J Mol Model (2004) 10:342–357
DOI 10.1007/s00894-004-0201-1

ORIGINAL PAPER

**Markus H. J. Seifert · Frank Schmitt · Thomas Herz · Bernd Kramer**

# ProPose: a docking engine based on a fully configurable protein–ligand interaction model

Received: 27 April 2004 / Accepted: 19 July 2004 / Published online: 8 October 2004


**Abstract** Virtual high-throughput screening of molecular databases and in particular high-throughput protein–ligand docking are both common methodologies that identify and enrich hits in the early stages of the drug design process. Current protein–ligand docking algorithms often implement a program-specific model for protein–ligand interaction geometries. However, in order to create a platform for arbitrary queries in molecular databases, a new program is desirable that allows more manual control of the modeling of molecular interactions. For that reason, ProPose, an advanced incremental construction docking engine, is presented here that implements a fast and fully configurable molecular interaction and scoring model. This program uses user-defined, discrete, pharmacophore-like representations of molecular interactions that are transformed on-the-fly into a continuous potential energy surface, allowing for the incorporation of target specific interaction mechanisms into docking protocols in a straightforward manner. A torsion angle library, based on semi-empirical quantum chemistry calculations, is used to provide minimum energy torsion angles for the incremental construction algorithm. Docking results of a diverse set of protein–ligand complexes from the Protein Data Bank demonstrate the feasibility of this new approach.

As a result, the seamless integration of pharmacophore-like interaction types into the docking and scoring scheme implemented in ProPose opens new opportunities for efficient, receptor-specific screening protocols.

**Keywords** Protein–ligand docking · Potential energy function · Virtual screening · Scoring

M. H. J. Seifert (✉) · F. Schmitt · T. Herz · B. Kramer
4SC AG, Am Klopferspitz 19a, 82152 Martinsried, Germany
e-mail: markus.seifert@4sc.com
Tel.: +49-89-700763-0
Fax: +49-89-700763-29

*Present address:*
F. Schmitt, quattro research GmbH,
Am Klopferspitz 19a, 82152 Martinsried, Germany

**Abbreviations** *AM1:* Austin Model 1 · *CCDC:* Cambridge Crystallographic Data Center · *COX2:* cyclooxygenase 2 · *IA:* interaction data structure · *NAPAP:* $N^{\alpha}$-($\beta$-naphtylsulfonyl-glycyl)-*p*-amidinophenylalanyl-piperidid · *QSAR:* quantitative structure activity relationships · *PDB:* Protein Data Bank · *RMSD:* root mean square deviation · *TDF:* target description file · *TPSA:* topological polar surface area · *vHTS:* virtual high-throughput screening

## Introduction

Molecular recognition, within protein–ligand complexes, represents a highly complex phenomenon. Its appropriate modeling is essential for rational, structure-based drug design. Despite limitations in today's knowledge about the biophysics and biochemistry of protein–ligand binding, it has been recognized that simplified models of molecular recognition provide a feasible and efficient approach for in silico ligand generation in the early stages of drug development. Generally, in silico screening can be viewed as a complex algorithm for the mining of molecular databases. The screening is usually based on either protein–ligand docking, ligand–ligand alignment or pharmacophore searches. In this paper, we focus on protein–ligand docking, although our software ProPose can be extended easily to both alignment- and pharmacophore-based screening. A variety of methodologies and algorithms for protein–ligand docking are described in the literature and have been reviewed extensively (e.g. [1, 2, 3, 4]). These methods range from the very simple rigid-body approaches to the detailed molecular dynamics simulations of protein–ligand binding. Virtual high-throughput screening (vHTS) can be achieved by accepting a compromise between a realistic representation of the molecular interactions and computational feasibility.

Docking software consists of two main parts: an algorithm for generating docked conformers of the ligand and a method for scoring these conformers. In order to

solve the first problem several algorithms have been devised and implemented into current docking software to avoid the combinatorial explosion of ligand conformers efficiently. AutoDock [5] and Gold, [6] for example, use genetic algorithms to optimize the bound ligand conformation. A rigid-body approach by shape comparison on multi-conformer libraries is performed by, for example, Dock [7] and FRED. [8] The program Glide [9] combines a variety of methods: a systematic search step for approximate positioning, is followed by a force field energy optimization and then a refinement by a Monte Carlo sampling of conformations. FlexX, [10, 11, 12, 13] PhDock, [14] Dock4, [15] and Hammerhead/Surflex, [16] for example, are based on incremental construction algorithms that first place base fragments in the binding site then build up the ligand fragment by fragment. For base fragment placement, the molecular interactions are transformed into a discrete representation: a set of discrete points is placed at locations where protein–ligand interactions are likely. This set of points, however, is used for fragment placement only, e.g. via triangle matching but not used for scoring. However, this kind of discrete representation offers the opportunity for a novel docking program design. After the base fragment placement, the complete ligand is constructed by attaching all the fragments. FlexX [10, 11, 12, 13] and Dock4 [15] implement a "greedy" algorithm that limits the conformational searches using a clustering algorithm based on the score and RMSD of the conformers. Slide [17] uses a discrete representation of the protein–ligand interactions to place the complete ligand by triangle matching for selected anchor fragments. A subsequent optimization step, for acceptable ligand and protein side-chain torsion angles, resolves the resulting collisions between protein and ligand atoms.

Scoring the generated ligand conformers represents a much more difficult problem. Essentially, three approaches are widely used today: force field-based scoring, "potentials of mean force" (for a recent example applied to protein–protein interaction see [18]), and QSAR regression based scoring. [19, 20] Each approach has its specific advantages and problems. According to Ferrara et al., [21] known scoring functions can discriminate between near-native and mis-docked conformations but do not reflect experimentally derived binding affinities. Efforts in improving established scoring methods are ongoing and novel scoring schemes are frequently proposed in the literature. In order to keep pace with the progress made in this area, users of established protein–ligand docking programs are often confronted with the major drawback of these programs, which limits their applicability in the long run: users can modify the interaction geometries and scoring functions only in a limited fashion and the implementation of new scoring and interaction models usually requires the re-compilation of the software. For example, to implement novel interaction geometries and types for screening for transition state analogs (e.g. in aspartic proteases), covalent inhibitors (e.g. in cysteine proteases), or ligands interacting with the heme oxygen in cytochrome P450, one either is limited to the geometries and functions provided by the respective docking software or may have to resort to modifying the source code, if available. Therefore, a program is required that allows a more flexible, user-defined treatment of molecular interactions. This will create a platform for queries in molecular databases that facilitates the adaptation to problem-specific interaction and scoring methods.

Most importantly, interaction types and geometries should be under full control of the user facilitating the handling of important, yet difficult screening targets. For example, for the targets mentioned in the last paragraph the modification of ligand structure and/or protonation state cannot be neglected in the course of its interaction with the receptor. Since the molecules in a database are usually stored with a standard protonation state and a fixed structure, modeling would require modifications of certain molecules. Some approaches for performing modifications before docking have been described using, for example, the docking program QXP/FLO (see [22] and references therein). The modifications may be feasible for a small number of molecules, but become increasingly time-consuming for the large data sets used in vHTS containing millions of molecules: (i) A molecule may contain more than one point susceptible to an, e.g. nucleophilic, attack or may be involved in several types of chemical reactions, all of which leads to a potentially large number of derivatives that have to be considered explicitly. (ii) It may not be desirable or practicable to modify a huge number of molecules in the database for different screening runs in different ways. A more efficient way to handle such cases is to define new, non-standard interaction types for the molecular substructures involved. For example, the point of molecular attack and the geometry of a nucleophilic reaction can be defined by a corresponding interaction. This increases the probability of a correct positioning of the reacting ligand within an active site, thereby simulating the contact pair just before formation of the transition state. This methodology is well suited, for example, to screen for covalent inhibitors where the balance between "molecular recognition" (affinity from non-covalent interactions) and "reactivity" (affinity from covalent bonding) is associated with important selectivity and toxicity considerations (see [22]). Moreover, our methodology is able to screen for non-covalent and potentially covalent ligands in parallel and, more importantly is able to evaluate different possible derivatives of a single molecule implicitly during the docking process without having to dock several of them explicitly.

In order to overcome those limitations of previously proposed approaches to protein–ligand docking we developed a novel docking program, "ProPose", designed for virtual high-throughput docking. The main design objectives were to include:

– A flexible interaction scheme, easy to configure for docking as well as for scoring

– A balanced treatment of polar and non-polar interactions
– Sufficient speed to support the integration into efficient database search technologies such as 4SCan [4]

This can be achieved by extending the established approaches as implemented in, for example, FlexX and Slide, and defining all interaction geometries between chemical moieties, for fragment placement as well as for scoring, as three-dimensional point sets associated with specific chemical substructures. This pharmacophore-like model of interactions allows for an easy-to-understand and flexible definition of arbitrary database queries. The discrete representations of interaction geometries are both computationally very efficient and easily stored in human-readable text files. An efficient weighting and averaging scheme allows the transformation of these sets of discrete points into a smooth potential used for the optimization and scoring of ligand positions. A smooth potential energy surface is necessary for (i) an efficient optimization of the ligand pose in the course of the docking process and (ii) a continuous scoring function.

This paper focuses on the description and validation of the docking engine. For this validation study, ProPose configuration files have been designed that implement Böhm's approach to protein–ligand scoring. [19] A diverse test set of 293 non-covalent protein–ligand complexes is docked in order to demonstrate the performance of ProPose. Additionally, screening runs against two targets, cyclooxygenase-2 and thrombin, are performed. ProPose, however, is not limited to a specific scoring approach and the user can define completely different scoring functions and interaction geometries. The efficiency of non-standard interactions is demonstrated by docking known inhibitors into the active site of cathepsin B.

## Methods

### System overview

The ProPose docking system is illustrated in Fig. 1. The design of the system was optimized for utmost applicability as well as for further development. First, the docking work-flow is split into preparation and screening. The preparation tool, PrepD, creates a target description file (TDF) which contains all information about receptor interactions and atoms needed for docking, including atom coordinates for the clash test and receptor interaction points. The screening engine itself, ProPose, uses the TDF to screen a database of ligands, resulting in a set of ligand poses and scores. During the incremental construction process of the ligand ProPose accesses a torsion angle library (see section Torsion angle library) and a plain text configuration file providing the interaction energies of all defined interaction types. Second, the library version of ProPose can be linked to other programs. This allows other programs to use the functionality of ProPose. For example, it can be linked by the interface wrapper SWIG [23] to create a shell interface to ProPose. ProPose itself is linked to three in-house libraries—libtools, libmath, and libchemistry—which provide algorithms of general, mathematical, and chemistry-related interest. Third, all configuration and data files are plain text files and may be easily altered by the user. For example, the complete information needed for ligand pose generation and scoring is defined in such plain text files. This allows for the convenient incorporation of problem-specific

