Predicting binding modes, binding affinities and 'hot spots' for protein-ligand complexes using a knowledge-based scoring function

HOLGER GOHLKE, MANFRED HENDLICH and GERHARD KLEBE*
Department of Pharmaceutical Chemistry, Philipps-University of Marburg, Marbacher Weg 6, D-35032 Marburg, Germany

Summary. The development of a new knowledge-based scoring function (DrugScore) and its power to recognize binding modes close to experiment, to predict binding affinities, and to identify 'hot spots' in binding pockets is presented. Structural information is extracted from crystallographically determined protein-ligand complexes using ReLiBase and converted into distance-dependent pair-preferences and solvent-accessible surface (SAS) dependent singlet preferences of protein and ligand atoms. The sum of the pair preferences and the singlet preferences is calculated using the 3D structure of protein-ligand complexes either taken directly from the X-ray structure or generated by the docking tool FlexX. DrugScore discriminates efficiently between well-docked ligand binding modes (root-mean-square deviation <2.0 Å with respect to a crystallographically determined reference complex) and computer-generated ones largely deviating from the native structure. For two test sets (91 and 68 protein-ligand complexes, taken from the PDB) the calculated score recognizes poses deviating <2 Å from the crystal structure on rank 1 in three quarters of all possible cases. Compared to the scoring function in FlexX, this is a substantial improvement. For five test sets of crystallographically determined protein-ligand complexes as well as for two sets of ligand geometries generated by FlexX, the calculated score is correlated with experimentally determined binding affinities. For a set of 16 crystallographically determined serine protease inhibitor complexes, a R^2 value of 0.86 and a standard deviation of 0.95 log units is achieved as best result; for a set of 64 thrombin and trypsin inhibitors docked into their target proteins, a R^2 value of 0.48 and a standard deviation of 0.7 log units is calculated. DrugScore performs better than other state-of-the-art scoring functions. To assess DrugScore's capability to reproduce the geometry of directional interactions correctly, 'hot spots' are identified and visualized in terms of isocontour surfaces inside the binding pocket. A data set of 159 X-ray protein-ligand complexes is used to reproduce and highlight the actually observed ligand atom positions. In 74% of all cases, the actually observed atom type corresponds to an atom type predicted by the most favorable score at the nearest grid point. The prediction rate increases to 85% if at least an atom type of the same class of interaction is suggested. DrugScore is fast to compute and includes implicitly solvation and entropy contributions. Small deviations in the 3D structure are tolerated and, since only contacts to non-hydrogen atoms are regarded, it does not require any assumptions on protonation states.

Key words: binding affinity, docking, knowledge-based, protein-ligand interactions, scoring function, virtual screening

* To whom correspondence should be addressed. E-mail: klebe@mailer.uni-marburg.de

Introduction

The process of finding novel leads for a new target is one of the most important steps in a drug development program. Today two complementary strategies are followed: experimental high-throughput screening and computational methods exploiting structural information of the protein binding site [1–4]. The latter approaches try to predict, e.g. via docking, the actual binding mode of a ligand at the binding site [5,6].

Several of the published docking procedures are fast enough to serve the purpose and suggest solutions approximating the native pose in up to 80% of the cases [7–9]. Nevertheless, the pose closest to the experimental situation is often not ranked as the energetically most favorable one within a set of decoy geometries, which indicates deficiencies in the applied ranking schemes [10]. Consequently, we embarked on the development of a new scoring function.

At best, binding affinity is determined by statistical thermodynamics resulting in a master equation that considers all contributing effects [11,12]. Although theoretically most convincing, elaborate methods such as free energy perturbation (FEP) or thermodynamic integration (TI) are computationally too demanding for the application described above [13].

Instead, the partitioning of binding affinity into several additive terms or descriptors is a widely accepted assumption for the development of empirical regression-based scoring functions [14]. Usually a number of empirically derived contributions is fitted to a data set of experimental observations [15–19]. Approaches such as VALIDATE [20] are based on the ideas of QSAR. These approaches achieve a precision of about 1.5 orders of magnitude while *predicting* K_i [15,20]. However, any regression analysis suffers from the fact that the obtained conclusions can only be as precise and generally valid as the data used covers all contributing and discriminating effects in protein-ligand complexes. The same arguments are true for penalty filters developed to discard computer-generated artifacts from a list of favorable ligand poses [21].

We decided to follow an alternative way to develop a scoring function based on empirical knowledge. During its development we decided not to assign protonation states to the various atom types, assuming that the derived statistical preferences implicitly reflect these influences. Furthermore, any binding feature not in agreement with the most frequently observed contact preferences will likely be penalized due to its minor occurrence.

Knowledge-based potentials have been applied to rank different solutions of the protein-folding problem [22–24]. Up to now, this approach has been applied to only five case studies for the ranking of different protein-ligand complexes. Except for a single test case in the most recently published work

[25], none of these, however, engaged in identifying near-native poses of *one* ligand with respect to *one* protein.

Wallqvist et al. [26,27] classified the surfaces of buried ligand atoms found in 38 complexes and developed a model to predict the Gibbs free energy of binding based on these observed atom-atom preferences. From an analysis of 10 HIV protease inhibitor complexes, they approximated the free energy of binding to within ± 1.5 kcal/mol.

Using a dataset of 30 HIV-1, HIV-2, and SIV proteases, Verkhivker et al. [28] compiled a distance-dependent knowledge-based pair-potential which was then combined with hydrophobicity [29] and conformational entropy scales [30].

DeWitte and Shaknovich [31] used a sample of 126 structures from the PDB [32] to develop a set of 'interatomic interaction free energies' for a variety of atom types.

Muegge and Martin [33] explored structural information of known protein-ligand complexes from the PDB and derived distance-dependent Helmholtz free interaction energies of protein-ligand atom pairs. Using 77 protein-ligand complexes for validation, the calculated score achieves a standard deviation from the observed binding affinities of 1.8 log K_i units. The scoring function was further evaluated by docking weak-binding ligands to the FK506-binding protein [34].

Mitchell et al. [25,35] developed a potential of mean force at atomic level using high-resolution X-ray structures from the PDB considering 820 possible atom-atom pairs. The performance to identify low-energy binding modes from decoy conformations is tested only for heparin binding to bFGF (PDB-code: 1bfc). While the crystal structure was ranked lowest, the best-scored of the generated structures deviates largely from the experimental situation. Evaluating a test set of 90 different PDB complexes, with respect to binding energies, a squared correlation coefficient of 0.55 is achieved as optimum.

In the present article, we describe the development of a new scoring function (called DrugScore) to predict protein-ligand interactions (see also [36]). It is based on the vast structural knowledge stored in the entire PDB retrievable using ReLiBase [37]. Knowledge-based probabilities, well adjusted to describe specific short-range distances between ligand and protein functional groups are combined with terms considering solvent-accessible-surface portions of both partners that become buried upon binding. For the first time, knowledge-based probabilities are used to discriminate and predict ligand-binding modes for 159 protein-ligand complexes. Multiple solutions, generated for these examples by FlexX, have been re-ranked to obtain a significantly improved scoring with respect to their deviation from the native pose. In addition, the power of this approach to predict binding affinities is tested by

analyzing sets of crystallographically determined protein-ligand complexes and protein-ligand geometries generated by FlexX. Compared to the results of commonly applied scoring functions, DrugScore reveals convincingly better predictions. Finally, the implicit consideration of directionality effects in intermolecular interactions described by the distance-dependent part of the function is demonstrated by visual inspection and a statistical evaluation. The results are encouraging compared to the recently published SuperStar method [38].

Theory and methods

In the following, we shortly summarize the theoretical background of our approach (for a detailed description see [36]).

The noncovalent complex formation between a ligand and a protein is usually performed in aqueous solution. An implicit description of the complex solute-solvent interactions and entropic solvent effects together with the involved enthalpic contributions resulting from interatomic forces (e.g. electrostatic or van der Waals) [39] is reflected by the formalism used to derive potentials of mean force from database knowledge [23,40].

Derivation of statistical distance-dependent pair-preferences and solvent-accessible surface-dependent singlet-preferences

Following an approach at atomic level [41,42], distance-dependent pair-potentials between ligand and protein atoms of type i and j are compiled by

$$\Delta W_{i,j}(r) = W_{i,j}(r) - W(r) = -\ln \frac{g_{i,j}(r)}{g(r)} \qquad (1)$$

where $g_{i,j}(r)$ is the normalized radial pair-distribution function for atoms of types i and j, separated by a distance in the interval of r and $r + dr$; $g(r)$ is the normalized mean radial pair distribution function for a distance between two atoms in the range r and $r + dr$. It incorporates all nonspecific information common to all atom pairs present in an environment typical for proteins.

The definition of an upper radius limit $r_{\max}$ for interactions between atoms i and j [43] determines the overall shape of the resulting potentials. Sampling over short distances will emphasize the specific interactions formed by a ligand functional group with the neighboring binding-site residues. To guarantee that these interactions will dominate, we restrict our sampling to distances between 1 and 6 Å with a bin size of 0.1 Å. The rationale for this upper limit arises from the fact that a 6 Å contact is short enough not to involve a water molecule as mutual mediator of a ligand-to-protein interaction. To avoid

the sampling over large distances but to include solvent-mediated effects, an alternative approach is required [42,44]. A combination of the short-distance sampling together with the findings from protein-fold prediction motivated us to derive a knowledge-based one-body potential scaled to the size of the solvent accessible surface (SAS) of the protein and the ligand that becomes buried upon complex formation:

$$\Delta W_i(SAS, SAS_0) = W_i(SAS) - W_i(SAS_0) = -\ln \frac{g_i(SAS)}{g_i(SAS_0)} \qquad (2)$$

In this equation, g_i is the normalized distribution function of the surface area of an atom i in the buried state (*SAS*) (considering ligand and protein individually) in comparison to the solvated state (SAS_0). It is calculated by an approximate cube-algorithm similar to the one introduced by Böhm [15]. In this assumption any *polar* portion of the SAS that becomes buried in the complex, however still facing a *polar* environment, is considered to remain in a condition equivalent to 'solvent accessible' [45].

