Journal of Computer-Aided Molecular Design, **14:** 435–448, 2000.
KLUWER/ESCOM

Evaluation of docking/scoring approaches: A comparative study based on MMP3 inhibitors

Sookhee Ha*, Romana Andreani, Arthur Robbins & Ingo Muegge**
Bayer Research Center, 400 Morgan Lane, West Haven, CT 06516, U.S.A.

Received 5 August 1999; Accepted 20 December 1999

Key words: DOCK, FlexX, PMF scoring, protein–ligand binding, stromelysin

Summary

An increasing number of docking/scoring programs are available that use different sampling and scoring algorithms. A reliable scoring function is the crucial element of such approaches. Comparative studies are needed to evaluate their current capabilities. DOCK4 with force field and PMF scoring as well as FlexX were used to evaluate the predictive power of these docking/scoring approaches to identify the correct binding mode of 61 MMP-3 inhibitors in a crystal structure of stromelysin and also to rank them according to their different binding affinities. It was found that DOCK4/PMF scoring performs significantly better than FlexX and DOCK4/FF in both ranking ligands and predicting their binding modes. Most notably, DOCK4/PMF was the only scoring/docking approach that found a significant correlation between binding affinity and predicted score of the docked inhibitors. However, comparing only those cases where the correct binding mode was identified (scoring highest among sampled poses), FlexX showed the best 'fine tuning' (lowest rmsd) in predicted binding modes. The results suggest that not so much the sampling procedure but rather the scoring function is the crucial element of a docking program.

Introduction

The fast and reliable prediction of binding affinities of large numbers of putative ligand–protein complexes is a long-standing and challenging issue in computational chemistry [1]. Solving this problem would enable us to virtually screen large databases and effectively compete with biological screening techniques such as high-throughput screening [2]. Unfortunately, the reliability of docking/scoring procedures that are used today is not sufficient [3]. The main problems lie in the accuracy of scoring functions that must be fast, produce no false negatives and a low rate of false positives even for weak binders in a virtual screening protocol. In addition, docking approaches need to allow for protein flexibility as well as water penetration of the binding site.

In order to evaluate the binding free energy of a putative protein–ligand complex, the ligand must be placed in the binding site of a target protein with known 3D structure. Several algorithms have been presented that facilitate this task (e.g., DOCK [4–6], FlexX [7], Hammerhead [8], GOLD [9], AutoDOCK [10], Flog [11]). They seem to be of similar efficiency and to have solved the so-called sampling problem – the effective sampling of the conformational space of the ligand molecule in the binding site [12]. From here, the challenge is to design a scoring function that is able to find the correct binding mode of the molecule and rank it highest among sampled poses during the docking process. Furthermore, the scoring function has to be able to identify a small number of biologically active molecules in a large molecular database. No scoring function has been reported to date that is able to solve this task with a high level of accuracy. Empirical scoring functions, knowledge-based scoring functions, as well as scoring functions developed from first principles, have been employed in several docking programs [13–22]. The fact that

*Present address: Merck Research Laboratory, WP42-3, West Point, PA 19486, U.S.A.
**To whom correspondence should be addressed. E-mail: ingo.mugge.b@bayer.com

Table 1. Cell parameters of tSL257/biphenyl inhibitor crystals

Inhibitor	a (Å)	b	c	α (°)	β	γ	Space group	Molecule/a.u[a]
A	62.90	47.43	55.43	90.0	98.6	90.0	$P2_1$	2
B	56.79	126.2	42.95	90.0	90.0	90.0	$P2_12_12$	2
C	137.3	137.3	124.7	90.0	90.0	120.0	$P6_522$	3
D	49.02	119.0	63.75	90.0	91.1	90.0	$P2_1$	4
E	56.67	126.3	43.25	90.0	90.0	90.0	$P2_12_12$	2
F	56.26	134.1	42.93	90.0	90.0	90.0	$P2_12_12_1$	2

[a]a.u. is the asymmetric unit of the crystallographic unit cell.

most of the docking programs mentioned above use a different scoring function illustrates that no scoring function seems to be optimal. Empirical scoring functions use multivariate regression methods to fit a training set of protein–ligand complexes to measured binding constants [13, 15, 16, 20, 23]. Force-field scoring functions use non-bonded interaction terms of force fields for calculating a ranking measure [24, 25]. Additional solvation terms are sometimes considered and chemical as well as contact scores are used. Recently, a general and fast, knowledge-based scoring function has been introduced by us. It uses the sum of potentials of mean force (PMF) between protein and ligand atoms, derived from the Brookhaven Protein Data Bank, as a measure for the protein–ligand binding affinity. The PMF scoring has been shown to outperform force-field scoring in finding weak binders for FKBP in a database of building blocks for SAR by NMR [22, 26]. It has also shown improved scoring of diverse sets of known protein–ligand complexes and modeled HIV-1 complexes. In addition, DOCK4/PMF has predicted the correct binding modes of neuraminidase inhibitors as highest scoring binding modes in different crystal structures [27].

A good concept for minimizing the number of false positives in a virtual screen is to combine several scoring functions and take only those compounds that score best in all measures. Therefore, the number of scoring functions available in one particular docking package should be increased. In order to find the most promising scoring functions we wanted to compare here a group of three different scoring functions, each one of the most prominent scoring function classes: force field score, empirical score and knowledge-based score. As a test set we chose the matrix metalloproteinase stromelysin (MMP-3) with a set of 61 biphenyl-type inhibitors.

Stromelysin-1 (MMP-3) is a matrix metalloproteinase that belongs to the family of zinc endopeptidases [28]. Functions of MMPs include proteolytic activity and degradation of connective tissues of articular cartilage or basement membrane. MMPs are secreted by fibroblasts, chondrocytes, and other various cells. In particular, a high level of stromelysin (MMP-3) is found in the synovial fluid and cartilage from rheumatoid and osteoarthritis patients, where proteolytic activity flourishes. MMP-3 is also known as an activator of other MMPs such as MMP-1 and MMP-9, by cleaving their propeptide domains. Therefore, stromelysin contributes to the progressive loss of articular cartilage and joint function [28, 29]. Inhibitors of stromelysin are expected to stop or possibly reverse the process of slow cartilage loss.

