Journal of Computer-Aided Molecular Design **18:** 635–650, 2004.
DOI 10.1007/s10822-004-5291-4

Unsupervised guided docking of covalently bound ligands

Xavier Fradera[1,*], Jasmit Kaur[1,2] & Jordi Mestres[1,3]
[1]*Department of Medicinal Chemistry, Organon Laboratories Ltd., Newhouse, Lanarkshire ML1 5SH, Scotland, UK;* [2]*Present address: OSI Pharmaceuticals, Watlington Road, Oxford OX46LT, UK;* [3]*Present address: Institut Municipal d'Investigació Mèdica, Universitat Pompeu Fabra, Passeig Marítim de la Barceloneta, 37–49, Barcelona, Spain*

Received 23 June 2004; accepted in revised form 22 October 2004

Key words: covalently bound ligands, molecular similarity, protein–ligand docking, thrombin, virtual screening

Summary

An approach for docking covalently bound ligands in protein enzymes or receptors was implemented in MacDOCK, a similarity-driven docking program based on DOCK 4.0. This approach was tested with a small number of covalent ligand–protein structures, using both native and non-native protein structures. In all cases, MacDOCK was able to generate orientations consistent with the known covalent binding mode of these complexes, with a performance similar to that of other docking programs. This method was also applied to search for known covalent thrombin inhibitors in a medium-sized molecular database (ca. 11,000 compounds). Detection of functional groups suitable for covalent docking was carried out automatically. A significant enrichment in known active molecules in the first 5% of the database was obtained, showing that MacDOCK can be used efficiently for the virtual screening of covalently bound ligands.

Introduction

In recent years, the number of experimentally determined protein structures has increased significantly with advances in NMR and X-ray crystallography [1–3] and this trend is expected to be surpassed in the near future with the consolidation of current global initiatives in structural genomics [4, 5]. Today, there are over 25,000 structures stored in the Protein Data Bank (PDB) [6], more than 2100 of them containing drug-like molecules bound to protein cavities [7]. Within this scenario, structure-based approaches to drug design and screening [8, 9] are being adapted to maximally exploit all the structural information being generated not only for proteins, but also for protein–ligand complexes, in order to be consolidated as part of the arsenal of technologies for lead discovery, generation, and optimisation.

The availability of protein structures led to the development of structure-based computational methods, essentially involving the docking of ligands into the target binding site. Docking approaches model the molecular recognition process between the protein and a ligand by sampling and scoring a number of orientations of the ligand in the binding site, resulting in a predicted binding mode and a measure of the quality of the fit. This scheme works well for modelling non-covalent interactions, as illustrated by the number of active ligands identified for a variety of targets reported recently in the literature [10]. However, many inhibitors are known to bind covalently to the target active site residues. Unfortunately, the ability to model adequately the reactivity of inhibitors with active site residues in a high-throughput, unsupervised manner remains a challenge in computational

*To whom correspondence should be addressed. Tel: +44-1698-73-6273; E-mail: X.Fradera@organon.co.uk

chemistry. Docking of covalently bound ligands using many of the public and commercial packages [11–13] is currently feasible but can often only be managed if the covalent bond between the ligand and the receptor is manually specified. Even though this process can eventually be automated, we are not aware of any publication reporting a virtual screening of covalently bound ligands using any of the publicly available docking programs.

To overcome this limitation, Gehlhaar et al. [14] developed a proprietary virtual screening docking approach to model the covalent binding of inhibitors to serine proteases in which, after a preprocessing step where functional groups in ligands susceptible to attack were converted to the two stereoisomeric tetrahedral forms, the covalent bond between protein and ligand was modelled by aligning the ligand with the catalytic serine and fixing the bond distance and angle between the serine oxygen and the ligand carbon. The approach was successful in reproducing the bound structures, in their correct tetrahedral adduct stereochemistry, of five ligands from three different serine proteases and in identifying a known elastase inhibitor when applied to virtual screening of a subset of the Available Chemicals Directory (ACD) database.

In our laboratory, we have been developing a guided docking approach (MacDOCK [15]) aiming at maximally exploiting all the structural information present in ligand-bound protein structures by combining a protein–ligand docking method (DOCK 4.0 [16]) and a ligand–ligand superposition method (MIMIC [17]). Within this approach, the docking of a ligand into the protein cavity is guided by the presence of functional groups or structural features from bound ligands, so that the final ligand orientations obtained are a balance between having a good interaction within the cavity and retaining a good overlap with the ligand-based structural information provided. Using the same strategy, the approach has been recently extended to allow the docking of covalently bound ligands in an unsupervised manner. The present study introduces this new approach, with examples on its applicability and performance in binding mode assessment and virtual screening.

Materials and methods

MacDOCK

MacDOCK is a hybrid package combining a docking program, DOCK 4.0 [16], and a superposition program, MIMIC [17]. DOCK is the main program, while MIMIC is used to calculate similarity corrections when needed. MacDOCK can perform standard docking calculations, as in DOCK 4.0, and several kinds of similarity-driven docking calculations [15].

DOCK module

DOCK 4.0 [16] is the basis of MacDOCK. Within DOCK, flexible ligand docking is performed using an incremental construction algorithm. Flexible ligands are considered as a collection of rigid fragments, joined by rotatable bonds. Both the segments and the allowed torsions of the rotatable bonds are assigned automatically by DOCK. One or more of these rigid fragments are used as initial base fragments (so-called anchors). Then, for each anchor, different orientations are tried into the protein cavity and assigned a docking energy score. A number of the best anchor orientations are kept for the next stage. Starting from these orientations, groups of one or more fragments (so-called layers) are progressively added and scored. At each stage, several orientations of the partially reconstructed ligand are scored, keeping a number of the best orientations (so-called seeds) for the next stage. This incremental construction process continues until the ligand is fully reconstructed. Final ligand orientations are optimized in order to minimize their docking scores.

Initial orientations of anchors or ligand fragments within the active site are generated by matching ligand atoms to a set of spheres complementary to the protein cavity. The docking score is based on the AMBER force field [18], and has intra- and inter-molecular components. In order to speed up the calculations, the electrostatic and Van der Waals potentials generated in the cavity of the protein are mapped into grid files. Regions where ligands cannot be placed because of close contacts with the protein (so-called bumps)

are also mapped into grids. Further speed-up is achieved by using a united-atom approach.

The DOCK grid files used in this work were generated with the CHEMGRID program, using a 0.3 Å grid spacing and a 4r distance-dependent dielectric, and a trilineal interpolation approximation was used to calculate force-field scores. Sphere sets in the active site were generated with SPHGEN. Both CHEMGRID and SPHGEN are distributed with DOCK 4.0. The number of spheres was manually reduced using Sybyl 6.8 (Tripos, Inc., St. Louis, MO), in order to obtain well-dispersed sets of 20–30 spheres covering the active site. The maximal number of bumps between the ligand and the protein was set to three. The maximal number of iterations for optimization of the docking score was set to 100, and the convergence was set to 0.1. The scope of the sampling is mainly related to two parameters: the number of initial orientations tried for each anchor, and the number of ligand orientations (seeds) that are kept at each stage of the incremental construction process. For low-sampling calculations, the number of initial orientations and seeds was set to 100 and 10, respectively, whereas for high-sampling calculations, it was set to 500 and 50, respectively.

MIMIC module

MIMIC [17] is used for all the similarity calculations necessary in MacDOCK. In general, a steric-volume field, an electrostatic potential field, or a combination of them can be used to describe the structures to be compared. Both molecular fields are represented by atom-centered Gaussian functions. In the present work, only a steric-volume field was used to account for shape overlap. A united-atom approach was applied to speed up calculations.

