Article No. jmbi.1999.3371 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 295, 337–356

JMB

Knowledge-based Scoring Function to Predict Protein-Ligand Interactions

Holger Gohlke, Manfred Hendlich and Gerhard Klebe*

Department of Pharmaceutical Chemistry, Philipps-University of Marburg, Marbacher Weg 6 D-35032 Marburg, Germany

The development and validation of a new knowledge-based scoring function (DrugScore) to describe the binding geometry of ligands in proteins is presented. It discriminates efficiently between well-docked ligand binding modes (root-mean-square deviation <2.0 Å with respect to a crystallographically determined reference complex) and those largely deviating from the native structure, e.g. generated by computer docking programs. 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 for protein and ligand atoms. Definition of an appropriate reference state and accounting for inaccuracies inherently present in experimental data is required to achieve good predictive power. The sum of the pair preferences and the singlet preferences is calculated based on the 3D structure of protein-ligand binding modes generated by docking tools. For two test sets of 91 and 68 protein-ligand complexes, taken from the Protein Data Bank (PDB), the calculated score recognizes poses generated by FlexX deviating <2 Å from the crystal structure on rank 1 in three quarters of all possible cases. Compared to FlexX, this is a substantial improvement. For ligand geometries generated by DOCK, DrugScore is superior to the ''chemical scoring'' implemented into this tool, while comparable results are obtained using the ''energy scoring'' in DOCK. None of the presently known scoring functions achieves comparable power to extract binding modes in agreement with experiment. It is fast to compute, regards implicitly solvation and entropy contributions and produces correctly the geometry of directional interactions. Small deviations in the 3D structure are tolerated and, since only contacts to non-hydrogen atoms are regarded, it is independent from assumptions of protonation states.

© 2000 Academic Press

Keywords: scoring function; knowledge-based; protein-ligand interactions; docking; virtual screening

*Corresponding author

Introduction

The process of finding novel leads for a new target is the most important and undoubtedly one of the most crucial steps in a drug development program. Today two complementary strategies are followed: experimental high-throughput screening to discover possible leads from large compound libraries, and computational methods exploiting structural information of the protein binding site aiming at the construction of a ligand *de novo* or their discovery by virtual screening of large databases (Kubinyi, 1998; Muller, 1995; Van Drie & Lajiness, 1998; Walters *et al.*, 1998). The latter approaches try to predict, e.g. *via* docking, the actual binding mode of a ligand at the binding site (Kuntz *et al.*, 1994; Lengauer & Rarey, 1996). They can only be applied in the present context if they are (1) fast enough to scan over several hundred to thousand compounds, (2) if they suggest reliable geometries in agreement with experimental knowledge, (3) if the generated multiple solutions (poses) are ranked correctly, i.e. those most closely resembling the experimental structures are scored best, and (4), as an extension, if the Gibbs free energy of binding is predicted reliably. Usually the perform-

Abbreviations used: SAS, solvent-accessible surface; FEP, free energy perturbation; TI, thermodynamic integration; rmsd, root-mean-square deviation.

E-mail address of the corresponding author: klebe@mailer.uni-marburg.de

0022-2836/00/020337–20 $35.00/0

© 2000 Academic Press

ance of such methods is determined by assessing whether the binding geometry of protein-ligand complexes resolved by X-ray crystallography or NMR is reproduced†. This latter validation criterion imposes some preconditions onto the methods being developed. Due to the limited resolution of the structure determination techniques applied, no precise data on protonation states of ligand functional groups and the binding site residues are given. The development of methods in computer-aided drug design has to cope with these inherent inaccuracies and shortcomings.

Meanwhile several of the published docking procedures are fast enough to serve the outlined purpose (Jones *et al*., 1997; Kuntz *et al*., 1982; Rarey *et al*., 1996). The recently performed CASP2 competition (Dixon, 1997) revealed that among the multiple solutions generated by different approaches, in nearly all cases several solutions were found approximating the native pose. Though found in most cases, the pose being most close to the experimentally given situation is often not ranked as the energetically most favorable one. This indicates that a near-native geometry cannot be recognized within a set of poses largely deviating from the crystal structure. Since this step is of utmost importance for the relevance of any computer-assisted lead finding process, we embarked into the development of a new scoring function. The purpose of the work presented here is to correctly identify those computer-generated poses which most closely resemble the native structure, since this is the crucial step prior to any estimation of the binding energy.

Usually binding affinity is quantified by the binding constant K_i assuming thermodynamic equilibrium conditions for the protein-ligand complex formation. Gibbs free energy of binding ΔG^0 is then related to the binding constant by:

$$\Delta G^0 = -RT \ln K_i \qquad (1)$$

At best, ΔG^0 is determined by statistical thermodynamics resulting in a master equation that considers all contributing effects (Beveridge & DiCapua, 1989; Kollman, 1993). Although being theoretically the most convincing approach, elaborate methods such as free energy perturbation (FEP) or thermodynamic integration (TI) are computationally too demanding for the application described above. Even so an explicit treatment of the solvent environment is included, the results obtained by FEP or TI can still suffer from insufficient sampling over contributing conformational states of the system or inaccuracies in the forcefield used (Kollman, 1996). Various levels of approximations have been considered, particularly the treatment of electrostatics, e.g. of the solvent, by continuum methods or the estimation of configurational entropy (Honig & Nicholls, 1995; Warshel & Aqvist, 1991). Generally, with increasing level of approximation, the methods try to partition binding affinity into several additive terms (Dill, 1997).

The partitioning into individual terms or descriptors is a widely accepted assumption for the development of empirical regression-based scoring functions. Usually a number of empirically derived contributions is fitted to a data set of experimental observations (Bohm, 1994, 1998; Jain, 1996; Murray *et al*., 1998; Rose, 1997). In some of these approaches, e.g. SCORE by (Bohm, 1994), the considered terms are selected following physical concepts and aimed at a fundamental understanding of the binding process. Approaches such as VALIDATE (Head *et al*., 1996) are based on the ideas of QSAR and use efficiently the information derived from protein-ligand complexes by combining a heuristic correlation analysis with various molecular 3D descriptors. These approaches achieve a precision of about 1 to 1.5 orders of magnitude when predicting K_i (Bohm, 1994; Head *et al*., 1996). However, any regression analysis suffers severely 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 concept introduced by SCORE has been implemented into the commonly used design tools LUDI (Bohm, 1992) and FlexX (Rarey *et al*., 1996). A detailed analysis of binding modes generated by FlexX for a test set of 200 examples suggests that in about 80 % of the cases a binding geometry close to the native pose is generated among the multiple solutions. Yet, in half of the cases, it is ranked less favorable than other obviously artificial solutions (B. Kramer, M. Rarey & T. Lengauer, unpublished results). Quite similar findings have been reported for the two popular docking tools DOCK and GOLD. As a first consequence, Stahl & Bohm (1998) developed several penalty filters to successfully discard computer-generated artifacts from the list of favorable ligand poses. However, such post-processing filters will only be as complete as the data set that is used for their development contains all possibly influencing effects.

We decided to follow an alternative way to develop a scoring function based on empirical knowledge. Following the ideas of a so-called ''inverse Boltzmann'' distribution (*vide infra*), it is assumed that only those binding modes are favorable that fit to the maxima of distributions of occurrence frequencies among interatomic contacts between particular atom pairs in experimentally determined structures. A scoring function based on this concept is supposed to rank best all ligand poses that are geometrically very similar to the native pose. During its development we decided not to assign proper protonation states to the considered atom types, assuming that the derived statistical preferences implicitly reflect these influences along with any favorable long-range interaction

† In the following the binding geometry found in the experimentally determined structure will be called native pose.

patterns between functional groups. 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 successfully to rank different solutions of the protein-folding problem (Jernigan & Bahar, 1996; Torda, 1997; Vajda *et al*., 1997). Up to now, this approach has only been applied to four case studies for the ranking of different protein-ligand complexes. None of these, however, engaged in identifying near-native poses of *one* ligand with respect to *one* protein. Wallqvist and co-workers (Wallqvist & Covell, 1996; Wallqvist *et al*., 1995) 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. Analyzing ten HIV protease inhibitor complexes, they approximated the free energy of binding to an accuracy of ±1.5 kcal/mol.

Using a data set of 30 HIV-1, HIV-2, and SIV proteases, Verkhivker *et al*. (1995) compiled a distance-dependent knowledge-based pair-potential which was then combined with the hydrophobicity (Sharp *et al*., 1991) and conformational entropy scales (Pickett & Sternberg, 1993) that originally had been developed to explain protein folding and stability. Applied to different HIV-1 protease complexes, differences in the binding affinity could be estimated.

DeWitte & Shaknovich (1996) used a sample of 126 structures from the PDB (Bernstein *et al*., 1977) to develop a set of ''interatomic interaction free energies'' for a variety of atom types. In combination with a Metropolis Monte Carlo molecular growth algorithm, ligands were gradually constructed in the binding site and energetically scored.