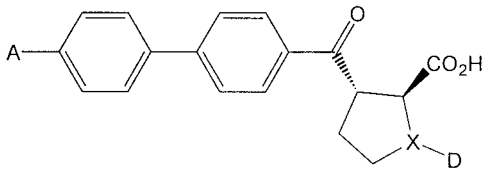
Crystal structures as well as solution structures of MMP-3 were solved [30, 31]. A biphenyl class of inhibitors was developed for a new therapy against osteoarthritis. Crystal structures of the complexes between the catalytic domain of MMP-3 and biphenyl compounds were solved in-house and were used for the docking experiments.

Materials and methods

Crystallization

Recombinant stromelysin, truncated at residues 257 (tSL257) or 284 (tSL284), was cloned and expressed in *E. coli*, and purified by a method similar to that reported by Marcy and co-workers [32]. Complexes of the biphenyl series of inhibitors were formed overnight at 37 °C, the mole ratio of inhibitor to enzyme being 2:1. Crystals of the complexes were grown by the hanging drop vapor diffusion method. The protein concentration was approximately 20 mg/ml, in 5 mM Tris buffer, 5 mM $CaCl_2$, 100 mM NaCl, pH 7.5. Two μl of the protein solution was mixed with an equal volume of polyethylene glycol precip-

Table 2. Root mean square deviations of the biphenyl core portion of 61 stromelysin inhibitors[a]



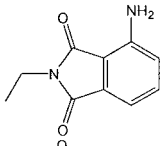
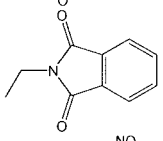
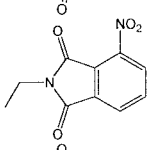
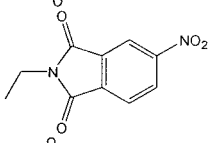
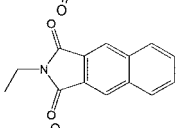
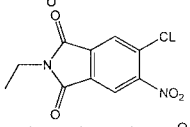
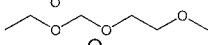
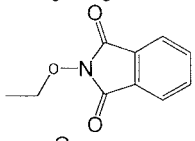
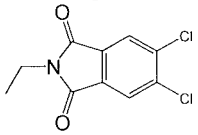
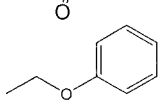
#	A	X	D	pIC_{50}	PMF	FF	FlexX
1	Cl	C		8.60	1.101	0.832	1.769
2	Cl	C		8.37	1.073	3.204	0.786
3	Cl	C		8.26	8.678	7.060	0.747
4	Cl	C		8.08	9.008	9.905	0.786
5	Cl	C		8.03	9.813	8.278	0.752
6	Cl	C		8.00	10.99	8.585	0.493
7	Cl	C		8.00	1.379	2.949	8.410
8	Cl	C		7.82	1.148	2.358	1.093
9	Cl	C		7.72	0.690	9.770	0.786
10	Cl	C		7.66	1.606	2.805	8.074

Table 2. (continued)

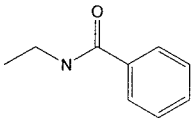
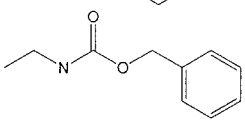
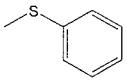
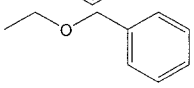
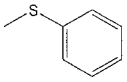
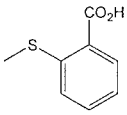
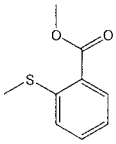
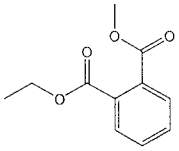
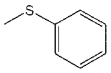
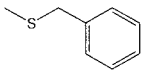
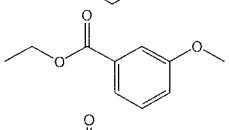
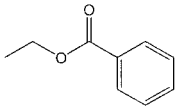
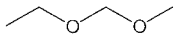
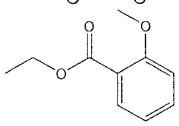
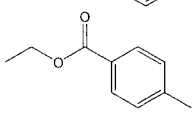
#	A	X	D	pIC_{50}	PMF	FF	FlexX
11	Cl	C		7.62	1.756	3.530	0.731
12	Cl	C		7.62	1.245	2.106	0.878
13	Cl	C		7.59	1.166	0.821	0.711
14	Cl	C		7.57	1.187	4.621	8.004
15	O-Pentyl	C		7.52	2.072	4.140	1.073
16	Cl	C		7.52	1.396	3.068	1.374
17	Cl	C		7.48	1.099	1.754	0.734
18	Cl	C		7.44	1.148	1.726	0.940
19	O-Ethyl	C		7.44	1.106	2.782	1.049
20	Cl	C		7.42	2.291	3.190	0.854
21	Cl	C		7.38	1.413	3.149	0.716
22	Cl	C		7.36	1.195	1.820	4.379
23	Cl	C		7.34	1.312	1.032	7.987
24	Cl	C		7.34	1.437	1.062	0.726
25	Cl	C		7.34	1.349	0.974	0.721

Table 2. (continued)

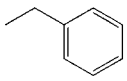
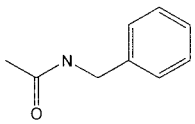
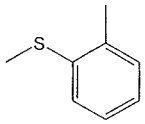
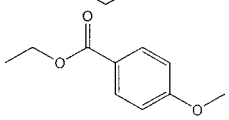
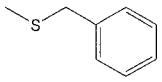
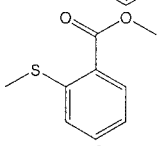
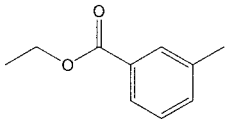
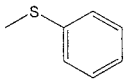
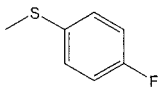
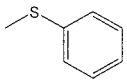
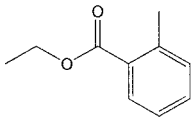
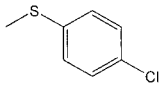
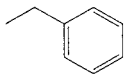
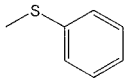
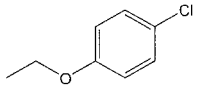
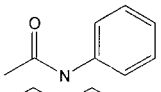
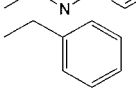
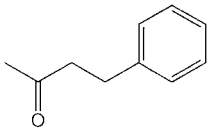
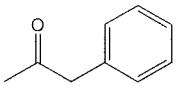
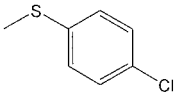
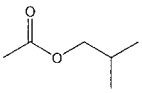
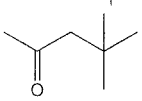
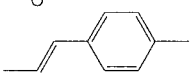
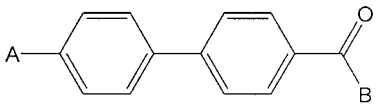
#	A	X	D	pIC_{50}	PMF	FF	FlexX
26	O-Pentyl	C		7.34	1.156	3.202	1.094
27	Cl	N		7.31	1.081	4.469	5.691
28	Cl	C		7.26	1.497	2.913	0.851
29	Cl	C		7.24	1.470	1.038	6.034
30	Cl	C		7.24	1.214	1.146	0.834
31	Cl	C		7.24	1.217	0.998	0.851
32	Cl	C		7.20	1.242	0.723	0.929
33	O-Ethyl	C		7.18	0.998	2.242	0.808
34	Cl	C		7.14	1.092	2.849	0.711
35	O-Ethyl	C		7.12	1.404	2.197	1.106
36	Cl	C		7.11	1.204	0.973	4.392
37	Cl	C		7.10	1.045	1.063	0.711
38	O-Pentyl	C		7.01	1.079	0.975	0.618
39	Cl	C		7.00	1.001	2.079	1.005
40	Cl	C		6.98	1.016	0.839	0.678