A field-based similarity measure between two molecules or fragments is defined as,

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

$F_A(\mathbf{r})$ and $F_B(\mathbf{r})$ being the molecular fields associated to the reference structure, A, and the target structure, B, respectively. A normalized cosine-like similarity index is defined as,

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

where Z_{AA} and Z_{BB} are the self-similarities of structures A and B, respectively. For steric field similarity calculations, S_{AB} can take values between 0 and 1. Assuming a rigid reference structure, the actual value of S_{AB} depends on three translational ($\mathbf{t}$) and three rotational degrees of freedom (Θ), plus a number of conformational degrees of freedom (rotatable bonds) for the target structure ($\boldsymbol{\tau}_B$), in flexible alignments. The stand-alone MIMIC program is designed to optimize the value of S_{AB} and obtain the best alignment between A and B. However, within MacDOCK the orientation and conformation of the target molecule is determined by the DOCK module, the MIMIC module being used solely to perform single-point similarity calculations on the coordinates provided by DOCK.

Alternatively, an asymmetric similarity index [19], defined as

$$S_{AB}^{(A)}(\mathbf{t}, \Theta, \boldsymbol{\tau}_B) = \frac{Z_{AB}(\mathbf{t}, \Theta, \boldsymbol{\tau}_B)}{Z_{AA}}, \tag{3}$$

can also be used. $S_{AB}^{(A)}$ measures the degree to which the reference structure, A, is overlapping the target structure, B. In turn, $S_{AB}^{(B)}$ measures the degree to which B is overlapping A. Note that $S_{AB} = \left(S_{AB}^{(A)} \cdot S_{AB}^{(B)}\right)^{1/2}$. Asymmetric similarity indices have been found to be especially useful when comparing structures of different size. In this work, they are used to calculate similarities between putative ligand molecules and reference *pharmacophoric* structures, usually based on fragments of known ligands.

Similarity-driven docking

The similarity-driven docking approach implemented in MacDOCK is a docking procedure in which, at certain steps, similarity values between the target ligand or fragment and a reference structure are used to correct the docking energy score. Two main variants have been developed. One such variant uses the structure of a known ligand or pharmacophore to drive the incremental construction of putative ligands towards a desired binding mode. The other uses structural information on the anchoring atoms for the docking of

covalently bound inhibitors, allowing orientations compatible with a covalent binding of the ligand and the protein to be promoted by maximizing the similarity between selected parts (so-called anchors) of the ligand and receptor. The two variants can be also applied simultaneously in a single calculation (*vide infra*).

For each of the two similarity-driven docking variants described above, two scoring approaches are available in MacDOCK [15]. Similarity-penalized docking applies similarity corrections only at the end of the ligand reconstruction process, whereas similarity-guided docking applies a similarity correction every time a docking score is calculated during the ligand incremental construction. Generally, similarity-guided docking shows a better performance in driving the ligand towards the desired binding mode, at the expense of being computationally more demanding. Therefore, even though similarity-penalized docking is available in all cases, the present work will focus on similarity-guided docking of the two variants mentioned above. A more detailed description of these approaches follows.

Ligand-guided docking

The implementation and assessment of a ligand-guided docking (LG-DOCK) approach have been reported previously [15]. Essentially, the similarity of the target ligand to a reference structure is used as a weighting factor correcting the energy docking score. The reference structure can be either a co-crystallized ligand, if a ligand-bound structure of the protein is available, or a set of pharmacophore-like substructures, obtained from relevant structure-activity relationships, the binding mode of related ligands, or the mapping of functional groups into the protein cavity.

Given a protein structure, R, a reference ligand structure, A, and a target ligand structure, B, one can obtain the docking score for the interaction between B and R, D_{RB}, and the similarity index between A and B, S_{AB}. Then, a similarity-weighted docking score, $T_{RB}^{(A)}$, can be calculated as

$$T_{RB}^{(A)} = D_{RB} \cdot S_{AB}. \quad (4)$$

Actually, since D_{RB} can have positive and negative values, the similarity factor is applied only to orientations with a negative docking score. Thus, LG-DOCK penalyzes those binding modes deviating from the reference, effectively driving the incremental construction algorithm towards orientations of the target ligand following the binding mode of the reference ligand or pharmacophore. The final best orientation is the one with minimal $T_{RB}^{(A)}$ score, which should be a compromise between minimizing the ligand-receptor docking energy score and maximizing the superposition between the reference and target structures.

Anchor-guided docking

The aim of anchor-guided docking (AG-DOCK) is to incorporate information on the location and directionality of the atoms known to be involved in covalent binding to drive the docking of ligands interacting covalently with a protein. In this case, the role of the similarity correction applied to the docking score is to guide the placement of the ligand towards an orientation compatible with the covalent bond formed between ligand and protein.

Previous to the docking process, functional groups in the ligand that can form covalent bonds with the protein are identified. If necessary, these groups are then transformed, so that the geometrical arrangement of all the atoms matches that of the product structure. When several of these functional groups are present in a ligand, or a single functional group can be transformed in different ways (e.g., two different stereoisomers), each of the resulting structures is docked independently, and the results of all docking calculations are combined to obtain the best solution. Figure 1 shows this process for a classical covalently-bound thrombin inhibitor (PPACK). After detection of the carbonyl group that is expected to suffer a nucleophilic attack from Ser195 in the protein, the hybridization of the C atom changes from sp^2 to sp^3 and the two possible tetrahedral stereoisomers are constructed.

For each molecule being docked, an *anchor* group containing all atoms necessary to define the ligand–protein bond is selected for the similarity calculations. This *similarity anchor* serves also to select the *docking anchor* used in the incremental construction process. Note that the *docking anchor* is required to contain the atoms used to define the orientation of the inter-molecular bond, but might contain additional atoms. In other words, the *similarity anchor* is a subset of the *docking anchor*.

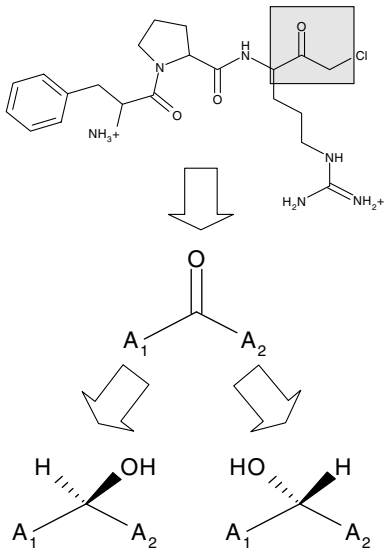

Figure 1. Schematic illustration of the process followed to prepare the PPACK ligand for covalent docking to thrombin (PDB: 1DWE). First, a suitable carbonyl group is located (in this case, a carbonyl bonded to two C atoms in sp^2 or sp^3 hybridization). Then, the carbonyl is transformed to hydroxyl. Both the R and the S stereoisomers are generated. Note that the H bound to the C in the hydroxyl structures works as a capping atom. For this molecule, the C and H atoms were selected as the *similarity anchor*, and all the atoms in the C–HCOH–C fragment were skipped in the calculation of the docking score.

A place suitable for covalent binding in the protein must also be identified. Then, a reference anchor structure mimicking the position of the *similarity anchor* of the ligand in the covalent complex is created. It is worth remarking that the same grid files and sets of spheres can be used for docking covalent and non-covalent ligands. Generally, the reference and target anchor structures have only two atoms. One of these atoms is a capping atom (usually H), which would not be present in a bonded structure. The capping atom is needed to fill all the valences in the ligand, and also to define the directionality of the ligand–protein covalent bond.