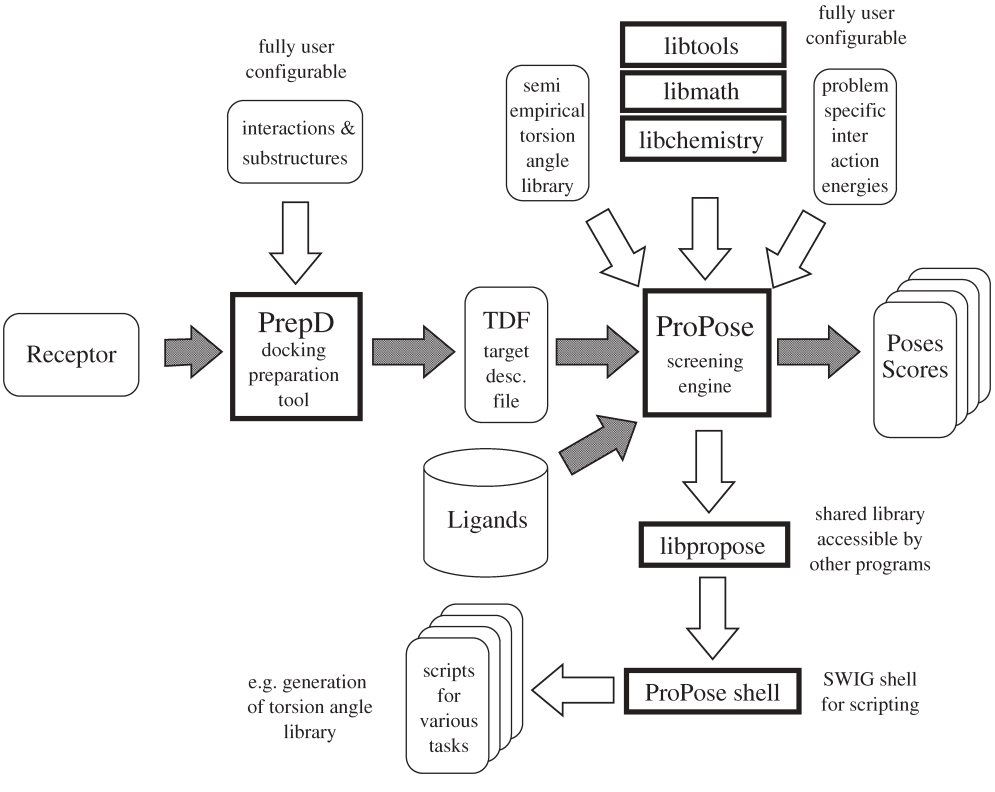


**Fig. 1** Work-flow using ProPose. Boxes with *thick* and *thin* lines depict program modules and text files, respectively. The ligands for screening are retrieved from a database which is symbolized by a cylinder. The work-flow for applying ProPose in a screening protocol is indicated by *gray arrows*. *White arrows* depict the dependencies of the ProPose modules

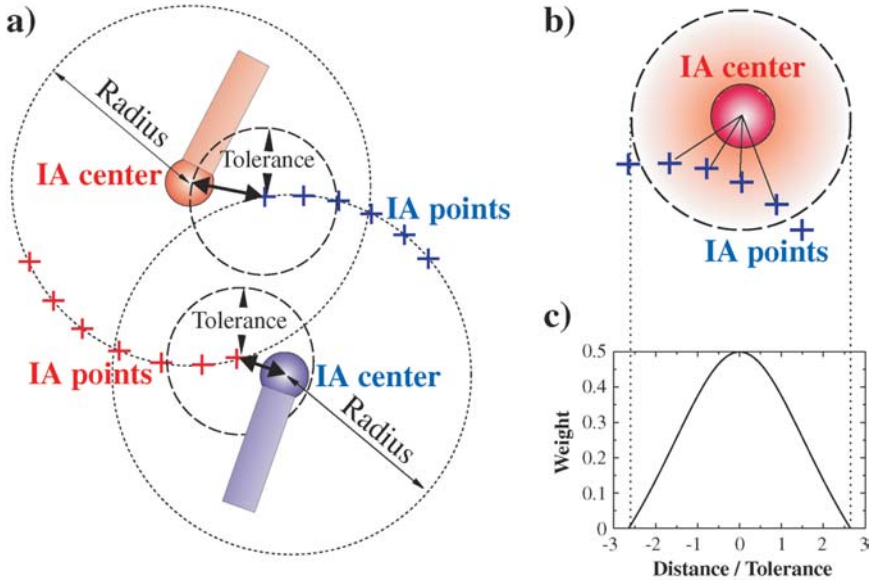


**Fig. 2a–c** Interaction geometry and scoring scheme. **a** An interaction between two chemical moieties (*red* and *blue*) requires the proximity of an interaction point (*cross*) of the first group to the interaction center (*colored sphere*) of the second group and vice versa within a certain tolerance radius (indicated by *bold arrows*). The interaction energy consists of a basic interaction energy multiplied by (i) two geometry factors describing the spatial proximity of points and centers and (ii) the respective geometric weights of the interaction points which account for e.g. angular dependencies of the interaction energies. **b** The score of interaction point–interaction center superposition is computed by averaging the IA point energies within a certain cutoff radius. **c** In order to create a smooth potential energy surface the point energies are weighted according to their point—center distance using a shifted Lorentzian

knowledge into a screening protocol either by manual optimization or by automated procedures.

## Interaction model

### *Discrete modeling of interactions*

The interactions between chemical moieties are modeled by a discrete interaction center–interaction point superposition algorithm, similar to the algorithm published in [10, 11, 12, 13]. More specifically, an interaction results when the center of one moiety is in proximity to an interaction point of the second group and vice versa (see Fig. 2). This description of chemical interactions may be regarded as "pharmacophore-like", in contrast to continuous models used, for example, in force fields. However, in contrast to established docking software an effective averaging function transforms our "pharmacophores" into a force-field-like potential energy, as described in section Scoring function and Lorentzian smoothing of potentials. This methodology is much more flexible than standard pharmacophore queries: the seamless integration of the pharmacophore-like interaction types into the scoring scheme allows the screening of substructures with specific modes of interaction without discarding other potentially high-scoring substructures completely. Additionally, arbitrary interaction geometries may be defined by creating a set of points encoding the desired geometry.

Chemical moieties that are able to interact with other moieties are associated with a data structure called "interaction" (IA). Each interaction is defined by a molecular substructure, a point set encoding the geometry of the interaction and a specific basic energy. The substructure definition consists of a SMARTS [24] string and a corresponding 3D substructure. The SMARTS string is used to identify the chemical moieties in the receptor as well as in the ligand. The 3D substructure definition includes hydrogens which have to be added to the protein as well as the ligand structure before docking. Therefore, protonation and orientation of hydrogen bond donor groups can be used to tune the interactions. The 3D substructure is superimposed on the corresponding receptor or ligand atoms to define the transformation of the point set into the local coordinate system of the specific moiety.

The point set describing the interaction geometry consists of interaction points and an interaction center. For example, to encode a spherical potential with a minimum at radius $R$, the interaction points are located on the surface of a sphere with radius $R$ and the interaction center corresponds to the center of the sphere. Additionally, each interaction point is associated with a point weight $W$ that allows the modeling of angular dependencies of the interaction energy without having to add an explicit angular term in the scoring function. The mean distance of the interaction points $d$ defines the $Tolerance = d/\sqrt{2}$ used in the scoring algorithm. Examples for interaction geometries are shown in Fig. 3: the amide moiety is characterized by (i) a hydrogen bond acceptor with the interaction center (red ball) located on the carbonyl oxygen, (ii) a hydrogen bond donor (yellow ball) on the hydrogen atom and (iii) a hydrophobic interaction (gray ball) parallel to the amide plane. The interaction points (depicted by crosses) of these interactions are located on spherical segments. The novel interaction simulating the nucleophilic attack of a thiolate group on the $\alpha$-carbon of an epoxide moiety is modeled by interaction centers located on the thiolate and on the $\alpha$-carbon. The interaction points outline a spherical segment with a radius of 3.0 Å.

ProPose will try to superimpose the interaction points and interaction centers resulting in a close contact of the epoxide ligand and the thiolate of a receptor. This simulates the contact pair of such a complex just before the transition state of the reaction is formed. This method therefore allows for a pseudo-covalent docking without having to modify the ligand before.

## Scoring function and Lorentzian smoothing of potentials

The protein–ligand interaction score is based on the summation of interaction base energies multiplied by point weights (see Fig. 2). In case of non-ideal superposition of points and centers the interaction energy drops off according to a distance-dependent interpolation

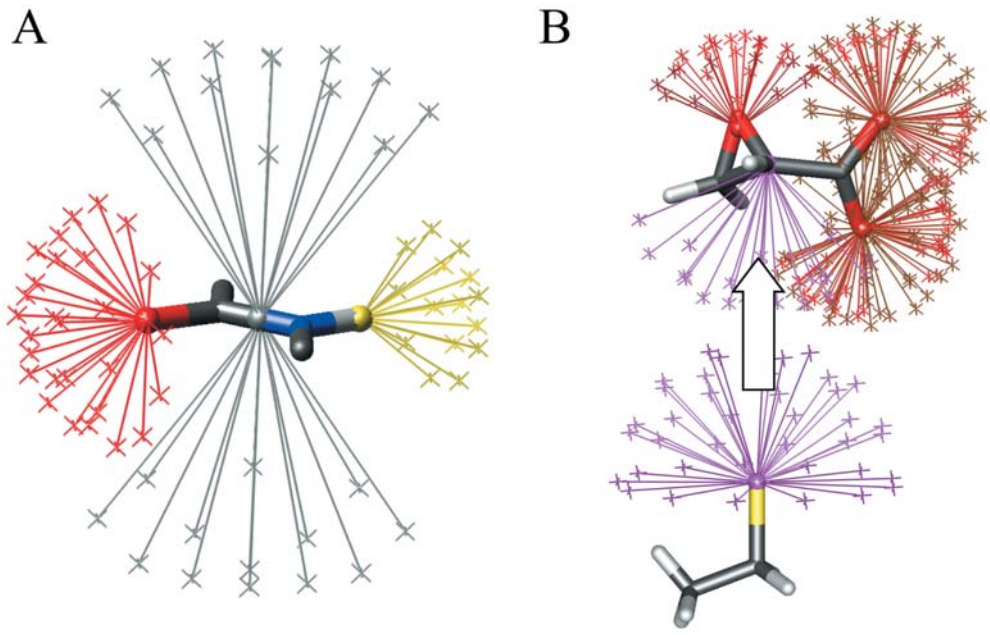


**Fig. 3a, b** Examples of interaction geometries. **a** Amide moiety: hydrogen bond donor (*yellow*), lone pairs (*red*), and hydrophobic interaction (*gray*). **b** Novel interaction type simulating the nucleophilic attack on the $\alpha$-carbon of epoxide by thiolate (*purple*, see text). Negative charge interactions are depicted in *brown*. The interaction points and centers are represented by *crosses* and *spheres*, respectively

scheme. A Lorentzian averaging of point weights $W_k$, provides a fast and efficient smoothing of the potential energy $E^{i,j}$

$$E_k^{i,j} = W_k \left( \frac{c^2}{d^2 + c^2} - \frac{1}{2} \right) \quad \forall \quad d = \left| \vec{R}_k^i - \vec{R}^j \right| < c \tag{1}$$

where $\left| \vec{R}_k^i - \vec{R}^j \right|$ is the distance between the interaction point $k$ of interaction $i$ and the interaction center of the interacting moiety $j$. $c$ is a cutoff distance. Tests showed that a value of $c \approx 2.6 \times$Tolerance is optimal to reduce the "wiggles" in the potential energy function (see Fig. 4). The Lorentzian smoothing function can be computed very efficiently: since all distances are calculated as squared distances, the evaluation of Eq (1) requires only one more expensive division operation. The total energy of an interaction $i$ with respect to a second interaction $j$, is given by summation over all interactions points $k$ of interaction $i$ multiplied by a scaling factor $a$

$$E^{i,j} = a \sum_k E_k^{i,j} \tag{2}$$

The value of $a$ was adjusted to 0.4 in order to calibrate the mean value of the potential to the basic interaction energy (see Results). The score $S$ for two interacting moieties $i$ and $j$ is given by

$$S_{i,j} = E_{\text{basic}} E^{i,j} E^{j,i} \tag{3}$$

where $E_{\text{basic}}$ is the basic interaction energy specified in a configuration file. The product of $E^{i,j}$ and $E^{j,i}$ ensures the proximity of interaction points of interaction $i$ and the interaction center of interaction $j$ and vice versa. The total score for two molecules is computed by the summation over all ligand and receptor interactions $i$ and $j$, respectively:

$$S = \sum_i \sum_j S_{i,j} \tag{4}$$

Only pairs of interaction with an interaction center distance smaller than a cutoff of 10 Å are taken into account. To speed up the computation of this sum, the interaction points of the receptor are hashed. The scoring algorithm is based on the definitions of the interactions only, therefore allowing full control by the user. There is no reference to any hard-coded geometric models, for example the spherical squares in FlexX, so the scoring algorithm can be called "model-free". An example of a specific implementation of a scoring function is given in the section Validation.