Table 2. (continued)

#	A	X	D	pIC_{50}	PMF	FF	FlexX
41	Cl	N		6.92	1.331	0.967	1.046
42	Cl	C		6.85	1.189	0.577	0.615
43	Cl	N		6.61	1.088	2.072	0.580
44	Cl	N		6.53	1.240	3.172	0.980
45	Cl	C		6.47	1.315	0.666	0.814
46	Cl	C	none	6.46	1.293	1.224	1.105
47	Cl	N		6.34	1.172	2.155	5.256
48	Cl	N		6.18	1.857	3.061	1.002
49	Cl	C		5.30	1.106	0.956	5.005



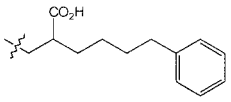
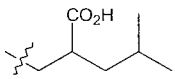
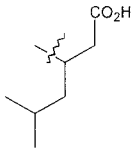
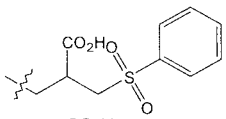
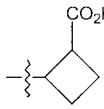
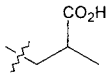
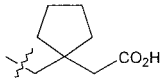
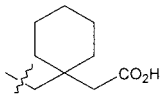
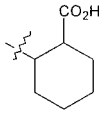
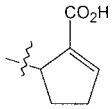
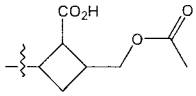
#	A	B	pIC_{50}	PMF	FF	FlexX
50	Br		6.82	1.124	1.377	0.962
51	Cl		6.57	1.400	4.026	1.264
52	Cl		6.42	1.153	1.058	5.820
53	Cl		6.40	1.676	1.247	9.710
54	Cl		6.36	1.347	0.941	0.717

Table 2. (continued)

#	A	B	pIC_{50}	PMF	FF	FlexX
55	Cl		6.02	1.440	2.829	1.145
56	Cl		5.85	1.315	1.204	0.811
57	Cl		5.77	1.400	1.300	0.871
58	Cl		5.62	1.098	3.004	0.824
59	Cl		5.60	1.166	2.286	0.819
60	O-Pentyl		5.56	1.211	2.256	0.567
61	Cl		5.49	1.336	2.187	1.034

[a]Root mean square deviations of the biphenyl core portions of 61 flexibly docked stromelysin inhibitors from the crystallographic binding mode of the core portion of compound 13 are given in Å. The origin of the 61 inhibitors and their assay data were taken from Kluender and co-workers [47]. If not indicated otherwise, racemic mixtures were used in the binding assay.

itant solution and suspended on a cover slip over 1 ml of the precipitant solution in 24-well tissue culture plates. Crystals were mounted in thin-walled glass capillaries. The stabilizing solutions for this operation were fortified with an additional 5% of the appropriate precipitant.

Data collection and processing

Crystals, mounted in capillaries with a small volume of the stabilizing solution, diffracted X-rays beyond 3.0 Å for all of the complexes. X-ray data were measured using dual Xuong–Hamlin multiwire detectors, mounted on a Rigaku RU-300 generator operated at 50 kV and 100 mA. Low angle screenless precession exposures were used to determine cell constants and systematic absences in the X-ray data. The unit cell constants for each of the complexes and other crystal data are given in Table 1. Intensity data were measured using the area detector system at room temperature in omega oscillations of 0.1° per frame. Typical data runs spanned 55° in omega, and a total exposure time of two to four days was required for a complete data set on one crystal. The crystals were remarkably insensitive to radiation damage. Intensity data were integrated, scaled and merged using software provided by San Diego Multiwire Systems [33].

tSL284 structure solution by Multiple Isomorphous Replacement (MIR)

The structure of activated stromelysin, truncated at residue 284 (tSL284), was solved in a complex with a peptide-based hydroxamic acid inhibitor. Three heavy atom compounds were used, K_2IrCl_6, K_2PtCl_4, and an iodinated modification to the hydroxamic acid inhibitor. Three binding sites each were found for the iridium and platinum complexes, and a single iodine site corresponding to a *para*-iodophenylalanine residue in the inhibitor complex. Heavy atom sites were refined against the isomorphous differences in the PHASES package [34]. Solvent flattening [35] was used to improve the electron density map, and the model was built into the MIR map on a Silicon Graphics workstation, running the program FRODO [36]. The atomic positions were refined against the

native diffraction data using the package XPLOR [37]. The final R-value was 22.4% at 2.1 Å, and the model deviated from ideal bond and angle values by 0.014 Å and 3.5°, respectively.

tSL 257/inhibitor structure solution by molecular replacement

The method of molecular replacement was used to solve the structures of the tSL257/biphenyl inhibitor complexes. The computer package XPLOR [37] was used throughout. The model used for molecular replacement consisted of amino acid residues 83–250, two zinc atoms, and two calcium atoms. In the structure solution of the first biphenyl complex, the model of the tSL284 structure, solved by MIR, was used. Subsequently, the model used was one of the independent molecules from the first tSL257/biphenyl complex.

Computational docking methods

The docking programs DOCK4 [6] and FlexX [7] were used to flexibly dock 61 biphenyl MMP3 inhibitors in a crystal structure of stromelysin. All methods used in this study have been described elsewhere and are therefore only briefly discussed.