In order to obtain meaningful matchings, some modifications have been performed to the standard similarity function (Equation 1). First, all anchor atoms are considered as C atoms within the MIMIC module. This ensures that both anchoring atoms (C and H in the example) are given the same weight in the calculation of the similarity index. Second, in the calculation of the similarity between the reference and anchor ligands, only the contributions arising from overlapping the atoms that need to be matched are considered. This avoids generating artificially high scores to clearly mismatched structures (e.g., a C–H/H–C alignment). However, it must be taken into account that the Gaussian functions used to represent the Van der Waals volume of the C atoms are relatively soft. Therefore, small deviations from an ideal atom–atom matching are not penalized heavily in terms of similarity. This allows some flexibility in the placement of the anchor atoms, which is essential when docking a ligand into a non-native receptor structure. One must be aware, though, that there might be cases where the orientation of the covalent bond in the *reference* structure is significantly different to that of the ligand to be docked. DOCK might fail to find a reasonable ligand orientation in these cases. It is worth to remark that the lack of protein flexibility is a general limitation of DOCK 4.0, which affects docking of non-covalent ligands as well.

The calculation of the docking score has also been modified (D'_{RB}). The difference between D_{RB} and D'_{RB} is that, in order to avoid severe Van der Waals clashes in the vicinity of the ligand–protein covalent bond, some atoms in the ligand are not taken into account in the calculation of the inter-molecular docking score. Typically, the atoms skipped are those involved in the covalent bond plus, optionally, their first neighbours.

Altogether, the similarity-weighted docking score, $T^{(a)}_{RB}$, is obtained as,

$$T^{(a)}_{RB} = D'_{RB} \cdot S_{ab}, \tag{5}$$

where D'_{RB} is the modified docking energy between the ligand and the receptor, and S_{ab} is the similarity between the reference and target anchors, a and b, respectively. As in LG-DOCK, only orientations with negative D'_{RB} score are actually corrected. Note that, if all the atoms in the docking anchor are skipped in the energy calculation, D'_{RB} is always zero for the initial anchor. In this case, the docking score is defined as $T^{(a)}_{RB} = -S_{ab}$, in order to optimize the location of the initial anchor. As more layers are added, the score defined in Equation 5 is used. In summary, use of $T^{(a)}_{RB}$ instead of the original docking score (D_{RB}) is expected to promote orientations with a favourable ligand-receptor interaction consistent with a covalent binding mode.

Anchor/ligand-guided docking (AG/LG-DOCK)

The combined use of ligand- and anchor-guided docking approaches gives rise to what will be referred to as anchor/ligand-guided docking (AG/LG-DOCK). This can be used to drive a ligand to form a covalent bond to the protein at the desired location and, at the same time, follow the binding mode of a known ligand or pharmacophore. In this case, the similarity-weighted docking score, $T_{RB}^{(A,a)}$, is calculated as

$$T_{RB}^{(A,a)} = D'_{RB} \cdot S_{AB} \cdot S_{ab}. \quad (6)$$

As usual, only orientations with negative D'_{RB} are corrected. There is also the possibility of using only the anchor similarity correction in the first phases of the docking process (Equation 5), and start considering the ligand similarity correction (Equation 6) when a given layer is added during the incremental construction process. This approach is useful in order to focus the first steps on finding an orientation compatible with a ligand–protein covalent bond.

Results and discussion

The AG-DOCK and AG/LG-DOCK approaches recently implemented in MacDOCK were evaluated for flexible docking of covalently bound ligands. As a first validation step, the performance in obtaining the correct binding mode for a series of six diverse ligand–protein complexes was assessed by comparing the results with those obtained with other commonly used docking programs. Then, MacDOCK was challenged for its ability to identify the correct binding mode of six covalent thrombin inhibitors, using non-native thrombin structures forming a complex with non-covalent inhibitors. Finally, MacDOCK was used for the screening of a molecular database of ca. 11,000 molecules among which 36 known covalent thrombin inhibitors were present.

Binding mode assessment for native ligands

The objective is to assess the binding mode of a set of ligands into their native protein structures, as obtained from experimental X-ray data, analyzing the dependence of the quality of the orientations produced on the docking methodology (AG-DOCK or AG/LG-DOCK) and level of sampling. The following complexes were extracted from the PDB: 1aec (actinidin), 1blh (β-lactamase), 1lpm (lipase), 1tpp (β -trypsin), 3gch (γ-chymotrypsin), and 4est (pancreatic elastase). DOCK grid files for each protein were obtained using the protocol described above. The ligands covalently bound into these proteins are depicted in Figure 2. Hydrogen atoms were added automatically using Sybyl 6.8, and atomic charges were calculated using the Gasteiger–Marsili algorithm [20]. For each ligand, the atom bound to the protein and a capping hydrogen atom were selected as the reference *similarity anchor*. First-neighbour atoms (with respect to the reference anchor) were flagged to be skipped for the calculation of docking energy scores. For the AG/LG-DOCK calculations, reference *pharmacophore-like* structures were constructed by taking a number of key structural fragments from each ligand in the experimentally observed orientation (enclosed in a box in Figure 2). The ligand fragments were not selected in terms of functionality; rather, we selected *terminal* fragments of each ligand, which could help growing the putative ligand in the right orientation. Starting conformations for each ligand were generated automatically, using CORINA [21]. All the calculations were performed using both low- and high-level orientational sampling options, as defined above. Furthermore, calculations with no orientation sampling were also performed, by minimizing the AG-DOCK score of each ligand from its bound orientation. Table 1 shows the root-mean-square deviation (RMSD) values of all predicted ligand orientations with respect to the experimental ones for this group of complexes, as well as previously published results obtained with FlexX [13] and GOLD [11].

MacDOCK minimizations are used to reveal inherent limitations of the scoring function used in the AG-DOCK approach. The orientations obtained are very close to the experimental ones, with RMSD values below 0.6 Å for all ligands. However, it is remarkable that the docking energy score for 4est is positive. Close inspection of the PDB structure reveals that the amide moiety in this ligand is not planar. This conformation is probably defective, but cannot be corrected with MacDOCK's local optimization routine. This is reflected in a large positive intra-molecular term in the docking score.

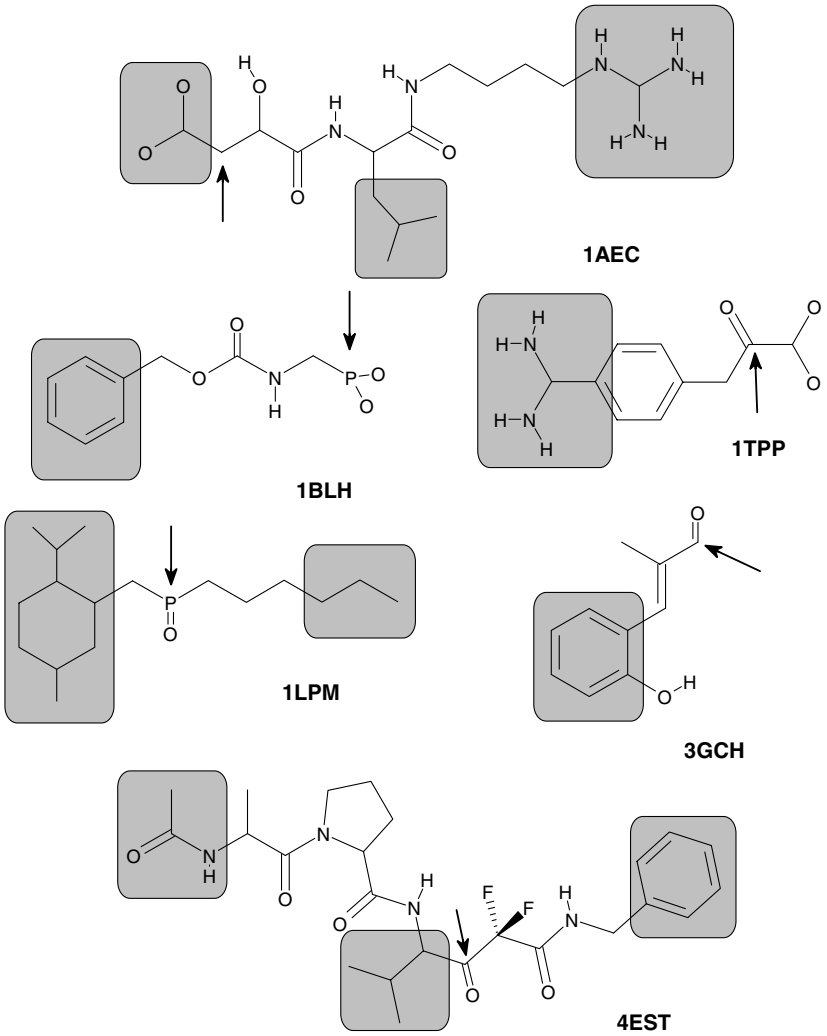

Figure 2. Ligand set used in the assessment of binding modes for native ligands. Reactive sites are marked with an arrow. Atoms forming part of the *pharmacophore* are enclosed within boxes.