As a first approximation, the ligand conformations found by X-ray crystallography or docking procedures are assumed to be identical to those adopted in the solvent. In this crude model, conformational changes experienced upon ligand binding [46] are not considered.

Both short-range pair- and SAS-potentials are derived using the ReLiBase system [37] for data extraction. For our purpose, we evaluated crystallographically determined complexes with resolutions better than 2.5 Å. Complexes with covalently bound ligands or ligands with less than 6 or more than 50 non-hydrogen atoms were discarded. Furthermore, we excluded all complexes that were subsequently used in the validation of the predictive power of the potentials to avoid any redundancy or training effects due to overfitting. Potentials were derived for 17 different atom types using the SYBYL atom type notation: C.3, C.2, C.ar, C.cat, N.3 (= N.4), N.ar (= N.2), N.am, N.pl3, O.3, O.2, O.co2, S.3 (= S.2), P.3, F, Cl, Br including metal atoms Met (= Ca, Zn, Ni, Fe).

Calculation of the total score for a given ligand pose

In our approach we assume that a reasonable description of the total preference ΔW of a particular binding geometry can be approximated by summing over all individual contributions (i.e. of k_i ligand atoms of type i and l_j protein atoms of type j).

$$\Delta W = \gamma \sum_{k_i} \sum_{l_j} \Delta W_{i,j}(r)+$$

$$(1 - \gamma) \left[\sum_{k_i} \Delta W_i (SAS, SAS_0) + \sum_{l_j} \Delta W_j (SAS, SAS_0) \right] \qquad (3)$$

γ is an adjustable parameter, optimized empirically to be 0.5.

Our approach does not incorporate explicitly additional contributions to the binding energy such as conformational, rotational, and translational entropy. Furthermore, energy contributions arising from intramolecular interactions (van der Waals and torsion potentials) are neglected. Since popular docking tools such as FlexX, DOCK, and GOLD generate only favorable ligand conformations, we believe that these terms can only be of minor importance in comparison to the solute state contributions.

The obtained scoring values are taken to rank different poses of one ligand in a single protein with respect to the rms deviation from the geometry as found in the crystal structure.

Calculation of binding affinities

As stated above, the obtained statistical preferences and in consequence the calculated scores are considered to implicitly contain not only enthalpic but also entropic contributions to binding. Although not a proof, but if the derived scoring values correlate with experimentally determined binding free energies, it appears evident that the important contributions are correctly and sufficiently covered.

For the calculation of scoring values, cofactors and metal atoms are taken as part of the protein whereas water molecules are omitted. The values obtained are related to the experimentally determined binding affinities using an adjustable parameter c_s. It is determined iteratively by scaling ΔW in a way that the standard deviations of these calculated values become equal to the observed ones (pK_i) according to a straight line with zero intercept.

$$pK_i = c_s * \Delta W \qquad (4)$$

The general applicability of this relationship depends on the fact whether this adjustable parameter can be transferred among different data sets.

Implicit consideration of directionality in radial-symmetrical pair-potentials

The strength of interactions, especially between polar functional groups but also e.g. between aromatic rings, depends on their mutual distance and relative orientation in space. Since a single pair-potential purely exhibits spherical symmetry, the question is how well directionality between pairs of atoms is

implicitly reflected if multiple pair-potentials are considered as a composite representation.

To assess the spatial resolution of the entire ensemble of the pair potentials, a cubic grid with 0.5 Å spacing is constructed in the binding pocket with a margin of 8 Å around a ligand. At every grid point not occupied by the protein, a scoring value is calculated considering all possible ligand-atom types (Equation 1). The obtained grid values are contoured for the individual ligand-atom types. The isopleths shown comprise 10% of the potential values above the global minima for each type. For a statistical evaluation, the type of a solvent-inaccessible ligand atom actually found in the analyzed crystal structure is compared to the type predicted by our function in this local area. For the analysis, the scoring values for a C.3, N.3, O.3, O.2, or O.co2 probe at the neighboring grid point are compared and the probe with the best scored value is selected. We decided to use a 0.5 Å grid because the largest possible distance between a ligand atom and its nearest grid point amounts to half of the through-space diagonal (≈ 0.43 Å). We believe the grid is detailed enough, since this distance is close to the mean positional errors in experimental structure determinations.

In principle a scoring value can be calculated by DrugScore at any position in space, also where precisely a ligand atom is found in a crystal structure. However, with the motivation to predict favorable ligand atom sites inside a binding pocket *de novo*, the grid-based approach appears most appropriate.

Results

Statistical preferences of ligand-protein atom pairs

Correlation functions for pairs of ligand-protein atoms were derived by counting their occurrence frequencies at discrete distances. Subsequently, statistical preferences were computed using Equation 1. In the following, we will focus on some illustrative examples, depicted in Figures 1 and 2, out of 289 possible atom-pair combinations. In all cases, the first atom-type index i is attributed to a ligand atom, the second j to a protein or cofactor atom.

The derived preferences can be divided into two main classes: the first contains interactions between polar and charged atoms and exhibits pronounced minima at distances between 2.5 and 3.0 Å. They correspond to hydrogen-bonds and salt-bridges (Figure 1). The second class comprises non-polar interactions and displays broader minima which reveal favorable interactions at distances >3.5 Å (Figure 2).

Going from O.3-O.3 to O.3-O.co2 and O.co2-N.pl3, the minima corresponding to the shell of next neighbors fall into a decreasing distance range,
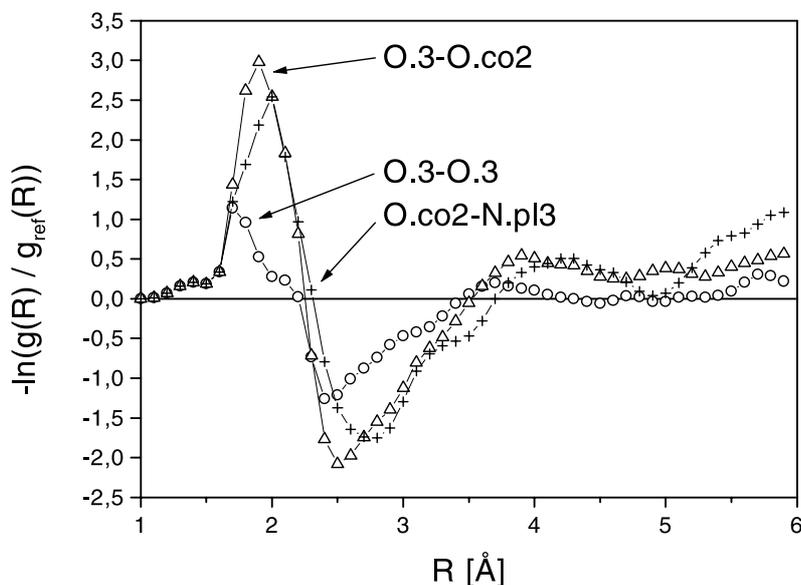
Figure 1. Statistical preferences for polar/charged pair interactions as a function of the distance R, calculated according to Equation 1.

thus exhibiting favorable interactions at lower distances. Contacts between the above-mentioned atom-types can be assigned to a 'normal' hydrogen-bond, a polar charge-assisted interaction and a salt-bridge [48]. Expressed in terms of statistical preferences (Equation 1), an ideal O.3-O.3 interaction is 2.5 times less favorable than a similar O.3-O.co2 interaction. For the O.co2-N.pl3 atom pair one has to take into account that in a bidentate salt bridge between carboxylate and amidinium/guanidinium this interaction is counted twice.