Muegge & Martin (1999) explored structural information of known protein-ligand complexes from the PDB and derived distance-dependent Helmholtz free interaction energies of protein-ligand atom pairs. Tested on 77 protein-ligand complexes, the calculated score displayed 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 (Muegge *et al*., 1999).

Here, we describe in detail the development of a new scoring function tailored to discriminate multiple ligand poses. It is based on the vast structural knowledge stored in the entire PDB and retrieved using ReLiBase (Hendlich, 1998). 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. The new function has been applied to data sets of 91 and 68 complexes of known crystal structure. 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. To test the performance of our new DrugScore approach in the context of other docking programs, a subset of 100 protein-ligand complexes was extracted from both data sets and used to generate ligand geometries with DOCK (Ewing & Kuntz, 1997; Makino & Kuntz, 1997). Both, the ''energy'' and ''chemical scoring'' together with our scoring function were applied to rank the resulting binding modes.

Theory

In the following, we shortly summarize the theoretical background of the selected approach.

The non-covalent binding of ligand $L_{aq.}$ to protein $R_{aq.}$ to form a complex $RL_{aq.}$ typically occurs in aqueous environment. The transfer from the solute state with complete separation of both reactants to the complexed state involves either enthalpic and entropic effects (Bohm & Klebe, 1996) that contribute to the Gibbs free energy of binding. In turn, it is related to the complex formation equilibrium:

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 = -RT \ln K_i \qquad (2)$$

Accordingly, for an appropriate description of the binding properties of a $RL_{aq.}$ complex, the consideration of effects resulting from the deliberation of solvent molecules from the binding site or the reorganization of the surrounding solvent must be taken into account (Blokzijl & Engberts, 1993). A treatment only in terms of enthalpic contributions will be insufficient and must, in general, lead to false predictions (Bohm & Klebe, 1996).

At first sight, it seems hardly possible to model entropic effects by solely considering atom pair, triplet, or higher-order interactions. Nevertheless, an implicit description of the complex solute-solvent interactions and solvent entropic effects along with the involved enthalpic contributions resulting from interatomic forces (e.g. electrostatic or van der Waals) is reflected by the formalism used to derive potentials of mean force from database knowledge (Sippl, 1995; Jernigan & Bahar, 1996).

As recently pointed out by Koppensteiner & Sippl (1998), there is a dissatisfying confusion in literature using terms such as ''potential of mean force'' and ''knowledge-based potentials'' when relating database-derived quantities to the Boltzmann equation. The term ''potential of mean force'' has its physically sound basis in the theory of liquids derived from statistical mechanics (Ben-Naim, 1987). There, a n-particle correlation function $g^{(n)}(\vec{r}_1, \ldots, \vec{r}_n)$ can be translated into a potential of mean force *via*:

$$W^{(n)}(\vec{r}_1, \ldots, \vec{r}_n) = -RT \ln g^{(n)}(\vec{r}_1, \ldots, \vec{r}_n). \qquad (3)$$

Taking a two-body case with spherical symmetry,

the correlation function $g^{(2)}(\vec{r}_1, \vec{r}_2)$ corresponds to the radial pair distribution function $g^{(2)}(r_{12})$, with r_{12} equal to $|\vec{r}_1 - \vec{r}_2|$. In principle, the latter function can be derived straight-forwardly from crystal data by simply sampling the frequencies of both atoms in a distance interval between r_{12} and $r_{12} + dr$. Although not obvious for physical reasons, the Boltzmann law has been applied successfully to translate such a quantity into a potential of mean force (Burgi & Dunitz, 1988).

Strictly speaking, the Boltzmann law is only applicable to an ensemble of equal particles being in thermodynamic equilibrium. In contrast, a database of protein structures consists of many different particles not necessarily residing in the same global energy minimum. In addition, it is not known if each member of this heterogeneous set resides in its own global energy minimum. Furthermore, it is by no means trivial to assign a fixed absolute temperature to a sample retrieved from a database of crystal structures (Finkelstein *et al.*, 1995).

In contrast to the folding problem, in the present study, we do not evaluate the properties of proteins in total but we are interested in the properties of a ''fragment-of-interest'' (e.g. an atom-atom pair contact) embedded into the molecular environment of various protein structures. Any statistical evaluation is focussed on this ''fragment-of-interest''. It is assumed that this fragment is exposed in a statistically representative way to all relevant states.

The physical ground appears even less solid if, instead of energies or potentials, the logarithm of occurrence frequencies is enumerated (Godzik, 1996):

$$quantity = -\ln \frac{g_{\text{observed}}}{g_{\text{expected}}} \tag{4}$$

Despite these assumptions and short-comings, we still believe our heuristic knowledge-based approach is appropriate for the problem to be solved. We note that the ''quantity'' above should be termed a ''knowledge-based quantity'' or a ''statistical preference'' (Koppensteiner & Sippl, 1998) instead of a ''potential'', and we want the expression ''potential'' to be understood in this sense in the following. As with any empirical approach, the selected evaluation procedure can only be justified by the results obtained and its predictive power to reproduce or estimate experimental data.

Distance-dependent pair-potentials

In the context of protein-fold predictions it has been shown that coarse-grained (i.e. low resolution) energy models (Bowie *et al.*, 1991; Hendlich *et al.*, 1990; Jones *et al.*, 1992; Kocher *et al.*, 1994; Miyazawa & Jernigan, 1996) can effectively discriminate experimentally observed (i.e. crystal structures) and near-native folds from decoy conformations. In addition, the incorporation of artificial distance-dependencies into Miyazawa-Jernigan contact potentials (Park & Levitt, 1996) as well as the use of smaller interaction radii for the development of such potentials (Bahar & Jernigan, 1997), corresponding to increased resolution, unravels more details contained within the data.

To correctly distinguish (near) native ligand poses from computer-generated artifacts, e.g. obtained by docking, an approach on an atomic level is followed by compiling distance-dependent pair-potentials between ligand and protein atoms of type i and j. In our approach, we have followed the formalism developed by Sippl (1990, 1993):

$$\Delta W_{i,j}(r) = W_{i,j}(r) - W(r) = -\ln \frac{g_{i,j}(r)}{g(r)} \tag{5}$$

$$g(r) = \frac{\sum_i \sum_j g_{i,j}(r)}{i^* j} \tag{6}$$

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 any two atoms in the range of r and $r + dr$. It corresponds to the reference state and incorporates all non-specific information common to all atom pairs present in an environment typical for proteins. Taken together, the net statistical preferences $\Delta W_{i,j}$ are obtained by comparing the mean statistical preferences of the subsystems i, j ($W_{i,j}$) to the reference system (W).

In addition, both radial distribution functions are normalized with respect to the volume $4\pi r^2 dr$ of the spherical shell associated with the interatomic distance r to achieve a faster convergence toward zero at large distances (Bahar & Jernigan, 1997).

The fixing of an upper radius limit $r_{\max}$ for interactions between atoms i and j (Godzik *et al.*, 1995) determines the overall shape of the resulting potentials. Sampling over short distances up to this limit will emphasize the specific interactions formed by a ligand functional group to the neighboring binding-site residues. An extension of this sampling region to much larger distances incorporates the influences of an averaged solvent contribution that is mainly entropy driven (DeWitte & Shaknovich, 1996). Recently, Muegge & Martin (1999) developed potentials for large distances of up to 12 Å. Since we want to focus on the geometrical discrimination of various ligand binding modes, we restrict our sampling to 6 Å, thereby guaranteeing that highly specific interactions will dominate. The rationale for this 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, an alternative approach to incorporate solvent effects is required. As discussed in the context of protein-

fold prediction, solvent effects can be considered by two-body potentials to a different extent (Godzik *et al*., 1995; Skolnick *et al*., 1997). As was noted first by Sippl (1993) and later by Miyazawa & Jernigan (1999), solvent-mediated effects are underestimated when solely applying pair potentials. Nevertheless, the recent study of Muegge & Martin has obviously demonstrated the opposite by finding correlations between the sum of atom pair potentials and binding free energies of protein-ligand complexes. The authors explain this behavior by the consideration of a large cutoff for their atom pair interactions. It thus converts the degree of ligand penetration into the protein in an implicit recognition of solvation effects.

Non-polar surface-dependent singlet-potential

A combination of the short-distances 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:

$$\Delta W_i(SAS, SAS_0) = W_i(SAS) - W_i(SAS_0) = -\ln\frac{g_i(SAS)}{g_i(SAS_0)} \quad (7)$$

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 (see Methods). In this assumption any polar portion of the SAS that becomes buried in the complex in a polar environment is considered to remain ''solvent-accessible'' (Koehl & Delarue, 1994). The latter strategy is based on the rationale that changes in the Gibbs free energy can only be expected if polar molecular portions are transferred from a polar solvent to a non-polar protein environment. In contrast, polar portions carried over from the polar solvent to a polar binding site environment can be neglected.

In contrast to the atom pair potential mentioned above (equation (6)), $g_i(SAS_0)$ is not an averaged distribution function over all atom types but refers to the atom of type i only. Thus, ΔW_i reflects the contribution arising from differences in the solvent-accessible surface between the protein-bound and fully solvated state.

