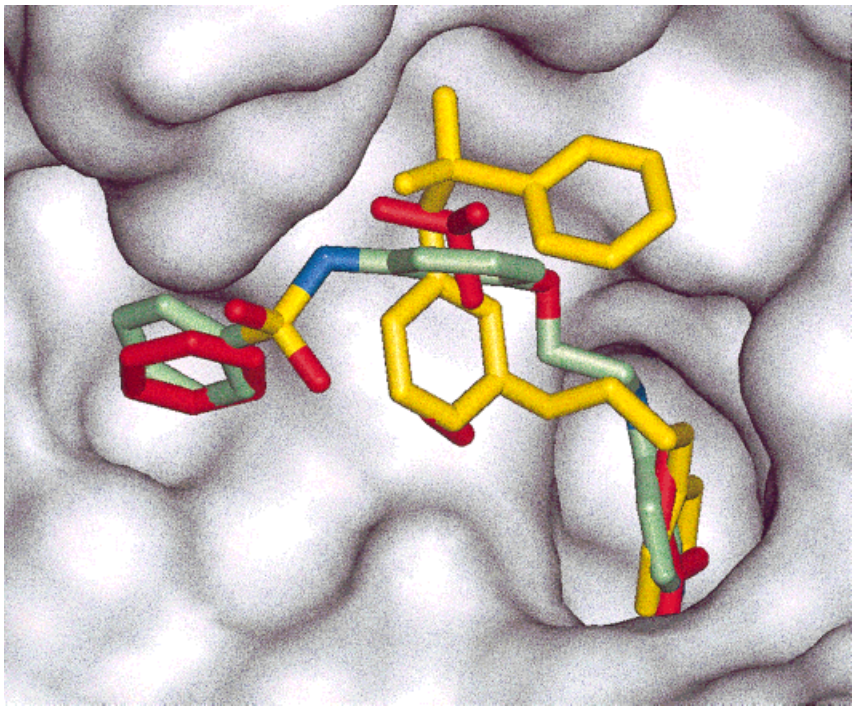

Fig. 4. Best-scoring solutions of BM14.1224 from SP-DOCK(phar) and DOCK. The pharmacophore used in SP-DOCK(phar) calculations is depicted in red, the best-scoring solution of BM14.1224 from SP-DOCK(phar) is colored by element type and that from DOCK is colored in yellow. For the sake of clarity, residues Trp60D and Glu192 were omitted for constructing the surface defined by the active site of thrombin, in white.