For nonpolar contacts (Figure 2), the C.ar-C.ar interaction shows a slightly more structured preference compared to the C.3-C.3 interaction and the minimum of the former resides at a shorter distance of 3.7 Å, in agreement with the well-known aromatic-aromatic interactions [49]. In contrast, C.3-C.3 interactions do not show any distinct preference of the atom pair distribution over the entire distance range of 4 to 6 Å of favorable interactions. This is clearly in agreement with the well-known fact that the latter type of interaction hardly exhibits any directional preferences.
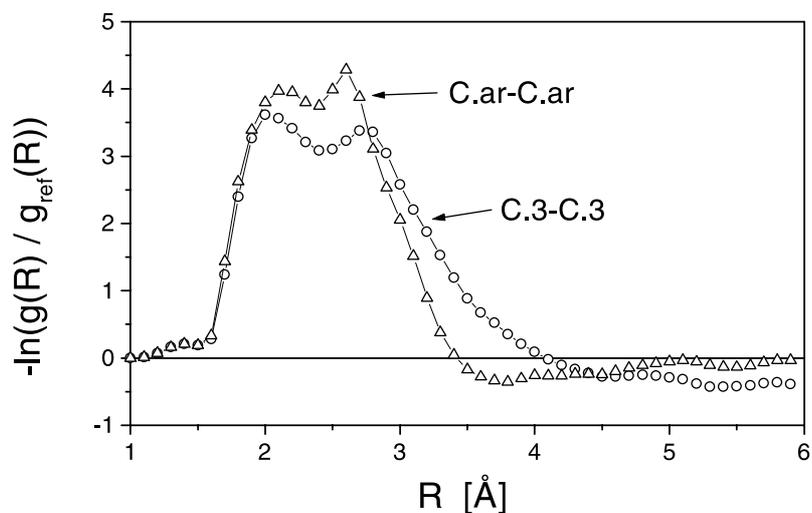
Figure 2. Statistical preferences for nonpolar/aromatic pair interactions as a function of the distance R, calculated according to Equation 1.
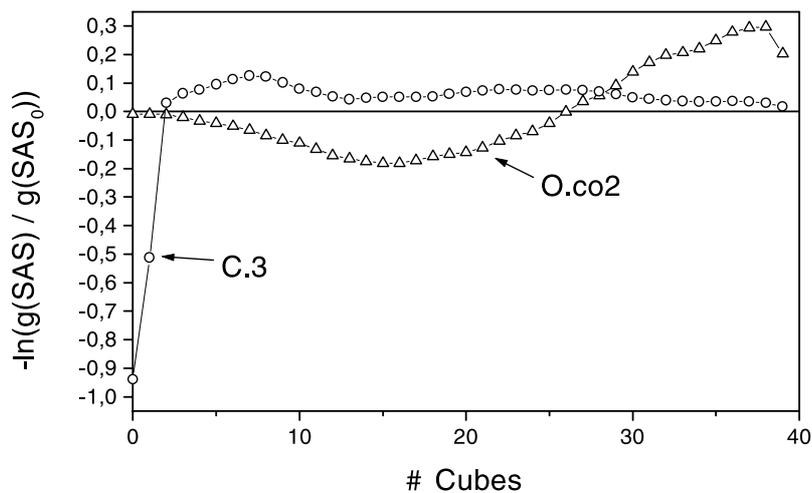

Figure 3. Statistical preferences for ligand atoms of type C.3 and O.co2 as calculated from the distribution functions for solvent accessibility of both atom types for complexed and separated state from the protein according to Equation 2. The number of cubes (#cubes) is an approximate measure for the solvent accessibility; zero cubes refer to complete burial.

Table 1. Results for scoring multiple docking solutions of 91 protein-ligand complexes generated by FlexX and DrugScore

	% of complexes with solutions exhibiting rmsd of the crystal structure			
	<1.0 Å	<1.5 Å	<2.0 Å	≥2.0 Å
All ranks[a]	65	76	84	16
1st rank[b]				
FlexX	20	37	54	46
DrugScore	39	66	73	27
Improvement[c]	95	78	35	–41

[a] All solutions of each docking experiment for the 91 complexes are considered. This number expresses the portion of all complexes for which at least one solution with the given rmsd was computed by FlexX.
[b] Only the ligand geometry scored to be on the 1st rank by either FlexX or DrugScore is considered. The numbers are related to the ones in the first line.
[c] Calculated by (%DrugScore – %FlexX)/%FlexX.

SAS-dependent preferences for solvent-exposure of ligand and protein atoms

To incorporate solvent effects we derived a SAS-dependent singlet preference either for ligand and protein atoms according to Equation 2. Figure 3 shows the statistical preferences for a C.3 or O.co2 ligand atom to be exposed in the protein-bound state compared to the unbound state.

Except for very small surface portions of remaining solvent exposed at the binding site, complete burial of C.3 atoms in protein-ligand complexes is strongly favorable for complex formation.

A completely different, although quite reasonable behavior is observed for O.co2-type atoms. In the complexed state, the distribution results as a compromise between several effects. On the one side, burial of O.co2 atoms diminishes the SAS. On the other side, surface portions contacting polar protein atoms are not considered as being buried. Hence, 'partial' burial for O.co2 results as best compromise for the complexed state. Expressed in terms of preferences, a 'partial' burial favors complex formation in contrast to a complete burial.

'Scoring scoring functions'

To assess the quality of a generated docking solution in terms of its deviation from the experimental structure, we define those ligand poses as 'well docked' that deviate by not more than 2.0 Å root-mean-square deviation (rmsd) from the crystal structure. Visual inspection of a number of test cases showed that within this limit the generated solutions resemble the native binding mode. (For detailed statistics on ligand poses with rmsd <1.0 Å and <1.5 Å from the crystal structure, see Table 1.)

To test the discriminatory power of a scoring function in rendering prominent well-docked solutions from obvious artifacts, we checked how often 'well-docked' solutions correspond to the *best* rank. This criterion meets the requirement for virtual screening of large databases: due to the immense data flow a detailed investigation of only the best hits will be feasible and has to focus on the best ranked solutions. Assuming that the total statistical preference renders prominent the native protein-ligand pose, the experimental solutions should be ranked best. While applying this most stringent criterion one has to keep in mind that crystal structures of proteins are determined to limited resolution. Accordingly, we alleviate our criterion and allow also solutions deviating by less than 2.0 Å from the experimentally given structure to be ranked better than the crystal structure.

The precision of predicted binding affinities is limited by the accuracy of the experimentally determined ones. For data obtained from different laboratories, an error of one order of magnitude with respect to the binding constant is assumed [50,51] whereas a mean error of perhaps 20% is found for sets of related ligands binding to the same protein measured in one laboratory by the same person in an assay with unchanged conditions.

Correlation of the calculated scores versus rmsd of the crystal structure

Total preferences were calculated (Equation 3) for docking solutions obtained by FlexX. In Figure 4 the correlations of calculated scores versus the rmsd from the crystal structure for four protein-ligand complexes are displayed; the individual scores were normalized to fall between 0 and 100.

In all cases, the crystal structure (rmsd = 0) corresponds to the best score. For 1bbp, 1lst and 2ada, the native geometry is well separated from any computed solution.
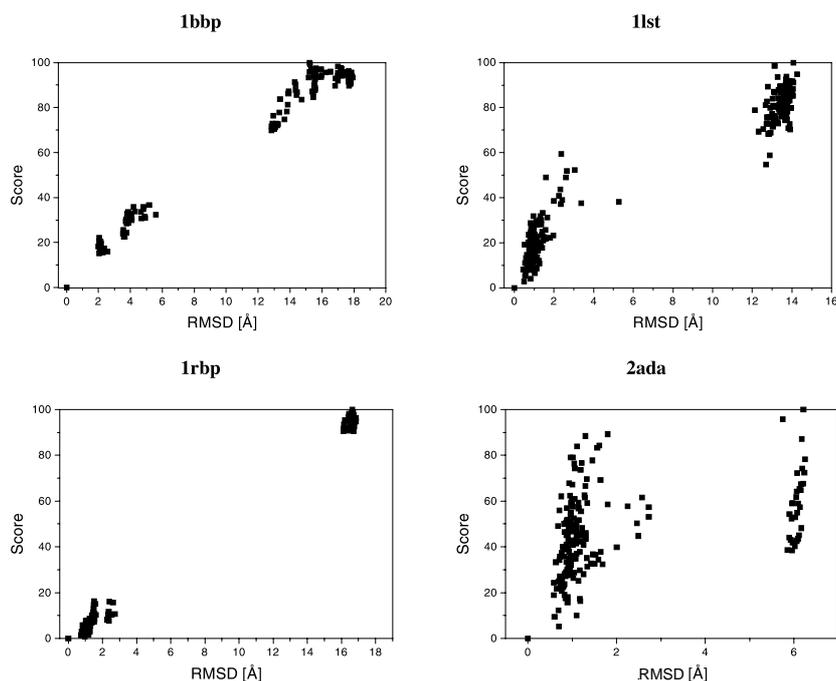
Figure 4. Correlations of scores, normalized to values between 0 and 100, versus the rmsd from the crystal structure for the protein-ligand complexes 1bbp, 1lst, 1rbp, and 2ada are depicted. The percentage of nonpolar (i.e. carbon) ligand atoms varies from 52% for 2ada to 95% for 1rbp, the number of rotatable bonds from 1 (1rbp, 2ada) to 6 (1bbp) and the number of solutions generated by FlexX from 99 (1rbp) to 289 (1lst). Favorable ligand geometries correspond to low scores on the ordinate.

Validation of the approach using two test sets of 91 and 68 protein-ligand complexes

The successful reproduction of some test cases indicates the scope of the method; however, to rigorously validate our method we studied two sets of protein-ligand complexes.

Both test sets are taken from the sample used to validate FlexX [52]. The considered ligands cover a broad range of chemical diversity. In the first case (91 complexes), FlexX has already recognized a generated solution with rmsd <2.0 Å on the first rank in 54% of the cases. For the remaining 46%, FlexX also generates a geometry with rmsd <2.0 Å, however, not ranked as best solution.