As a first approximation, the ligand conformations as found by X-ray crystallography or docking procedures are assumed to be identical to those adopted in the solvent. In this rough model, conformational changes, e.g. due to a ''hydrophobic collapse'' of the ligand (Testa *et al*., 1996), are not considered.

Calculation of the total score to rank ligand poses

Individual atom-pair potentials sampled for protein-ligand complexes are intercorrelated and depend on specific features arising from the direct molecular environment embedding a particular contact. However, in our approach we assume that a reasonable description of the total preference ΔW for a particular binding geometry can be approximated by summing all individual contributions (i.e. of k_i ligand atoms of type i and l_j protein atoms of type j):

$$\Delta W = \gamma \sum_{k_i}\sum_{l_j} \Delta W_{i,j}(r) + (1-\gamma) \times \left[\sum_{k_i} \Delta W_i(SAS, SAS_0) + \sum_{l_j} \Delta W_j(SAS, SAS_0)\right] \quad (8)$$

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

The scorings thus obtained will only be compared among different poses of the same ligand in one given protein. Additional contributions to the binding energy such as conformational, rotational, and translational entropy are not required, since they cancel out in a relative comparison of different poses. In accordance to our goal to rank multiple binding modes suggested by a docking run, this relative estimate is therefore satisfactory. Interestingly enough Muegge & Martin (1999) showed in their approach that a pure consideration of knowledge-based potentials is sufficient to derive absolute binding constants. This suggests that the above-mentioned entropy-related ordering parameters are implicitly accounted for in this approach, or (unlikely) that they hardly matter in the ligand-binding process.

Our approach does not incorporate any energy contribution arising from intramolecular interactions (van der Waals and torsion potentials). 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. This assumption is supported by a recent study by Bostrom *et al*. (1998). The authors demonstrate that energies resulting from conformational transitions between a bound and unbound state usually amount to less than 3 kcal/mol. Contributions resulting from the inherent flexibility of the protein are supposed to have negligible impact on the calculated scoring, in particular since the protein is assumed rigid during docking runs.

Methods

Protein-ligand complexes for derivation of potentials

Both, short-range pair and SAS-potentials are derived using for data extraction the ReLiBase system (Hendlich, 1998), which contains 6026 PDB protein structures holding all ligands with bond and atom-types notation according to the SYBYL type convention.

For our purpose, we evaluated crystallographically determined complexes only with resolution better than 2.5 Å. Complexes with covalently bound ligands or ligands with less than six or more than 50 non-hydrogen atoms were excluded. The latter restriction is used to regard ligands only with a size of typical drug molecules. Due to the first pass effect in the liver, a rather strict upper limit of about 600 Daltons can be applied for a typical organic molecule. Examples with frequently occurring prosthetic groups such as haemoglobin, flavin-adenine-dinucleotide or nicotinamide-adenine-dinucleotide were not considered as explicit ligands but as part of the protein.

Furthermore, we excluded all complexes that were subsequently used in the validation of the predictive power of the potentials in order to avoid any redundancy or training effects due to overfitting. To check for further redundancy bias, we compiled additional data sets by considering only proteins with a mutual sequence homology of less than 30 % compared to the protein used to test the predictive capability of the approach. Though, this was not found to introduce a significant change either in terms of the depths or the position of the minima. To account for the volume occupied by the ligand, a correction term similar to the one proposed by Muegge & Martin (1999) was investigated. However, on finding that it did not effect the overall shape of the potentials significantly, it was omitted in further calculations.

Potentials were derived for the following atom types: C.3 (carbon sp^3), C.2 (carbon sp^2), C.ar (carbon in aromatic rings), C.cat (carbon in amidinium and guanidinium groups), N.3 (nitrogen sp^3), N.ar (nitrogen in aromatic rings), N.am (nitrogen in amid bonds), N.pl3 (nitrogen in amidinium and guanidinium groups), O.3 (oxygen sp^3), O.2 (oxygen sp^2), O.co2 (oxygen in carboxylate groups), S.3 (sulfur sp^3), P.3 (phosphorus sp^3), F (fluorine), Cl (chlorine), Br (bromine), Met (Ca, Zn, Ni, Fe).

Due to their low occurrence frequency or non-unique assignment criteria, the following atom types were grouped together: S.2 (sulfur sp^2) and S.3, N.4 (positively charged nitrogen) and N.3.

Compilation of distance-dependent pair-potentials

A normalized distance-dependent radial pair distribution function for atom pairs with types i and j is given by:

$$g_{i,j}(r) = \frac{N_{i,j}(r)/4\pi r^2}{\sum_r (N_{i,j}(r)/4\pi r^2)} \tag{9}$$

Scaling to $4\pi r^2$ accounts for the volume of the spherical shell of the radius r and the thickness dr.

The number $N_{i,j}(r)$ of atom pairs i,j at a distance between r and $r + dr$ is obtained by counting the occurrences:

$$N_{i,j}(r) = \sum_i \sum_j \delta(|\vec{r}_i - \vec{r}_j|, r) \tag{10}$$

where the double summation runs over all atom types i and j present in the database, respectively. The delta function δ equals 1, if $r \leqslant |\vec{r}_i - \vec{r}_j| \leqslant r + dr$, otherwise 0.

Using $g_{i,j}(r)$ (equation (9)), the mean radial pair distribution function $g(r)$ and the distance dependent pair-potentials $W_{i,j}(r)$ and $W(r)$ (equations (5) and (6)) and finally the net potential $\Delta W_{i,j}(r)$ (equation (8)) are computed (see Theory).

The size of the bins dr was chosen in a way that a sufficiently high resolution is guaranteed and an adequate amount of data are sampled within each bin. Thus, statistically significant results can be obtained. A bin size dr of 0.1 Å was found to reveal satisfactory results.

The minimal distance boundary r_{min} was set to 1 Å, especially since metal-to-oxygen/nitrogen contacts occur at short distances around 1.8 Å; the maximal distance bound r_{max} was set to 6 Å. Any special treatment of ''short contact distances'' (with no occurrences) to reflect repulsive terms are not required, since these will not be present in either a crystal structure or a computer-docked complex. Both techniques inherently avoid short atom-to-atom contacts due to the consideration of van der Waals repulsion terms.

To account for uncertainties inherently present in experimental data (for a resolution of 2.5 Å, inaccuracies of atom positions may be as large as 0.4 Å (Kossiakoff *et al.*, 1992)), a smoothing function is applied to the distribution data. It effectively distributes a single hit over more than one bin according to a triangular weighting scheme. This scheme causes one hit to be assigned to all neighboring bins within a 0.2 Å distance from the central bin, with a weight linearly falling off from one to zero within this range.

Compilation of solvent-accessible surface-dependent singlet-potentials

The solvent-accessible surface-dependent singlet-potentials are calculated separately for ligand and protein atoms from the normalized distribution functions (equation (11)):

$$g_i(SAS) = \frac{N_i(SAS)}{\sum_{SAS} N_i(SAS)} \tag{11}$$

and:

$$g_i(SAS_0) = \frac{N_i(SAS_0)}{\sum_{SAS_0} N_i(SAS_0)} \quad (12)$$

$g_i(SAS)$ is the probability to find an atom of type i with an exposed solvent-accessible surface SAS in a complexed state, while $g_i(SAS_0)$ is the probability to find the same atom with the same solvent-accessible surface in a state totally separated from the complex. The normalization (denominator) accounts for the total number of occurrences in both cases.

The solvent-accessible surface is calculated by a fast and approximate cube-algorithm adapted from Bohm (1994). A cubic grid with a spacing of 1 Å is constructed around a ligand and embedded into the active site of the protein. All cubes with centers within a distance of less than $r_{vdW} + 1.4$ Å of a ligand or a protein atom are defined as occupied in the first step. Subsequently, those cubes being simultaneously located in the neighborhood of at least one O or N atom in the protein and in the ligand are excluded. In the next step, for those of the unoccupied cubes that fall next to occupied ones, the closest atom is determined and these cubes are labeled to be ''surface portion'' of the neighboring atom. Summing up all cubes assigned to atoms finally yields a value which is approximately proportional to the considered surface portion (Bohm, 1994).

Van der Waals radii are used according to the Tripos force field (Clark *et al.*, 1989), those of O and N atoms are reduced by 0.2 Å (Li & Nussinov, 1998) to account for a potential involvement in hydrogen bonding.

The surface attributed to the solvated state is calculated similarly without considering any particular counterpart (ligand/protein). The conformation in solution is assumed to be identical to the one adopted in the crystal structure or the docked complex.

Test data to evaluate ranking performance

We have tested the performance of the derived potentials by assessing for two different data sets their capability to rank best those ligand poses that closely approximate the native one. The first data set† is a subset of 91 protein-ligand complexes extracted from the 200 examples that have been used in the validation of the FlexX docking program (Rarey *et al.*, 1996). These 91 complexes cover a broad range of ligand diversity, e.g. considering the number of rotatable bonds (ranging from 0 to 27), the number of hydrogen-bond donors and acceptors (ranging from 1 to 17) or the number of non-polar atoms (from 2 to 39). In half of the cases, FlexX finds solutions with an rmsd value of less than 2 Å to the crystal structure on the first rank, while for the other portion FlexX fails to select these ''well-docked'' cases as best solutions. The ligand poses are generated using standard input files for FlexX (B. Kramer *et al.*, unpublished results) with default parameter settings.