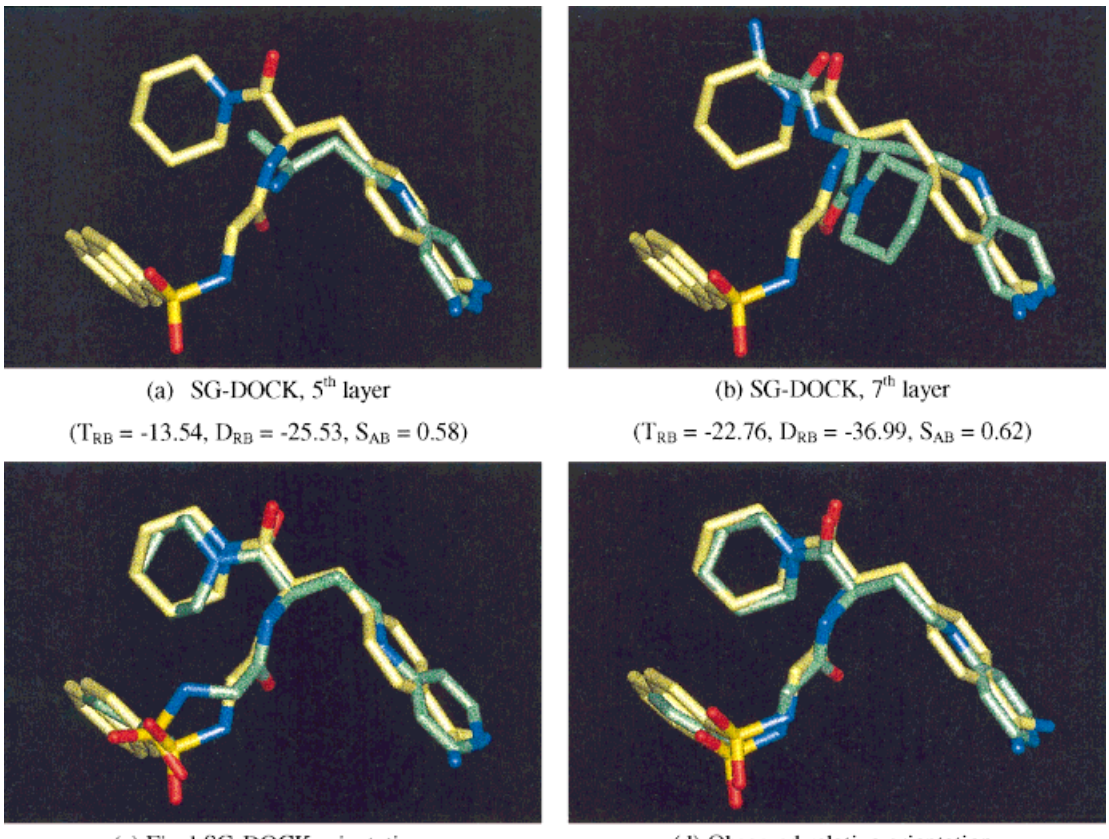

Fig. 5. Best-scoring solutions at different stages of the ligand construction process of BM51.0986, by using the 1ETS structure in SG-DOCK. The experimentally observed relative orientation is also shown. The carbon atoms of the reference ligand (NAPAP) are colored in light yellow and those of the BM51.0986 fragments in green.

tal ligand construction can be followed in detail in Figure 5. Figure 5 shows three steps (5th layer, 7th layer, and final) of the construction process for the BM51.0986 ligand by using 1ETS as reference structure. The experimentally observed relative orientation is shown in Figure 5d. The BM51.0986 ligand was initially divided in 10 different fragments. After orientational sampling of the anchor fragments at different sites of the binding cavity, some of them were ultimately kept as anchors from which incremental construction of the ligand structure evolved. Figure 5a shows the best-scoring seed of BM51.0986 after the 5th layer has been added. Note that the segment of BM51.0986 constructed so far follows closely the structure of the reference ligand, with the pyridine ring correctly selected as the base anchor. However, the following layers were added incorrectly. Figure 5b shows that at the 7th layer the best-scoring seed deviates strongly from the observed orientation in Figure 5d. This finding is due to an incorrect matching of two pairs of carbonyl oxygens at this stage. The correct conformation was recovered at the end of the ligand construction and minimization. As illustrated in Figure 5c, the two pairs of carbonyl oxygens are correctly matched and the orientation obtained follows closely the observed orientation in Figure 5d. It is evident from the final outcome that an orientation, correctly matching the two pairs of carbonyl oxygens, was among the 25 seeds retained for the next step of ligand construction. Eventually, that orientation became the best-scoring orientation after the final stage of ligand construction and energy minimization. The RMSD between the BM51.0986 orientations in Figure 5c and d is 0.99 Å. The low RMSD obtained by SG-DOCK is in part due to the fact that the observed binding mode for BM51.0986 is similar to that of NAPAP (both ligands belong to the same cluster in Figure 2). For the same BM51.0986 ligand, the orientation obtained by DOCK gives a RMSD value of 7.16 Å. The high RMSD value produced by DOCK is because the pyridine ring was not selected as the best anchor to be placed in the P1 pocket and a seed containing the correct anchor was never promoted to the best-scoring position in the successive incremental construction steps. Therefore, the SG-DOCK method seems to be a good strategy to model the binding mode of structurally related compounds during a lead optimization program.

Molecular Database Screening

For all methods, flexible calculations for the molecular database screening applications were carried out with "low-sampling" parameters (vide supra). The performance of each method is measured by its ability to identify the 32 thrombin ligands within a given percentage of the best-scoring molecules from three different docked databases: a database of 1,000 diverse compounds selected from the Available Chemicals Directory (ACD), a database of 1000 diverse compounds selected from a corporate library containing inactive compounds for thrombin (DIV), and a database of 1,000 diverse compounds selected from the World Drug Index (WDI). Diverse selections were obtained by clustering the BCI fingerprints of the respective databases.[56] These three databases provide good test cases for assessing the performance of SP-DOCK and SG-DOCK in virtual screening when compared with DOCK and MIMIC. Also, the different nature of the compounds present in each database will allow for analyzing the dependence of the virtual screening results on the degree of drug-likeness of the database.[57,58] For all molecules in the three databases, 3-dimensional structures were generated automatically by CORINA[54] and Gasteiger-Marsili atomic charges[55] were calculated with Sybyl 6.5. Both 1DWC and 1ETS reference structures were used to study the dependence of the virtual screening results on the reference structure.

TABLE II. Percentages of Active Molecules Identified Within the Top-Ranked 10% of Molecules for Each Database, Depending on the Method and the Reference Structure†

Method	1DWC (%)	1ETS (%)	Time (hours)
(a) ACD database			
DOCK	65.6	65.6	7.3
SP-DOCK (phar)	81.3	75.0	7.5
SP-DOCK	78.1	81.3	7.9
SG-DOCK	78.1	84.4	25.3
MIMIC	68.8	100.0	3.8
(b) DIV database			
DOCK	59.4	65.6	9.5
SP-DOCK (phar)	78.1	71.9	9.7
SP-DOCK	75.0	75.0	10.1
SG-DOCK	75.0	84.4	31.2
MIMIC	65.6	87.5	4.3
(c) WDI database			
DOCK	31.3	31.3	17.9
SP-DOCK (phar)	68.9	65.6	18.2
SP-DOCK	71.9	62.5	18.6
SG-DOCK	62.5	75.0	47.2
MIMIC	59.4	84.4	6.7

†Average screening timings (in hours) are also given.