The second test set contains 68 additional complexes. Out of these, for 28 cases a computed solution with an rmsd <2.0 Å is found by FlexX on rank 1.
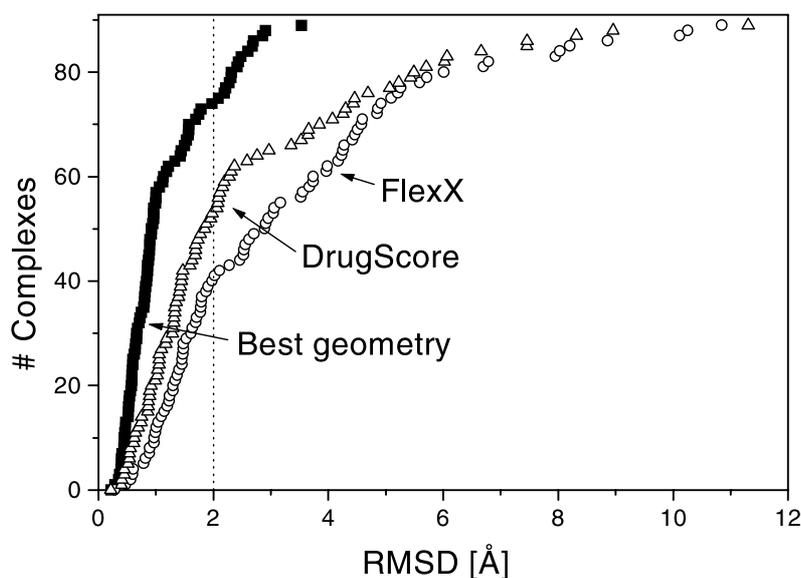
Figure 5. Accumulated number of complexes as a function of the rmsd from the crystal structure found for ligand poses on rank 1 scored by FlexX (○) and DrugScore (△), respectively. The number of complexes with the best geometry found on any rank by FlexX is depicted as solid squares.

For 38 remaining cases, FlexX did not generate a pose with rmsd <2.0 Å using default settings. This second set was not involved in the development and parameter adjustment and served as a validation set for the scoring function.

All complexes used for validating have been excluded from the database used to derive the probability distributions and statistical preferences.

Recognition of 'well-docked' solutions

Figure 5 summarizes the accumulated number of complexes plotted versus the rmsd with respect to the crystal structure of the best ranked ligand pose either determined by FlexX or by DrugScore. In addition, the FlexX solution with the smallest rmsd disregarding its actual rank is plotted. It gives an idea how well an ideal scoring function could perform and shows how often FlexX generates well-docked solutions for the present data set.

Apparently, the new scoring function performs significantly better than the one implemented in FlexX. Table 1 gives the percentages of cases found on rank 1 with an rmsd of <1.0, <1.5, <2.0 and ≥2.0 Å compared to the best approximating geometry found on any rank. With respect to identifying well-
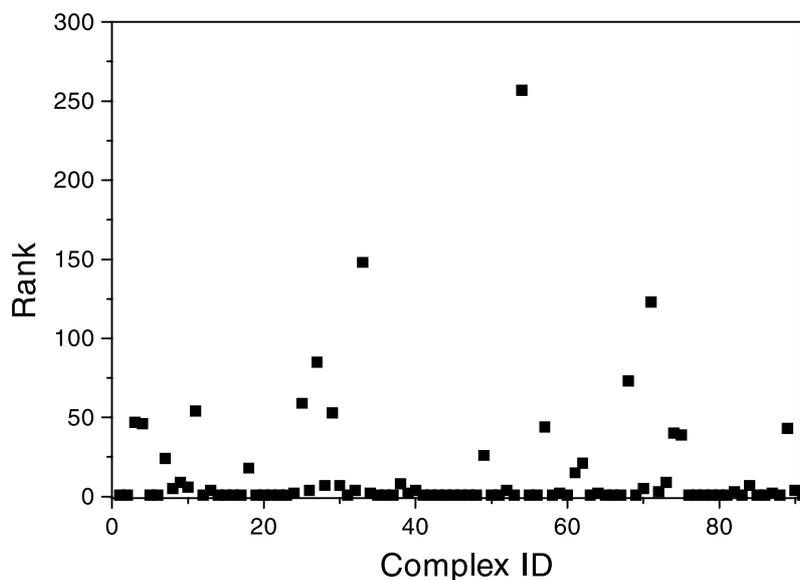
128

Figure 6. The rank of the crystal structure calculated by DrugScore among all decoy geo-
metries generated by FlexX for each of the 91 protein-ligand complexes of the first test set is
shown.

docked solutions on rank 1, FlexX succeeds in 54% of the cases whereas
DrugScore detects 73%.

For the second test set (68 complexes), out of the 30 satisfactorily docked
solutions, FlexX and DrugScore are equally successful (93 and 92%, respect-
ively).

Recognition of the crystal structures
As mentioned above, assuming that the crystallographically determined struc-
tures are 'optimal' solutions, they should obtain the best scoring compared to
any computer-generated ligand pose. As shown in Figure 6, this is actually
the case for 54% of the examples of the first test set (91 complexes). If we
alleviate the criterion to well-docked solutions, in 71% of the cases the crystal
structure or a close-by solution is ranked best by DrugScore. Applying the
scoring function to the second test set, 65% of all crystal structures are found
on rank 1. Here the alleviated conditions are even fulfilled in 90% of all cases.

Prediction of binding affinities

To assess the predictive power of DrugScore to estimate binding affinities, a scoring value is calculated (Equation 3) for different sets of protein-ligand complexes taken from literature [15,16,50]. The obtained values are scaled versus the experimental results (Equation 4). and the squared regression coefficient, the standard deviation and the maximal deviation are determined.

Two different tests are performed for reason of comparison. DrugScore is applied to complexes of known X-ray structure and binding affinities. Additionally, to asses how well the near-native binding modes are recognized and their affinities are predicted, a series of ligands is docked with FlexX.

Predicting binding affinities for data sets of PDB crystal structures
Data sets of PDB protein-ligand complexes are taken from Eldridge et al. [50] and Böhm [15,16] comprising serine proteases, metalloproteases, L-arabinose binding proteins, endothiapepsins and mixed samples composed of different proteins. Similar data have recently been used by Muegge and Martin [33] to compare their scoring function to SCORE1 [15] and SMoG score [31]. Table 2 and Figure 7 summarize the statistics on the performance of DrugScore on six different data sets.

The set of serine proteases comprises 16 trypsin and thrombin protein-ligand complexes (Figure 7a). The best correlation with $R^2 = 0.86$ is achieved for this test set. Examples of low-binding affinities (pK_i's <2) (1bra, 1tni, 1tnj, 1tnk, 1tnl) are predicted to bind tighter by about 1.5 orders of magnitude.

The second set of metalloproteases contains 15 complexes of carboxypeptidase A, thermolysin, and collagenase (Figure 7b) ($R^2 = 0.70$). The largest deviation is obtained for 6tmn with 3.3 orders of magnitude (see also section on 'Visualization of 'hot spots' in protein binding pockets').

The third set comprises 11 endothiapepsin complexes. A weak correlation of $R^2 = 0.30$ is found. Nevertheless, the weakest binding ligand (PD125754 in 1eed) as well as the strongest one (H-261 in 2er7) are correctly recognized compared to examples with intermediate affinity (pK_i's between 6 and 8).

Set no. 4 consists of 9 arabinose-binding protein-ligand complexes. Since each of the nine crystal structures contains two epimers of the sugars, for both of them scoring values are calculated separately. However, the computed values for each pair of epimers are nearly identical. Thus, even though different amounts of α- or β-form of each sugar have been present during experimental affinity determinations, due to the mutarotation equilibrium this fact has no consequences for our computer experiment. The correlation obtained is the worst of all data sets with $R^2 = 0.22$, although the standard deviation only amounts to 0.75 log units. Including the crystallographically determined water molecules 309, 310 and 311 does not yield any improvement. These

Table 2. Statistical parameters of the correlations between experimentally determined binding affinities and those calculated by DrugScore

Data set	No. of complexes	pK_i range	R^2	SD[a]	MD[a]	$\log(-c_s)$[b]
Serine proteases	16	7	0.86	0.95	1.50	−1.55
Metalloproteases	15	10	0.70	1.53	3.32	−1.49
Endothiapepsins	17	4	0.30	0.94	1.67	−1.86
Arabinose-binding proteins	18 (9)[c]	3	0.22	0.75	1.30	−1.23
'Others'[d]	17	8	0.43	1.85	3.39	−1.53
Böhm1998[e]	71	13	0.33	2.21	7.22	−1.48
Böhm1998 (I)[f]	49	10	0.44	1.79	4.52	−1.44
Böhm1998 (II)[g]	46	10	0.56	1.53	4.36	−1.44
'Mixed set'[h]	55	10	0.44	1.80	4.27	−1.43
Thrombin/trypsin inhibitors[i]	64	5	0.48	0.71	1.25	−1.59
Thermolysin[j]	61	10	0.35	1.70	4.00	−1.59
Thermolysin (I)[k]	43	10	0.43	1.68	3.90	−1.58
Thermolysin (II)[l]	15	5	0.36	1.53	3.00	−1.58
Thermolysin (III)[m]	14	5	0.50	1.39	3.27	−1.60