The second set‡ consists of 68 protein-ligand complexes matching the same criteria as those applied to the first set. Again ligand poses are generated using FlexX. However, for only 30 examples FlexX finds a solution deviating by less than 2.0 Å from the X-ray reference. Out of these 30 cases, for 28 (93%) a computed solution <2.0 Å is found by FlexX on rank 1. In our study, this second set was used for a cross-validation of our scoring function in conjunction with FlexX, since it was not involved in the parameter adjustment during development of the function.

To test DrugScore in context with DOCK, 100 protein-ligand complexes§ are extracted from both data sets to generate input files following the procedure described by Ewing (1997). Flexible docking with dihedral and rigid body minimization is applied using default parameter settings except for the ''peripheral seeds'' parameter which is set to 50. All generated solutions using ''uniform sampling'' are ranked and stored by either applying the ''energy'' or ''chemical score''. Since the two scorings are used during structure generation, the resulting solutions in either cases are distinct. In the first case, DOCK generates in 61% of all cases a solution that deviates by less than 2.0 Å from the crystal structure and recognizes in 54% out of these cases a geometry <2.0 Å on the first rank. Applying the ''chemical score'', in 43% of all cases a solution with less than 2.0 Å deviation from experiment is generated; in 46% out of these cases such a ligand geometry is ranked best. Thus, in the present case and different from recently published work of Knegtel *et al.* (1999) who applied

† 1abe 1abf 1atl 1azm 1bbp 1cbx 1cde 1cil 1com 1cps 1ctr 1did 1die 1dr1 1dwc 1dwd 1ela 1epb 1frp 1ghb 1hfc 1hgj 1hsl 1hyt 1icn 1imb 1ivc 1ivd 1ive 1ivf 1lah 1lcp 1lic 1lna 1lst 1mld 1mrg 1mrk 1nis 1nsc 1pbd 1phf 1poc 1pph 1ppl 1pso 1rbp 1rds 1rnt 1rob 1slt 1snc 1srj 1tlp 1tng 1tnh 1tni 1tpp 1ukz 1wap 1xid 1xie 2ada 2ak3 2cgr 2cht 2cmd 2cpp 2gbp 2mth 2pk4 2sim 2tmn 2xis 2ypi 3aa 3cpa 3hvt 4fbp 4hmg 4phv 4tim 4tln 4ts1 5abp 5p2p 6abp 6rnt 6tmn 7tim 8atc

‡ 121p 1aaq 1acm 1aco 1aec 1aha 1ake 1apt 1avd 1bma 1byb 1cbs 1cdg 1coy 1dbb 1eap 1eed 1elb 1elc 1eld 1ele 1etr 1fen 1fkg 1glp 1glq 1hdc 1hef 1hvr 1ida 1igj 1ivb 1ldm 1lmo 1lpm 1mbi 1mdr 1mmq 1nco 1phd 1phg 1ppc 1ppi 1ppk 1ppm 1rne 1tnk 1tnl 1tph 1trk 2ctc 2er6 3cla 3gch 3ptb 4dfr 4fxn 4hvp 4phv 4tmn 5cts 5tim 5tmn 6cpa 6tim 7cpa 8gch 9hvp

§ 1abe 1abf 1acj 1ack 1ase 1azm 1blh 1cbx 1cde 1cil 1cps 1ctr 1dbm 1did 1die 1dr1 1dwd 1ela 1frp 1ghb 1hfc 1hgi 1hgj 1hsl 1hti 1hyt 1icn 1imb 1ivc 1ivd 1ive 1ivf 1lah 1lcp 1lic 1lna 1lst 1mld 1mrg 1mrk 1mup 1nis 1nsc 1pbd 1phf 1poc 1pph 1ppl 1pso 1rds 1rnt 1rob 1snc 1srj 1tdb 1thy 1tlp 1tng 1tnh 1tni 1tpp 1ukz 1ulb 1wap 1xid 1xie 2ada 2ak3 2cgr 2cht 2cmd 2cpp 2gbp 2lgs 2mcp 2mth 2pk4 2r04 2r07 2sim 2tin 2xis 2yhx 2ypi 3aah 3cpa 4cts 4est 4fab 4fbp 4phv 4tim 4tln 5abp 5p2p 6abp 6rnt 6tmn 7tim 8atc

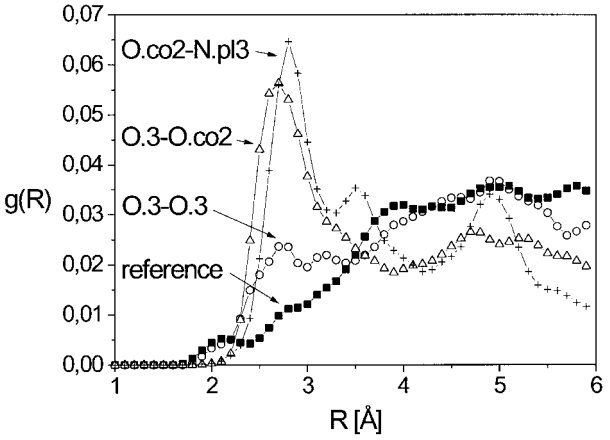

Figure 1. Pair distribution functions of polar/charged interactions as computed by equation (9). The first atom-type label refers to the ligand atoms, the second to protein atoms. The reference state (mean distribution function over all pairs) as calculated by equation (6) is depicted by solid squares.

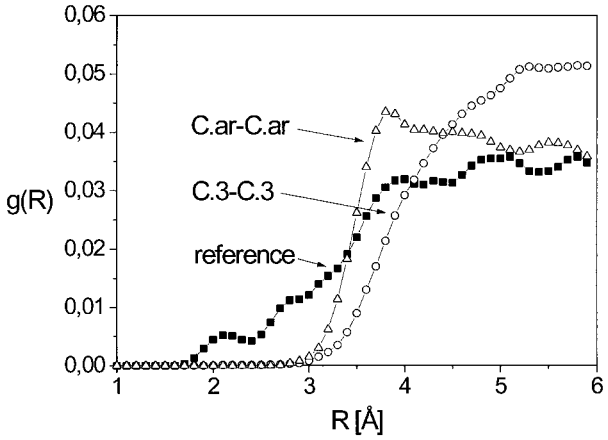

Figure 2. Pair distribution functions of nonpolar and aromatic interactions as computed by equation (9). Atom-type characterization and reference state are displayed as given in Figure 1.

DOCK to the flexible docking of 32 thrombin inhibitors, the ''energy score'' reveals better results compared to the ''chemical score''.

Results

Ligand-protein atom pair correlation functions and statistical preferences

Pair correlation functions for ligand-protein atom pairs were derived by counting the occurrence frequencies of distinct pairs at discrete distances using equation (10). The pair correlation functions for polar/charged interactions are presented in Figure 1, those of non-polar interactions in Figure 2. Statistical preferences were computed by equation (5). Utilizing 17 different atom types yields 289 possible pair combinations depicted in Figures 3 and 4, respectively. In all cases, the first atom-type index i is attributed to a ligand atom, the second j to a protein or cofactor atom. In the following we will focus on some illustrative examples.

Frequency of occurrences of pair interactions

Based on 1376 protein structures considered the minimum number of hits integrated over a distance range from 1 to 6 Å for the examples presented here amounts to 8028 in the case of the O.co2-N.pl3 pair; the highest count was found for the C.3-C.3 pair to be 120,533. Thus, in the case of O.co2-N.pl3, on average more than 160 hits are collected within one bin of width 0.1 Å. Although this average number gives only a rough estimate on the statistical significance of this distribution, we note that for the distance range below 2.4 Å the number of counts is markedly lower. It thus yields less significance in this area. Of all pair distributions, 173 (60 % of 289) contain less than 500 hits, i.e. less than ten hits per bin on average. For 156 (i.e. 54 % of all) of these, one of the atom types of either the ligand or the protein belongs to S.3, P.3, C.cat, metal, F, Cl, or Br.

The rare occurrence of S.3, F, Cl, and Br in the X-ray structures used to compile the statistical preferences is also reflected in the test data set. However, we anticipate that as long as interactions involving S.3/F/Cl/Br-X contacts do not dominate the energetics of ligand binding, a reliable scoring can still be estimated based on the contributions resulting from the remaining more frequently populated atom pairs.

Atoms of type P.3 are usually connected to oxygen, carbon and nitrogen atoms, resulting in phosphate, phosphonate or phosphinate derivatives. Being not exposed to the molecular surface, the major contributions of these functional groups are determined by the preferences arising from the highly populated distributions of the neighboring oxygen and nitrogen atoms. As a first approximation, the same holds for C.cat coding for the carbon atom in amidinium and guanidinium groups. These atoms are at least in the molecular plane shielded by the surrounding nitrogen atoms.