After the molecules have been ranked according to the scoring used by each method, the percentage of the 32 thrombin ligands included within the 10% best-ranked molecules of each database is used arbitrarily to measure the capacity of a method to identify molecules with appropriate steric and electrostatic characteristics likely to inhibit thrombin. The percentages of active molecules retrieved for all the combinations of database, method and reference structure are listed in Table II. Comparison of the results obtained for the three different databases by each of the methods reveals that the percentage of active molecules retrieved within the 10% best-ranked molecules tends to decrease from the ACD, to the DIV, and to the WDI database. This trend correlates well with the recognized degree of drug-likeness of the molecules in each database.[57,58] On the one hand, the ACD database is assumed to contain very dissimilar molecules with non-drug-like characteristics. On the other hand, the WDI database is assumed to contain molecules showing more drug-like characteristics. In between, the DIV database tries to combine the high molecular diversity of molecules in the ACD with the drug-like character of molecules in the

WDI. Results presented in Table II indicate that, for all methods tested in this study, the difficulty in recognizing a set of 32 thrombin ligands among 1,000 molecules of a database tends to increase with the degree of drug-likeness of the molecules in the database. The main implication for database screening applications is that with the growing interest in compiling and designing databases of molecules with a good drug-like profile, the performance of 3-dimensional virtual screening methods can be below the original expectations.

With regard to the comparison between different combinations of method and reference structure, similar trends are found for each database. The percentage of thrombin ligands identified by DOCK within the 10% of best-ranked molecules in a database is quite satisfactory for the ACD and DIV databases, for which 65.6% of the thrombin ligands are identified within the top-ranked 10% of the database. However, this value drops to 31.3% for the WDI database. MIMIC gives similar results to DOCK when 1DWC is the reference structure but clearly outperforms DOCK when taking 1ETS as reference structure, with retrieved percentages of actives within the best 10% of the database between 84.4% for the WDI and 100% for the ACD. This having been said, a good performance of a flexible ligand-ligand superposition method was somehow expected a priori because, if the similarity scoring used is reasonably discriminative, on average the 32 thrombin ligands should be relatively more similar to the reference compound than to a selection of diverse compounds from databases. The performance of the similarity-driven docking methods shows some dependence on the reference structure used for the similarity evaluation. Whereas the use of a pharmacophore in SP-DOCK(phar) introduces a slight bias toward the 1DWC reference structure compared with DOCK, more active molecules are retrieved by using the 1ETS system as the emphasis on similarity increases from SP-DOCK (no clear preference), to SG-DOCK (slight preference toward 1ETS), to MIMIC (clear preference toward 1ETS). This result is consistent with the similarity analysis of the 32 thrombin ligands performed above (see Fig. 2). SG-DOCK, although approximately three times slower, performs comparable to SP-DOCK on 1DWC and slightly better on 1ETS. Interestingly, even though SG-DOCK was found to produce better ligand conformations (see Table I), this finding is not reflected in a large improvement in the retrieval of actives from the databases. Apparently, obtaining conformations closer to the experimentally determined bound state does not guarantee improved ranking of databases and a scoring function that discriminates between conformations does not necessarily discriminate between different molecules. Finally, it is important to stress that in the case of 1DWC as reference, where DOCK and MIMIC show a similar performance, all similarity-driven docking methods are able to produce better results than the two original docking and alignment methods. This finding indicates that consideration of both docking and similarity terms into a scoring function can be, in some instances, an advantage over the use of the two terms alone.

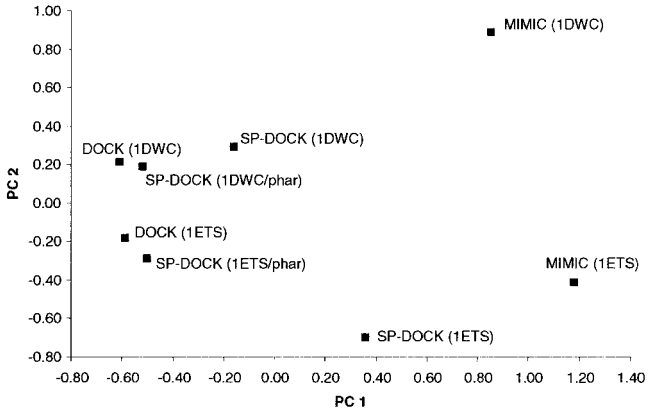

Fig. 6. Principal component analysis of the overlap matrix of the molecules ranked within the first 10% of the ACD database for the different methods used. For the sake of clarity, only the first and second components are shown.