Results obtained from flexible docking under the AG-DOCK approach are illustrative to highlight the dependency of the quality of the orientations produced on the degree of sampling used. At first, visual inspection of the orientations obtained confirm that, in all cases, the *similarity anchor* of the ligand is correctly superimposed onto the reference *similarity anchor*, in such a way that a proper ligand–protein covalent bond is formed. Then, the overall quality of the results depends mainly on the orientation of the rest of the ligand structure. Using low sampling, only the 1lpm and 4est complexes are reasonably well predicted by the best-scoring orientation obtained. Note that, in contrast to the results obtained from minimizing the experimentally observed structure, orientations derived for 4est now do have a negative energy score since the internal structural stress has been relieved by having a planar amide bond in the CORINA-generated structure. Increasing the sampling does not have a big impact on the quality of the best-scoring orientations obtained already using a lower sampling, the only exception being 1tpp for which a remarkable improvement is obtained in both energy score and RMSD. However, using a higher sampling does have a significant effect in the ability of MacDOCK in finding an acceptable orientation among all solutions found. Under these sampling conditions, AG-DOCK is able to identify a good

Table 1. RMSD values for the docking of 6 diverse ligands into their respective native proteins.

Complex	FlexX	Gold	MacDOCK											
					AG-DOCK				AG/LG-DOCK					
	Best RMSD	Best scoring	Best RMSD	Minimization	Low sampling		High sampling		Low sampling		High sampling			
					Best scoring	Best RMSD	Best scoring	Best RMSD	Best scoring	Best RMSD	Best scoring	Best RMSD		
1aec	> 3.0	1.11	0.35	0.30	*−35.2*	5.13	*−45.8*	2.78 (3)	5.06	*−47.0*	1.20 (4)	1.47	*−39.3*	1.42 (4)	1.89	*−39.8*	0.96 (20)
1blh	(1.0, 1.5]	1.95	0.53	0.52	*−12.6*	3.55	*−12.7*	3.44 (5)	6.96	*−11.7*	1.29 (29)	1.25	*−9.3*	1.25 (1)	1.56	*−10.7*	0.59 (15)
1lpm	(1.5, 2.0]	–	–	0.53	*−23.0*	1.21	*−10.1*	1.20 (5)	0.88	*−16.8*	0.62 (13)	1.05	*−9.6*	1.05 (1)	0.56	*−16.3*	0.48 (4)
1tpp	(1.0, 1.5]	0.43	0.37	0.26	*−32.0*	5.08	*−6.7*	5.08 (1)	2.08	*−21.5*	1.82 (3)	5.96	*−3.1*	5.15 (4)	1.87	*−23.8*	1.79 (9)
3gch	(1.5, 2.0]	2.64	1.67	0.60	*−17.2*	2.88	*−8.8*	1.77 (3)	2.77	*−9.4*	1.18 (15)	2.25	*−4.1*	2.25 (1)	1.08	*−10.2*	1.08 (1)
4est	(1.0, 1.5]	1.38	1.04	0.60	*2.4*	0.95	*−14.5*	0.95 (1)	1.10	*−15.4*	0.99 (10)	0.77	*−12.0*	0.77 (1)	1.24	*−15.6*	1.05 (2)

Results from FlexX and GOLD were extracted directly from Ref. [11] and [13], respectively. AG-DOCK minimizations were performed by optimizing the docking score of the bound ligand in the observed conformation within the protein. Results for the flexible ligand docking with MacDOCK are reported for the best-scoring orientation and the one with the lowest RMSD value, using two levels of orientation sampling. Energy scores are given in italics and the ranking of the orientation with lowest RMSD in parentheses.

quality orientation under 1.30 Å RMSD for 5 of the 6 ligands and under 1.90 Å RMSD for 1tpp. It is thus encouraging to see that an approach including information only on the location of the anchor provides already good estimates of the experimentally observed binding mode for these ligands among the solutions generated.

Incorporation of a similarity term to account for the overlap between the putative ligand and the *pharmacophore*-like structure under the AG/LG-DOCK approach increases significantly the quality of the orientations produced by MacDOCK. At low sampling, only 1tpp and 3gch are unable to deliver a best-scoring solution under 2 Å RMSD. Increasing the sampling remedies this situation and the best-scoring orientation obtained for all 6 cases is found within 1.90 Å RMSD. The use of high-sampling parameters has also a noticeable effect in the quality of the orientations, with the lowest RMSD found among all solutions retrieved by MacDOCK. The best-RMSD orientation matches the one observed experimentally with an error under 1.10 Å, with the exception of 1tpp for which a value of 1.79 Å RMSD is obtained.

The performance of MacDOCK for modelling the binding mode of covalently bound ligands can be assessed by comparing the results obtained for these 6 protein–ligand complexes with those reported by two of the most widely used docking programs, namely, FlexX [13] and GOLD [11], and included also in Table 1. Results on the quality of the best-scoring orientations are only available for GOLD. As can be observed, GOLD is able to produce a best-scoring solution under 2 Å for 4 out of the 5 cases for which results are available. These results are clearly better than the ones obtained with high-sampling AG-DOCK calculations but are on average comparable to those generated with high-sampling AG/LG-DOCK calculations. With respect to comparing the RMSD values obtained for the orientations having the lowest RMSD among all solutions generated, it is found that MacDOCK (with both high-sampling AG-DOCK and AG/LG-DOCK calculations) provides comparable results to those obtained with GOLD and appreciably better than those reported by FlexX. No detailed data is available on the ranking of the best-RMSD solutions obtained with both FlexX and GOLD. Although not exhaustive, this small benchmarking against FlexX and GOLD serves the purpose of assessing at this stage

the expected performance of MacDOCK for generating acceptable models of the binding mode of covalently bound native ligands.

Binding mode assessment for non-native ligands

The next validation step should explore MacDOCK's ability to reproduce the correct binding mode orientation of a covalently bound ligand when docked into a protein structure originally co-crystallized with another ligand. Docking of these non-native ligands represents a challenging test to any docking program not accounting properly for the flexibility of the protein, since minor changes in the conformation of the residues forming the active site could impede the right accommodation of those ligands into the cavity. In addition, we wanted to investigate also to which extent a protein structure co-crystallised with a non-covalent ligand could be employed for docking covalently bound ligands as well without having to recalculate all energy and bump grid files.