DOCK4/FF

The docking program DOCK4, which was first introduced in 1982 [4], offers several scoring functions including contact scoring, chemical scoring and energy scoring. The latter is based on calculating the non-bonded interaction terms as defined in the AMBER force field [24, 25]. Therefore we refer to this scoring function as force field (FF) scoring. This standard scoring function consists of a Coulomb and a van der Waals term that describe the electrostatic and dispersion potentials of the protein–ligand interaction as

$$FF = \sum_{i=1}^{lig} \sum_{j=1}^{prot} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + 332 \frac{q_i q_j}{\epsilon r_{ij}} \right] \qquad (1)$$

where A_{ij} and B_{ij} are the van der Waals repulsion and attraction parameters of a ligand atom i and a protein atom j, r is the distance between i and j and ϵ is the dielectric constant. In addition, intra-ligand interactions are calculated and can be added to the above score.

DOCK4/PMF

Recently, a knowledge-based scoring function was introduced by us and implemented into the DOCK4 program [22, 26, 27]. This scoring function uses Helmholtz free interaction energies (potentials of mean force, PMF) between protein and ligand atoms derived by applying the inverse Boltzmann principle akin to scoring functions used for protein folding [38–40]. In contrast to other PMF scoring functions, however, an additional correction term that accounts for the volume taken by the ligand was found to be crucial for deriving meaningful interaction potentials [22]. Also, the definition of a relevant reference state and the choice of statistically relevant and still functionally distinguishing atom types were found to be important in creating a scoring function of high predictive power. The conceptual advantage of the PMF scoring functions lies in its generality, as no measured binding constants are used to derive the potentials. Also, by treating solvation and entropic effects implicitly, it thereby avoids the challenging task of calculating them and balancing their large contributions to predict small binding affinities. This task is prone to large systematic and statistical errors [41–43]. One conceptual disadvantage of the PMF score is that there is no unique thermodynamic cycle that leads from atom pair potentials to protein–ligand binding free energies. It was shown, however, that the sum of the pair potentials served as a good measure for ranking ligands and binding modes against a protein target [44].

The PMF score is calculated as sum over all protein–ligand atom pair interaction free energies $A_{ij}(r)$ as function of the atom pair distance r by

$$PMF\_score = \sum_{\substack{kl \\ ij \\ r<r_{cut-off}}} A_{ij}(r) \qquad (2)$$

where kl is a protein-ligand atom pair of type ij. The potential of mean force for an atom pair of type ij can be written as

$$A_{ij}(r) = -k_B T \ln \left[f^{j}_{Vol\_corr}(r) \frac{\rho^{ij}_{seg}(r)}{\rho^{ij}_{bulk}} \right] \qquad (3)$$

where k_B is the Boltzmann factor, T is the absolute temperature, $f^{j}_{Volcorr}$(r) is a ligand volume correction factor, ρ^{ij}_{seg} is the number density of atom pair ij occurrences at a certain distance, and ρ^{ij}_{bulk} is the number density of pair ij in an appropriate reference state.

van der Waals (VDW) interactions are added to the PMF between i and j for distances shorter than the

longest unoccupied distance for the respective atom type found in the PDB. Furthermore, if the VDW interaction for a particular distance (regardless if pair occurrences are found in the PDB) is larger than 4 kcal/mol, the PMF is overwritten by the VDW term. The VDW term is calculated using a 6–12 Lennard-Jones potential following the implementation of VDW interactions in the FF scoring (AMBER [24,25]) of DOCK4. A united atom model is applied for protein and ligand atoms. In order to prevent VDW collisions within the ligand, we add intra-ligand VDW contributions to the PMF score during the flexible ligand-docking process.

FLEXX

The docking package FlexX was introduced about three years ago [7]. It uses a scoring function similar to that developed by Böhm [13, 23]. The free binding energy of a protein–ligand complex is estimated in FlexX as the sum of free energy contributions from hydrogen bonds, ion-pair interactions, hydrophobic and π-stacking interactions of aromatic groups, and lipophilic interactions:

$$\begin{aligned}\Delta G = \; & \Delta G_0 + \Delta G_{rot} N_{rot} + \\ & \Delta G_{hb} \sum_{neutral\_Hbonds} f(\Delta R, \Delta\alpha) + \\ & \Delta G_{io} \sum_{ionic\_int.} f(\Delta R, \Delta\alpha) + \\ & \Delta G_{aro} \sum_{aro\_int} f(\Delta R, \Delta\alpha) + \\ & \Delta G_{lipo} \sum_{lipo.cont.} f^*(\Delta R)\end{aligned} \tag{4}$$

where ΔG_0, ΔG_{hb}, ΔG_{io}, ΔG_{aro}, and ΔG_{lipo} are adjustable parameters that were fitted to reproduce measured binding affinities of a training set of known 3D protein–ligand complexes by using a multivariate regression method. $f(\Delta R, \Delta\alpha)$ is a scaling function penalizing deviations from the ideal geometry and N_{rot} is the number of free rotatable bonds. The interaction of aromatic groups is an addition to Böhm's original force field design [13,23]. The lipophilic contributions are calculated as sum of atom pair contacts, in contrast to evaluating a surface grid as in Böhm's scoring function. For more details we refer the reader to the original publication [7].

Docking conditions

Standard sampling conditions of DOCK4 were used for the flexible docking runs for both FF scoring and PMF scoring with the following exceptions: An anchor search with standard peripheral seeds was performed and the anchor was minimized. Torsion drive, torsion minimization, and a clash overlap of 0.3 were used. The ligand was re-minimized. 100 minimization iterations, 500 orientations, and a bump filter with a bump maximum of 1 were chosen. FlexX was used under standard conditions through a Sybyl interface in a menu-driven process.

Results and discussion

Binding mode of the biphenyl series of MMP-3 inhibitors

Like other MMPs, the structure of human stromelysin-1 consists of three α-helices and five β-strands. The catalytic site is defined by the fourth β-strand and the second helix. The zinc ion at the catalytic site is coordinated with His201, His205, His211, as well as the carboxylic acid moiety of the biphenyl inhibitors (Figure 1). Sixty-one biphenyl inhibitors were used in this study (Table 2). They were chemically diverse. The majority of the compounds used in this study were conformationally restricted around the carboxylic acid moiety by a cyclopentyl or pyrrolidine group. The compounds showed structural variation, including phthalimides, naphthalimides, phenones, ether linkers of different lengths, esters, and toluene (Table 2).