The computational cost of each method when applied to the three different databases, averaged for the 1DWC and 1ETS reference structures, is also reported in Table II. Although one might be willing to invest more time for modelling the binding mode of a limited set of compounds, speed is a key issue for flexible screening of large compound databases. As can be seen, DOCK and SP-DOCK have approximately the same computational cost, whereas SG-DOCK calculations are approximately three times slower than SP-DOCK runs. This makes SG-DOCK a computationally expensive method to be applied in molecular database screening. Another interesting observation is that timings are extremely dependent on the database screened. Again, a correlation is found with the degree of drug-likeness of the molecules in a database. For all methods, calculations were found to be much more expensive when screening 1,000 diverse molecules from the WDI database than from the ACD database. As the molecular weight of the WDI compounds is on average 37% higher than those of the ACD compounds and they contain approximately twice as many rotatable bonds, the more "drug-like" databases warrant more elaborate sampling. Due to the timings obtained for SG-DOCK, which make it less suitable for large numbers of molecules, in the remainder of the article, the discussion will focus on the other four methods.

In addition to the ability of each method to distinguish between thrombin-like and unrelated molecules, it is also interesting to establish if all methods retrieve similar sets of highest-ranked molecules. The compounds retrieved within the 10% of the top-ranked molecules of each database were compared for each combination of receptor and method. A principal component analysis was then carried out on the number of identical compounds found between each pair of method-reference calculation. Figure 6 shows the position of the eight possible combinations of method and reference structure for the ACD database, in the plane defined by the first and second principal components, which can be mainly associated to method and reference structure, respectively. For each of the two reference

structures, the different calculations appear ordered according to the importance of the similarity term in each method, namely, DOCK, SP-DOCK(phar), SP-DOCK, and MIMIC. DOCK and SP-DOCK(phar) are very close, MIMIC is found very far from all the docking-based methods, and SP-DOCK occupies an intermediate position, depending on the reference structure. Interestingly, when going from DOCK to MIMIC along the first principal components, the difference between a given method with the two reference structures increases considerably. These trends are in accordance with those already found when analyzing results from Table II and reveal once more the higher dependency on the reference structure of similarity-based methods with respect to docking-based methods. Similar trends were found for the DIV and WDI databases. Furthermore, Figure 6 reveals also that, even in cases where the percentage of actives identified within the 10% of highest-ranked molecules in a database is similar for two different methods, the other molecules selected from the database within that 10% of the database can be quite different. For instance, although DOCK and MIMIC are able to retrieve 65.6% and 68.8% of the thrombin ligands within the 10% of highest-ranked molecules in the ACD database (see Table II), the percentage of identical molecules identified within that 10% of the ACD database from DOCK and MIMIC is 47.5%. Therefore, the list of molecules present in this 10% of highest-ranked molecules of a database is found to be strongly dependent on the methodology used. The more important the similarity term in a similarity-driven docking method is, the more dissimilar/similar to the original docking/similarity methods the molecules of the 10% of highest-ranked molecules of a database are.

Another important issue often overlooked in database screening applications is the selection of the optimal percentage of highest-ranked molecules that should be kept for further analysis after virtual screening. The number of potential hits present in a database is not known a priori in a practical screening process. Therefore, the scores obtained by the screening method present the only available metric to make an objective selection of a fraction of a database for further evaluation or experimental testing. Plots of the number of molecules versus normalized score for each method, reference structure, and database are presented in Figure 7. For each calculation, the normalized score is defined by assigning a value of 1.0 to the highest-ranked molecule and a value of 0.0 to the lowest-ranked molecule. The scores of the rest of the molecules are scaled accordingly. As can be observed, in all the cases the normalized score decays quickly for the highest-ranked molecules, until a plateau is reached and the normalized score decays more uniformly. At the region of lowest-ranked molecules, there is again a steep decrease in the score values, until a value of zero is reached for the worst-ranked molecule(s) in the database. To select the optimal region for discriminating highest-ranked molecules from medium-ranked molecules, a gradient criterion is used to identify the point along the ranking of molecules where the plateau starts. To smooth out irregularities in score that arise when comparing adjacent molecules in the ranked list, differences in normalized score are considered between each molecule and the following 10 molecules. When the difference in score between a given molecule and each of the subsequent 10 molecules is below a threshold of 0.005 units in normalized score (an average gradient of 0.0005 normalized score units per molecule), it will be considered that the molecular plateau has been reached. These points are marked with crosses in Figure 7 and determine objectively the "optimal" percentage of highest-ranked molecules that should be considered for further analysis, depending on the database, method, and reference structure used.