In order to cover the two aspects highlighted above, we needed a protein test case for which structures containing both non-covalent and covalently bound inhibitors were available. A protein target that has been widely used as a validation test in many computational approaches to molecular docking [15, 22–25] and fitting well into those criteria was thrombin. Following previous validation analyses [15], two thrombin structures were selected, namely, human α-thrombin (1dwc, 3.0 Å resolution) and bovine ϵ-thrombin (1ets, 2.3 Å resolution), originally co-crystallised with the non-covalent inhibitors Argatroban and NAPAP, respectively. The same energy and bump grid files used for docking non-covalent ligands to these structures [15] were used in this work for docking covalently bound inhibitors. In addition, a set of six known covalently bound inhibitors for which X-ray crystal structures complexed with thrombin were available, were also selected. The structures of these ligands are compiled in Figure 3, labelled according to their respective PDB code (1ad8, 1awh, 1ba8, 1dwe, 1doj, and 1qj1). Conformations of the ligands, including H atoms, were generated automatically with the program CORINA. For the ligands with a chiral center in the ligand-protein bond (1ad8, 1ba8, 1dwe, 1doj), the two stereoisomers were generated. In order to compare predicted and observed binding modes, all the complexes used in this study were first superimposed by aligning the backbone of all the proteins to that of 1dwc. A likely position for the *similarity* anchor was derived by averaging the positions of the C atom bonded to the protein and the H atom capping the broken bond in the observed orientations of all six ligands.

For AG/LG-DOCK calculations, a pharmacophore-like reference structure is required. The *pharmacophore* used in the present work was derived in an objective manner from the structure of a native substrate analogue bound to thrombin (1ucy). It includes some substructural features occupying the major specificity pockets in thrombin and thus representing a good set of seeds for guiding the docking of ligands into the active site. The list of fragments contains a guanidino group binding at P1, a leucine side-chain binding at P2, and a phenyl group binding at P3, together with a carbonyl group making a key polar interaction with the backbone of Gly216. Figure 4a shows the active site of thrombin (1dwc), as set up for the AG/LG-DOCK calculation. The solvent-accessible surface of the protein is in grey, and only the sidechain of the catalytic Ser195 is depicted. The reference anchor mimicking the formation of a covalent bond between the ligand and thrombin is indicated by two purple spheres located close to Ser195. The set of fragments forming the *pharmacophore* is depicted in yellow. Figure 4b shows how the six covalently bound ligands fit onto those features. This figure reveals that the *pharmacophore* used for these calculations is not fully consistent with the binding mode observed for all the ligands. Obviously, this will have an impact on the performance of AG/LG-DOCK calculations, but this is a situation likely to be encountered in real-life projects and thus we decided to keep it as such.

The results of AG-DOCK and AG/LG-DOCK calculations for the nine stereoisomers derived from the six ligands are collected in Table 2. Following the conclusions obtained from the previous validation exercise on native ligands, only high-sampling parameters were used for the docking of all stereoisomers into the two protein structures. As can be observed, good quality results are obtained from AG-DOCK calculations using a reference anchor only, irrespective of the protein structure used for docking. Best-scoring orientations under 2 Å are obtained for at least one stereoisomer for

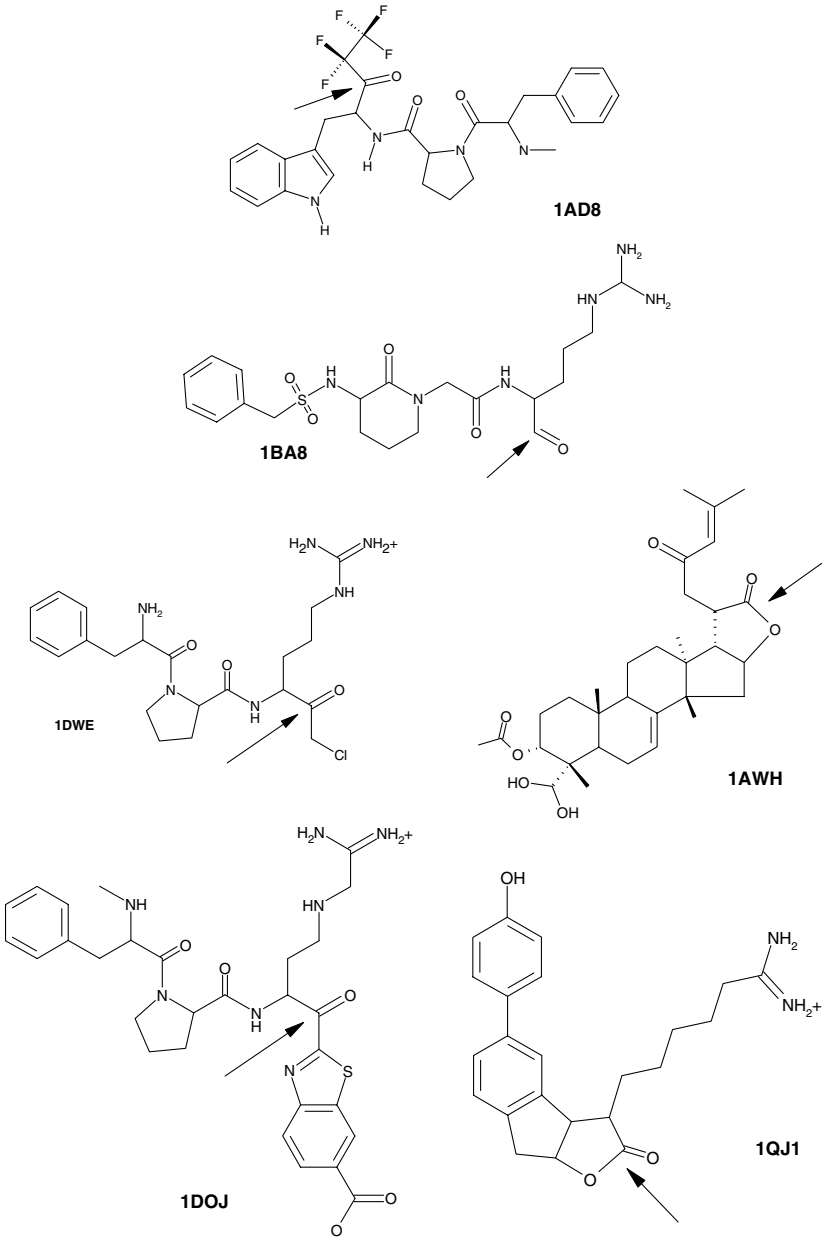

Figure 3. Set of thrombin inhibitors used for assessment of molecular docking with non-native receptors. Reactive sites are marked with an arrow. 1dwe loses the Cl atom after binding to thrombin. 1awh and 1qj1 open the lactone ring.

the 1ad8, 1ba8, 1dwe and 1doj ligands, using either 1dwc or 1ets protein structures. However, best scoring solutions for 1awh and 1qjl have RMSD values over 3.5 Å. A comparison of the structures of all the PDB complexes involved in this study reveals that there are significant conformational differences in some of the residues defining the active site, which explains why cross-docking calculations can be quite difficult sometimes.

In addition, it is worth noting that a decent degree of discrimination between stereoisomers is retrieved, with lower RMSD values found for the stereoisomer present in the X-ray crystal structure in 1ad8 and 1doj. The most apparent discrimination