*Clustering of hydrophobic interaction points*

Hydrophobic interactions require a special treatment: usually hydrophobic interaction points are modeled onto a sphere around a hydrophobic group, e.g. methyl. This leads to an enormous number of interaction points with a relatively unspecific spatial distribution, therefore being unsuitable for base fragment placement as well as for incremental construction. Therefore, the hydrophobic interaction points are clustered according to one of the following methods, which is specified in a configuration parameter:

(i) Pocket mode: only interaction points with a sufficient hydrophobic interaction point density in their neighborhood are retained (minimum 2 within 0.5 Å and minimum 4 within 1.0 Å, in order to focus on regions with a high density), all others are deleted (see Fig. 5). This procedure creates interaction point clusters specifically located in hydrophobic pockets and ProPose will subsequently try to fill these pockets with hydrophobic ligands.
(ii) Surface mode: only interaction points that represent the solvent accessible surface of hydrophobic moieties are retained (see Fig. 5). This method is suitable for rather shallow ligand binding sites where the pocket mode does not work efficiently due to a low density of hydrophobic interaction points.

## Preparation tool

The ProPose system decouples input processing and incremental construction (see Fig. 1). Therefore, the input processing for the target only takes place once, even if a large number of ligands are to be screened. Additionally, the docking engine does not access the specific receptor PDB or mol2 files, thereby reducing the network load in computer clusters. The preparation tool, PrepD reads the receptor structure and the interaction substructure definitions and generates the target description file (TDF) containing only the minimum information necessary for incremental construction: the receptor interactions including their respective interaction points and the atom coordinates needed to evaluate clashes of ligand and receptor atoms. The file format of the TDF does not contain any protein specific data. This additional abstraction layer of ProPose allows for a straightforward extension of the range of applications by writing new preparation tools which transform the specific input into a TDF.

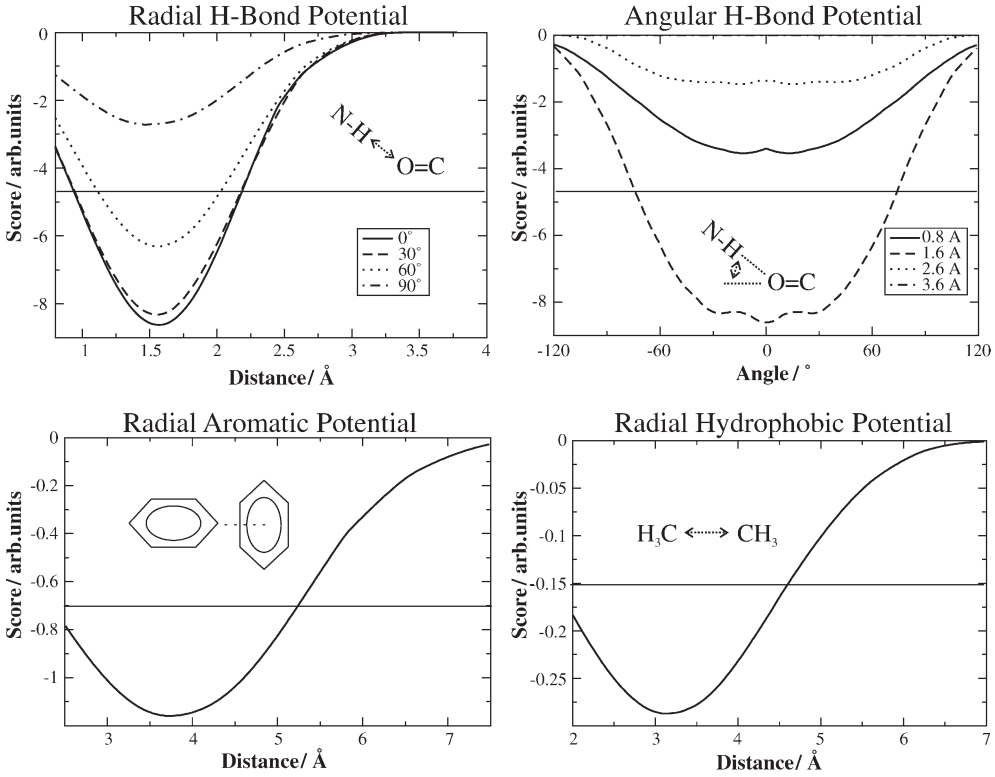


**Fig. 4** Potential energy functions. The interaction scheme is based on discrete points, but in order to achieve a smooth potential energy surface the different weights of the interactions points are averaged using a shifted Lorentzian function. Four examples derived from Böhm's scoring function are shown in this figure: radial H-bond potential (*upper left corner*), angular H-bond potential (*upper right corner*), radial potential for aromatic interactions (*lower left corner*), and the radial hydrophobic potential (*lower right corner*). The score is given in arbitrary units. At short distances the potentials are truncated by a clash test routine that removes all ligand poses with a significant overlap with receptor atoms (not shown in this figure). The basic interaction energies $E_{basic}$(lone_pair/donor)=−4.7, $E_{basic}$(aro_ring/aro_center)=−0.7, and $E_{basic}$(hydrophobic)=−0.15 are indicated as *horizontal lines*

Docking engine

The ProPose docking engine handles the conformational flexibility of the ligand by using a fragment-based approach. [10, 11, 12, 13, 14, 15, 16] A step-by-step breakdown of the docking process is given here:

- Processing of the target description file
- Fragmentation of the ligand by cutting at rotatable bonds
- Base fragment selection and placement
- Incremental construction inside the protein's binding site

After reading the TDF and the ligand molecule, the latter is split into fragments by cutting at the rotatable single bonds, interaction points are generated, and ring conformations are computed using Corina. [25] Subsequently, two base fragments are selected according to their potential interaction energies and placed into the target active site. The selection criteria for base fragments are low conformational flexibility and strong interaction possibilities:

- Base fragments with less than four atoms are not allowed.
- At least one triangle of interaction centers exists in the fragment where the sum of possible interaction energies is below a certain threshold (default: −8.0).
- Two fragments may be combined to one base fragment, if the possible interaction energy does not exceed a certain cutoff (default: −30).
- The base fragments are ranked according to their possible interaction energies, including a penalty for fragments with more than 12 conformations.

The base fragment placement is accomplished by aligning compatible triangles of interaction points and centers, which are identified by geometric hashing. [26] In order to allow for an energy-weighted triangle superposition, a Quaternion algorithm [27] is used to align the compatible triangles. In contrast to, for example FlexX, no interaction hierarchy is used for base fragment placement. In combination with the clustering of hydrophobic interaction points even base fragments without polar interactions can be positioned favorably leading to a more balanced treatment of polar and apolar ligands for database screening. In the next step, the ligand is constructed within the binding or active site of the target. The ligand fragments are attached to the base fragment using torsion angles from a torsion angle library (see below). A rigid-body and/or a torsion angle optimization of the resulting ligand pose is applied. Ligand poses that do not interfere sterically with the receptor atoms are clustered according to a heuristic algorithm to reduce the number of possible poses so as to keep their number within a certain range for successful docking. The clustering algorithm uses the similarity between the interaction pattern with the receptor atoms as its measure:

(i) Starting with a list sorted by score of partial ligand placements, the poses that share a certain number of interactions with the receptor are clustered into groups. As a standard, poses which share a minimum of 60% of their interactions are compiled into one cluster. Poses with a RMSD below a certain threshold with respect to poses already present in the corresponding cluster are removed.

(ii) A number of representatives (at least one) is chosen from each group. The number of representatives depends exponentially on

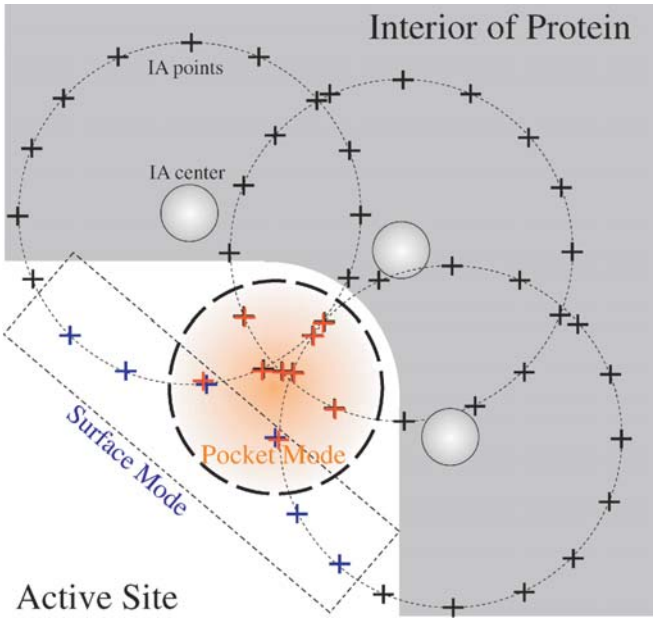


**Fig. 5** Clustering of hydrophobic interaction points. Using the pocket mode clustering, hydrophobic interaction points are selected according to their neighbor density to identify hydrophobic pockets. Only the interaction points marked as an area of *orange crosses* are retained. Each interaction point gets a weighting factor depending on the interaction point density in its vicinity. Using the surface mode, only non-redundant surface-exposed interaction points are retained (*blue crosses*). All other hydrophobic interaction points and centers are depicted by *black crosses* and *gray spheres*, respectively

the score of the best-scoring pose within a group. A MaxMin Algorithm [28] is used to extract the representatives from each cluster.

In contrast to using a ligand RMSD-based distance measure, this clustering algorithm does not require the computation of a full distance matrix. Additionally, the clusters do not depend on an arbitrary starting point since the ligand poses are referenced only to the fixed coordinate system defined by the receptor.

## Torsion angle library

A torsion angle library was generated based on semi-empirical geometry optimization of molecules created by linking molecular fragments. First, a set of fragments was selected based on an empirical analysis of an in-house database of 3.3 million commercially available compounds. In order to get a first idea of the fragments necessary for the torsion angle library, a random sample ($N$=10,000) of molecules from the database was split into fragments by cleaving all single bonds. The resulting fragment statistics was used to choose 50 fragments for the torsion angle fragment library (see Table 1 and Fig. 6). Purely aliphatic rings were not considered here for the torsion angle library since the ring conformations were handled using Corina [25] and the torsion angle for fragments attached to such a ring modeled by a torsion angle of a corresponding acyclic fragment. For example, the torsion angles of a isopropyl fragment attached to hexane were determined from the torsion angle library entries for a isopropyl–isopropyl combination. Some torsion angles lead to steric clashes for certain ring conformations, but these cases could be removed by an internal clash test routine for the ligand. Furthermore, the fragments were linked by formation of a single bond between two fragments resulting in a set of 1,225 generated molecules. A set of 24 conformations for each of the molecules was generated by rotating the torsion at the newly formed single bond in steps of 15°. The resulting set of 29,400 conformations was optimized using VAMP [29] and the AM1 Hamiltonian. This procedure identified local minima in the potential energy surface for the rotation around the single bond. The minima for each molecule were mapped onto a 15° grid with a ±5° flexibility window (±10° and ±20° for C=C–C=C and similar compounds). The atoms along the rotatable single bond and their neighbors within a maximum distance of two bonds were used to create a SMARTS pattern for this particular molecule. According to the number of bonds between an atom and the rotatable bond atom, the SMARTS definition becomes more general. For example for dimethyl-amide, a combination of fragments **15** and **28** (see Fig. 6) is transformed into [#7](–[#6,#7,#8])[#6](–[#6,#7,#8])(=[#8]). The SMARTS pattern, corresponding atom and bond types, heat of formation, and minimum energy torsion angles were stored in the ProPose torsion angle library. In the current version of ProPose the heat of formation is not taken into account during the ligand scoring procedure. This will be implemented in later versions of ProPose.