[a] The standard deviation (SD) and the maximum deviation (MD) are given in units of pK_i.
[b] c_s is calculated as described in the text.
[c] Each epimer found in the crystal structure of 9 arabinose-binding proteins is treated separately.
[d] This set contains complexes of the combined training and test sets of Böhm's work [15] that were not found in the four data sets above and show a resolution of less than or equal to 2.5 Å.
[e] Only protein-ligand complexes found in the PDB are taken from the training and test set of Böhm's work [16].
[f] Subset of Böhm1998 that contains only complexes that show a resolution less than or equal to 2.5 Å and comprise ligands with less than or equal to 40 non-hydrogen atoms.
[g] Subset of Böhm1998(I) excluding 3 outliers: 1cil, 1sbp, 1tnk. See text for explanation.
[h] Subset of both validation sets of DrugScore for recognition of near-native geometries for which inhibition constants were found in literature. Binding affinities were calculated for the best ranked ligand geometries generated by FlexX of each ligand.
[i] Set of related inhibitors of thrombin and trypsin, taken from [54] for which ligand geometries are generated by FlexX via docking the compounds into the protein structure taken from Obst et al. [55] or from 1pph.
[j] Sixty-one thermolysin inhibitors taken from the training set reported in [56] for which ligand geometries are generated by FlexX via docking the compounds into the protein structure 1tlp.
[k] Subset of 'Thermolysin' containing only those ligands for which a reasonable geometry was generated by FlexX.
[l] Fifteen thermolysin inhibitors taken from the test set in [56] for which ligand geometries were generated as described (j).
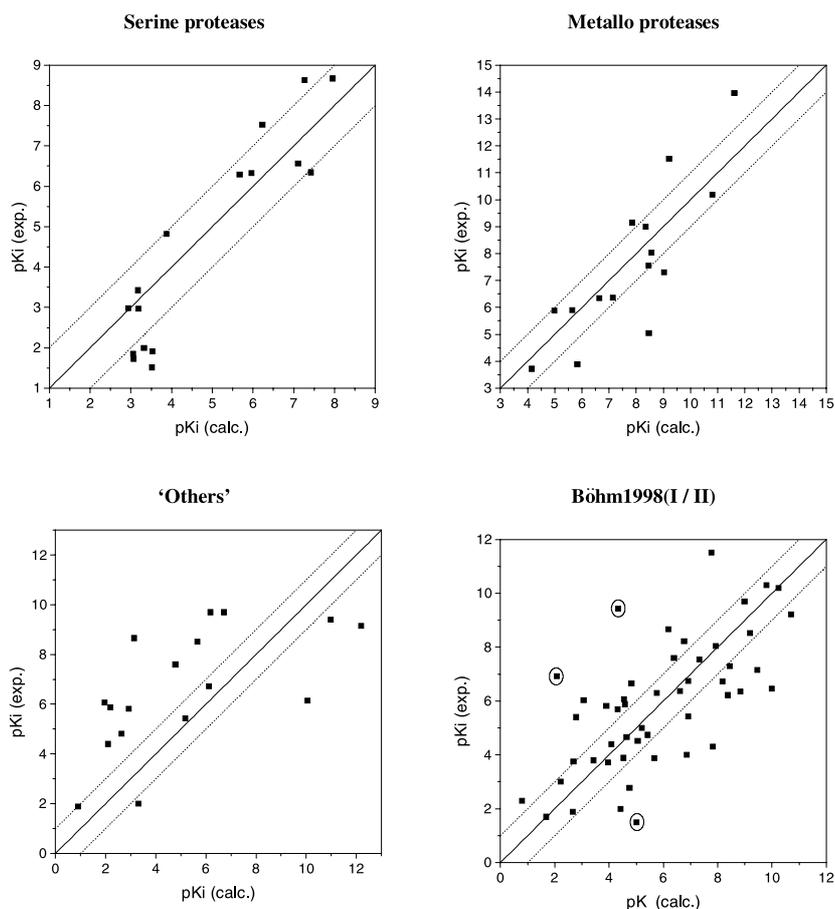[m] Subset of Thermolysin(II) excluding the outlier ZGPLNH2.

Figure 7. Correlation plots of experimentally determined pK_i values versus calculated ones by DrugScore for four different sets of X-ray protein-ligand complexes as defined in the text: serine proteases (a), metallo-proteases (b), 'others' (c), and Böhm1998(I/II) (d). For the latter, the three outliers 1cil, 1sbp, 1tnk are highlighted (circles). Together with the ideal correlation line, deviations by ± 1 log units are depicted.

waters are considered to contribute to the specificity of the protein towards the binding of L-arabinose, D-fucose and D-galactose [53].

Data set 5 ('others') comprises 17 protein-ligand complexes formed with different proteins. These data were also used by Muegge and Martin (Figure 7c). Including cofactors as parts of the protein, we obtain a correlation of $R^2 = 0.43$. Obviously, three examples are predicted too high: the renin-complex 1rne as well as the HIV-proteases 4hvp and 4phv with 51, 54 and

46 non-H atoms in the ligands, respectively. Since DrugScore has been parameterized only for ligands composed of less than 50 non-hydrogen atoms, which is considered to be an upper limit for the size of drug-like molecules, information concerning ligands with a size close to or beyond this limit could be regarded as incompletely covered in our analysis.

Finally, a composed data set has been selected considering the X-ray data used by Böhm in his SCORE2 study [16] (Figure 7d). Any modeled complexes used by Böhm have been discarded. This set (Böhm1998) of 71 complexes yields a correlation of $R^2 = 0.33$. Excluding all complexes with a resolution >2.5 Å or possessing ligands with more than 40 non-H atoms (*vide supra*) improved the correlation to $R^2 = 0.44$ (Böhm1998 (I)). Three remaining outliers can be explained: in 1sbp the ligand is a sulfate, thus falling beyond the lower limit of 6 non-hydrogen atoms used to parameterize DrugScore. In 1cil (inhibitor ETS bound to carbonic anhydrase II), the sulfonamide nitrogen bound to the zinc ion is deprotonated and yields a strong ionic interaction to the metal. DrugScore treats this interaction as a contact between a normal amide nitrogen and a zinc, thus underestimating its contribution to the binding. Finally, the ligand in 1tnk shows quite unusual *intra*molecular dimensions supposedly resulting from some deficiencies in the final refinement process. Removing these three entries results in a significantly improved regression with $R^2 = 0.56$ and SD = 1.53 (Böhm1998 (II)).

The performance of DrugScore is compared to other scoring functions in terms of some statistical descriptors determined for the first five test sets (data for PMFScore, SCORE1, SMoGScore taken from [33]; for ProteusScore taken from [50]). Figure 8 depicts the standard deviation for each of these test sets; for arabinose-binding proteins, Muegge and Martin report a value of 69.7 in the case of SCORE1 while there are no results given by Eldridge *et al.* for the 'others' test set. Compared to both knowledge-based scoring functions (PMFScore and SMoGScore), DrugScore performs better, resulting in a lower standard deviation for all four test sets considering one particular protein target. For the mixed data set ('others') PMFScore performs better. Comparing the regression-based scoring functions (SCORE1 and ProteusScore) with respect to DrugScore, the latter is superior to SCORE1 in all reported cases while ProteusScore obviously performs better in the case of the arabinose-binding proteins.

Predicting binding affinities for ligand geometries generated with FlexX
The examples used so far are all based on experimentally determined binding geometries. Obviously, the scoring function performs satisfactorily for experimentally given geometries. However, in virtual screening, affinity predictions are attempted based on modeled binding modes. This is in partic-
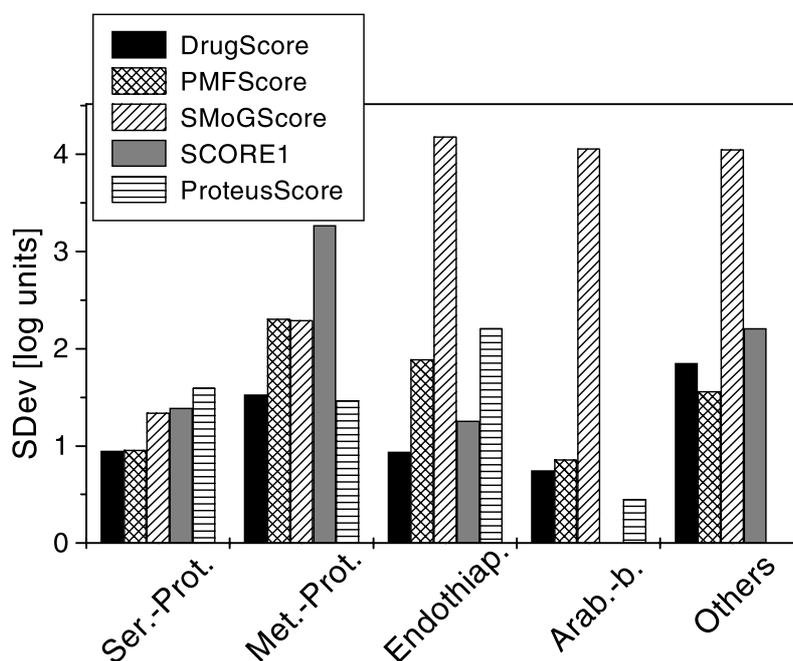
Figure 8. Comparison of different currently developed scoring functions with DrugScore's performance to predict binding affinities in terms of their standard deviation. Five sets of X-ray protein-ligand complexes are used as defined in the text: serine proteases, metallo-proteases, endothiapepsins, arabinose-binding proteins, 'others'. Data for PMFScore, SCORE1, and SMoGScore are taken from Muegge and Martin [33], for ProteusScore from Eldridge et al. [50] (except for test set 'others'). In the case of arabinose-binding proteins, scored by SCORE1, a value of 69.7 is reported [33].

ular the case if a docking tool is applied. Here, the scoring function has to render prominent the most likely geometry from a large sample of generated geometries.

To test the performance of DrugScore with respect to virtual screening, we calculated scoring values for ligand binding geometries generated by FlexX for (a) a sample of 55 complexes taken from the combined validation sets for DrugScore to render prominent well-docked solutions, (b) a congeneric series of 32 inhibitors [54] docked into thrombin [55] or trypsin (1pph), as well as (c) a series of 61 and 15 thermolysin inhibitors (taken from the training and test sets of [56], respectively) docked into the protein structure 1tlp.