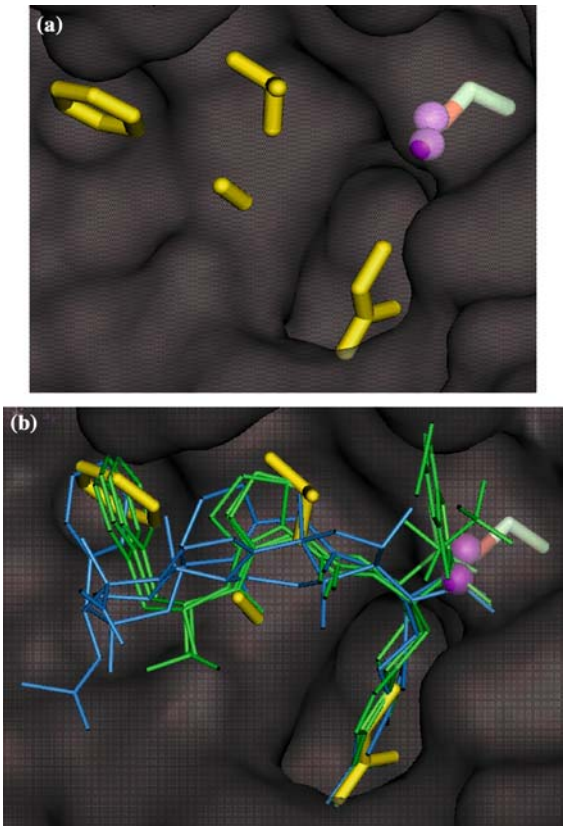

Figure 4. (a) Set-up of the 1dwc receptor for an AG or AG/LG-DOCK calculation. The solvent-accessible protein surface is depicted in grey. For the sake of clarity, residues Trp60D and Glu192 were omitted for constructing the surface. No receptor atoms are drawn, except for the sidechain of Ser195. The atoms in the *similarity anchor* reference are represented as purple spheres, and the *pharmacophore* is depicted in yellow. (b) The thrombin ligands used as a validation test are depicted in their observed orientation in the active site. Ligands with a good overlap with the pharmacophore structure are coloured in green (1ad8, 1doj, 1dwe), whereas those with a poor overlap are depicted in blue (1awh, 1ba8, 1qj1). The anchor similarity reference and the pharmacophore are depicted as in (a).

is observed for 1ad8. In this case, MacDOCK generates best-scoring solutions under 1 Å RMSD for the 'correct' stereoisomer, while it is unable to produce a sensible solution among all the ones generated for the 'wrong' stereoisomer. In contrast, MacDOCK is unable to discriminate correctly between the two stereoisomers in the 1dwe and 1ba8 ligands. For these two ligands, the two stereoisomers are actually equivalent, with respect to the AG-DOCK algorithm, because all the atoms that would be affected by a chiral inversion are actually first-neighbours of the functional group involved in the ligand–protein bond (note that 1dwc loses the Cl atom when bound to the protein). Thus, these atoms are not taken into account in the DOCK energy score calculation. The small differences in RMSD between the stereoisomers of these ligands are due to the stochastic nature of the incremental construction algorithm in DOCK.

Analysis of the results obtained with AG/LG-DOCK calculations reveals that not much improvement is attained compared to AG-DOCK calculations. This means that, for this particular set of ligands, incorporation of structural information in the different binding pockets to guide the docking of ligands does not have a significant impact in performance. In fact, in some cases (e.g., 1ba8) clearly worse results are obtained with AG/LG-DOCK when compared to AG-DOCK. The reason for this trend can be found in the fact that, as mentioned above, some of these ligands (1ba8, 1awh, and 1qj1, in blue in Figure 4b) do not overlap optimally with the structural features of the pharmacophore fragments used as reference structure. In these cases, AG/LG-DOCK calculations guide these ligands towards having a binding mode orientation in agreement with the location of

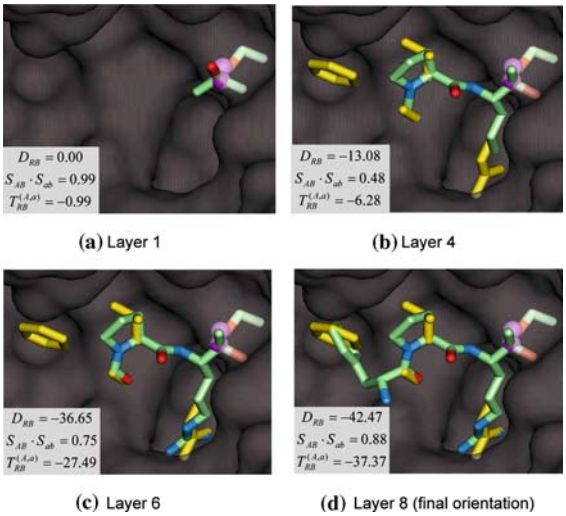

Figure 5. Best-scoring solutions at different stages of the AG/LG-DOCK ligand incremental construction of PPACK within the 1dwc receptor. The PPACK ligand is colored according to atom typing, the *pharmacophore* reference in yellow and the *similarity anchor* reference in purple. Docking score, combined anchor and ligand similarity values, and similarity-guided docking score are reported at each step. (a) Layer 1. The pharmacophore is not depicted, because it is not used yet at this stage. (b) Layer 4. (c) Layer 6. (d) Layer 8 (final orientation).

Table 2. RMSD values for several non-native ligands within two thrombin enzyme structures, obtained from AG-DOCK and AG/LG-DOCK flexible sampling of CORINA-generated structures.

Ligand	1dwc				1ets			
	AG-DOCK		AG/LG-DOCK		AG-DOCK		AG/LG-DOCK	
	Best scoring	Best RMSD	Best scoring	Best RMSD	Best scoring	Best RMSD	Best scoring	Best RMSD
1ad8_1	10.31	6.58 (29)	5.71	4.90 (26)	9.76	7.94 (41)	9.09	6.78 (38)
1ad8_2	0.95	0.95 (1)	1.19	0.99 (4)	1.00	0.94 (5)	1.19	0.94 (11)
1ba8_1	1.99	1.83 (2)	3.05	3.01 (28)	1.93	1.93 (1)	3.37	3.12 (29)
1ba8_2	2.41	2.18 (32)	2.46	2.43 (2)	2.40	2.24 (2)	2.43	2.43 (1)
1dwe_1	1.45	1.00 (12)	1.12	1.02 (9)	1.10	1.06 (4)	1.04	1.04 (2)
1dwe_2	1.68	1.19 (10)	0.96	0.92 (3)	1.83	1.83 (1)	1.80	1.14 (21)
1awh	5.73	1.80 (26)	5.64	1.73 (28)	3.77	3.30 (21)	3.17	3.07 (6)
1doj_1	1.17	0.94 (19)	0.86	0.70 (9)	1.16	0.89 (4)	0.84	0.66 (5)
1doj_2	2.15	1.83 (5)	2.15	1.92 (3)	2.30	2.30 (1)	1.77	1.54 (14)
1qj1	3.88	3.55 (39)	5.13	3.95 (13)	4.66	2.46 (28)	5.02	2.87 (40)

Two stereoisomers were generated for each chiral compound; stereoisomers in bold type have the same chirality as the ligand in the observed X-ray structure.

the pharmacophore fragments and thus driving them away from the orientation observed experimentally in their native structures. This highlights some of the limitations of incorporating additional structural information in the AG/LG-DOCK approach and emphasizes that this will be a successful strategy only if the pharmacophore reference structure is representative of the true binding mode of the ligands being modelled.

The docking of PPACK using the AG/LG-DOCK approach is illustrated in detail in Figure 5, where some stages of the incremental construction algorithm are shown for the stereoisomer having the chirality observed in the experimental structure. Figure 5a shows the best orientation obtained in the 1st layer. Note that the pharmacophore reference structure is not depicted because, as explained above, ligand similarity is not taken into account yet at this stage. The starting docking anchor is a CC(OH)C moiety (hydrogen atoms are not depicted in Figure 5a). The similarity anchor is formed by the O and H atoms in the target ligand and is perfectly superimposed to the reference anchor (S_{ab} = 0.99). To avoid excessive ligand–protein repulsion, none of the atoms in this layer is scored; therefore, the anchor similarity is the only criterion used for selecting the best orientation. Figure 5b shows the best ligand orientation for the 4th layer. At this stage, ligand similarity with the pharmacophore is already taken into account. Note that the anchor atoms continue to be perfectly superimposed, while the ligand is overlapping only part of the pharmacophoric structure. The combined AG/LG similarity factor is 0.48. Figure 5c shows the 6th layer of the incremental construction process. The two additional layers added with respect to Figure 5b allow the ligand to have a better overlap with the structural features of the pharmacophore, the recently added guanidino group in the ligand now superimposing well with the reference fragment in the S1 pocket. This is reflected in a much higher combined similarity index (0.75). Finally, Figure 5d shows the ligand fully reconstructed, after 8 layers. Both the anchor and the ligand are fairly well superimposed to the corresponding reference structures, with a combined similarity index of 0.88. At the same time, the interaction between ligand and receptor is also favourable, with a docking energy score of −42.5, which is reduced to −37.4 after applying the similarity correction.