We used an in-house crystal structure of stromelysin co-crystallized with compound 13 (Figure 1, Table 2) to conduct our docking studies. The biphenyl group of the inhibitor occupies the S1′ pocket, which is the most prominent binding pocket in the MMP catalytic domain. The carbonyl oxygen of the inhibitor forms bifurcated hydrogen bonds with the backbone nitrogen atoms of Leu164 and Ala165 in the fourth β-strand. This carbonyl group is out of plane with respect to the proximal phenyl with a dihedral angle of approximately 40°. The phenyl group of the D moiety (Table 2) is located in the S2 pocket. It interacts with the side chains of His211 and Phe210 whereas the biphenyl, the carboxylic acid, and the carbonyl moieties satisfy specific interactions with the enzyme.

Docking results

The use of closely related inhibitors in a docking study is of great interest since, if a docking/scoring protocol is used for lead optimization, the binding modes of similar compounds must be predicted and the inhibitors must be ranked according to their binding affinities. On the other hand, there are not many data available for a large number of similar inhibitors that were co-crystallized with the same protein and

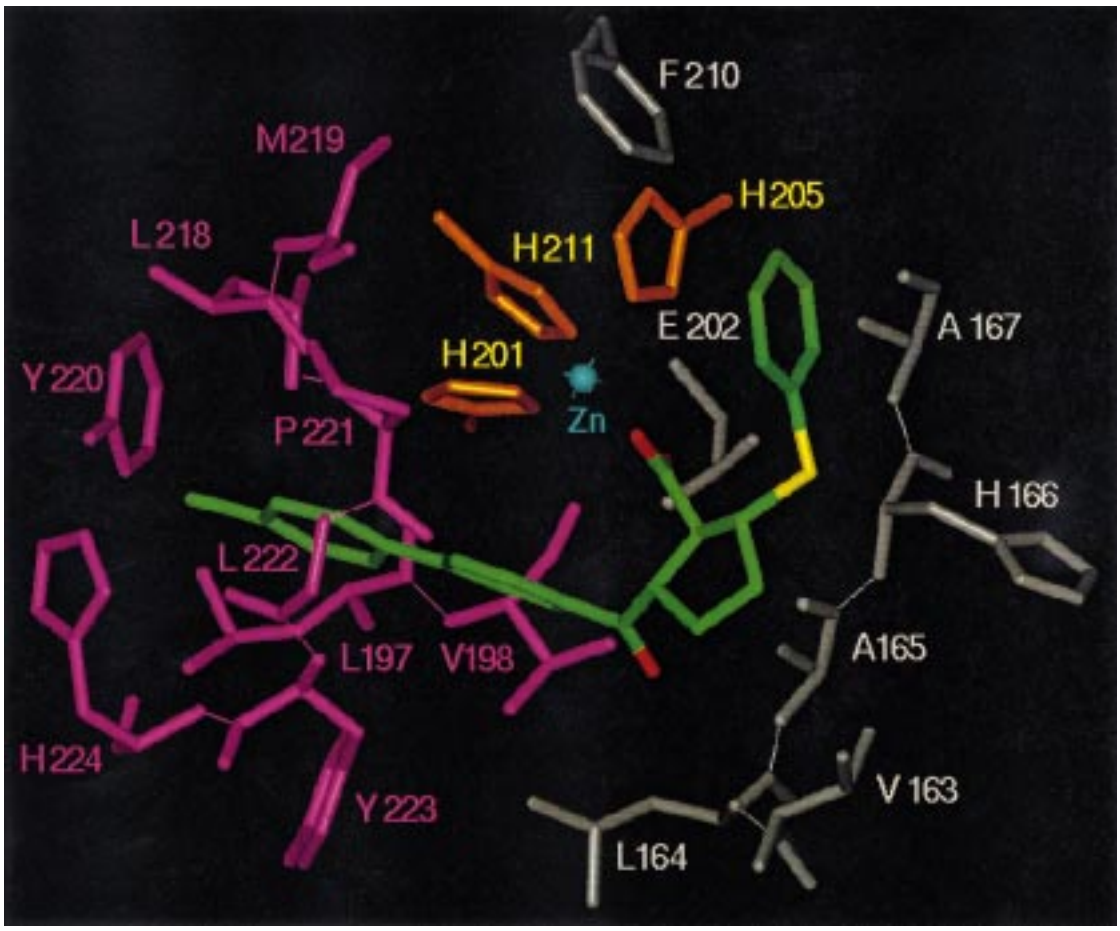


Figure 1. Crystal structure of stromelysin and co-crystallized compound 13 (green). The biphenyl moiety of the inhibitor binds in the S1′ pocket (magenta). H201, H205, and H211 (orange) form the ionization sphere of the Zn ion together with the carboxylic acid of the inhibitor. V163, L164, A165, H166 and A167 belong to the β4 strand of stromelysin (grey). These residues as well as F210 are in close contact with the inhibitor.

could be used in such a docking study. We therefore felt that the biphenyl MMP-3 inhibitor test set with a relatively high number of biologically active compounds was an interesting case for such a study. In order to compare the capabilities of different docking/scoring approaches to find the correct binding mode, we docked 61 biphenyl inhibitors into the binding site of stromelysin. The binding mode of 6 biphenyl compounds was crystallographically determined. The orientations of the biphenyl moieties as well as of the carboxylic acid and the carbonyl group (when present) in the binding site were found to be highly conserved (rms fluctuations < 0.2 Å). Therefore, we make the general assumption in this paper that this is also the case for all other biphenyl inhibitors used here. This assumption appears to be reasonable, since the interactions of the biphenyl (S1′ pocket) as well as of the carboxylic acid (Zn coordination) and the carbonyl group (hydrogen bonds to backbone nitrogens of Leu164 and Ala165) are highly favorable and specific. Keep in mind, however, that when we refer below to 'correct' or 'incorrect' binding modes identified in the docking experiments, the underlying assumption is that the biphenyl moieties bind in the same place for all 61 inhibitors.