**Table 1** Fragment count in a random sample ($N$=10,000) extracted from a 3.3 milion molecule database of commercially available compounds. In total, 91,813 fragments were created by cleaving the molecules at single bonds. Purely aliphatic ring systems have been discarded for the fragment library setup

| SMILES code | Count | SMILES code | Count |
|---|---|---|---|
| C | 31760 | C1CCCC1 | 285 |
| c1ccccc1 | 9761 | HN1CCCCC1 | 280 |
| C=O | 8167 | HN1CCN(H)CC1 | 255 |
| NH | 7567 | HN1CCNCC1 | 251 |
| O | 6035 | c1cc2ccccc2cc1 | 244 |
| S | 1884 | C1Oc2ccccc2O1 | 237 |
| O=S=O | 1660 | C1COCC1 | 196 |
| Clc1ccccc1 | 1550 | Clc1cc(Cl)ccc1 | 182 |
| N | 1276 | HN1CCOCC1 | 180 |
| OH | 1097 | S1C=NN=C1 | 147 |
| Fc1ccccc1 | 1026 | C1CC1 | 127 |
| C=N | 725 | c1cc2cccnc2cc1 | 125 |
| C1CCNCC1 | 645 | c1cncnc1 | 120 |
| N(H)H | 624 | Clc1ccccc1Cl | 117 |
| O1C=CC=C1 | 616 | C=C1SC(=S)NC1=O | 115 |
| C1CCCCC1 | 545 | S2C=Nc1ccccc21 | 115 |
| Brc1ccccc1 | 527 | C1COCCN1 | 114 |
| c1ccncc1 | 525 | N(H)(H)H | 101 |
| FC(F)F | 490 | C=CC#N | 98 |
| C=C | 459 | O1C=NN=C1 | 95 |
| S1C=CC=C1 | 431 | C1CNCC1 | 87 |
| S1C=NC=C1 | 428 | c1ccc2ncccc2c1 | 80 |
| HN1c2ccccc2C=CC1=O | 380 | O=C2NC=Nc1ccccc12 | 75 |
| C=S | 332 | C1C2CC3CC1CC(C3)C2 | 72 |
| OC=O | 321 | O=C2NC(=O)c1ccccc12 | 70 |

## Implementation

ProPose is written in ANSI C++ and compiles with Linux gcc 3.2 and Intel compilers on x86 platforms. An in-house library of chemistry related algorithms, libchemistry, was used for input/output of molecules, atom typing, and several other tasks. The SMARTS matcher implemented in libchemistry is based on the subgraph monomorphism algorithm VF2. [30] The SWIG interface compiler [23] is used to wrap the functionality of ProPose to a Ruby shell [31] interface. Therefore, docking and other molecular modeling tasks can be performed by high-level language scripts. For example, ProPose/Ruby scripts in combination with Perl [32] scripts were used to generate the torsion angle library.

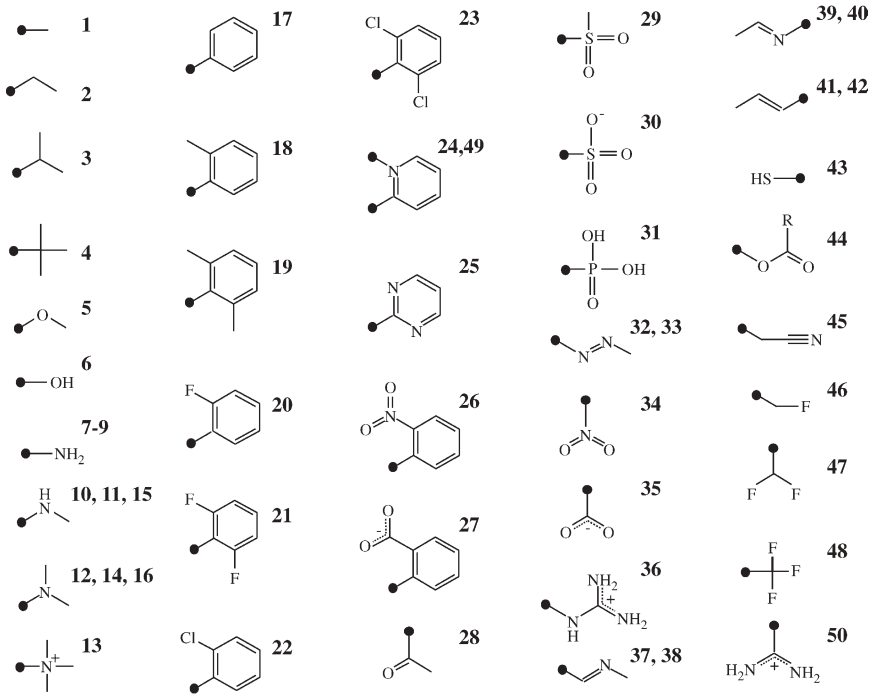


**Fig. 6** Fragments used to generate the torsion angle library. The atom being connected during fragment merge is indicated by a *filled circle*. Optional fragments are marked with an *asterisk*. Where necessary, the Sybyl atom type that was used to create the 3D structure of the fragments is printed in brackets. **1** methyl, **2** ethyl, **3** isopropyl, **4** *tert*-butyl, **5** methoxy, **6** hydroxy, **7** primary amine (N.4), **8** primary amine (N.pl3), **9** primary amine (N.3), **10** secondary amine (N.3), **11** secondary amine (N.4), **12** tertiary amine (N.4), **13** quaternary amine (N.4), **14** tertiary amine (N.pl3), **15** secondary amine (N.pl3), **16** tertiary amine (N.3), **17** benzene, **18** toluene, **19** *m*-xylene, **20** *o*-fluorobenzene, **21** 1,5-difluorobenzene, **22** *o*-chlorobenzene, **23** 1,5-dichlorobenzene, **24** pyridine [C], **25** pyrimidine, **26** *o*-nitro benzene, **27** benzoic acid, **28** methylketone, **29** sulfone, **30** sulfonic acid, **31** phosphate, **32*** azo (*cis*), **33** azo (*trans*), **34** nitro, **35** carboxylate, **36** guanidine, **37*** imine [C] (*cis*), **38** imine [C] (*trans*), **39** imine [N] (trans), **40*** imine [N], **41** alkene (*trans*), **42*** alkene (*cis*), **43** sulfide, **44** ester (R=H), **45*** nitrile, **46** fluoromethyl, **47** difluoromethyl, **48*** trifluoromethyl, **49** pyridine [N], **50** amidine

## Validation

In order to validate ProPose, we created configuration files that defined interaction types and geometries for the empirical approach to scoring derived by Böhm. [19] This, however, is only one possible model for interaction geometries and energies and should not implicate any loss of generality for ProPose's "model-free" approach.

The interaction energies are slightly modified (see Table 2) compared to the original publication of Böhm in order to compensate for implementation specific details, e.g. the clustering of hydrophobic interaction points. Additionally, the ProPose scoring scheme does not only favor attractive interactions, but also penalizes repulsive ones. The interaction geometries were designed to resemble the geometries described for FlexX, [10] which are appropriate in order to use the Böhm approach. The original Böhm scoring function contains a basic energy $\Delta G_0$, a penalty for rotatable bonds $\Delta G_{rot} N_{rot}$, a hydrogen bonding energy $\Delta G_{hb} f_{hb}$, an additional term for ionic hydrogen bonds $\Delta G_{io} f_{io}$, and terms for lipophilic interactions $\Delta G_{lipo} A_{lipo}$ and aromatic interactions $\Delta G_{aro} f_{aro}$, with $N_{rot}$ being the number of rotatable bonds, $f_{hb}$, $f_{io}$, and $f_{aro}$ the respective geometry factors and $A_{lipo}$ the lipophilic surface area.

One major difference of ProPose, especially when compared to FlexX, is the introduction of interaction point weights and the Lorentzian averaging to model the angular and radial dependencies of hydrogen bonds and aromatic interactions. This completely replaces the geometric factors $f_{hb}$, $f_{io}$, and $f_{aro}$ of the original formula. In the ProPose definition of Böhm's model, salt bridges are regarded as ordinary hydrogen bonds (h_don – lone_pair, $\Delta G$=–4.7) with an additive term for the electrostatic interaction (pos_charge–neg_charge interaction, $\Delta G$=–3.0), resulting in total maximum interaction energy of –7.7 that is comparable to Böhm's value of –8.3. Aromatic atoms carry an extra hydrophobic interaction, therefore the aromatic interaction energy had to be reduced to –0.3 (Böhm's original value –0.7). For $\Delta G_0$ (–5.4), $\Delta G_{rot}$ (–1.4), and $\Delta G_{lipo} A_{lipo}$ ("close" contacts in Table 2), the values and methodology used in FlexX have been adapted. [10, 11, 12, 13] However, to avoid unreasonable docking results, penalties for repulsive interaction have been added to the scoring function: a repulsion for close contacts of lone pairs, hydrogen bond donors, and charges of the same sign (see Table 2). Clashes between atoms are penalized with a certain energy ("clash" energy in Table 2). Further tuning of the docking results can be achieved by setting a non-zero value for the penalties for unsaturated polar and/or apolar interactions (default: 0.0).

In order to evaluate the performance of ProPose the CCDC-Astex Gold test set, [33] comprising 293 non-covalent protein–ligand complexes, was docked using ProPose. The Gold test set was used as a reference to allow an objective evaluation of the docking performance without designing a test set specifically optimized for our program. The target definitions were generated automatically according to the binding site specified in the gold.conf file provided with each complex. All water molecules were removed from the binding site and no further receptor-specific optimization was performed. ProPose was configured to use 400 base fragment placements, at least 100 poses, and a maximum of 800 poses during incremental construction. The minimum root mean square deviation (RMSD) of all final solutions with respect to the reference ligand pose was calculated including all corresponding heavy atom

**Table 2** Interaction types and scores for implementing Böhm's approach to scoring in the framework of ProPose. Positive and negative values indicate repulsive and attractive interactions, respectively. These values can be optimized for specific targets. A reasonable range of values is provided for aromatic, hydrophobic, and clash interactions; the default values are given in bold letters. The values of the original Böhm scoring function were slightly adjusted to fit to the novel features of ProPose, for example, the clustering of hydrophobic interaction points. The interaction of organic fluorine is regarded to be mainly hydrophobic since fluorine usually occurs in substituted aromatic rings or, for example, trifluoromethyl (see [47] and Table 1). No unit of the score (such as kJ/mol) is given in order to indicate that the scoring is solely intended to give a ranking of the ligands, but no precise prediction of the binding energy

| Type 1 | Type 2 | Score | Remark |
|---|---|---|---|
| h_don | lone_pair | **–4.7** | Hydrogen bonding |
| pos_charge | neg_charge | **–3.0** | Charge–charge interaction (attractive) |
| pos_charge | lone_pair | **–2.0** | Charge–partial charge interaction (attractive) |
| aro_center | aro_ring | –0.7 **... –0.3** | Aromatic interaction |
| hydrophobic | hydrophobic | –0.2 **... –0.1** | Hydrophobic interaction |
| fluorine | hydrophobic | **–0.11 ...** –0.1 | Hydrophobic interaction |
| close | – | **–0.1 ...** –0.05 | Close contacts |
| polar_unsat | – | **0.0** | Penalty for unsaturated polar interactions |
| apolar_unsat | – | **0.0** | Penalty for unsaturated apolar interactions |
| clash | – | **+0.1 ...** +1.0 | Penalty for steric clash |
| lone_pair | lone_pair | **+0.5** | Partial charge interaction (repulsive) |
| h_don | h_don | **+1.0** | Partial charge interaction (repulsive) |
| pos_charge | pos_charge | **+3.0** | Charge–charge interaction (repulsive) |
| neg_charge | neg_charge | **+3.0** | Charge–charge interaction (repulsive) |

positions. Docking is regarded as successful if the minimum RMSD is below 2.0 Å, as it was suggested by Gohlke et al. [34]

To further explore the performance of ProPose, especially when docking mainly hydrophobic ligands, additional complexes were docked using optimized target descriptions and interaction energies: camphor (PDB 2cpp), adamantane (PDB 4cpp), adamantone (PDB 5cpp), camphane (PDB 6cpp), norchamphor (PDB 7cpp), and thiocamphor (PDB 8cpp) were docked into the active site of cytochrome P450 camphor monooxygenase (CYP450CAM, PDB 7cpp [35]). A stronger clash penalty energy of +1.0, and an increased hydrophobic interaction energy of –0.2 were used for docking. Additionally, the position of the tyrosine 87 hydroxyl hydrogen interacting with the carbonyl of norcamphor was corrected in order to form a hydrogen bond with the carbonyl. The hydrophobic interaction points were clustered according to the pocket mode scheme. A second test set to evaluate the docking of hydrophobic ligands contained retinol (PDB 1rbp), axerophthene (PDB 1fen), and retinoic acid (PDB 1cbs,1epb). These ligands were docked into the active site of retinol binding protein (PDB 1rbp).