Comparison of the final best scoring orientation obtained for PPACK into the thrombin structure in 1dwc with the one observed experimentally into its native structure in 1dwe gives an RMSD value of 0.96 Å (see Figure 6). It is worth emphasizing that the receptor model was build using a non-covalent complex structure and the position of the *similarity anchor* was derived by averaging the ligand–protein

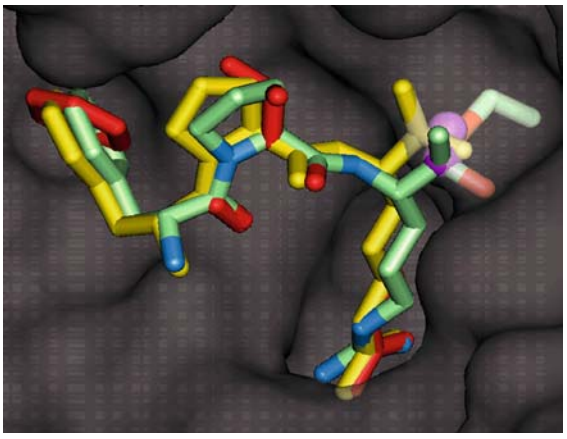

Figure 6. Final orientation of PPACK docked in 1dwc, using AG/LG-DOCK (colored according to atom type), and experimentally observed orientation (colored in yellow). The pharmacophore reference structure is depicted in red, and the *similarity anchor* reference in purple.

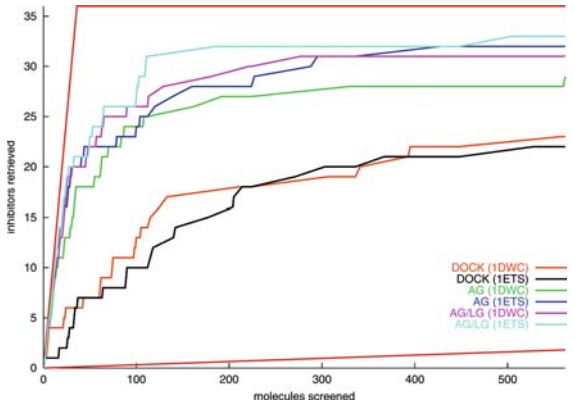

Figure 7. Enrichment graphs for the first 5% of the molecular database, using DOCK, AG-DOCK and AG/LG-DOCK with the 1dwc and 1ets receptors. Ideal and random enrichment graphs are depicted in red.

bonds in five unrelated complexes. Close inspection of the superposition presented in Figure 6 reveals that there is indeed an appreciable deviation between the position of the ligand-protein bond assumed in the docking calculation and the one observed in the experimental complex. In spite of these limitations, MacDOCK was able to generate a good model of the binding mode of PPACK into thrombin. Having established its performance in binding mode assessment, the next step will be to test its applicability for the virtual screening of covalently bound ligands.

Molecular database screening

One of the advantages of our approach to covalent docking is the possibility of carrying out unsupervised virtual screening of molecular databases, provided that suitable functional groups in the ligands are appropriately located and automatically transformed. As an example, we screened a medium-sized molecular database (about 11,000 ligands containing at least one carbonyl functionality), looking for covalently bound thrombin inhibitors. Virtual screenings carried out with AG-DOCK and AG/LG-DOCK were compared to standard DOCK calculations without properly considering the covalent bond between the ligand and protein. Note that standard DOCK should not be considered a proper method for screening of covalently bound ligands. However, we found that, even though the binding modes obtained must be qualitatively wrong, DOCK still obtains a significant enrichment (see below). Therefore, we decided to keep DOCK, rather than a random enrichment, as a reference for our calculations.

The molecular database included all molecules having one or more suitable carbonyl groups in the ACD database (containing over 230,000 compounds). A suitable carbonyl group was defined such that the C atom of the carbonyl group is bound to C atoms, in sp^2 or sp^3 hybridisations, and/or to H atoms. This is a very conservative estimate: there might be many carbonyl groups or related functional groups which are not included in this definition. Indeed, some reactive groups in known active molecules are missed with this approach. However, we decided to keep the reactivity rules simple, and not tailor them according to the active reference set. In a real case scenario, one could define a wider set of rules, trying to include any functional reactive group that could be considered attractive in a hit.

Duplicated molecules, those already in the reference set, those with more than 20 flexible bonds, and those with less than 5 or more than 70 heavy atoms were removed. This left a total of 11,242 compounds. The reference set was a group of 36 known active covalent thrombin inhibitors (see [22, 23]). The carbonyl groups of the molecules in the two sets were transformed to hydroxyls, generating one and two stereoisomers for aldehyde and ketone carbonyl groups, respectively. Altogether, this process yielded a total of

25,997 and 68 structures for the screening database and the reference set, respectively. Note that some molecules in the screening set have more than one suitable carbonyl group, and generated more than two hybridized structures. Whilst actual AG-DOCK and AG/LG-DOCK calculations were performed using the transformed stereoisomers, the results refer to the original structures: each parent molecule is related to one or more transformed molecules, and the docking score of the best scoring one is taken as the score for the parent molecule. The rehybridization of the molecules in the screening database and reference sets, and the collection of the results were performed automatically, using a perl script. Standard DOCK calculations were performed on the original database of 11,242 compounds with carbonyl functionality. All calculations were carried out using the 1dwc and 1ets structures used in the previous section. Based on the quality of the results obtained from the validation cases presented above, high sampling options were used. For the AG/LG-DOCK calculations, the reference structure selected for the LG similarity score was the same as that for the assessment of binding modes of non-native ligands (see above).

Enrichment graphs for the first 5% of the database are presented in Figure 7. The enrichment graphs reveal that all the combinations of method and receptor achieve a significant enrichment in active molecules, especially in the first 1% of the database. Indeed, it is remarkable that DOCK performs reasonably well, in spite of assuming a non-covalent binding mode for all the ligands. However, it is clear that AG-DOCK performs significantly much better than DOCK, with further improvements achieved with AG/LG-DOCK. For instance, within the best ranked 1% of the database, DOCK retrieves 14 known ligands, using 1dwc, or 10 ligands, using 1ets; AG-DOCK retrieves 25 ligands, using any receptor, and AG/LG-DOCK retrieves 26 ligands, using 1dwc, or 31, using 1ets. The performance of the AG and AG/LG methods is especially strong for the best ranked compounds. Indeed, Figure 7 reveals that the enrichment graphs for the AG and AG/LG calculations are very close to ideal for approx. the first 50 compounds. Figure 7 also shows that all the methods perform similarly with the two enzyme structures; however, it is worth noting that, while DOCK tends to perform better using 1dwc, AG-DOCK and AG/LG-DOCK perform slightly better with 1ets.

Using the AG/LG-DOCK method, and the 1dwc protein, there are only 4 reference structures which are not ranked within the first 5% of the database. One of them is actually retrieved within the first 1% when using the 1ets protein, but the other 3 structures are not even ranked within the first 10%. Two of these molecules are similar to PPACK, the ligand in the 1dwe complex, in that a Cl atom should be removed for the ligand to bind in the protein. The program used to process the ligands before the virtual screening did not take this into account, and the ligands were docked with the Cl atom included. When the Cl group is removed and the two ligands are redocked, both structures are ranked within the best 5%, using any of the 1dwc or 1ets receptors.