Table 2 shows the flexible ligand docking results (highest scoring binding modes) for 61 biphenyl inhibitors in stromelysin in the form of rmsd values between the crystallographic binding mode and the docked binding mode of the core fragment of the biphenyl compounds. The average rmsd for all 61 compounds was found to be best for the DOCK4/PMF approach (1.8 Å), followed by FlexX (2.0 Å) and DOCK4/FF (2.5 Å). This is largely due to the differ-

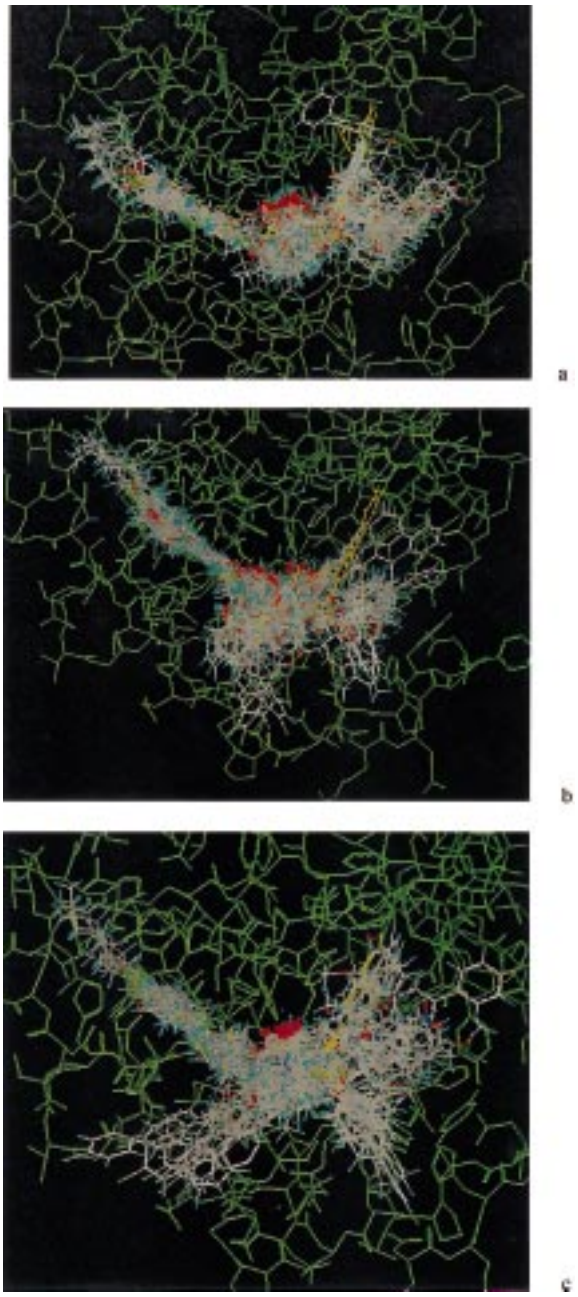


Figure 2. Best scoring binding modes of 61 biphenyl inhibitors in the stromelysin crystal structure co-crystallized with compound 13 docked by using (a) DOCK4/PMF, (b) DOCK4/FF, (c) FlexX.

ent number of incorrectly docked compounds in each docking/scoring protocol. DOCK4/PMF docked four compounds incorrectly (rmsd of core fragment (Table 2) $>$ 4 Å), DOCK4/FF docked nine compounds incorrectly, and FlexX docked 12 compounds incorrectly. Figure 2 shows the stromelysin crystal structure co-crystallized with compound 13 and the docked biphenyl inhibitors in their best scoring poses for the three docking approaches used. One can see from the figure that DOCK4/PMF generated the tightest set of docked binding geometries (scoring highest among sampled poses) that occupy the smallest number of possible binding pockets in MMP-3. This is a non-trivial finding since three other pockets of the binding site of stromelysin (S1, S2′ and S3′) remain unoccupied, although they offer sufficient volume to host the biphenyl inhibitors. Identifying the correct binding mode in a case of competing binding pockets in a binding site is a good test case for docking/scoring approaches. It is encouraging to find that DOCK4/PMF identified the correct placement of the biphenyl inhibitors in the majority of cases. Some of the compounds docked by DOCK4/FF occupied the S2′ and S3 pockets. Compounds docked by FlexX occupied all binding pockets of the stromelysin binding site. The carboxylic acid moiety of the biphenyl inhibitors coordinated best with the zinc ion for inhibitors docked with FlexX. Using DOCK/PMF, the zinc coordination with the carboxylic acid moiety of the biphenyl inhibitors was also preserved. However, DOCK/FF calculations resulted in a widely scattered distribution of the carboxylic acid moiety of the biphenyl inhibitors around the zinc that did not retain the coordination sphere of the zinc ion. Table 2 shows that the rmsd for compounds that were docked correctly is much lower in FlexX (0.9 Å) than in both DOCK4 scoring approaches (1.3 Å for DOCK4/PMF and 1.8 Å for DOCK4/FF). This result suggests that the ‘fine tuning’ of a binding mode may be better achieved by the empirical scoring function implemented in FlexX. It will be interesting to see how a new knowledge-based scoring function, planned to be implemented in FlexX, will be able to further improve this ‘fine tuning’ [45].

For the four compounds that were docked incorrectly in DOCK4/PMF we compared the scores of the best scoring conformation found in the docking process with the score of the anticipated (correct) binding mode. For DOCK4/PMF it was found that in all cases the anticipated binding mode scored higher than the incorrect binding mode found in the docking study (Table 3). This result suggests that insufficient

Table 3. Comparison between binding scores of anticipated and found binding modes of four incorrectly docked biphenyl compounds[a]

Compound #	Anticipated binding mode score	Docked binding mode score	Docking/scoring approach
3	−148.9	−140.6	DOCK4/PMF
4	−147.2	−146.7	DOCK4/PMF
5	−152.9	−148.2	DOCK4/PMF
6	−153.6	−142.6	DOCK4/PMF
3	−25.2	−43.0	DOCK4/FF
4	−34.8	−51.3	DOCK4/FF
5	−41.0	−46.8	DOCK4/FF
6	−39.6	−47.1	DOCK4/FF

[a]The anticipated binding mode was generated by overlaying the given compound with the crystal structure of compound 13 that was used for the docking studies and represented the correct binding mode for the biphenyl portion of all molecules. The given compound was then minimized into the next local minimum by using the scoring function specified in the table.