Reference state of pair interactions

The reference state is calculated as arithmetic mean over all normalized pair correlation functions of atom types ij. It is represented by filled squares in Figures 1 and 2. Defined in this way, it may be regarded as a mean interaction preference between ''averaged atom-types'', thus mainly representing non-specific contributions from dense packing effects.

The nature of certain features expressed in the distributions or the subsequently derived statistical

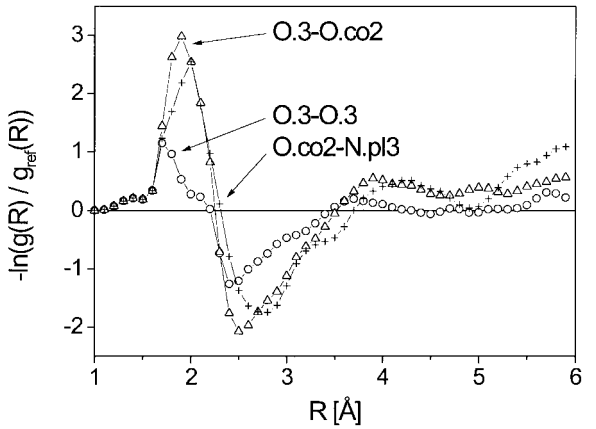

Figure 3. Statistical preferences for polar/charged pair interactions as a function of the distance R, calculated according to equation (5).

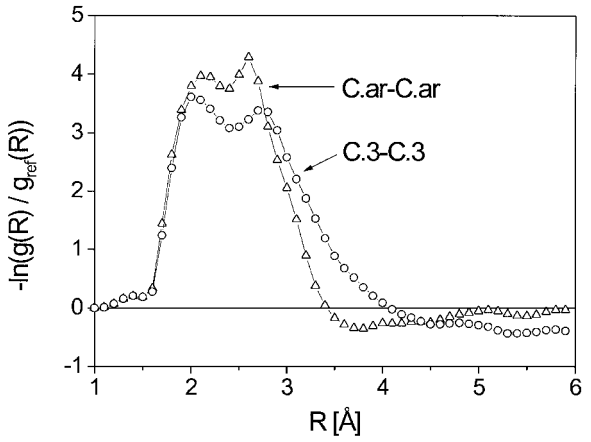

Figure 4. Statistical preferences for nonpolar/aromatic pair interactions as a function of the distance R, calculated according to equation (5).

preferences are difficult to partition into types of interaction that operate between isolated pairs of atoms, since the compiled pair-correlation functions are averages of the ensemble, comprising intercorrelated many-body interactions occurring in a dense packing and arising from several physical effects (electrostatic, steric, solvent, etc.).

Our reference state is derived from a large sample of protein-ligand complexes, accordingly some structuring of this distribution is observed. Local fluctuations in the distribution supposedly arise at low distances (≈2 Å) from metal contacts, around 2.7 Å from polar and charge-assisted interactions. The elevated probability around 4 Å is presumably caused by aromatic contacts and structuring at larger distances may be attributed to contacts resulting from pattern formation in the second coordination sphere.

Individual distributions and statistical preferences of pair interactions

To distinguish ligand binding modes approximating the native structure from inappropriate ones generated by docking tools, it is important that the type-specific distributions differ substantially from each other.

Based on a visual inspection of the computed preferences we decided to refrain from averaging *i-j* and *j-i* atom-type contacts (first letter corresponds to ligand atom, second one to protein atom). If our data originated from isolated atoms or a homogenous non-structured molecular distribution, symmetrical conditions would be expected. However, for the data selected the applied mean-field approach reveals clear differences due to a distinct molecular embedding of the considered atom-types in ligand and protein environments.

The derived distributions can be divided into two main classes: the first contains interactions between polar and charged atoms and exhibits pronounced maxima 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 distributions. The latter reveal higher probabilities compared to the reference state at distances >3.5 Å (Figure 2).

The distributions within one group show differences to which a physical meaning can be attributed. Going from O.3-O.3 to O.3-O.co2 and O.co2-N.pl3, the distribution maxima corresponding to the shell of next neighbors fall into a decreasing distance range, thus exhibiting higher probability 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 (Davis & Teague, 1999). Expressed in terms of statistical preferences (equation (5)), an ideal O.3-O.3 interaction is 2.5 times less probable than a similar O.3-O.co2 interaction.

Surprisingly, at a first glance, O.3-O.co2 and O.co2-N.pl3 interactions are equally probable contradicting to the assumption that salt bridges with both partners bearing opposite charges contribute more to the stability of a protein-ligand complex than a single charge-assisted H-bond (Hossain & Schneider, 1999). According to our definition, N.pl3 atom types only occur in amidinium and guanidinium groups. In consequence, a carboxylate group and an amidinium/guanidinium group in ideal bidentate H-bonding geometry show two O.co2-N.pl3 interactions instead of one ideally oriented O.3-O.co2 interaction. Thus, the contribution of a bidentate salt-bridge, considering only interactions at the contact distance, amounts to about twice that of the polar-charged interaction.

For non-polar contacts (Figures 2 and 4), the C.ar-C.ar interaction shows a slightly more structured distribution compared to the C.3-C.3 interaction and the maximum of the former resides at a

shorter distance of 3.7 Å. Accordingly, the C.ar-C.ar contact exhibits an elevated preference compared to the all-atom distribution at this distance, in agreement with the well-known aromatic-aromatic interactions (Burley & Petsko, 1985). In contrast, C.3-C.3 interactions do not show any preference of the atom pair distribution over the entire distance range of 1 to 6 Å sampled here. This is clearly in agreement with the well-known fact that the latter type of interaction hardly exhibits any directional preferences.

In comparing the depth of the preference minima, one has to consider that at short distances (<3 Å), on average, only one direct partner will be involved. With increasing distance additional next-neighbors get into contact and the coordination number roughly scales with the square of the distances. Estimating the total contribution of one specific contact to the stability of a ligand-protein complex, both the individual strength and its occurrence frequency have to be considered.

The derived statistical preferences display differences when compared to pure van der Waals interactions operating between isolated atom pairs in two aspects. The occurrence of multiple minima at larger distances results from particular packing patterns present in a protein-type environment and reflects organization in high-order coordination shells. In addition, our preferences fall off below distances of 2 Å and display minima at 1 Å. This is bound to occur, since for good reasons, no experimental observations are recorded at such short distances. We believe there is no need for a specific correction term to cope with this non-defined distance range, since the computed binding modes from FlexX or other docking programs will not contain geometries with mutually penetrating protein and ligand atoms.

The compiled preferences for a given atom pair *i*, *j* are calculated only as a function of the distance between the atoms, any directional preference patterns are implicitly contained and reflected; e.g. let us consider a hydrogen-bond between a carbonyl group and an O.3-type oxygen (Figure 5). The minimum of an O.2-O.3 interaction (not shown here) is located at 2.55 Å, the C.2-O.3 interaction has a minimum at 3.45 Å. Assuming a C.2-O.2 bond length of 1.22 Å, a mean spatial C.2-O.2-O.3 angle of 128° is calculated. Taking the preferred distances O.2-O.3 and C.2-O.3 together, these conditions clearly constrain the frequently observed hydrogen-bonding geometry of a carbonyl group with an alcohol-type oxygen atom. The same orientational preference is observed for representative fragments stored in ISOSTAR (Bruno *et al.*, 1997). Additional contact preferences formed by the neighboring atoms will further define the spatial arrangement of a specific directional interaction.

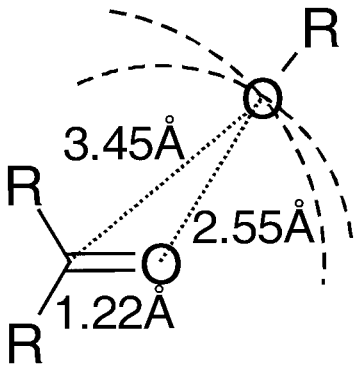

Figure 5. Intrinsic geometrical constraints reflected by the atom pair preferences of O.2-O.3 and C.2-O.3. Given the minima of the statistical pair preferences (O.2-O.3: 2.55 Å; C.2-O.3: 3.45 Å) and the bond length (C.2-O.2: 1.22 Å), the C.2-O.2-O.3 angle is calculated to be 128°.

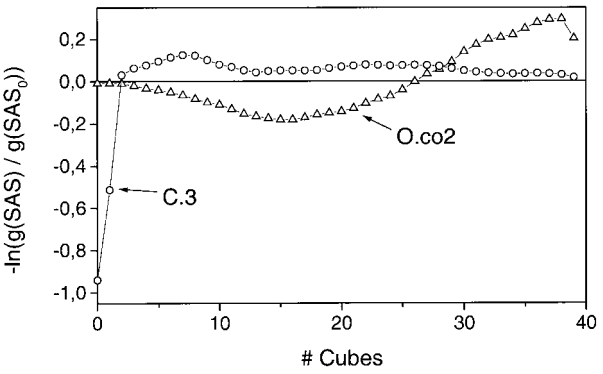

Figure 6. Statistical preferences for ligand atoms of type C.3 and O.co2 as calculated from the distribution functions for solvent accessibility of both atom types for complexed and separated state from the protein according to equation (7). The number of cubes (#cubes) is an approximate measure for the solvent accessibility; zero cubes refer to complete burial.