Screening runs against two targets selected from the Gold test set were performed using ProPose. The PDB structures 1cx2 (COX2/SC-558 [36]) and 1uvt (thrombin/BM14.1248 [37]) were chosen as targets. A random set of 9995 molecules from our database of 3.3 million commercially available compounds (selection criteria: 0...12 aromatic atoms, 0...5 h-bond donors, 2...14 rotatable bonds, –3...+6 *S* log *P*, 20...200 Å$^2$ topological polar surface area (TPSA), 250...600 D weight, and no reactive groups) was seeded with the co-crystallized ligand in the X-ray structure of the respective target. The active site was defined automatically as described before. All molecules were docked with the same parameters as in the validation docking run. The same set of random molecules was used for both screening runs. Only the top scoring pose of each molecule was considered in these screening runs.

Finally, a set of 39 known epoxysuccinyl inhibitors of the cysteine protease cathepsin B was docked using ProPose. [38, 39, 40, 41] These inhibitors bind covalently to the cysteine residue in the active site of cathepsin B. The irreversible inactivation of cysteine proteases by an active-site directed covalent inhibitor usually proceeds via the rapid formation of a reversible enzyme–inhibitor complex before the transition state of the $S_N2$ nucleophilic reaction is formed. [40] Therefore, in silico screening for active-site directed irreversible inhibitors seems to be feasible using docking methods. However, an increase in docking efficiency is expected if the specific characteristics of the cysteine thiolate interaction with the ligand can be incorporated in the docking protocol. A new interaction was designed to model the nucleophilic attack of the thiolate to the epoxide moiety of the inhibitor (see Fig. 3d). Scoring of these interactions, of course, is difficult, but an empirical adjustment of the associated interaction energy with respect to the other interaction energies allows for first experiments. More specifically, an energy score between –2.0 and –3.0 leads to a better positioning of the epoxide moiety without overruling strong interactions such as hydrogens bonds. This value can be rationalized as a $\delta^+$–lone pair interaction similar to a positive charge–lone pair interaction defined in Table 2. A more reliable parameterization of this interaction will be obtained by future VAMP calculations. The hydrophobic interaction points were clustered according to the surface mode scheme.

## Results and discussion

### Torsion angle library

The AM1 calculations not only resulted in geometry-optimized minima for the various conformations of all examined molecules but additionally identified chemically unreasonable combinations of fragments by dissociating them. These "molecules" were then removed from the torsion angle library. It is well known that the geometry of amides may cause problems in AM1 calculations. However, our calculations reproduced the planar geometry of the amide group quite well (e.g. dimethyl amide, a combination of fragments **15** and **28**, see Fig. 6). Starting with 24 equidistant torsion angles in the range of –180° ... +180°, the VAMP geometry optimization resulted in the following minimum energy torsion angles: 1.2°, 4.9°, 6.7°, 7.7°, 8.4°, 6.4°, 6.8°, 6.5°, 178.2°, 179.0°, –179.7°, –180.0°, –179.5°, –179.2°, –178.4°, –178.7°, –178.7°, –179.2°, –179.3°, –179.7°, –6.5°, –6.2°, –6.4°, and –4.5°. These torsion angles were mapped onto 0°, ±5°, 10°, 15°, 20°, ±175°, and 180°, which correspond to the *cis* and *trans* conformation with a limited flexibility. A tendency to deviate from planarity is observable for the *cis* con-

formation, which reflected the interaction of the methyl groups (shortest distance of the hydrogens 2.39 Å). In general, the results clearly demonstrated the applicability of the AM1/VAMP calculations for generating a torsion angle library.

## Effective potentials

ProPose transforms the interaction points defined in the TDF into an effective potential for scoring of ligand poses. Usually, the number of receptor interaction points in the active sites is approximately $10^3$. This number is considerably larger than for example the number of 100 ... 150 cited for SLIDE, [17] providing a much more detailed representation of the active site. Despite the large number of interaction points the time needed for docking a single ligand is usually in the order of seconds to minutes on a present day personal computer. For example, the single fragment ligand molecules from, e.g., PDB 1c1e or 1hdy were docked in 4.7 s on a 1.5-GHz x86 computer. The mean time per fragment (including base fragment placement) was approximately 2 min for the 198 successfully docked molecules of the Gold test set using 800 clustered poses per iteration while evaluating on average 6×800= 4,800 ligand poses per iteration. However, this value for the mean time has to be interpreted with care: depending on the conformational flexibility of the fragment, the number of ligand poses per iteration reached values $>10^4$ for some ligands, thereby shifting the mean time to the cited value. Usually the times are much shorter: for example, the 10 fragments of the ligand from PDB 1hsb are docked within 6.6 min, i.e. the mean time per fragment is 40 s. Due to the integration of ProPose into the intelligent database screening method "4SCan" [4] running on a Linux cluster system, this time scale does not impose a restriction to small molecule screening applications.

Figure 4 shows the smooth potential energy functions resulting from the Lorentzian averaging of the discrete interaction points. The radial potentials for hydrogen bonding, hydrophobic and aromatic interactions exhibit a shape similar to a Lennard-Jones potential with a minimum at 1.6 Å, 2.7 Å, and 3.2 Å, respectively. The angular potential for hydrogen bonding has a broad minimum in the range of ±30° and a full width at half minimum (FWHM) of 150°. In order to reflect the uncertainties inherent in X-ray structures used for docking, these potentials are geometrically less restrictive than potentials of established force fields.

## Validation

The CCDC-Astex test set—comprising 293 non-covalent protein–ligand complexes—was docked using ProPose. 198 (67.6%) of the non-covalent complexes had an RMSD≤2 Å (see Table 3). 235 (80.2%) docked complexes reproduced the reference pose with a RMSD≤3 Å. For 20 (6.8%) complexes no solution could be found by ProPose; most of these complexes include peptide ligands (e.g. 1rne), larger cyclic molecules (e.g. 1b6n, 1fki), or sugars (e.g. 1byb, 1cdg, 1ppi) with a large number of rotatable bonds. This may be attributed to the incremental construction method in general. [12] Examples of successfully docked complexes are shown in Fig. 7. The binding mode of benzamidine in the active site of thrombin is reproduced very accurately (0.18 Å). This is not unexpected due to the small size of the ligand and its well defined hydrophilic interactions with the receptor. Ligands containing more rotatable bonds are docked with a slight loss of accuracy but the binding mode is well reproduced (human endothelial nitric oxide synthase/ *N*-omega-hydroxy-L-arginine: 0.30 Å, cyclooxygenase-2 (COX2)/arachidonic acid: 0.49 Å, protein tyrosine phosphatase 1B/6-(oxalyl-amino)-1*H*-indole-5-carboxylic acid: 0.59 Å, streptavidin/biotin: 0.85 Å). *N*-Phosphoryl-L-leucinamide docked to thermolysin shows a rotation of the isopropyl group compared to the reference binding mode, leading to a larger rms deviation (1.00 Å). Similarly, the *tert*-butyl phenol moiety of 3'-*tert*-butyl-haba docked into streptavidin is rotated by 180° (RMSD 1.89 Å). The base fragment, however, is positioned correctly. In the docked complex of transketolase with 3'-deazo-thiamin diphosphate a rotation of the thiazol ring and a slightly shifted base fragment are observable (RMSD 1.36 Å). These examples in conjunction with Table 3 demonstrate that ProPose produces reasonable binding modes for most of the complexes within the test set.

A variety of papers in the literature analyze the performance of docking programs. For example, Erickson et al. recently summarized the performance of known programs like Dock, FlexX, and Gold, resulting in success rates between 50 and 70% [42] for a test set of 41 protein–ligand complexes. Similarly, Kontoyianni et al. evaluated FlexX (38% success rate), Dock (6%), Gold (68%), LigandFit (35%) and Glide (57%) on 69 targets. [3] Kramer et al. reproduced 46.5% of 200 complexes within 2 Å RMSD (rank 1 solution only) using FlexX. [12] This rate rose to 70% when looking at the entire solution set. Nissink et al. reported that Gold is able to dock between 65 and 85% of the Gold test set ligands with less than 16 rotatable bonds including the covalent complexes which have been neglected in our study. [33] Schulz-Gasch et al. examined the enrichment rates for various targets using FRED, Glide, and FlexX, [2] which are, however, not directly comparable to our result due to the different validation methodology. It is obvious that all these values are hardly comparable since they depend on the targets, the parameters used for docking, the classification scheme for docking poses, and the effort put in optimizing the docking runs. Consequently, we decided to follow a more objective approach using a test set not specifically compiled for our program, and an automated active site definition without further optimization. Under these conditions ProPose achieved a success rate of 68% (≤2 Å minimum RMSD). This value may be compared to the 70% achieved by FlexX on a substantially smaller set of