On average, among the four methods compared, the scoring functions used in MIMIC and SP-DOCK(phar) tend to reach the molecular plateau within the lowest (6.2%) and highest (12.6%) percentages of the top-ranked molecules in the databases, respectively. However, discrimination (difference in scoring between the highest-ranked molecules and the molecular plateau) and selectivity (identification of true thrombin ligands among the highest-ranked molecules) are two concepts that do not often go together. This is also illustrated in Figure 7. As can be observed, the scoring function in all docking-based methods is much more discriminative (the decay of the normalized score with the number of molecules is steeper) than the scoring function in MIMIC, especially for the 1DWC reference structure. This is not surprising if one realizes that MIMIC is based on the overlap between molecular fields; hence, two molecules will always overlap to some extent. However, the degree of selectivity (the percentage of true thrombin ligands retrieved within a given percentage of the highest-ranked molecules in a database) provided by the scoring function of each method does not follow the same trend. For example, when taking 1DWC as the reference structure for screening 1000 diverse molecules of the ACD database, DOCK is much more discriminative than MIMIC (Figure 7a), but the location of the plateau and the percentage of actives at that point are 8.1% and 59.4% for DOCK and 5.2% and 62.5% for MIMIC, respectively. These values are comparable to the 65.6% and 68.8% for DOCK and MIMIC, respectively, when the top-ranked 10% of the database is considered (see Table II). Therefore, a better ability of the scoring function to discriminate between promising molecules from the bulk does not necessarily lead to improved enrichment of actives within the top-ranked molecules of the database.

CONCLUSIONS

Two similarity-driven flexible docking approaches have been implemented by interfacing the DOCK 4.0 and MIMIC programs. The resulting program can be considered a modified version of DOCK, which uses MIMIC as a module to assess the field-based similarity of docked ligands with respect to a reference ligand or pharmacophore structure. With respect to a standard docking calculation, the only additional data needed for a similarity-driven calculation is the structure of a reference ligand or pharmacophore. The two similarity-driven docking algo-