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. As defined by equation (7), this preference is calculated as the logarithm of the probability for a ligand or protein atom of type *i* to expose parts of its SAS in the complexed state compared to a state where both partners are completely separated from each other (see Methods). We note that by defining the reference state in this way, the derived preferences are independent from each other and not a result of a mean-field approach.

Figure 6 shows the statistical preferences for a C.3-type or O.co2-type ligand atom to be exposed in the protein-bound state compared to the unbound state. Here the ''number of cubes'' is approximately proportional to the SAS area.

Except for very small surface portions remaining solvent-exposed at the binding site, complete bur-

ial of C.3 atoms in protein-ligand complexes is strongly favorable for complex formation which is expressed by a deep minimum near zero cubes.

A completely different although quite reasonable behavior is observed for O.co2-type atoms: in the isolated state, carboxylate oxygen atoms strongly prefer to be solvent-exposed. In complexed state, the distribution is the result of a compromise between several effects. On the one hand, burial of O.co2 atoms diminishes the SAS (shift to a smaller numbers of cubes). On the other hand, 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 stabilizes complex formation in contrast to a complete burial; additionally a complete solvent exposure is also slightly unfavorable.

''Scoring the scoring function''

Usually the root-mean-square deviation of a ligand pose, generated by a docking tool, is determined with respect to the corresponding crystal structure. This is a generally accepted quality measure for the obtained docking result. We define, based on this measure, those ligand poses as ''well-docked'' that do not deviate more than 2.0 Å from the crystal structure. Visual inspection of a number of test cases showed that within this limit the generated solutions usually resemble the native binding mode.

To assess the discriminatory power of a scoring function, i.e. its capability to distinguish between well-docked and clearly different solutions, we selected the criterion that one of the ''well-docked'' solutions has to end up on the best scoring rank. Since this condition is more stringent than just demanding one of the ''well-docked'' solutions to be among, i.e. the ten best rankings, our criterion meets the requirement for virtual screening of large databases.

As an even more stringent criterion, we can demand that the experimentally given solution is ranked better than any other computer-generated ligand binding pose (including the ''well-docked'' ones). However, we have to remember that even these ''ideal'' reference solutions are affected by experimental deficiencies. Taking this limited accuracy into account, we also consider our scoring function to operate satisfactorily if solutions, deviating not further than 2.0 Å from the experimentally given structure (''well-docked''

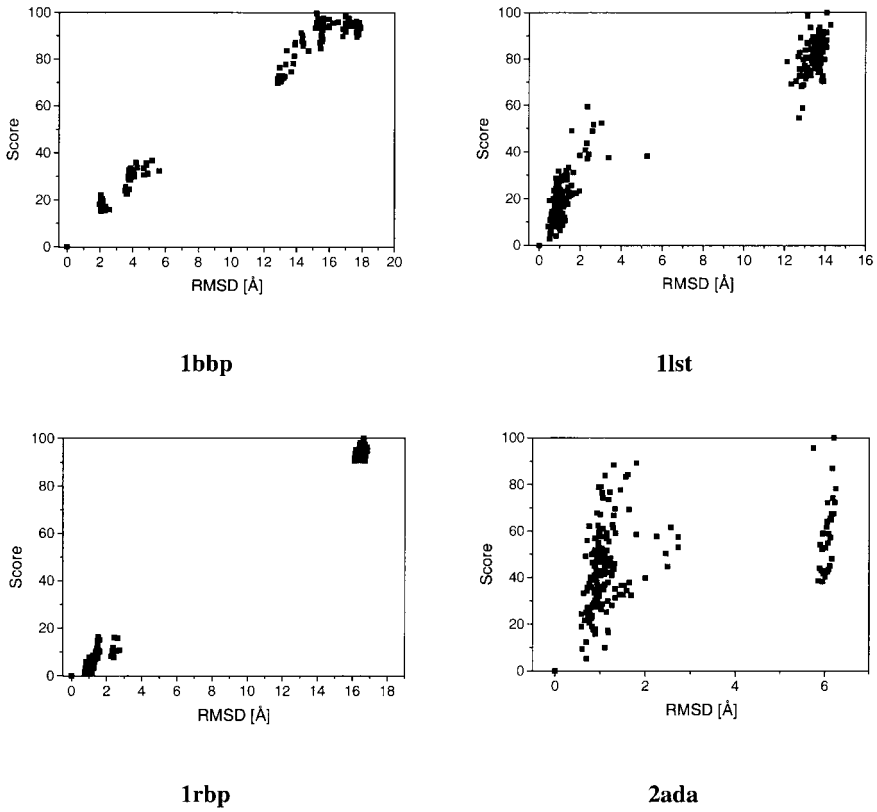

Figure 7. Correlations of scores, normalized to values between 0 and 100, *versus* the rmsd value from the crystal structure for the protein-ligand complexes 1bbp, 1lst, 1rbp, and 2ada are depicted. The percentage of non-polar (i.e. carbon) ligand atoms amounts from 52% for 2ada to 95% for 1rbp, the 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.

solutions), are ranked better than the crystal structure.

Correlation of the calculated scores *versus* rmsd of the crystal structure

Total preferences were calculated according to equation (8) for the different solutions obtained by docking experiments. They were plotted *versus* the rmsd value from the crystal structure. This procedure involves the projection of the total preference score, multifactorily composed from many individual contributions, onto the single geometrical rmsd descriptor. It is difficult to estimate what kind of correlation will be present; however, one can expect that the crystal structure and approximate ligand poses should be ranked better than strongly deviating solutions. Furthermore, the lower envelope of the resulting scatter plot may exhibit multiple local minima (due to the roughness of the potential hypersurface).

We have to recall that similar rmsd values do not necessarily represent similar ligand poses, especially for large values. Accordingly, poses falling into the same range of larger rmsd values can obtain quite different scorings. In contrast, similar ligand binding modes should reveal similar scores if the underlying function is sufficiently ''soft'' to tolerate some geometric deviations.

In Figure 7 the correlations of calculated scores for ligand geometries generated by FlexX *versus* the rmsd value from the crystal structure for four protein-ligand complexes are displayed; the individual scores were normalized to values between 0 and 100. Favorable solutions correspond to low scores. The examples shown were selected to represent quite distinct cases (e.g. spread of non-polar surface contribution, number of rotatable bonds, number of generated FlexX solutions).

In all cases, the crystal structure (rmsd = 0) is represented by the best score. For 1bbp, 1lst and 2ada, the native geometry is well separated from any computed solution. Docked geometries being clearly apart from the native structure are satisfactorily separated from the approximating poses. These latter solutions tend to obtain a better scoring with decreasing rmsd.

In conclusion, for all four cases the scoring function, defined by equation (8), successfully recognizes the experimental and ''well-docked'' poses among a set of up to 289 distinct ligand poses (Figure 7).

Validation of the approach on large test sets of protein-ligand complexes

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

Two test sets are taken from a sample used to validate FlexX (B. Kramer *et al.*, unpublished results). In the first case, comprising 91 protein-ligand 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, theses poses are attributed to a worse rank.

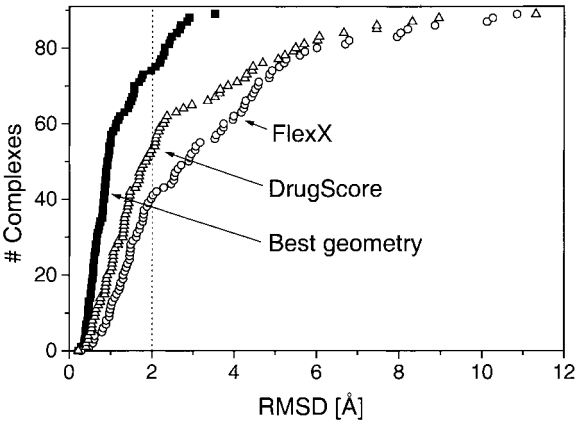

Figure 8. Accumulated number of complexes as a function of the rmsd value from the crystal structure found for ligand poses on rank 1 scored by FlexX (○) and DrugScore (△), respectively. The number of complexes with the best geometry found on any rank by FlexX is depicted as filled squares.

The second test set contains 68 additional complexes. Out of these, for 28 cases a computed solution with rmsd <2.0 Å is found by FlexX on rank 1. 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 of the scoring function, i.e. parameter adjustment was solely performed on the basis of data from the first set. We used this set for a cross-validation of our scoring function in conjunction with FlexX.

Out of both data sets, 100 protein-ligand complexes were selected to evaluate the performance of DrugScore with respect to docking solutions generated by DOCK. Applying the DOCK specific ''energy scoring'' (or ''chemical scoring''), a generated solution with rmsd <2.0 Å is recognized on the first rank in 54 % (46 %) of the cases by DOCK.