**Table 3** Results of docking the Gold test set. The PDB code, the corresponding minimum heavy atom RMSD (Å) of docking and reference pose, and the rank of the minimum RMSD pose within a set of 800 final poses are shown. The list is sorted according to the RMSD. A dash indicates ligands that were not docked by ProPose using the standard protocol

| PDB | rmsd | # | PDB | rmsd | # | PDB | rmsd | # | PDB | rmsd | # | PDB | rmsd | # | PDB | rmsd | # |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1dwb | 0.18 | 28 | 4fab | 0.65 | 4 | 4fbp | 0.98 | 133 | 1srh | 1.46 | 7 | 1glq | 1.99 | 5 | 6cpa | 3.26 | 13 |
| 1mbi | 0.20 | 37 | 4cts | 0.67 | 273 | 1aqw | 0.98 | 76 | 1dg5 | 1.46 | 350 | 1lmo | 2.00 | 664 | 1poc | 3.51 | 28 |
| 1a28 | 0.21 | 62 | 1cle | 0.69 | 267 | 2ifb | 1.00 | 214 | 1srj | 1.48 | 11 | 1cil | 2.02 | 505 | 1ctr | 3.54 | 30 |
| 3nos | 0.30 | 18 | 2ypi | 0.69 | 38 | 2tmn | 1.00 | 17 | 1uvs | 1.48 | 46 | 1etr | 2.03 | 105 | 1sln | 3.60 | 31 |
| 1qh7 | 0.35 | 3 | 1wap | 0.69 | 3 | 1imb | 1.01 | 411 | 1xid | 1.48 | 11 | 1ydt | 2.03 | 164 | 1kel | 3.66 | 164 |
| 1mdr | 0.36 | 17 | 1ivd | 0.70 | 22 | 1lst | 1.01 | 162 | 1azm | 1.49 | 117 | 1hfc | 2.04 | 35 | 1eed | 3.75 | 146 |
| 1lyl | 0.37 | 260 | 1dbj | 0.71 | 10 | 1ivb | 1.02 | 89 | 1gpy | 1.50 | 6 | 1dwc | 2.05 | 65 | 1elc | 3.76 | 283 |
| 2ack | 0.37 | 2 | 1did | 0.72 | 316 | 1lah | 1.03 | 427 | 1rt2 | 1.51 | 2 | 1qpe | 2.06 | 13 | 4cox | 3.79 | 27 |
| 1coy | 0.39 | 8 | 2pk4 | 0.72 | 6 | 1a6w | 1.04 | 3 | 1hri | 1.51 | 15 | 1lic | 2.10 | 106 | 1bma | 3.92 | 17 |
| 1cps | 0.40 | 267 | 6abp | 0.73 | 13 | 1lna | 1.04 | 100 | 1mtw | 1.52 | 1 | 1a1b | 2.10 | 483 | 3ert | 4.01 | 181 |
| 1ngp | 0.41 | 1 | 1xkb | 0.73 | 3 | 1jap | 1.05 | 9 | 1frp | 1.53 | 81 | 1bbp | 2.13 | 55 | 2plv | 4.12 | 55 |
| 1ctt | 0.41 | 5 | 1tni | 0.75 | 86 | 1srg | 1.05 | 207 | 1baf | 1.53 | 20 | 1lkk | 2.14 | 445 | 1a4 k | 4.36 | 278 |
| 3mth | 0.42 | 64 | 1b9v | 0.75 | 50 | 1a4 g | 1.06 | 93 | 1ppc | 1.53 | 3 | 1hdc | 2.15 | 120 | 1hvr | 4.45 | 181 |
| 1fgi | 0.42 | 16 | 1cx2 | 0.75 | 1 | 6rnt | 1.07 | 65 | 1vrh | 1.54 | 267 | 1a9u | 2.15 | 20 | 1ake | 4.62 | 456 |
| 1c5c | 0.43 | 26 | 2r04 | 0.75 | 180 | 1ive | 1.08 | 9 | 1pso | 1.55 | 122 | 1mcr | 2.21 | 153 | 1qbr | 5.00 | 27 |
| 1aoe | 0.44 | 22 | 1xie | 0.76 | 208 | 1mcq | 1.09 | 93 | 2cmd | 1.56 | 192 | 1ppl | 2.34 | 191 | 1hpv | 5.10 | 68 |
| 1tyl | 0.46 | 2 | 1ydr | 0.77 | 22 | 7tim | 1.09 | 223 | 1jao | 1.57 | 13 | 1slt | 2.35 | 152 | 4lbd | 5.25 | 1 |
| 2pcp | 0.46 | 30 | 1acl | 0.78 | 109 | 1a0q | 1.10 | 101 | 1die | 1.58 | 330 | 1snc | 2.46 | 83 | 1hef | 5.28 | 363 |
| 2ada | 0.47 | 1 | 1com | 0.79 | 4 | 1tph | 1.12 | 279 | 2h4n | 1.60 | 40 | 1ets | 2.47 | 255 | 2er7 | 5.37 | 3 |
| 1hdy | 0.48 | 4 | 1fig | 0.80 | 51 | 5p2p | 1.12 | 2 | 1ele | 1.61 | 38 | 3cla | 2.54 | 586 | 5er1 | 5.41 | 127 |
| 1cvu | 0.49 | 75 | 1c5x | 0.81 | 142 | 1yds | 1.13 | 85 | 1aha | 1.61 | 45 | 2tsc | 2.54 | 35 | 1ivq | 5.61 | 211 |
| 4aah | 0.50 | 81 | 1b58 | 0.81 | 256 | 1pph | 1.14 | 49 | 1mld | 1.62 | 135 | 4er2 | 2.59 | 15 | 1fl3 | 5.88 | 6 |
| 3erd | 0.50 | 137 | 1acj | 0.82 | 41 | 1f0r | 1.14 | 106 | 1pgp | 1.62 | 519 | 25c8 | 2.59 | 199 | 1qbt | 5.94 | 7 |
| 1tnl | 0.50 | 6 | 1pdz | 0.82 | 446 | 1mts | 1.16 | 16 | 1a42 | 1.62 | 269 | 1phg | 2.59 | 17 | 1qbu | 6.01 | 289 |
| 1hti | 0.52 | 83 | 2aad | 0.82 | 3 | 7cpa | 1.17 | 79 | 2sim | 1.63 | 346 | 1elb | 2.67 | 75 | 1hos | 6.56 | 113 |
| 1c1e | 0.53 | 35 | 2fox | 0.82 | 8 | 1nis | 1.18 | 137 | 1icn | 1.64 | 120 | 1eap | 2.71 | 76 | 1dmp | 7.50 | 1 |
| 1tnh | 0.53 | 79 | 1dbm | 0.83 | 11 | 1cbx | 1.19 | 95 | 4tpi | 1.69 | 5 | 1aaq | 2.72 | 141 | 2mip | 8.05 | 57 |
| 1abe | 0.54 | 322 | 1flr | 0.83 | 28 | 1byg | 1.19 | 2 | 2dbl | 1.69 | 4 | 1apt | 2.73 | 178 | 1hiv | 8.20 | 5 |
| 1d4p | 0.54 | 12 | 2qwk | 0.85 | 200 | 1d0 1 | 1.19 | 50 | 1ukz | 1.69 | 34 | 1yee | 2.74 | 16 | 1b6n | – | – |
| 1mrg | 0.54 | 11 | 1stp | 0.85 | 14 | 1ckp | 1.20 | 7 | 1lyb | 1.73 | 118 | 1mmq | 2.74 | 461 | 1byb | – | – |
| 1pbd | 0.55 | 6 | 1hsl | 0.85 | 6 | 1acm | 1.21 | 17 | 1bl7 | 1.74 | 55 | 1dwd | 2.77 | 51 | 1cdg | – | – |
| 1ejn | 0.56 | 17 | 1cin | 0.86 | 59 | 1trk | 1.23 | 76 | 1fbl | 1.76 | 230 | 1eld | 2.81 | 124 | 1ei1 | – | – |
| 1c12 | 0.57 | 75 | 2cht | 0.87 | 2 | 2mcp | 1.23 | 251 | 1dog | 1.77 | 11 | 1ela | 2.82 | 29 | 1eoc | –[b] | – |
| 2ctc | 0.57 | 68 | 1lcp | 0.88 | 454 | 1epb | 1.25 | 30 | 1phd | 1.78 | 165 | 4dfr | 2.83 | 14 | 1fki | – | – |
| 1ibg | 0.58 | 6 | 1ptv | 0.88 | 31 | 1ai5 | 1.26 | 74 | 1qcf | 1.79 | 35 | 1apu | 2.86 | 202 | 1hak | – | – |
| 1ebg | 0.58 | 126 | 3tpi | 0.89 | 8 | 1okl | 1.26 | 271 | 1etz | 1.79 | 22 | 1c2 t | 2.88 | 86 | 1ida | – | – |
| 1c83 | 0.59 | 1 | 1aco | 0.90 | 3 | 1fkg | 1.27 | 177 | 1dbb | 1.80 | 61 | 1dd7 | 2.90 | 565 | 1nco | – | – |
| 1abf | 0.59 | 128 | 1kno | 0.91 | 192 | 1f0 s | 1.30 | 21 | 1ghb | 1.82 | 55 | 2r07 | 2.90 | 111 | 1nsd | – | – |
| 1fen | 0.59 | 35 | 1a4q | 0.92 | 141 | 3cpa | 1.33 | 389 | 1mnc | 1.82 | 129 | 8gch | 2.96 | 63 | 1pha | – | – |
| 1ml1 | 0.59 | 47 | 1uvt | 0.93 | 3 | 5abp | 1.33 | 20 | 1dhf | 1.84 | 313 | 1epo | 3.03 | 154 | 1phf | – | – |
| 1hsb | 0.60 | 9 | 1tng | 0.93 | 145 | 1ql7 | 1.34 | 15 | 1a1e | 1.84 | 393 | 1a07 | 3.04 | 286 | 1ppi | – | – |
| 2lgs | 0.60 | 2 | 1hyt | 0.93 | 49 | 1tka | 1.36 | 12 | 1cqp | 1.85 | 2 | 1igj | 3.07 | 69 | 1rds | – | – |
| 1ulb | 0.61 | 9 | 6rsa | 0.93 | 65 | 2yhx | 1.36 | 40 | 1cbs | 1.87 | 24 | 1glp | 3.08 | 493 | 1rne | – | – |
| 1f3d | 0.61 | 3 | 1rnt | 0.94 | 172 | 1fax | 1.37 | 1 | 1ett | 1.88 | 7 | 1mmb | 3.12 | 597 | 2cpp | –[a] | – |
| 1ldm | 0.61 | 10 | 3gpb | 0.94 | 13 | 1mrk | 1.41 | 72 | 1tdb | 1.88 | 178 | 2cgr | 3.12 | 36 | 3hvt | – | – |
| 1d3 h | 0.62 | 423 | 1aj7 | 0.95 | 3 | 1rob | 1.42 | 354 | 1srf | 1.89 | 4 | 1cf8 | 3.14 | 394 | 3pgh | – | – |
| 3ptb | 0.62 | 79 | 1dr1 | 0.95 | 60 | 1okm | 1.42 | 33 | 2ak3 | 1.92 | 148 | 1tmn | 3.15 | 3 | 4phv | – | – |
| 2phh | 0.63 | 116 | 1ivc | 0.95 | 157 | 1mup | 1.44 | 411 | 1atl | 1.93 | 38 | 1htf | 3.22 | 62 | 5cpp | –[a] | – |
| 1rbp | 0.65 | 219 | 2gbp | 0.97 | 1 | 1qpq | 1.45 | 225 | 1tlp | 1.95 | 312 | 1eta | 3.25 | 254 | | | |

[a] These ligands can be docked by using a reduced value for the minimum base fragment energy (see text)

[b] The standard protocol used for docking does not contain an interaction for a Fe(III) ion, therefore docking 4-nitrocatechol into protocatechuate 3,4-dioxygenase failed

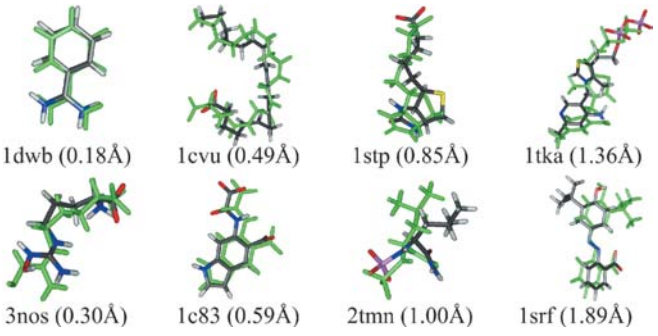


**Fig. 7** Examples of ligand poses and crystal structures. A diverse set of docked complexes is shown with RSMDs ranging from 0.18 Å to 1.89 Å: 1dwb (thrombin/benzamidine), 3nos (human endothelial nitric oxide synthase/*N*-omega-hydroxy-L-arginine), 1cvu (cox-2/arachidonic acid), 1c83 (protein tyrosine phosphatase 1B/6-(oxalyl-amino)-1H-indole-5-carboxylic acid), 1stp (streptavidin/biotin), 2tmn (thermolysin/*N*-phosphoryl-L-leucinamide), 1tka (transketolase/3'-deazo-thiamin diphosphate), and 1srf (streptavidin/3'-*tert*-butyl-haba). The reference binding modes are shown in *green*. The RMSD is specified in brackets (Å)