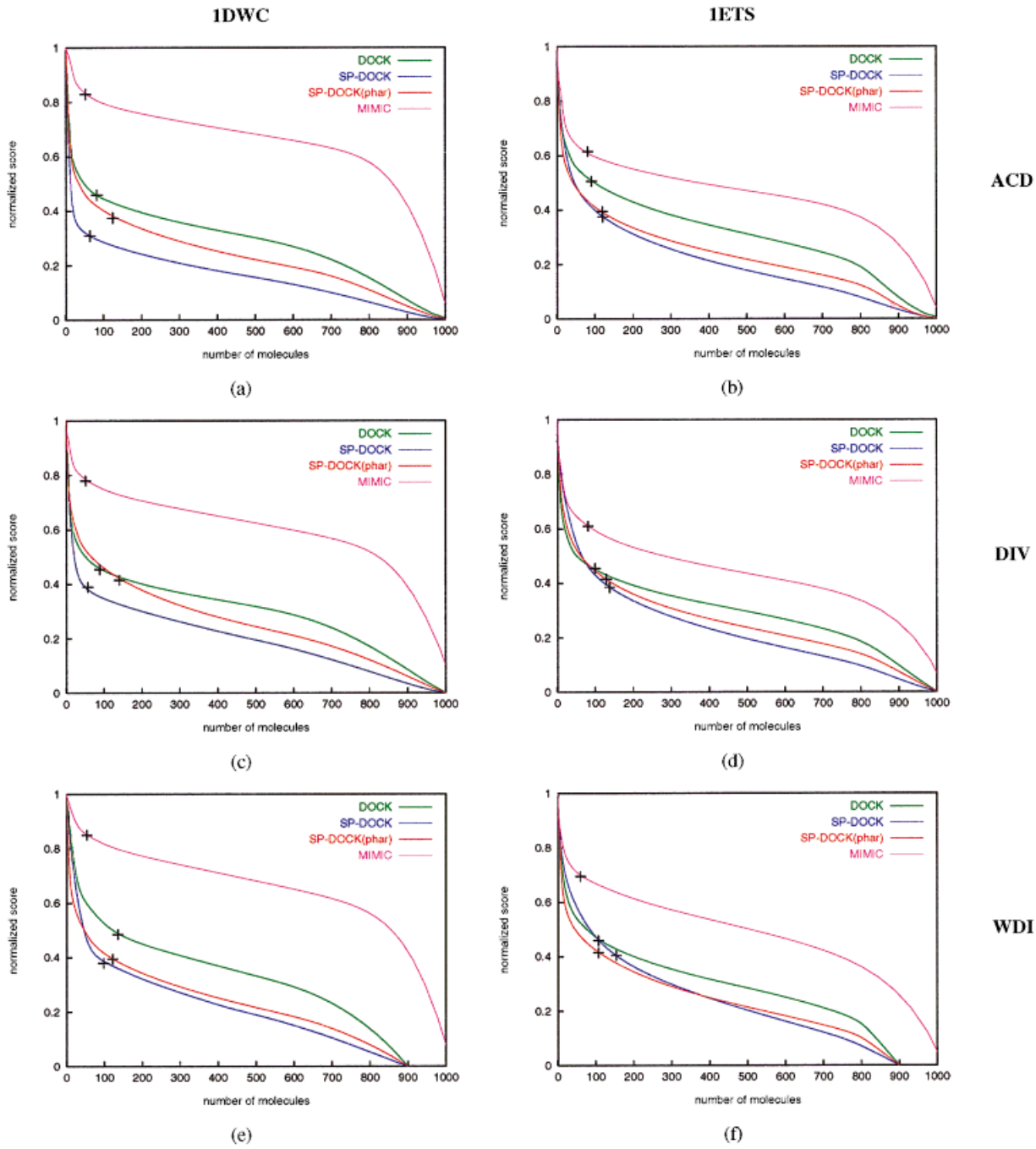

Fig. 7. Discriminative power of the different virtual screening methods depending on the reference structure and databases used. The scores obtained from each method were normalized between 0 and 1. For each method, a cross indicates where the "hit" region can be objectively separated from the "noise" region.

rithms implemented, similarity-penalized docking (SP-DOCK) and similarity-guided docking (SG-DOCK), differ in the way they use similarity measures to modify docking energy scores. In SP-DOCK similarity corrections are taken into account only at the end of the incremental construction process, whereas in SG-DOCK similarity corrections are applied during the entire ligand reconstruction process.

For the test set of 32 thrombin inhibitors used in this study, both similarity-driven docking approaches are able to improve the prediction of binding modes with respect to DOCK. SP-DOCK yields only a modest improvement in terms of RMSD from the experimentally determined binding modes but adds no significant computational load with respect to the original DOCK algorithm. Improvements in the order of 1 Å RMSD are on average obtained with

SG-DOCK, although it is computationally more demanding than the other docking protocols. Nevertheless, a method that has a higher ability of producing correct orientations and conformations for bound ligands is still a valuable computational tool for binding-mode modeling of a limited set of compounds. In this respect, and in the line of recently developed flexible-alignment methods as FlexS[39] and SQ,[40] flexible alignment with MIMIC[41,48] yields also very encouraging results. However, results obtained with MIMIC, as well as those obtained with SP-DOCK and SG-DOCK, are found to be more dependent on the choice of the reference structure than results obtained with DOCK. In particular, SG-DOCK and MIMIC methods tend to obtain better results with 1ETS than with 1DWC. The reason for the poor performance of the Argatroban ligand of 1DWC as a similarity reference is that it is not representative for the majority of the ligands in the test set selected. This is evidenced by a similarity analysis of the 32 ligands in their bound conformations. On the other hand, the smaller dependency of the results obtained by DOCK and SP-DOCK on the reference protein structure is because thrombin has a fairly rigid active site. Stronger dependencies on the reference protein structure will appear in cases where the protein active site experiences some degree of conformational flexibility.

For the evaluation of the usefulness of similarity corrections in virtual screening, three different databases of 1,000 diverse molecules, including the 32 thrombin inhibitors, were considered. As expected, MIMIC obtains better results than DOCK in terms of percentage of active molecules retrieved, for all the databases and reference structures used. However, results obtained with MIMIC remain rather dependent on the choice of the reference ligand. Similarity-driven docking methods are found to give results that either improve the percentage of thrombin ligands identified within the top-ranked molecules with respect to the original docking and similarity methods (when taking 1DWC as reference) or provide intermediate percentages of thrombin ligands between those obtained by the original docking and similarity methods (when taking 1ETS as reference). In this respect, the use of a pharmacophore as a reference structure, which removes some of the bias introduced when a particular ligand structure is used as a reference, was identified as a promising strategy. Finally, an objective way for selecting the set of most promising molecules from a database search is proposed.

Although the results reported here relate only to a single test system, the thrombin case can still be considered as a representative example for proteases, to which much of the current structure-based design efforts have been applied. Especially, given the often flexible and peptidic nature of protease inhibitors, the use of 3D similarity to guide the difficult docking of large flexible inhibitors is expected to be of value for similar systems and should be extendable to other protein families.

At the sampling levels used for database screening, MIMIC is considerably faster than both DOCK and SP-DOCK methods, as can be expected for a method that does not require intensive computational scoring and sampling of protein-ligand interactions. Nevertheless, for all methods calculation times are found to be quite dependent on the nature of the molecules present in the databases. According to timings presented here on the different databases, DOCK and SP-DOCK allow for flexible docking of approximately 1,500–3,000 molecules per CPU per day, and MIMIC allows for flexible alignment of approximately 3,500–6,000 molecules per CPU per day, on a current workstation. This finding suggests that both approaches can be used competitively for flexible screening of medium-sized databases. The timings and the general concept of enforcing field-based similarity make similarity-driven flexible ligand docking a well-suited strategy for the virtual screening of both medium-sized general or targeted libraries during a lead discovery program, where diverse scaffolds with similar steric and electrostatic characteristics need to be identified, and medium-sized focused libraries during a lead optimization program, where a common scaffold is often already in place.