We want to stress that all complexes used for validating our approach (including those shown in Figure 7) have been excluded from the database used to compute the probability distributions and to derive the statistical preferences.

Recognition of ''well-docked'' solutions

Figure 8 summarizes the accumulated number of complexes plotted *versus* the rmsd value with respect to the crystal structure of the best ranked ligand pose either determined by FlexX or the scoring function described here (DrugScore).

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 value was computed by FlexX.
[b] Only the ligand geometry scored to be on the first rank by either FlexX or DrugScore is considered. The numbers are related to the ones in the first line.
[c] The improvement is calculated by (%DrugScore − %FlexX)/%FlexX.

In addition, for all examples the solution generated by FlexX that obtains the smallest rmsd value, disregarding its actual scoring rank, is also plotted. These values give an idea how well an ideal scoring function could perform. It also indicates by how much the generated poses with the best matching geometry deviate from the X-ray reference. Note that all cumulated values are sorted independently. Thus, the ordinate displays a counter for the complexes and not an identification number.

As is obvious from the diagram, the new scoring function performs significantly better than the one implemented in FlexX. Table 1 gives the percentage of test cases found on rank 1 with rmsd values of <1.0 Å, <1.5 Å, <2.0 Å and ⩾2.0 Å compared to the best approximating geometry found on any rank. With respect to the recognition of poses deviating by <2 Å on rank 1, FlexX succeeds in 54 % of the cases whereas DrugScore detects 73 %.

The difference in rmsd value, exhibited by the solutions ranked best according to FlexX or DrugScore, respectively, are shown in Figure 9. Whereas in ten cases, DrugScore selects a solution deviating more strongly in geometry (by more than 1 Å) on the first rank compared to FlexX, in 28 cases a pose is selected being more than 1 Å closer to the X-ray reference. Averaged over all 91 complexes, the mean improvement in reducing the rmsd value for rank 1 is 0.7 Å.

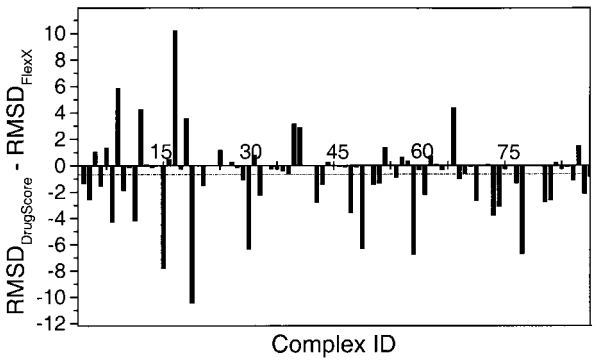

Figure 9. The differences between rmsd values exhibited by solutions found on rank 1 by FlexX and DrugScore are displayed, respectively, for each protein-ligand complex. The broken line shows the mean value of all differences of −0.7 Å.

For the second test set (68 complexes), out of the 30 cases satisfactorily docked by FlexX (below 2 Å rmsd), the scoring in FlexX and DrugScore obtain nearly identical success rates (93 % and 92 %, respectively).

In Figure 10, the accumulated number of complexes is plotted *versus* the rmsd value from the crystal structure of the best ranked ligand pose either selected by the ''chemical scoring'' in DOCK or by DrugScore. Furthermore, the generated solutions with the smallest rmsd values are accumulated disregarding their actual scoring rank.

Table 2 gives the percentage of test cases found on rank 1 with rmsd values of <1.0 Å, <1.5 Å, <2.0 Å and ⩾2.0 Å compared to the best approximating geometry found on any rank. With respect to the recognition of ''well-docked'' solutions (<2 Å) on rank 1, the ''chemical scoring'' in DOCK succeeds in 46 % of the cases whereas DrugScore detects 70 %. When using the ''energy scoring'' during structure generation (data not shown here), the scoring of DOCK and DrugScore yields similar success rates (54 % and 51 %, respectively).

Recognition of the crystal structures

As mentioned above, assuming the crystallographically determined structures to be ''optimal'' solutions, they should obtain the best scoring compared to any computer-generated ligand pose. As shown in Figure 11, this is actually the case for 54 % of the examples of the first test set for FlexX (91 complexes). If we alleviate, as described above, the criterion to complexes falling into a window of 2 Å deviation, in even 71 % of the cases the crystal structure or a ''well-docked'' geometry will be ranked by DrugScore as favorable.

What are the reasons that some cases do not perform well? Visual inspection of some of the unsuccessful examples provides some explanations. In the case of 1icn (resolution: 1.74 Å; Figure 12), the crystal structure is scored on rank 35. The ligand oleate is oriented with its carboxylate group towards the interior of the protein, thereby expos-

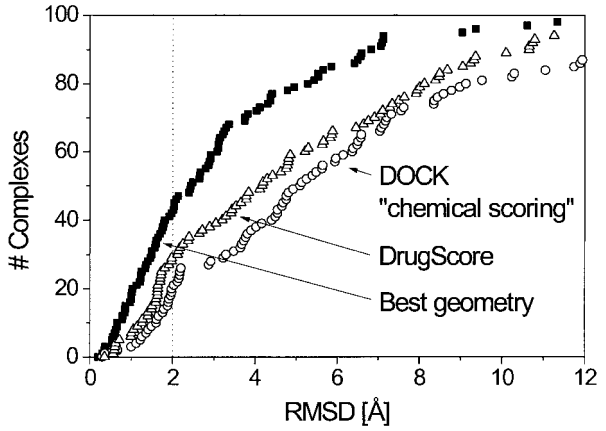

Figure 10. Accumulated number of complexes as a function of the rmsd from the crystal structure found for ligand poses on rank 1 scored by DOCK (''chemical scoring'') (○) and DrugScore (△), respectively. The number of complexes with the best geometry found on any rank by FlexX is depicted as solid squares.

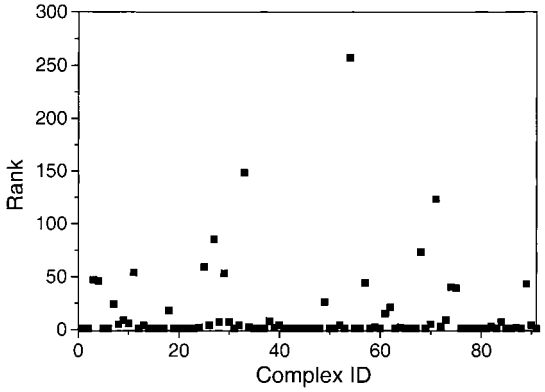

Figure 11. The rank of the crystal structure calculated by DrugScore among all decoy geometries generated by FlexX for each of the 91 protein-ligand complexes of the first test set is shown.

ing the non-polar end of the hydrocarbon chain towards the solvent. The carboxylate group is not involved in any directional interaction to any protein functional group within a distance of 3.5 Å; only a water molecule occurs in its next neighborhood (3.3 Å). Instead, a ligand pose with reversed orientation of the molecule (rmsd = 11.3 Å) is ranked best. In this mode, two hydrogen bonds to a protein amide group are formed and the non-polar terminus is buried inside the pocket, resulting in a much better score. Interestingly, in the crystal structure the carboxylate group is disordered and three alternative positions are given in the PDB. Additionally, a weak but continuous, *J*-shaped electron density is reported beyond the location of the terminal methyl of oleate in the wild-type holo protein (Eads *et al.*, 1993), which has an arginine at the bottom of the binding pocket. In contrast, this amino acid has been replaced by a glutamine in the mutated protein used for 1icn. It thus can not form a salt bridge with the carboxylate of the fatty acid. With some care, these findings could also allow for an interpretation of the reported electron density with a reversed orientation of the ligand. Then, the disordered part of the electron density would correspond to the hydrocarbon tail and this inverted orientation matches with our predictions. In the complex 2pk4 (resolution: 2.25 Å; Figure 13), ω-amino-hexanoic acid is placed into a shallow groove close to the protein surface. While a solution with rmsd 1.3 Å is ranked best, the crystal structure (rank 129) itself shows an interaction between the amino group of the ligand and a protein carboxylate group of 2.1 Å. This remarkably short distance is scored as repulsive by the corresponding atom pair preference.

Applying the scoring function to the second test set for FlexX, 65% of all crystal structures are found on rank 1. In 90% of the 68 cases, only docked solutions with rmsd <2 Å reveal a higher rank than those of the native structure, i.e. these cases fulfill the alleviated conditions (results not shown here).

For the set of complexes generated by DOCK using the ''chemical scoring'' (''energy scoring''), DrugScore finds in 76% (57%) of all cases the crystal structure on rank 1. Using the alleviated condition, in 83% (62%) of all cases the crystal

Table 2. Results for scoring multiple docking solutions of 100 protein-ligand complexes generated by DOCK (applying ''chemical scoring'') 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]		17	31	43	57
1st rank[b]	DOCK	18	33	46	54
	DrugScore	41	48	70	30
Improvement[c]		128	45	52	−44

[a] All solutions of each docking experiment for the 100 complexes are considered. This number expresses the portion of all complexes for which at least one solution with the given rmsd value was computed by DOCK.