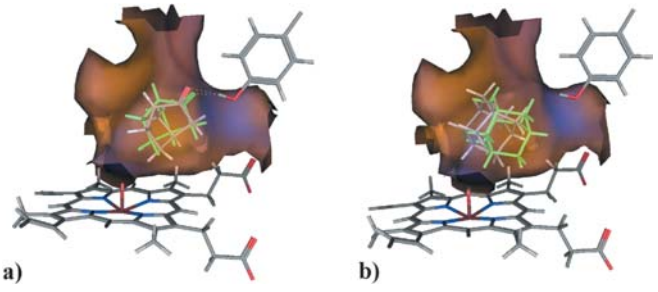


**Fig. 8a, b** Comparison of ligand poses and crystal structures in Cytochrome P450 camphor monooxygenase. Norcamphor (a) and adamantane (b) were docked with an RMSD of 0.32 Å and 0.49 Å, respectively. The position of the ligand poses relative to the heme and tyrosine 87 of Cyp450 CAM are shown. The reference binding mode is depicted in green. The active site is outlined by its inner surface. *Orange* and *blue* indicate hydrophobic and hydrophilic interactions, respectively. The hydrogen bond connecting the hydroxyl group of tyrosine 87 with the carbonyl of norchamphor is indicated as a *dotted line*

complexes but with a similar setup. However, our result has to be regarded as a lower limit since no receptor specific optimization was performed. Additionally, Böhm's approach to scoring, as used in these tests, is rather generic. Using a more sophisticated scheme, one can expect a substantial improvement of this success rate. Therefore, ProPose is certainly competitive with the success rate of other docking programs.

To explore the performance when docking mainly hydrophobic ligands, some more complexes were studied using optimized parameters: first, camphor (minimum RMSD 0.18 Å), norchamphor (0.32 Å), adamantane (0.49 Å), adamantone (0.32 Å), camphane (0.27 Å), and thiocamphor (0.54 Å) were successfully docked into the active site of cytochrome P450 camphor monooxygenase (CAM) (see Fig. 8). This was achieved using only minor modifications for clash energies, minimum base fragment energy and hydrophobic interaction energies. Due to the rotational symmetry of the ligands some of the top scoring poses show higher RMSD values; however, nearly all of the ligand poses are located at the correct position within the active site. Hence, ProPose is clearly able to handle partially or even purely hydrophobic ligands. This is underlined by successfully docking a retinol analog into the active site of retinol binding protein (see Fig. 9) using this protocol. Remarkably, all ligands including the purely hydrophobic axerophthene are docked by ProPose without preselecting the base fragment. FlexX, in contrast, uses a hierarchical interaction scheme, where the strongest (i.e. hydrogen bonding and aromatic) interactions are utilized for base fragment placement due to the possibly large number of geometrically less restrictive hydrophobic interaction points. This method, however, discriminates fragments with mainly weak hydrophobic interactions.

The screening for COX2 (PDB 1cx2) and thrombin (PDB 1uvt) inhibitors shows that ProPose is able to correctly rank molecules which are similar to the native ligands within the top scoring docked molecules (Fig. 10). The distribution of scores approximately resembles a Gaussian function, as is expected for a random set of molecules. For COX2, 6874 molecules were docked successfully into the active site, with the native ligand being ranked at number 16. Among the 20 top scoring molecules, seven carry a sulfonamide moiety, six contain fluorine (either fluorinated aromatic rings or trifluoromethyl groups), two features also present in the native ligand. An example is given Fig. 10a, molecule **2** (#5 of the top scoring molecules): the binding mode is similar to the native ligand with a sulfur–sulfur distance of 1.57 Å and a F–$CF_3$ distance of 1.66 Å. In the case of thrombin, ProPose docked 8,431 molecules into the active site with the native 4-amino-pyridine ligand found at rank #208, i.e. within the first 3% of the docked molecules. The example shown in Fig. 10b, molecule **2**, fills exactly the same cavity as the native ligand. It, however, lacks an interaction comparable with the amino-pyridine–carboxylate of Asp189 interaction in the native ligand. The first molecule forming a salt bridge with Asp189 is ranked at #100, and the first 4-amino-pyridine can be found at rank #143. In order to further illustrate the results some known inhibitors of COX2 and thrombin have been docked and the score is shown in Fig. 10. The COX2 inhibitors diclofenac ($IC_{50}$=1.17–8.9 nM in a cell assay [43]) and flurbiprofen ($IC_{50}\approx$1nM in a cell assay [44]) exhibit mid-range scores, −34.5 and −34.6 respectively, which indicates the preference of this docking model for the native ligand and similar molecules. In the case of thrombin, the lactam derivative of $N^{\alpha}$-($\beta$-naphtylsulfonyl-glycyl)-*p*-amidino-phenylalanyl-piperidid (NAPAP) [45] and benzamidine achieve scores of −39.5 and −18.5, respectively. The NAPAP derivative is known to bind to thrombin with an $IC_{50}$ of 1.6 nM, in contrast to $K_i$=23 nM for the native ligand of 1uvt. These two molecules are scored just with the opposite ranking, but both are on the top scoring slope of the score distribution. As expected, a relatively weak micromolar ligand such as benzamidine (PDB 1dwb) is

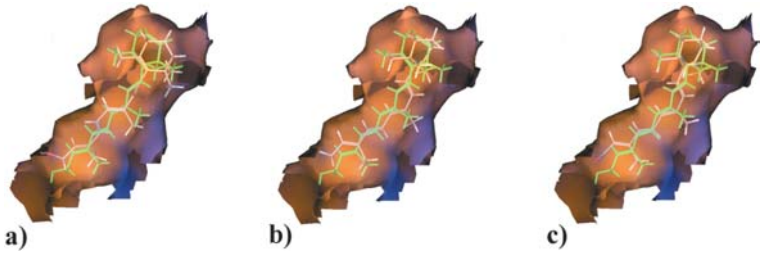


**Fig. 9a–c** Comparison of ligand poses and crystal structures in retinol binding protein. Retinol (a), axerophthene (b), and retinoic acid (c) were docked into retinol binding protein. The reference ligand (retinol extracted from crystal structure PDB 1rbp) is shown in *green*. The active site is outlined by its inner surface. *Orange* and *blue* indicate hydrophobic and hydrophilic interactions, respectively. Remarkably, all ligands including the purely hydrophobic axerophthene are docked by ProPose without preselecting the base fragment

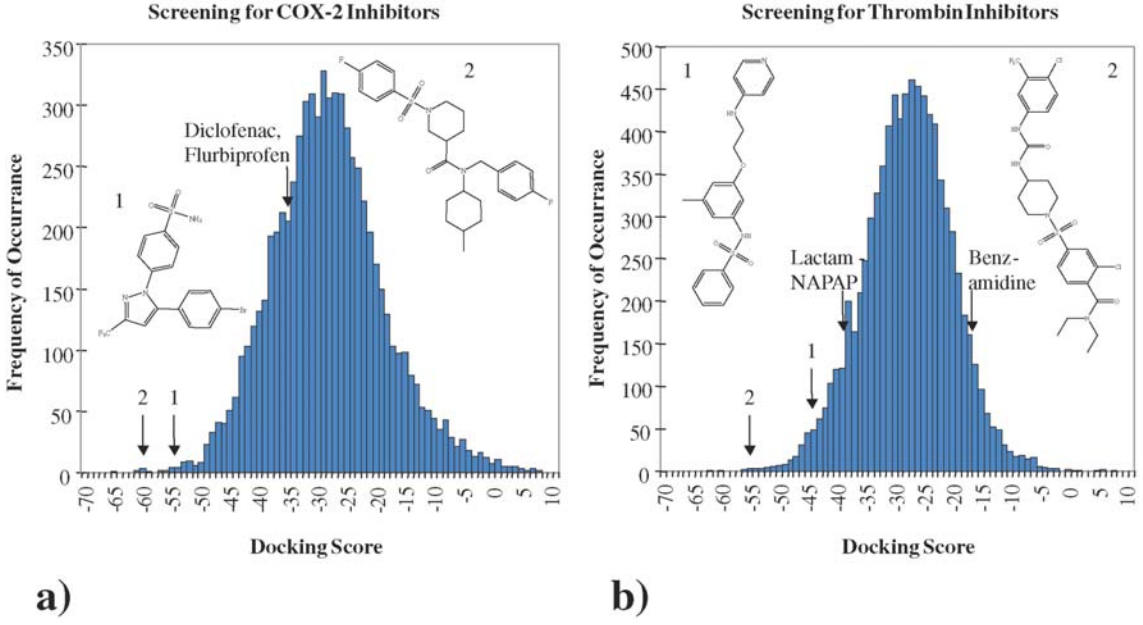


**Fig. 10a, b** Screening results for cyclooxygenase-2 (**A**, PDB 1cx2) and thrombin (**B**, PDB 1uvt). A set of ~10,000 random molecules was screened and the distribution of docking scores is shown. The native ligands (**1**) are among the top scoring molecules for both targets. The chemical structure and the score of another top scoring molecule is marked by (**2**). Additionally, diclofenac (COX2), flurbiprofen (COX2), a lactam derivative of NAPAP (thrombin), and benzamidine (thrombin) have been docked and the respective score is indicated by an *arrow*

well separated on the low scoring slope of the score distribution. The results clearly indicate that ProPose is applicable for screening. However, it is evident that docking molecules into a rigid active site, where the co-crystallized ligand was removed, will prefer molecules similar to the native ligand. Additionally, a relatively simple scoring function, as used in this test setup, cannot explain all important details of protein–ligand binding. This again underlines the necessity for advanced scoring methods, for which the docking engine ProPose was designed to act as a framework.

Finally, the docking of epoxysuccinyl inhibitors of the cathepsin B protease showed that using the novel interaction type leads to a significantly better positioning of epoxysuccinyl inhibitors inside the active site (see Fig. 11 and Table 4). For 38% of the ligands the introduction of the novel interaction improves the ranking within the entire solution set as well as the RMSD with respect to the reference placement of the oxirane moiety, for 85% either ranking or RMSD improved, and for 15% no improvement could be achieved. This indicates that the novel interaction leads to a higher probability for correctly docking these inhibitors, despite the complete neglect of the structural changes due to the configuration inversion in course of the $S_N2$ reaction. However, it is clear that the most important step in screening for covalent inhibitors is the optimization of the non-covalent molecular recognition in order to avoid or at least reduce selectivity and toxicity problems and biochemical assay artifacts (see [22, 46]). Therefore it is appropriate to model the complex based on non-covalent interactions that are accompanied by an extra "pseudo-covalent" interaction. Further studies, however, are necessary to parameterize the scoring of such an interaction in a more rigorous manner.