REFERENCES

1. Blaney JM, Dixon JS. A good ligand is hard to find: automatic docking methods. Perspect Drug Discov Design 1993;1:301–319.
2. Jones G, Willett P. Docking small molecule ligands into active sites. Curr Opin Biotech 1995;6:652–656.
3. Lybrand TP. Ligand-protein docking and rational drug design. Curr Opin Struct Biol 1995;5:224–228.
4. Gschwend DA, Good AC, Kuntz ID. Molecular docking towards drug discovery. J Mol Recognit 1996;8:175–186.
5. Lengauer T, Rarey M. Computational methods for biomolecular docking. Curr Opin Struct Biol 1996;6:402–406.
6. Kuntz ID, Blaney JM, Oatley SJ, Langridge RL, Ferrin TE. A geometric approach to macromolecule-ligand interaction. J Mol Biol 1982;161:269–288.
7. Meng EC, Gschwend DA, Blaney JM, Kuntz ID. Orientational sampling and rigid-body minimization in molecular docking. Proteins 1993;17:266–278.
8. Gschwend DA, Kuntz ID. Orientational sampling and rigid-body minimization in molecular docking revisited: on-the-fly optimization and degeneracy removal. J Comput Aided Mol Des 1996;10: 123–132.
9. 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.
10. Welch W, Ruppert J, Jain AN. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. Chem Biol 1996;3:449–462.
11. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727–748.
12. Ewing T, Kuntz ID. Critical evaluation of search algorithms for automated molecular docking and database screening. J Comput Chem 1997;18:1175–1189.
13. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 1998;19:1639–1662.
14. Makino S, Kuntz ID. ELECT++: faster computational search method for docking flexible molecules using molecular similarity. J Comput Chem 1998;19:1834–1852.
15. Baxter CA, Murray CW, Clark DE, Westhead DR, Eldridge MD. Flexible docking using tabu search and an empirical estimate of binding affinity. Proteins 1998;33:367–382.
16. Wang J, Kollman PA, Kuntz ID. Flexible ligand docking: a multistep strategy approach. Proteins 1999;36:1–19.
17. Leach AR. Ligand docking to proteins with discrete side-chain flexibility. J Mol Biol 1994;235:345–356.
18. Knegtel RMA, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. J Mol Biol 1997;266:424–440.