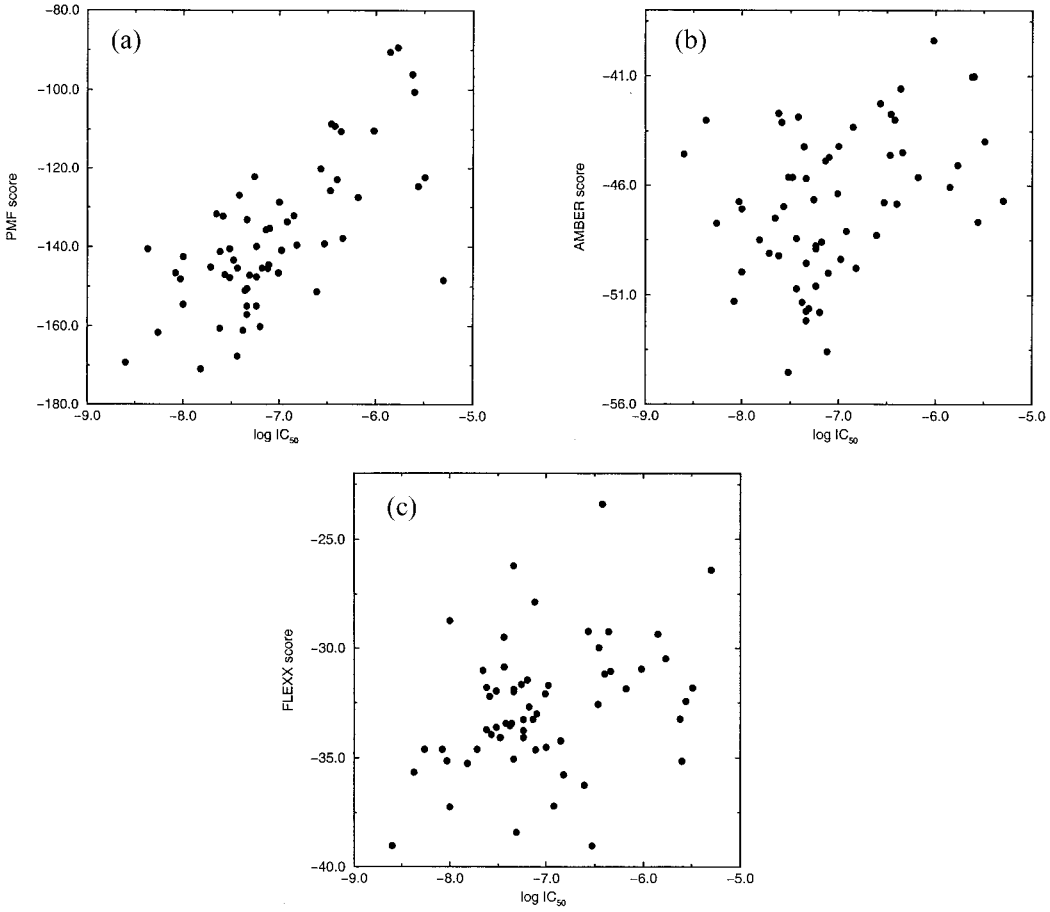


Figure 3. Calculated binding score as a function of the observed binding affinity for 61 biphenyl inhibitors docked into the crystal structure of stromelysin co-crystallized with compound 13 by (a) DOCK4/PMF, (b) DOCK4/FF, (c) FlexX.

sampling rather than incorrect scoring is responsible for the incorrect docking of these compounds. This is a very encouraging finding that shows the robustness of the PMF scoring function. Note that with increased sampling (1500 accepted initial orientations) the correct binding mode still could not be found in the docking experiment. In case of FF scoring all incorrect binding modes were scored significantly higher than the anticipated binding mode (Table 3).

Scoring results

In addition to identifying the correct binding mode of a ligand in a protein binding site, a scoring/docking approach should also be able to rank different ligands bound to the same protein according to their binding affinities. This task is of paramount importance in a virtual screening protocol, where few biologically active compounds must be identified in a pool of non-binders. Figure 3 shows the correlation between the measured binding affinities of the biphenyl inhibitors and the calculated scores of the docked conformations. Only the PMF scoring function was able to correlate the measured binding affinities and the calculated scores with an R^2 of 0.49 and a standard error of 0.80 log IC_{50} units. Removing one outlier (compound 61) improved the correlation to an R^2 of 0.56 and a standard error of 0.61 log units. This is about 20% of the overall IC_{50} range of 3.3 log units in the test set. For FlexX (DOCK4/FF) there was virtually no correlation found with an R^2 of 0.15 (0.12) and a standard deviation of 1.82 (2.07).

Figure 3 reveals another interesting aspect of scoring putative protein–ligand complexes in a docking approach that may be used in virtual screening (VS). In a VS approach one is interested in filtering out compounds of a database by computational means that are likely not to be biologically active against a certain target [46]. It is therefore crucial that no 'false negatives' are generated in the VS screen (active compounds that are scored too low). Figure 3c suggests that the DOCK4/PMF approach does not produce drastic false negatives that would lie in the upper left quarter of the graph. Deviations from the optimal correlation between IC_{50} and PMF score mainly lie in the lower right quarter of the figure, indicating 'false positives' – inactive compounds that show high scores. While a high rate of false positives is a general problem in VS, it is important to point out that false negatives have been practically avoided in the DOCK4/PMF docking experiment. This behavior of the DOCK4/PMF approach has been seen earlier in case of neuraminidase inhibitors docked to different crystal structures [27]. This result further supports the recent finding that DOCK4/PMF can be used as a VS tool to exclude large numbers of molecules in a database from biological testing [26].

Conclusions

Docking/scoring experiments with 61 MMP-3 inhibitors showed that DOCK4/PMF is a robust docking/scoring approach that found the correct binding mode scoring highest in 93% of the cases. DOCK4/FF identified the correct binding mode in 85% of the cases and FlexX in 80%. The binding modes of the four inhibitors, docked incorrectly with DOCK4/PMF are due to sampling insufficiencies rather than an unreliable scoring function since the anticipated binding modes scored higher than the best scoring binding modes suggested by the program. DOCK4/PMF performed better than FlexX and DOCK4/FF in finding the correct binding mode of the majority of the compounds. However, comparing only those inhibitors for which the correct binding mode was identified in the docking/scoring approach, FlexX showed on average a significantly lower rmsd to the expected binding mode.

DOCK/PMF was the only docking/scoring approach tested that was able to significantly correlate measured binding data with the calculated binding score. Moreover, the calculations did not lead to drastic false negatives. This result shows the robustness of the PMF scoring function and its utility in virtual screening protocols.

Two different scoring functions in the same docking program (DOCK4) used the same sampling procedure. They performed significantly different, both in predicting binding modes of protein–inhibitor complexes and in ranking molecules according to their binding affinities. This finding suggests that the accuracy of the scoring function of a docking/scoring approach is crucial for generating correct docking modes of putative protein–ligand complexes and for predicting their binding free energies.

Acknowledgements

The authors thank Dr. Sarah Heald for critically reading the manuscript. S.H. thanks Betsy Murray (Tripos Inc.) for providing a FlexX demo license. I.M. thanks Dr. Harold Kluender for helpful discussions.