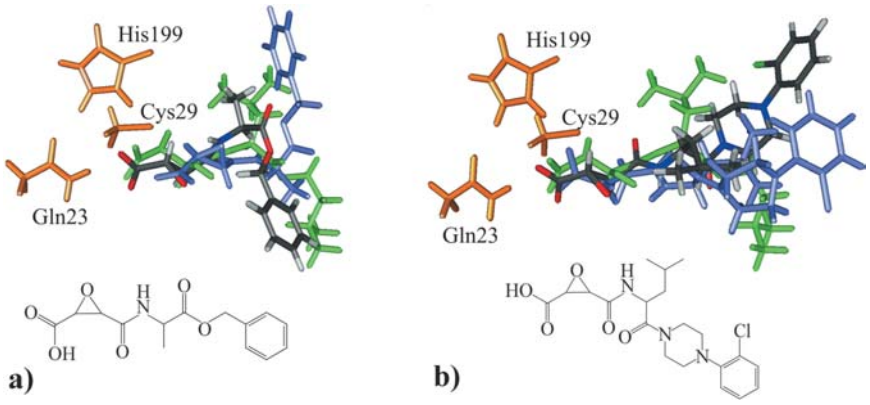


**Fig. 11a, b** Binding mode and docking poses of two epoxysuccinyl inhibitors: The anchor residues of the epoxide moiety are shown in orange: histidine 199, cysteine 29, and glutamine 23 of cathepsin B. The reference ligand (*N*-[1-hydroxycarboxyethylcarbonyl]-leucylamino-2-methyl-butane) crystallized in the PDB structure 1ito is depicted in *green* as a reference. This reference characterizes the ligand placement after the chemical reaction has taken place. The chemical structures of the inhibitors are shown below the corresponding binding mode. Using a specific interaction for the nucleophilic interaction between thiolate and epoxide (ligands in element colors) improves the docked binding modes significantly compared to a standard docking protocol (*blue* ligands). Without the nucleophilic interaction the oxirane ring of ligand A (no. 23 in Table 4) is shifted with respect to the reference, but is correctly positioned when using this interaction. Similarly, the oxirane ring of ligand B (no. 36 in Table 4) is rotated compared to the reference binding mode without the novel interaction

**Table 4** Results of docking the cathepsin B test set with and without the novel interaction which simulates the nucleophilic attack. The rank of the minimum RMSD solution within the entire solution set and the RMSD with respect to a reference placement of the oxirane-2-carboxy moiety (in Å) are shown. (++) indicates a better ranking as well as a smaller RMSD when using the novel interaction, (+ −) a better ranking, but an increased RMSD, (− +) a worse ranking, but a decreased RMSD, and (− −) a worsening of both parameters. (0) indicates no change

| Ligand No. | Rank/w | RMSD/w | Rank/wo | RMSD/wo | Remark |
|---|---|---|---|---|---|
| 1 | 81 | 1.076 | 3 | 1.356 | − + |
| 2 | 32 | 0.877 | 110 | 0.822 | + − |
| 3 | 133 | 0.939 | 71 | 1.434 | − + |
| 4 | 236 | 1.125 | 114 | 0.891 | − − |
| 5 | 63 | 2.117 | 151 | 1.129 | + − |
| 6 | 1 | 3.845 | 1 | 3.845 | 0 0 |
| 7 | 1 | 0.939 | 25 | 1.405 | + + |
| 8 | 41 | 1.553 | 5 | 1.716 | − + |
| 9 | 143 | 1.219 | 479 | 1.218 | + 0 |
| 10 | 3 | 1.092 | 138 | 1.324 | + + |
| 11 | 321 | 3.331 | 321 | 3.331 | 0 0 |
| 12 | 181 | 1.936 | 162 | 1.824 | − − |
| 13 | 1 | 0.979 | 248 | 1.068 | + + |
| 14 | 3 | 0.822 | 82 | 1.356 | + + |
| 15 | 1 | 0.942 | 145 | 1.264 | + + |
| 16 | 10 | 0.940 | 535 | 1.068 | + + |
| 17 | 52 | 0.939 | 5 | 1.405 | − + |
| 18 | 1 | 0.939 | 1 | 1.461 | 0 + |
| 19 | 32 | 0.939 | 338 | 1.555 | + + |
| 20 | 1 | 1.724 | 1 | 0.783 | 0 − |
| 21 | 10 | 0.930 | 168 | 1.207 | + + |
| 22 | 1 | 0.940 | 445 | 1.264 | + + |
| 23 | 1 | 0.940 | 275 | 0.999 | + + |
| 24 | 14 | 0.939 | 1 | 1.460 | − + |
| 25 | 77 | 1.825 | 3 | 1.748 | − − |
| 26 | 24 | 0.964 | 41 | 0.799 | + − |
| 27 | 18 | 2.169 | 101 | 1.174 | + − |
| 28 | 130 | 1.189 | 337 | 1.034 | + − |
| 29 | 1 | 0.939 | 34 | 1.405 | + + |
| 30 | 2 | 0.939 | 65 | 1.166 | + + |
| 31 | 19 | 1.553 | 2 | 1.716 | − + |
| 32 | 54 | 1.553 | 245 | 1.899 | + + |
| 33 | 90 | 1.554 | 479 | 1.499 | + − |
| 34 | 2 | 0.782 | 421 | 1.068 | + + |
| 35 | 16 | 1.554 | 148 | 1.166 | + − |
| 36 | 2 | 0.979 | 168 | 1.389 | + + |
| 37 | 128 | 1.357 | 166 | 1.353 | + − |
| 38 | 226 | 1.328 | 107 | 1.347 | − + |
| 39 | 49 | 0.939 | 128 | 0.783 | + − |

The introduction of a novel interaction type certainly is similar to a standard pharmacophore constraint during docking. However, our methodology is much more flexible: the seamless integration of such "pharmacophores" into both the interaction and scoring scheme, allows us to screen for certain interacting substructures without discarding other potentially high-scoring substructures completely. Therefore this methodology may supplement the screening for irreversible inhibitors with the aid of a technology that is usually focused on the reversible binding of a ligand. It is evident from the literature that scoring functions must be improved for protein–ligand docking. ProPose was explicitly designed to support this venture by its open architecture. Every interaction or scoring model is fully configurable according to user's needs and the user can intervene at any step without having to worry about details of the implementation into the program.

## Summary

ProPose offers a major advantage compared to other docking software: its utmost flexibility for application and development. It does not contain any hard-coded interaction geometries, energies, or substructures. Everything needed for docking as well as for scoring is defined in plain text configuration files; it may thus be called "model-free". The interaction geometries may be derived from experimental geometries, knowledge-based potentials or even quantum chemistry calculations, depending on the respective needs. The principal applicability of this approach has been demonstrated by defining pseudo-covalent interactions that simulate the nucleophilic attack on the epoxide moiety by a cysteine protease. This unified approach to docking and scoring by using a transformation of discrete interaction points to a continuous potential energy function is not limited to protein–ligand docking: the target description file may as well be generated using a reference ligand and an appropriate set of customized interaction geometries. This will turn the subsequent "docking" step effectively into a ligand–ligand alignment procedure simply by using a different configuration file. ProPose thus offers a flexible platform for arbitrary, not only docking-based, queries in molecular databases that can be extended easily to a wider range of virtual screening applications.

**Acknowledgment** We thank Matthias Busemann for providing the cathepsin B ligand data, Kristina Wolf for her innovative ideas about the program name and proof-reading, and Bernhard Schirm and Daniel Vitt for their unwavering support.

## References

1. Gohlke H, Klebe G (2002) Angew Chem Int Ed Engl 41:2644–2676
2. Schulz-Gasch T, Stahl M (2003) J Mol Model 9:47–57
3. Kontoyianni M, McClellan LM, Sokol GS (2004) J Med Chem 47:558–565
4. Seifert MHJ, Wolf K, Vitt D (2003) Biosilico 1:143–149
5. Goodsell DS, Morris GM, Olson AJ (1996) J Mol Recognit 9:1–5
6. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) J Mol Biol 267:727–748
7. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) J Mol Biol 161:269–288
8. McGann M, Almond H, Nicholls A, Grant JA, Brown F (2003) Biopolymers 68:76–90 (http://www.eyesopen.com/products/applications/fred.html)
9. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) J Med Chem 47:1739–1749
10. Rarey M, Wefing S, Lengauer T (1996) J Comput Aid Mol Des 10:41–54
11. Rarey M, Kramer B, Lengauer T (1999) Bioinformatics 15: 243–250
12. Kramer B, Rarey M, Lengauer T (1999) Proteins 37:228–241
13. Hindle SA, Rarey M, Buning C, Lengauer T (2002) J Comput Aid Mol Des 16:129–149
14. Joseph-McCarthy D, Thomas IV BE, Belmarsh M, Moustakas D, Alvarez JC (2003) Proteins 51:172–188
15. Ewing TJA, Kuntz ID (1997) J Comput Chem 18:1175–1189
16. Jain AN (2003) J Med Chem 46:499–511
17. Zavodszky MI, Sanschagrin PC, Korde RS, Kuhn LA (2002) J Comput Aid Mol Des 16:883–902
18. Liu S, Zhang C, Zhou H, Zhou Y (2004) Proteins 56:93–101
19. Böhm HJ (1998) J Comput Aided Mol Des 12:309–323
20. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) J Comput Aided Mol Des 11:425–445
21. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL (2004) J Med Chem 47:3032–3047
22. Lindvall MK (2002) Curr Pharm Design 8:1673–1681
23. Ousterhout JK (1998) IEEE Comput 31:23–30
24. Daylight Chemical Information Systems Inc, 27401 Los Altos, Mission Viejo, CA 92691, USA (http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html)
25. Gasteiger J, Rudolph C, Sadowski J (1990) Tetrahedron Comput Methods 3:537–547
26. Fischer D, Lin SL, Wolfson HL, Nussinov R (1995) J Mol Biol 248:459–477
27. Flower DR (1999) J Mol Graph Model 17:238–244
28. Snarey M, Terrett NK, Willett P, Wilton DJ (1997) J Mol Graph Model 15:372–385
29. Accelrys Inc, 9685 Scranton Road, San Diego, CA 92121–3752, USA (http://www.accelrys.com/products/tsar/vamp.html)
30. Foggia P, Sansone C, Vento M (2001) 3rd IAPR-TC15 Workshop on Graph-based Representations, Ischia
31. Matsumoto Y (2002) Ruby in a Nutshell. O'Reilly & Associates Inc, Cambridge, UK (http://www.ruby-lang.org)
32. Wall L, Christiansen T, Orwant J (2000) Programming Perl, 3rd edn. O'Reilly & Associates Inc, Cambridge, UK (http://www.perl.org)
33. Nissink JW, Murray C, Hartshorn M, Verdonk ML, Cole JC, Taylor R (2002) Proteins 49:457–471 (http://www.ccdc.cam.ac.uk/products/validate/gold.html)
34. Gohlke H, Hendlich M, Klebe G (2000) J Mol Biol 295:337–356
35. Raag R, Poulos TL (1989) Biochemistry 28:917–922
36. Kurumbail RG, Stevens AM, Gierse JK, McDonald JJ, Stegeman RA, Pak JY, Gildehaus D, Miyashiro JM, Penning TD, Seibert K, Isakson PC, Stallings WC Nature (1996) 384:644–648
37. Engh RA, Brandstetter H, Sucher G, Eichinger A, Baumann U, Bode W, Huber R, Poll T, Rudolph R, von der Saal W (1996) Structure 4:1353–1362
38. Gour-Salin B, Lachance P, Magny MC, Plouffe C, Menard R, Storer A (1994) Biochem J 299:389–392
39. Roush WR, Hernandez AA, McKerrow JH, Selzer PM, Hansell E, Engel JC (2000) Tetrahedron 56:9747–9762

40. Powers JC, Asgian JL, Ekici OD, James KE (2002) Chem Rev 102:4639–4750
41. Busemann M (2003) Diploma thesis, University of Würzburg, Germany
42. Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M (2004) J Med Chem 47:45–55
43. Klein T, Nusing RM, Pfeilschifter J, Ullrich V (1994) Biochem Pharmacol 48:1605–1610
44. Riendeau D, Charleson S, Cromlish W, Mancini JA, Wong E, Guay J (1997) Can J Physiol Pharmacol 75:1088–1095
45. Mack H, Pfeiffer T, Hornberger W, Böhm HJ, Hoffken HW (1995) J Enzyme Inhib 9:73–86
46. Rishton GM (2003) Drug Discovery Today 8:86–96
47. Dunitz JD, Taylor R (1997) Chemistry 3:89–98