19. Apostolakis J, Plückthun A, Caflisch A. Docking small ligands in flexible binding sites. J Comput Chem 1998;19:21–37.
20. Sandak B, Wolfson HJ, Nussinov R. Flexible docking allowing induced fit in proteins: insights from an open to closed conformational isomers. Proteins 1998;32:159–174.
21. Schnecke V, Swanson CA, Getzoff ED, Tainer JA, Kuhn LA. Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. Proteins 1998;33:74–87.
22. Zacharias M, Sklenar H. Harmonic modes as variables to approximately account for receptor flexibility in ligand-receptor docking simulations: application to DNA minor groove ligand complex. J Comput Chem 1999;20:287–300.
23. Trosset JY, Scheraga HA. Prodock: software package for protein modeling and docking. J Comput Chem 1999;20:412–427.
24. Mangoni M, Roccatano D, Di Nola A. Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation. Proteins 1999;35:153–162.
25. Vieth M, Hirst JD, Dominy BN, Daigler H, Brooks CL III. Assessing search strategies for flexible docking. J Comput Chem 1998;19:1623–1631.
26. Vieth M, Hirst JD, Kolinski A, Brooks CL III. Assessing energy functions for flexible docking. J Comput Chem 1998;19:1612–1622.
27. Johnson MA, Maggiora GM, editors. Concepts and applications of molecular similarity. New York: John Wiley and Sons; 1990.
28. Dean PM, editor. Molecular similarity in drug design. London: Blackie Academic; 1995.
29. Keble G. Structural alignment of molecules. In: Kubinyi H, editor. 3D QSAR in drug design: theory, methods, and applications. Leiden: ESCOM Science; 1993. p 173–199.
30. Kearsley SK, Smith GM. An alternative method for the alignment of molecular structures: maximizing electrostatic and steric overlap. Tetrahedron Comput Methods 1990;3:615–633.
31. Hermann RB, Herron DK. OVID and SUPER: two overlap programs for drug design. J Comput Aided Mol Des 1991;5:511–524.
32. Good AC, Hodgkin EE, Richards WG. Utilization of Gaussian functions for the rapid evaluation of molecular similarity. J Chem Inf Comput Sci 1992;32:188–191.
33. Jain AN, Dietterich TG, Lathrop RH, Chapman D, Critchlow RE Jr, Bauer BE, Webster TA, Lozano-Perez T. Compass: a shape-based machine learning tool for drug design. J Comput Aided Mol Des 1994;8:635–652.
34. Klebe G, Mietzner T, Weber F. Different approaches toward an automatic structural alignment of drug molecules: applications to sterol mimics, thrombin and thermolysin inhibitors. J Comput Aided Mol Des 1994;8:751–778.
35. McMartin C, Bohacek RS. Flexible matching of test ligands to a 3D pharmacophore using a molecular superposition force field: comparison of predicted and experimental conformations of inhibitors in three enzymes. J Comput Aided Mol Des 1995;9:237–250.
36. Perkins TDJ, Mills JEJ, Dean PM. Molecular surface-volume and property matching to superpose flexible dissimilar molecules. J Comput Aided Mol Des 1995;9:479–490.
37. Jones G, Willett P, Glen RC. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. J Comput Aided Mol Des 1995;9:532–549.
38. Mestres J, Rohrer DC, Maggiora GM. MIMIC: a molecular-field matching program. Exploiting the applicability of molecular similarity approaches. J Comput Chem 1997;18:934–954.
39. Lemmen C, Lengauer T, Klebe G. FlexS: a method for fast flexible ligand superposition. J Med Chem 1998;41:4502–4520.
40. Miller MD, Sheridan RP, Kearsley SK. SQ: a program for rapidly producing pharmacophorically relevant molecular superpositions. J Med Chem 1999;42:1505–1514.
41. Mestres J, Knegtel RMA. Similarity versus docking in 3D virtual screening. Perspect Drug Discov Design (in press).
42. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The protein data bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
43. Weiner SJ, Kollman PA, Hguyen DT, Case DA. An all atom force field for simulations of proteins and nucleic acids. J Comput Chem 1986;7:230–252.
44. Meng EC, Shoichet BK, Kuntz ID. Automated docking with grid-based energy evaluation. J Comput Chem 1992;13:505–524.
45. Rohrer DC. 3D molecular similarity modelling in computational drug design. In: Carbó R. editor. Molecular similarity and reactivity: from quantum chemical to phenomenological approaches. Amsterdam: Kluwer; 1995. p 141–161.
46. Mestres J, Rohrer DC, Maggiora GM. A molecular field-based similarity approach to pharmacophoric pattern recognition. J Mol Graphics Mod 1997;15:114–121.
47. Mestres J, Rohrer DC, Maggiora GM. A molecular-field-based similarity study of non-nucleoside HIV-1 reverse transcriptase inhibitors. J Comput Aided Mol Des 1999;13:79–93.
48. Mestres J, Rohrer DC, Maggiora GM. A molecular-field-based similarity study of non-nucleoside HIV-1 reverse transcriptase inhibitors. 2. The relationship between alignment solutions obtained from conformationally rigid and flexible matching. J Comput Aided Mol Des 2000;14:39–51.
49. Deleted in proofs.
50. Knegtel RMA, Bayada DM, Engh RA, von der Saal W, van Geerestein VJ, Grootenhuis PDJ. Comparison of two implementations of the incremental construction algorithm in flexible docking of thrombin inhibitors. J Comput Aided Mol Des 1999;13:167–183.
51. Knegtel RMA, Wagener M. Efficacy and selectivity in flexible database docking. Proteins 1999;37:334–345.
52. Engh RA, Brandstetter H, Sucher G, Eichinger A, Baumann U, Bode W, Huber R, Poll T, Rudolph R, von der Saal W. Enzyme flexibility, solvent and "weak" interactions characterize thrombin-ligand interactions: implications for drug design. Structure 1996; 4:1353–1362.
53. Von der Saal W, Kucznierz R, Leinert H, Engh RA. Derivatives of 4-amino-pyridine as selective thrombin inhibitors. Bioorg Med Chem Lett 1997;7:1283–1288.
54. Sadowski J, Gasteiger J, Klebe G. Comparison of automatic three-dimensional model builders using 639 X-ray structures. J Chem Inf Comput Sci 1994;34:1000–1008.
55. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges. Tetrahedron 1980;36:3219–3288.
56. Barnard JM, Downes GM. Chemical fragment generation and clustering software. J Chem Inf Comput Sci 1997;37:141–142.
57. Ajay, Walters WP, Murcko MA. Can we learn to distinguish between "drug-like" and "nondrug-like" molecules? J Med Chem 1998;41:3314–3324.
58. Sadowski J, Kubinyi H. A scoring scheme for discriminating between drugs and nondrugs. J Med Chem 1998;41:3325–3329.