For the sample (a), a standard deviation of 1.8 log units and a squared correlation coefficient of 0.44 (data not shown) is achieved. For comparison,
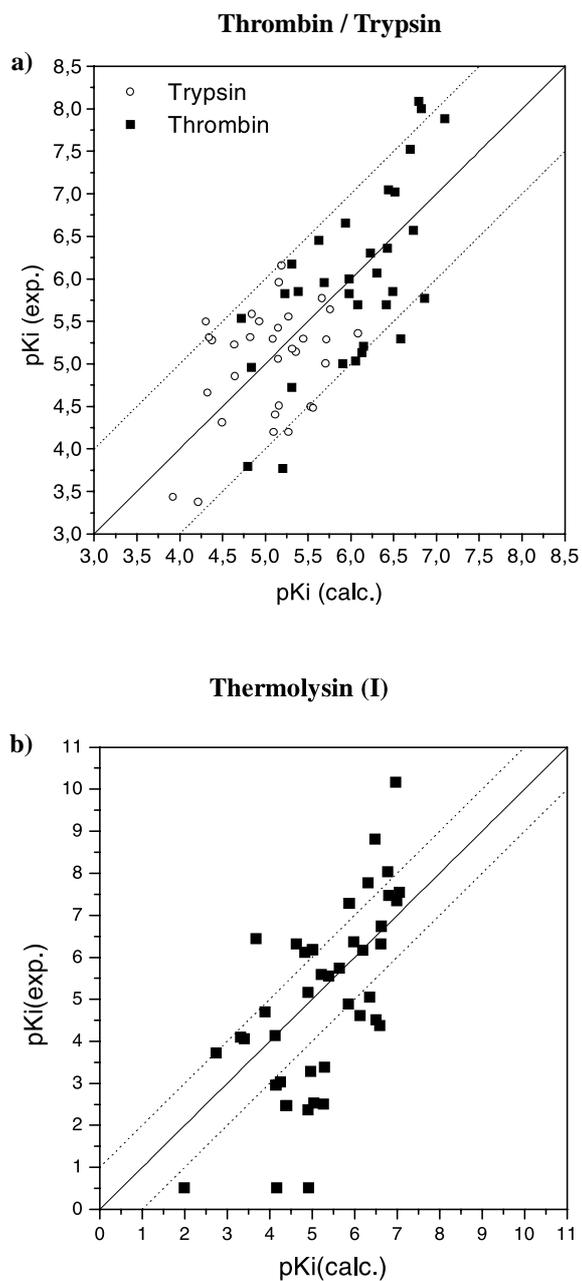
Thrombin / Trypsin

Thermolysin (I)

Figure 9. Correlation of experimentally determined pK_i values versus calculated ones by DrugScore for (a) a set of thrombin and trypsin inhibitors docked into thrombin (taken from [55]) and trypsin (taken from 1pph) and (b) a set of thermolysin inhibitors [56] docked into the protein structure 1tlp, respectively, by FlexX. In the latter case, only those ligands are displayed for which FlexX finds a reasonable geometry (Thermolysin(I)). Together with the ideal correlation line, deviations by ±1 log units are depicted.
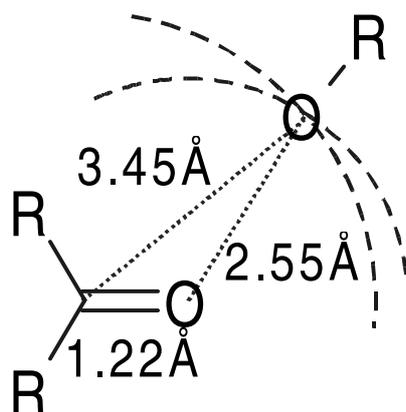
Figure 10. Intrinsic geometrical constraints reflected by the atom pair preferences of O.2-O.3 and C.2-O.3. Given the minima of the statistical pair preferences (O.2-O.3: 2.55 Å; C.2-O.3: 3.45 Å) and the bond length (C.2-O.2: 1.22 Å), the C.2-O.2-O.3 angle is calculated to be 128°.

the squared correlation coefficient for scoring values calculated for the *crystal* geometries of this set amounts to 0.34. Thus, in the present case the affinity prediction based on computer-generated geometries is more precise than using the crystal coordinates. However, we anticipate that is not usually the case. The affinity predictions for the thrombin and trypsin inhibitors (sample (b), Figure 9) deviate from the experimental *pKi* values by 0.7 log units and yield a squared correlation coefficient of 0.56. The predictions are of the same quality for thrombin and trypsin. In the case of the thermolysin inhibitors (c), a squared correlation coefficient of 0.35 is calculated for the total set of 61 ligands ('Thermolysin') (data not shown). If the set is restricted to only those ligands where FlexX at least predicts a geometry with less than 3 Å rmsd from modeled reference structures (using crystal geometries as templates for the modeling), the R^2 value increases to 0.42 (standard deviation: 1.68 log units) (Thermolysin(I), 43 cases, Figure 9). Considering the set of 15 thermolysin inhibitors (Thermolysin(II)), an R^2 value of 0.36 is calculated. Excluding ZGPLNH2 as outlier from the latter set (Thermolysin(III)), a squared correlation coefficient of 0.50 (standard deviation: 1.39) is found.

Implicit directionality of spherical-symmetric pair-potentials

Since the compiled preferences for a given atom pair are calculated solely as a function of the mutual *distance*, any directional features can only be implicitly contained in the derived pair-potentials. As an example, the hydrogen-bond between a carbonyl group and an O.3-type oxygen should be considered

(Figure 10). The most favorable interaction for O.2-O.3 occurs at a mutual distance of 2.55 Å, for the C.2-O.3 interaction at 3.45 Å. Starting with these values and assuming a C.2-O.2 bond length of 1.22 Å, a C.2-O.2-O.3 angle of 128° is calculated, well in agreement with the expected bond angle. A similar orientational preference is observed if representative fragments stored in ISOSTAR are consulted [57]. Additional contact preferences formed by the neighboring atoms will further constrain the spatial arrangement of a specific directional interaction.

Visualization of 'hot spots' in protein binding pockets
A regularly spaced grid is generated inside the binding site and scoring values are calculated at every grid point using different ligand-atom types. The results are contoured individually for each atom type.

Arabinose-binding protein (1abe) binds both epimers of arabinose preferentially to other sugar ligands [53]. The isocontour surface for C.3 (comprising values 10% above the minimum) encompasses all ligand-atom positions of this type as found in the crystal structure (Figure 11). The O.3 contour (10% level) depicts three favorable regions in space; all are occupied by hydroxyl groups of the ligand in the crystal structure. Interestingly, the contour for oxygen O-1 extends over a range where actually the oxygens of the α- and β-epimer bind. O-2 and O-5 do not coincide with regions contoured as most favorable. The fact can be explained because O-2 orients towards the solvent and no favorable interactions with the protein can be determined.

Contours encompassing scoring values calculated for C.3, O.3, and N.am (10% level) inside the binding pocket of thermolysin (5tmn) are displayed together with the phosphor-analogue (ZG^{P}LL) of the peptide carbobenzoxy-Gly-Leu-Leu (Figure 11). For the phosphonamidate, Bartlett and Marlowe [58] reported a K_i value of 9.1 nM while the phosphonate isomer (substitution of P-NH by P-O: ZG^P(O)LL) resulted in 990 times weaker binding. In contrast, for the phosphinate analogue (substitution of P-NH by P-CH$_2$: ZG^P(C)LL) an inhibition constant of 180 nM is reported [59], only 20 times weaker than the phosphonamidate. These findings were attributed mainly to solvation effects and the potential to form a hydrogen bond between the ligand atom adjacent to phosphorus and the carbonyl oxygen of Ala 113. Interestingly, DrugScore highlights favorable binding both for an N.am and a C.3 atom at a position next to the ligand atom adjacent to P. O.3 is not favorable at this position. However, for this atom-type, promising regions coincide with the terminal oxygens of the $- PO_2^-$ – moiety binding to zinc.