References

1. Greer, J., Erickson, J.W., Baldwin, J.J. and Varney, M.D., J. Med. Chem., 37 (1994) 1035.
2. Lutz, M.W., Menius, J.A., Choi, T.D., Laskody, R.G., Domanico, P.L., Goetz, A.S. and Saussy, D.L., Drug Discov. Today, 1 (1996) 277.
3. Dixon, J.S., Proteins, Suppl., 1 (1997) 198.
4. Kuntz, I.D., Blaney, J.M., Oatley, S.J. and Langridge, R.L., J. Mol. Biol., 161 (1982) 269.
5. Kuntz, I.D., Science, 257 (1992) 1078.
6. Ewing, T. and Kuntz, I.D., J. Comput. Chem., 18 (1997) 1175.
7. Rarey, M., Kramer, B., Lengauer, T. and Klebe, G., J. Mol. Biol., 261 (1996) 470.
8. Welch, W., Ruppert, J. and Jain, A.N., Chem. Biol., 3 (1996) 449.
9. Jones, G., Willett, P., Glen, R.C. and Leach, A.R., J. Mol. Biol., 267 (1997) 727.
10. Goodsell, D.S. and Olson, A.J., Proteins, 8 (1990) 195.
11. Miller, M.D., Kearsley, S.K., Underwood, D.J. and Sheridan, R.P., J. Comput-Aided Mol. Design, 8 (1994) 153.
12. Dixon, J.S. and Blaney, J.M., In Martin, Y.C. and Willett, P. (Eds.) Designing Bioactive Molecules. Three-dimensional Techniques and Applications, American Chemical Society, Washington, DC, 1998, p. 175.
13. Böhm, H.-J., J. Comput.-Aided Mol. Design, 8 (1994) 243.
14. Aqvist, J., Medina, C. and Samuelsson, J.E., Protein Eng., 7 (1994) 386.
15. Head, R.D., Smythe, M.L., Oprea, T.L., Waller, C.L., Green, S.M. and Marshall, G.M., J. Am. Chem. Soc., 118 (1996) 3959.
16. Jain, A.N., J. Comput.-Aided Mol. Design, 10 (1996) 427.
17. Kollman, P., Chem. Rev., 7 (1993) 2395.
18. Wallqvist, A., Jernigan, R.L. and Covell, D.G., Protein Sci., 4 (1995) 1881.
19. Williams, D.H., Cox, J.P.L., Doig, A.J., Gardner, M., Gerhard, U., Kaye, P.T., Lai, A.R., Nicholls, I.A., Salter, C.J. and Mitchell, R.C., J. Am. Chem. Soc., 113 (1991) 7020.
20. Eldridge, M.D., Murray, C.W., Auton, T.R., Paolini, G.V. and Mee, R.P., J. Comput.-Aided Mol. Design, 11 (1997) 425.
21. DeWitte, R.S. and Shakhnovich, E.I., J. Am. Chem. Soc., 118 (1996) 11733.
22. Muegge, I. and Martin, Y.C., J. Med. Chem., 42 (1999) 791.
23. Böhm, H.-J., J. Comput.-Aided Mol. Design, 8 (1998) 243.
24. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta Jr., S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
25. Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A., J. Comput. Chem., 7 (1986) 230.
26. Muegge, I., Martin, Y.C., Hajduk, P.J. and Fesik, S.W., J. Med. Chem., 42 (1999) 2498.
27. Muegge, I., Med. Chem. Res., 9 (1999) 490.
28. Birkedal-Hansen, H., Moore, W.G.I., Bodden, M.K., Windsor, L.J., Birkedal-Hansen, B., DeCarlo, A. and Engler, J.A., Crit. Rev. Oral Biol. Med., 4 (1993) 197.
29. Woessner, J.F.J., FASEB J., 5 (1991) 2145.
30. Dhanaraj, V., Ye, Q.Z., Johnson, L.L., Hupe, D.J., Ortwine, D.F., Dunbar, J.B., Rubin, J.R., Pavolvski, A., Humblet, C. and Blundell, T.L., Structure, 4 (1996) 466.
31. Stockman, B.J., Watson, D.J., Gates, J.A., Scahill, T.A., Kloosterman, D.A., Miszak, S.A., Jacobsen, E.J., Belonga, K.L., Mitchell, M.A., Mao, B., Petke, J.D., Goodman, L. and Powers, E.A., Protein Sci., 7 (1998) 2281.
32. Marcy, A.I., Eiberger, L.L., Harrison, R., Chan, H.K., Hutchinson, N.L., Hagman, W.K., Cameron, P.M., Boulton, D.A. and Hermes, J.D., Biochemistry, 30 (1991) 6476.
33. Howard, A.J., Nielsen, C. and Xuong, N.H., Methods Enzymol., 114 (1985) 452.
34. Furey, W.F. and Swaminathan, S., Methods Enzymol., 227 (1997) 590.
35. Wang, B.C., Methods Enzymol., 114 (1985) 90.
36. Jones, T.A., J. Appl. Crystallogr., 115 (1975) 157.
37. Brünger, A.T., X-PLOR Manual, Version 3.1, Yale University Press, New Haven, CT, 1992.
38. Sippl, M.J., J. Mol. Biol., 213 (1990) 859.
39. Sippl, M.J., J. Comput.-Aided Mol. Design, 7 (1993) 473.
40. Sippl, M.J., Ortner, M., Jaritz, M., Lackner, P. and Flöckner, H., Folding Design, 1 (1996) 289.
41. Warshel, A., Papazyan, A. and Muegge, I., J. Biol. Inorg. Chem., 2 (1997) 143.
42. Muegge, I., Qi, X.P., Wand, A.J., Chu, Z.T. and Warshel, A., J. Phys. Chem. B, 101 (1997) 825.
43. Muegge, I., Tao, H. and Warshel, A., Protein Eng., 10 (1997) 1363.
44. Muegge, I., Perspect. Drug Discov. Des., accepted for publication.
45. Gohlke, H., Hendlich, M. and Klebe, G., J. Mol. Biol., 295 (2000) 337.
46. Walters, W.P., Stahl, M.T. and Murcko, M.A., Drug Discov. Today, 3 (1998) 160.
47. Kluender, H.C.E., Dixon, B.R., VanZandt, M.C., Willhelm, S.M., Wolanin, D.J. and Wood, J.E., U.S. Patent No. 5,861,428, 1999.