The third *problematic* ligand is the steroid ligand from the 1awh complex in the PDB. In this case, the lactone ring should be opened in the bound ligand, a situation that was not adequately handled in the pre-processing step. Instead, another carbonyl group was found and selected to react with the protein. The ligand docked in this way is ranked approx. within the first 60% and 75% of the database, for the calculations with 1dwc and 1ets, respectively. Redocking of the ligand with the correct structure improved significantly its ranking in the molecular database; however, it is only ranked within the best 15%, using either of the 1dwc or 1ets structures. Indeed, it is not surprising that this ligand ranks poorly in the database screening, because it was already found to be a difficult case in the previous binding mode validation study.

An analysis of the best-ranking non-reference structures in the AG-DOCK and AG/LG-DOCK calculations reveals that there is a large number of flexible ligands, many of them with long alkyl chains which are folded into the active site. These compounds have no pharmacological interest, and could easily be removed from the database before or after the screening by using simple chemoinformatic filters. Interestingly, a number of steroid structures achieve good rankings. For instance, for the AG/LG-DOCK screening using the 1ets protein, there are 9 steroid structures in the best-ranked 1% of the database (112 molecules), the best ones ranked 11 and 14, respectively. This is quite remarkable, especially considering that the

only steroid in the reference set is not retrieved within the first 10% of the database in any of the calculations (see discussion above). Furthermore, some of these steroids are docked in orientations similar to the observed binding mode of the known steroidal inhibitor.

Altogether, these results show that automatic unsupervised docking of covalently bound ligands can be performed easily with MacDOCK. With respect to a standard DOCK calculation, consideration of the covalent bond between protein and ligand increases dramatically the efficiency of the screening. Further consideration of a pharmacophore-like structure to guide ligand docking leads to a further improvement. The efficiency of the screening depends also on the set of rules used to find and transform reactive functional groups in the database.

Conclusions

An approach for docking covalently bound ligands in protein enzymes or receptors has been implemented in MacDOCK, which is an in-house modified version of DOCK 4.0, combined with MIMIC. MacDOCK uses similarity corrections to drive the ligand reconstruction process towards a binding mode compatible with the structural information available in the similarity reference. The current version of MacDOCK can use two Similarity-Guided Docking approaches: (i) LG-DOCK, where structural information from a known ligand or a pharmacophore-like structure is used to guide the docking of a non-covalent ligand, and (ii) the recently implemented AG-DOCK, where the similarity correction is used to dock covalently bound ligands by matching selected atoms in the ligand and protein. Furthermore, the two approaches can be combined (AG/LG-DOCK).

In this paper, we have validated the AG-DOCK algorithm by docking a set of six diverse ligands. MacDOCK found a binding mode compatible with a ligand–protein bond in all cases, although the resulting binding mode still deviates considerably from the observed one for some ligands. Taking into account a pharmacophoric structure improves the accuracy of the docking process, obtaining correct orientations for the six ligands. Comparison with literature results reveals that the performance of MacDOCK is comparable to that of FlexX and GOLD.

A second test involves binding a group of ligands to a non-native structure. We considered six thrombin inhibitors and two different thrombin enzyme structures. It is worth to remark that both enzyme structures were obtained from non-covalent complexes. Since the orientation of the ligand–protein bond is not identical in all the complexes, the reference anchor structure was built by averaging the ligand–protein bonds in the observed orientations of the six thrombin inhibitors. MacDOCK generated binding modes compatible with a ligand–protein bond in all cases. The quality of the orientations generated was found to depend on the combination of ligand and enzyme. In general, increasing the level of sampling improved the quality of the orientations generated. For the ligands with a chiral center in the atom bound to the protein, both stereoisomers were docked. It was found that the stereoisomer corresponding to the observed one generally gets an orientation which is also closer to the observed one. Altogether, this second validation test reveals that this covalent docking approach is robust and generally applicable, even when a covalently bound complex is not available as reference.

A third test was carried out to assess the applicability of this approach for the virtual database screening of covalently bound ligands. The main advantage of MacDOCK with respect to existing programs is that covalent docking calculations can be performed automatically, after relevant functional groups are identified in the ligands. We screened a database of ca. 11,000 compounds including 36 known thrombin inhibitors. In general, MacDOCK obtained a significant enrichment of known inhibitors within the highest ranked molecules of the database, performing much better than the standard DOCK method, which considers all ligands binding non-covalently.

References

1. Hajduk, P.J., Gerfin, T., Boehlen, J.-M., Häberli, M., Marek, D. and Fesik, S.W., J. Med. Chem., 42 (1999) 2315.
2. Blundell, T.L., Jhoti, H. and Abell, C., Nat. Rev. Drug Discov., 1 (2002) 45.
3. Savchenko, A., Yee, A., Khachatryan, A., Skarina, T., Evdokimova, E., Pavlova, M., Semesi, A., Northey, J.,

Beasley, S., Lan, N., Das, R., Gerstein, M., Arrowmith, C.H. and Edwards, A.M., Proteins, 50 (2003) 392.
4. Stevens, R.C., Yokoyama, S. and Wilson, I.A., Science, 294 (2001) 89.
5. Heinemann, U., Büssow, K., Mueller, U. and Umbach, P., Acc. Chem. Res., 36 (2003) 157.
6. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E., Nucleic Acids Res., 28 (2000) 235.
7. Hendlich, M., Bergner, A., Günther, J. and Klebe, G., J. Mol. Biol., 326 (2003) 607.
8. Fradera, X. and Mestres, J., Curr. Top. Med. Chem., 4 (2004) 687.
9. Halperin, I., Ma, B., Wolfson, H. and Nussinov, R., Proteins, 47 (2002) 409.
10. Lyne, P.D., Drug Discov. Today, 7 (2002) 1047.
11 Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor, R., J. Mol. Biol., 267 (1997) 727.
12. McMartin, C. and Bohacek, R.S., J. Comput.-Aided Mol. Design, 11 (1997) 333.
13. Kramer, B., Rarey, M. and Lengauer, T., Proteins, 37 (1999) 228.
14. Gehlhaar, D.K., Bouzida, D. and Rejto, P.A., ACS Symp. Ser., 719 (1999) 292.
15. Fradera, X., Knegtel, R.M.A. and Mestres J., Proteins, 40 (2000) 623.
16. Ewing, T. and Kuntz, I.D., J. Comput. Chem., 18 (1997) 1175.
17. Mestres, J., Rohrer, D.C. and Maggiora, G.M., J. Comput. Chem., 18 (1997) 934.
18. Weiner, S.J., Kollman, P.A., Hguyen, D.T. and Case, D.A., J. Comput. Chem., 7 (1986) 230.
19. Maggiora, G.M., Petke, J.D. and Mestres, J., J. Math. Chem., 31 (2002) 251.
20. Gasteiger, J. and Marsili, M., Tetrahedron, 36 (1980) 3219.
21. Sadowski, J., Gasteiger, J. and Klebe, G., J. Chem. Inf. Comput. Sci., 34 (1994) 1000.
22. Knegtel, R.M.A., Bayada, D.M., Engh, R.A., von der Saal, W., van Geerestein V.J. and Grootenhuis, P.D.J., J. Comput.-Aided Mol. Design, 13 (1999) 167.
23. Knegtel, R.M.A. and Wagener, M., Proteins, 37 (1999) 334.
24. Baxter, C.A., Murray, C.W., Waszkowycz, B., Li, J., Sykes, R.A., Bone, R.G.A., Perkins, T.D.J. and Wylie, W., J. Chem. Inf. Comput. Sci., 40 (2000) 254.
25. Erickson, J.A., Jalaie, M., Robertson, D.H., Lewis, R.A. and Vieth, M., J. Med. Chem., 47 (2004) 45.