The contours described are displayed on a 10% level for each atom type thus resulting in a *relative* description. However, a per-atom contribution to the total score using C.3, O.3, and N.am, respectively, yields an *absolute* affin-
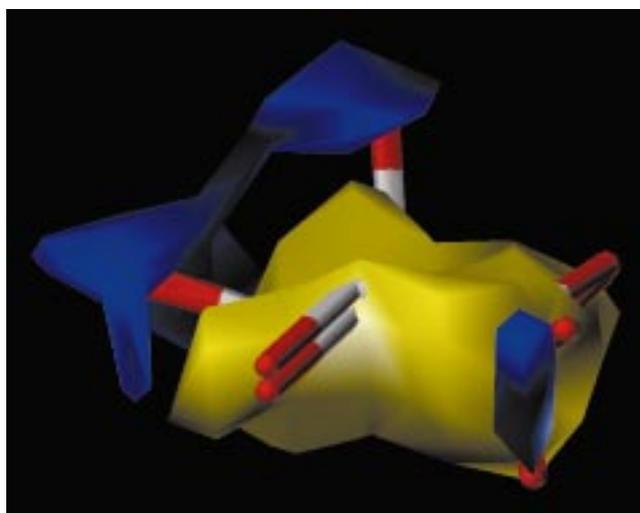
Figure 11. Isocontour surfaces encompassing values 10% above the global minimum of all scoring values on a 0.5 Å spaced grid for several atom types are depicted. Grid values are calculated by DrugScore inside the binding pockets of 1abe (top) and 5tmn (bottom) for different ligand probe atoms. In the case of 1abe, the α- and β-epimers of arabinose are shown. The surfaces are color-coded as: dark blue (sp^3-hybridized oxygen), yellow (aliphatic carbon) [1abe] and cyan (sp^3-hybridized oxygen), yellow (amide nitrogen), magenta (aliphatic carbon) [5tmn]. For 5tmn, the arrow indicates the phosphonamidate nitrogen of ZG^{P}LL, which is substituted by oxygen in the phosphonate ZG^P(O)LL and by carbon in the phosphinate ZG^P(C)LL.

ity description. The values for C.3 (-11.11) and N.am (–9.33) at the positions adjacent to P deviate by only 15%, however, that of an O.3 (–4.34) at the same position contributes much less to binding affinity. As a consequence, the phosphonamidate ZG^PLL (5tmn) and the phosphinate $ZG^P(C)LL$ (modeled from 5tmn by N.am/C.3 exchange) obtain comparable total scores while the phosphonate $ZG^P(O)LL$ (6tmn) is predicted to bind weakest. The rank ordering is qualitatively predicted correctly, however, the phosphonate is yet predicted too high in affinity (*vide supra*). Supposedly, this can be attributed to the simultaneous consideration of hydroxyl- and ether-oxygens in the general O.3 atom-type used in our compilation of the statistical potentials. Accordingly, in our potentials, the unfavorable ester-oxygen to carbonyl-oxygen interactions are overwhelmed by other, more favorable O.3/O.2 contacts.

Correspondence of 'hot spots' and observed ligand atom types
In order to determine how often 'hot spots' detected by DrugScore actually match with ligand atom types, 159 crystallographically determined protein-ligand complexes were analyzed. In principle, DrugScore can be computed analytically at every position of the binding site, accordingly also at the crystallographically determined positions of the ligands under investigation. However, to use a more general approach in particular with respect to the analysis of unoccupied binding sites, we estimate the scoring by extrapolating from precalculated values assigned to the intersections of a 0.5 Å grid.

In a first step, we focus on fully buried ligand atoms only and test how frequent these atoms fall next to local minima in DrugScore for the five atom types C.3, O.3, O.2, O.co2, and N.3. By selecting these atom types, we intended to consider a small set of typical representatives for distinct types of interactions (hydrophobic, H-bond donors/acceptors, positively/negatively charged). In addition, this choice is close to the one used for the validation of SuperStar [38] and allows a direct comparison (*vide infra*). The results are summarized in Table 3.

Assuming equal weights for all five atom types and full coincidence of ligand atomic positions with local minima of DrugScore, in the worst case a chance prediction of 20% would result. However, significantly higher prediction rates are observed. In particular, aliphatic carbon and amino nitrogen atoms (also implicitly including ammonium groups) are correctly predicted in 92 and 73% of all cases, respectively, while a carbonyl oxygen atom is recognized only in 27% of the cases. For O.3 and O.co2 types, the rates amount to 37 and 46%, respectively. In total, an overall prediction rate of 74% is achieved. Grouping similar atom types together ('hydrophobic': C.3; 'hydrophilic': O.3, O.2, O.co2, N.3), in 86% of all cases the correct or a similar atom type is recognized. Assuming that most of the N.3-type atoms (amino

Table 3. Statistics on the prediction rates of buried ligand atom types suggested by Drug-Score compared to atom types observed in crystallographically determined protein-ligand complexes at the same spatial positions

Actual	Predicted						Correct	H. phob./ h. phil.[b]	Sim. Interact.[c]
	No.	C.3[a]	O.3[a]	O.2[a]	O.co2[a]	N.3[a]			
C.3[a]	745	92%	3%	<1%	<1%	4%		92%	92%
O.3[a]	168	40%	37%	5%	7%	11%		60%	60%
O.2[a]	124	18%	26%	27%	23%	6%		82%	76%
O.co2[a]	67	6%	27%	17%	44%	6%		94%	88%
N.3[a]	15	20%	7%	0%	0%	73%		80%	80%
Overall	1119						74%	86%	85%

[a] Atom types are according to the SYBYL notation: aliphatic carbon (C.3), sp^3-hybridized oxygen (O.3), carbonyl-oxygen (O.2), carboxyl(ate)-oxygen (O.co2), (protonated) amino-nitrogen (N.3).
[b] Atoms are grouped separating hydrophobic – hydrophilic properties: C.3 versus O.3/O.2/O.co2/N.3.
[c] Atoms are grouped showing similar type of interaction (N.3 is considered to be protonated): C.3 (hydrophobic); O.3, N.3 (hydrogen-bond donors); O.3, O.2, O.co2 (hydrogen-bond acceptors).

nitrogen) are protonated in proteins, atom types can be grouped according to a possible interaction type of the atom under investigation (C.3 (hydrophobic); O.3 (donor + acceptor); O.2, O.co2 (only acceptor); N.3 (only donor)). With this classification, in 85% of all cases the correct interaction type is predicted.

The recently described tool SuperStar [38] only considers four distinct probes (aliphatic carbon; hydroxyl-, carbonyl-oxygen; protonated amino nitrogen) for a similar validation study. Following the above-described arguments, a chance prediction of 25% has to be assumed. For a test set of 122 protein-ligand complexes, Superstar detects the correct atom type in 82% of the cases if only solvent-inaccessible ligand atoms are considered. This figure increases to 90% if 'similar' atom types are admitted. We included the carboxylate oxygen to consider an atom type most likely bearing a negative charge. It is the contrast to an amino nitrogen most likely being positively charged. Indeed, if our approach suggests a carboxylate oxygen as the most favorable type, an amino nitrogen is the atom type found the least frequent at this position by experiment. The same holds for the reverse.

Discussion and conclusions

In this study, distance-dependent pair preferences and SAS-dependent singlet preferences are derived from crystallographically determined protein-ligand complexes. The scoring function DrugScore incorporates both terms and shows very promising results. It discriminates satisfactorily between well-docked (rmsd <2.0 Å) ligand binding modes and largely deviating ones generated by the docking tool FlexX. This is demonstrated for two test sets comprising 91 and 68 complexes, respectively. A substantial improvement of 35% is achieved compared to the original FlexX scoring.

DrugScore's ability to predict binding affinities is assessed by correlating experimentally determined pK_i values with the computed scores. Protein-ligand complexes taken from the PDB as well as sets of docked ligands were investigated. Compared to currently applied scoring functions, DrugScore reveals lower standard deviations.

Most remarkably, the composite picture of spherical pair-potentials in DrugScore exhibits implicitly information about the directionality of interactions. It possesses predictive power to suggest positions in space that are most favorable for particular ligand-atom types.

Knowledge-based approaches are assumed to be general since they implicitly incorporate even those effects that are yet not fully understood. Converting structural database information into statistical preferences considers entropy effects arising from cooperativity and changes of solvation due to their mean-field character. Moreover, less frequently populated states are considered with lower statistical preferences, thus implicitly penalizing computer-generated artifacts. Additionally, since no explicit training set is used in contrast to the derivation of e.g. regression-based scoring functions, our scoring function should be generally applicable.

Hydrogen atoms are not explicitly considered in our scoring function. Most complexes in the PDB either lack or contain only force-field assumed hydrogen atoms. However, in particular the positions of polar hydrogen atoms strongly depend on the influences of their molecular environment. Changes of the electrostatic field of a protein might result in substantial pK_a shifts of ionizable groups upon ligand binding. In consequence, defining protonation states *a priori* e.g. during a docking experiment is by no means straightforward. Although at first glance, the neglect of H-atom positions appears to imply the loss of information about the directionality of polar interactions, the composite consideration of many-fold pair-preferences in a compact molecular environment recovers these features (Figure 10).

By visualization of calculated hot spots, we definitely demonstrated this directionality in protein-ligand interactions to be implicitly included in our

distance-dependent pair-potentials. Similar considerations about anisotropic interactions resulting from the summation of individually isotropic contributions led to the correct description of e.g. directional hydrogen bonds by taking into account only Lennard-Jones-type and electrostatic interactions in force fields [60]. We believe that this important property of our approach can be attributed to the comparatively short upper limit of 6 Å considered during the compilation of our potentials. While binding affinity is largely determined by the amount of buried non-polar surface, the specificity of ligand binding is mainly attributed to directional interactions such as hydrogen bonds [48].

The graphical display of a knowledge-based scoring function in terms of 'hot spots' in a binding pocket suggests further applications. Highlighting the regions of a binding site where a particular type of ligand atom appears to be most favorable allows one to use them as an interactive design tool. *Nota bene*, these 'hot spots' do not simply represent regions of favorable *energy* but also include *entropical* contributions. Additionally, this information should be used in a docking tool to drive the initial ligand placement.

For comparison, data in the Cambridge Structural Database (CSD) [61] can be used to derive statistical preferences of intermolecular interactions [57]. They allow to develop a scoring function to discriminate different computer-generated crystal packings [62]. Furthermore, additional data call upon a more sophisticated consideration of atom types. However, mixing data from the PDB with those from the CSD involves the following fundamental problem: protein-ligand complexes are usually crystallized from water whereas the overwhelming part of organic small molecules are crystallized from organic solvents. As a consequence, for a quantitative correlation as anticipated in our analysis the influence of the hydrophobic effect is expected to be smaller in the data derived from the CSD compared to those from the PDB. This influence was first recognized by Verdonk et al. during the development of SuperStar [38].

The recent studies of Verkhivker et al. [28], Muegge et al. [33] and Mitchell et al. [35] use a similar formalism to derive potentials. These approaches are difficult to compare since hardly any of these studies elucidate the discriminative power to render prominent the native pose. In our opinion, this is the most crucial prerequisite *prior* to an estimation of binding affinities in virtual screening. A scoring function, demonstrated to operate satisfactorily on crystal structures, does not necessarily handle computer-generated, often artificial and incorrect binding modes equally well.

To compare different scoring functions with respect to the prediction of absolute binding affinities, one has to remember that R^2 values (but not the standard deviations) heavily depend on the composition of the data sets considered. Accordingly, we compared DrugScore's performance with respect

to other scoring functions in terms of their achieved standard deviations. Nevertheless, the trends observed in Figure 8 are similarly reflected once the R^2 values are computed, since identical data sets were used for comparison. Noteworthy, except for the mixed set ('others'), DrugScore performs better than the other available knowledge-based approaches. Supposedly, for virtual screening applications, it is more important to correctly predict the binding affinity of *different* ligands with respect to *one* selected protein than to rank correctly mixed sets of various protein-ligand complexes. In this study, the standard deviation for the docked thrombin and trypsin inhibitors falls below one log unit. This demonstrates DrugScore's power to predict binding energies for computer generated ligand geometries. Nevertheless, since the measured pK_i values are assumed to be affected by an even smaller error limit, the experimental accuracy is not yet matched. The larger standard deviations of 1.5 to 1.7 log units in the case of the docked thermolysin inhibitors have to be discussed in view of the supposedly much higher experimental data scatter, since these data were collected from several different sources with deviating assay conditions. In addition, due to an increasing conformational complexity of the ligands, FlexX does not detect reasonable geometries in all of the cases. Scoring unlikely geometries, however, cannot be expected to predict affinity reliably.

Currently, DrugScore is implemented into FlexX. We expect an improvement of the incremental ligand build-up and placement procedure mainly in reducing the number of generated solutions. With respect to the affinity predictions we expect further enhancement once water molecules are included in our considerations.

Acknowledgements

References

1. Muller, K., Perspect. Drug Discov. Design, 3 (1995) v.
2. Walters, W.P., Stahl, M.T. and Murcko, M.A., Drug Discov. Today, 3 (1998) 160.
3. Van Drie, J.H. and Lajiness, M.S., Drug Discov. Today, 3 (1998) 274.
4. Kubinyi, H., Curr. Opin. Drug Discov. Develop., 1 (1998) 4.

5.   Lengauer, T. and Rarey, M., Curr. Opin. Struct. Biol., 6 (1996) 402.
6.   Kuntz, I.D., Meng, E.C. and Shoichet, B.K., Acc. Chem. Res., 27 (1994) 117.
7.   Rarey, M., Kramer, B., Lengauer, T. and Klebe, G., J. Mol. Biol., 261 (1996) 470.
8.   Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., J. Mol. Biol., 161 (1982) 269.
9.   Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor, R., J. Mol. Biol., 267 (1997) 727.
10. Dixon, J.S., Proteins, Suppl. 1, (1997) 198.
11. Beveridge, D.L. and DiCapua, F.M., Annu. Rev. Biophys. Biophys. Chem., 18 (1989) 431.
12. Kollman, P., Chem. Rev., 93 (1993) 2395.
13. Kollman, P.A., Acc. Chem. Res., 29 (1996) 461.
14. Dill, K.A., J. Biol. Chem., 272 (1997) 701.
15. Böhm, H.J., J. Comput.-Aided Mol. Design, 8 (1994) 243.
16. Böhm, H.J., J. Comput.-Aided Mol. Design, 12 (1998) 309.
17. Jain, A.N., J. Comput.-Aided Mol. Design, 10 (1996) 427.
18. Murray, C.W., Auton, T.R. and Eldridge, M.D., J. Comput.-Aided Mol. Design, 12 (1998) 503.
19. Rose, P. W., Scoring methods in ligand design, Proceedings of 2nd UCSF Course in Computer-Aided Molecular Design, San Francisco, CA, 1997.
20. Head, R.D., Smythe, M.L., Oprea, T.I., Waller, C.L., Green, S.M. and Marshall, G.R., J. Am. Chem. Soc., 118 (1996) 3959.
21. Stahl, M. and Böhm, H.-J., J. Mol. Graph. Model, 16 (1998) 121.
22. Vajda, S., Sippl, M. and Novotny, J., Curr. Opin. Struct. Biol., 7 (1997) 222.
23. Jernigan, R.L. and Bahar, I., Curr. Opin. Struct. Biol., 6 (1996) 195.
24. Torda, A.E., Curr. Opin. Struct. Biol., 7 (1997) 200.
25. Mitchell, J.B.O., Laskowski, R.A., Alex, A., Forster, M.J. and Thornton, J.M., J. Comput. Chem., 20 (1999) 1177.
26. Wallqvist, A. and Covell, D.G., Proteins, 25 (1996) 403.
27. Wallqvist, A., Jernigan, R.L. and Covell, D.G., Protein Sci., 4 (1995) 1881.
28. Verkhivker, K., Appelt, K., Freer, S.T. and Villafranca, J.E., Protein Eng., 8 (1995) 677.
29. Sharp, K.A., Nicholls, A., Friedman, R. and Honig, B., Biochemistry, 30 (1991) 9686.
30. Pickett, S.D. and Sternberg, M.J., J. Mol. Biol., 231 (1993) 825.
31. DeWitte, R.S. and Shaknovich, E.I., J. Am. Chem. Soc., 118 (1996) 11733.
32. Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer Jr., E.E., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
33. Muegge, I. and Martin, Y.C., J. Med. Chem., 42 (1999) 791.
34. Muegge, I., Martin, Y.C., Hajduk, P.J. and Fesik, S.W., J. Med. Chem., 42 (1999) 2498.
35. Mitchell, J.B.O., Laskowski, R.A., Alex, A. and Thornton, J.M., J. Comput. Chem., 20 (1999) 1165.
36. Gohlke, H., Hendlich, K. and Klebe, G., J. Mol. Biol., 295 (2000) 337.
37. Hendlich, M., Acta Crystallogr., D 54 (1998) 1178.
38. Verdonk, M.L., Cole, J.C. and Taylor, R., J. Mol. Biol., 289 (1999) 1093.
39. Böhm, H.-J. and Klebe, G., Angew. Chem. Int. Ed. Engl., 35 (1996) 2566.
40. Sippl, M.J., Curr. Opin. Struct. Biol., 5 (1995) 229.
41. Sippl, M.J., J. Mol. Biol., 213 (1990) 859.
42. Sippl, M.J., J. Comput.-Aided Mol. Design, 7 (1993) 473.
43. Godzik, A., Kolinski, A. and Skolnick, J., Protein Sci., 4 (1995) 2107.
44. Miyazawa, S. and Jernigan, R.L., Proteins, 34 (1999) 49.

144

45. Koehl, P. and Delarue, M., Proteins, 20 (1994) 264.
46. Testa, B., Carrupt, P.A., Gaillard, P., Billois, F. and Weber, P., Pharm. Res., 13 (1996) 335.
47. SYBYL, Tripos Inc., St. Louis, MO.
48. Davis, A.M. and Teague, S.J., Angew. Chem. Int. Ed. Engl., 38 (1999) 736.
49. Burley, S.K. and Petsko, G.A., Science, 229 (1985) 23.
50. Eldridge, M.D., Murray, C.W., Auton, T.R., Paolini, G.V. and Mee, R.P., J. Comput.-Aided Mol. Design, 11 (1997) 425.
51. Hosur, M.V., Bhat, T.N., Kempf, D.J., Baldwin, E.T., Liu, B., Gulnik, S., Wideburg, N.E., Norbeck, D.W., Appelt, K. and Erickson, J.W., J. Am. Chem. Soc., 116 (1994) 847.
52. Kramer, B., Rarey, M. and Lengauer, T., Proteins, 37 (1999) 145.
53. Quiocho, F.A., Wilson, D.K. and Vyas, N.K., Nature, 340 (1989) 404.
54. Obst, U., *De novo*-Design und Synthese neuartiger, nichtpeptidischer Thrombin-Inhibitoren, Ph.D. Thesis, ETH Zürich, Zürich, 1997.
55. Obst, U., Banner, D.W., Weber, L. and Diederich, F., Chem. Biol., 4 (1997) 287.
56. De Priest, S.A., Mayer, D., Naylor, C.B. and Marshall, G.R., J. Am. Chem. Soc., 115 (1993) 5372.
57. Bruno, I.J., Cole, J.C., Lommerse, J.P., Rowland, R.S., Taylor, R. and Verdonk, M.L., J. Comput.-Aided Mol. Design, 11 (1997) 525.
58. Bartlett, P.A. and Marlowe, C.K., Science, 235 (1987) 569.
59. Grobelny, D., Goli, U.B. and Galardy, R.E., Biochemistry, 28 (1989) 4948.
60. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
61. Allen, F.H., Davies, J.E., Galloy, J.J., Johnson, O., Kennard, O., Macrae, C.F., Mitchell, E.M., Mitchell, G.F., Smith, J.M. and Watson, D.G., J. Chem. Inf. Comput. Sci., 31 (1991) 187.
62. Hofmann, D.W.M. and Lengauer, T., J. Mol. Model, 4 (1998) 132.