[b] Only the ligand geometry scored to be on the first rank by either DOCK or DrugScore is considered. The numbers are related to the ones in the first line.

[c] The improvement is calculated by (%DrugScore − %DOCK)/%DOCK.

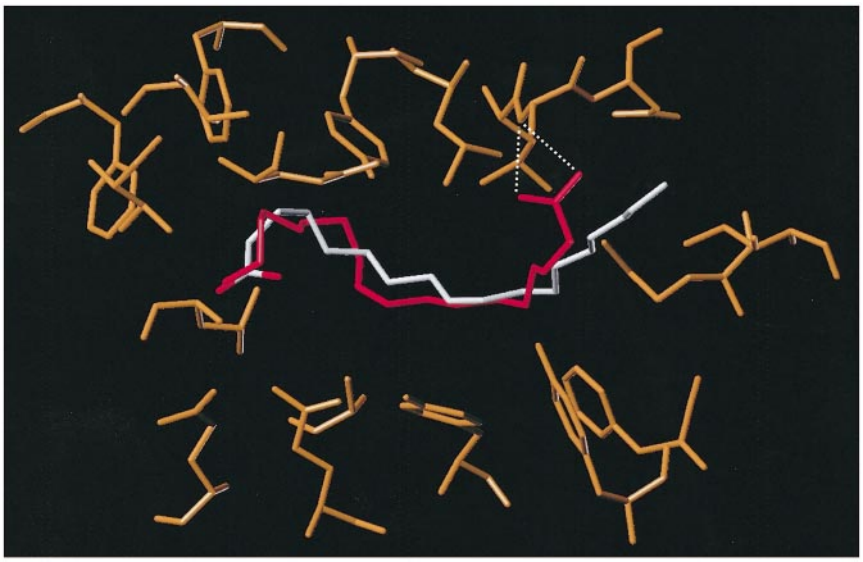

Figure 12. A section of the binding pocket of 1icn is displayed together with the X-ray structure of the ligand (color-coded by atom type) found on rank 35 by DrugScore and the geometry with rmsd = 11.3 Å ranked best (red colored). The entrance to the pocket orients toward the right, bottom and top parts are clipped for clarity. The carboxylate group of the crystal structure has no adequate interaction partner except a water molecule (not shown here) at 3.3 Å distance. In contrary, the better ranked, however geometrically deviating pose does not only form two hydrogen bonds to an amide-NH (distances 2.8/2.9 Å, depicted as dotted lines), but also buries its non-polar terminus towards the interior of the protein. See the text for explanations.

structure or a ''well-docked'' solution is ranked best (data not shown here).

Investigation of factors that might influence the recognition rate of solutions approximating the experimentally given structures

Considering two knowledge-based quantities that reflect either pair distributions of atom-atom contacts or solvent-accessible surface portions raises the question, how much redundancy is expressed by both terms? In Figure 14, we summarize the performance of both types of statistical preferences separately applied to the first test set used for FlexX in order to detect the ''well-docked'' solutions. Again the accumulated number of complexes is plotted *versus* the rmsd value observed for the solution on rank 1. While the score based on atom-atom preferences alone clearly performs better than the FlexX score (showed as filled circles), the SAS-dependent score shows only a slight improvement at low rmsd. Interestingly, the latter score does not contain any information on the nature of a molecular counterpart of either the ligand or of the binding site residue. It only takes into account whether polar or non-polar molecular portions are found in a polar or non-polar environment, respectively. Moreover, there is no implicit consideration of detailed atom pair interactions nor is there any information about the geometry of the interactions.

Although the statistics using the total preferences, calculated as a sum of the atom-pair and SAS preferences, only show a slight improvement compared to the consideration of atom-pair preferences alone, in 5 % of all 91 cases a correct recognition of a ''well-docked'' solution on rank 1 would have failed without the SAS-dependent score. An illustrative example exhibits the protein-ligand complex 1ela (Figure 15). The ligand coded by atom types represents the crystal structure, the molecule shown in blue corresponds to the best solution found by only considering the atom-pair preferences (rmsd = 11.9 Å) and the one in red results from using both preference scores (rmsd = 2.5 Å). For the incorrect, strongly deviating mode (blue) a deep hydrophobic pocket, composed by methyl groups of Thr221, Thr236 and Val224, remains accessible to the solvent. Furthermore, a hydrophobic CF_3 group of the ligand also exposes its surface to the solvent. Both features can only be treated correctly if the approach including the SAS-dependent score is applied.

To obtain some insight whether the developed approach depends on the detailed composition of the ligands, we plotted the rmsd values of the best solutions *versus* the percentage of non-polar atoms, the percentage of hydrogen bond donor and acceptor atoms (all related to the total number of heavy atoms), the absolute number of rotatable bonds or the resolution of the experimental structure determinations. No significant correlation could be found for either of these features, apart from a possibly very slight dependence on the number of rotatable bonds. It is likely that the latter points toward the complexity of the problem. The larger the number of rotatable bonds, the more difficult the prediction of a reasonable binding conformation will be. Accordingly, it is not surprising

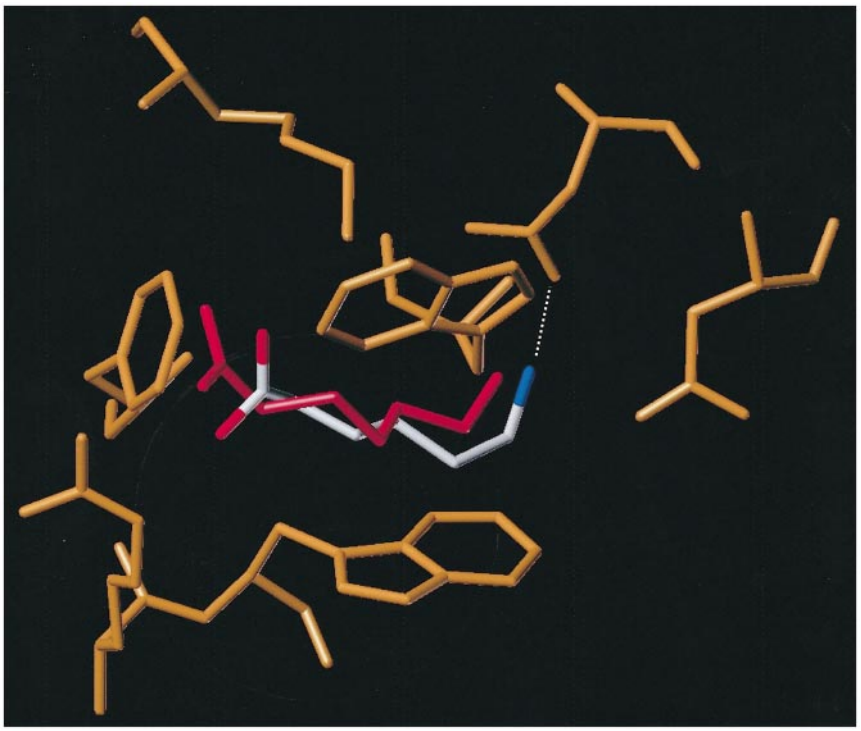

Figure 13. A section of the binding pocket of 2pk4 is shown together with the conformation of the ligand as found by X-ray crystallography (ranked as no. 129 by DrugScore, color-coded by atom type). The geometry of the best scored docking solution deviating by 1.3 Å from the crystal structure is depicted in red. In the crystal, the ligand forms a hydrogen bond between the terminal amino group and a carboxylate group of the protein that measures to only 2.1 Å length (shown as a dotted line).

that a more rigid ligand usually obtains a smaller rmsd value.

Discussion and Conclusion

In this study, distance-dependent pair preferences and SAS-dependent singlet preferences are derived from crystallographically determined protein-ligand complexes. A scoring function incorporating both terms has been shown to be very promising†. It discriminates satisfactorily between well-docked (rmsd <2.0 Å) ligand binding modes and those largely deviating from the native structure generated by the docking tools FlexX and DOCK. This has been shown for test sets comprising 91 and 68 complexes (FlexX) and a mixed subset of 100 complexes (DOCK). A substantial improvement is achieved compared to the original scoring in FlexX as well as the ''chemical scoring'' in DOCK. In comparison with the ''energy scoring'' in DOCK, similar results are obtained.

Apart from computationally demanding methods based on first principles such as FEP or TI, knowledge-based as well as regression-based approaches are primarily applied to predict protein-ligand interactions. While for the latter regression-based approaches the partitioning of the total ligand-to-protein binding features into several additive terms increases our understanding of the contributions to the binding process arising from different physical origin, it is difficult to assess whether this set of considered terms is complete. In addition, only anticipated effects can be considered during the regression analysis and it is by no means clear how strongly these terms are intercorrelated. As a consequence, including even more sophisticated terms to existing state-of-the-art regression-based scoring functions only revealed minor improvements (Bohm, 1998; Stahl & Bohm, 1998). Furthermore, the predictive power of these relationships towards protein-ligand complexes, distinct from all examples used in the training set, is difficult to estimate. Although PLS-derived relations (Head *et al.*, 1996) are in agreement with

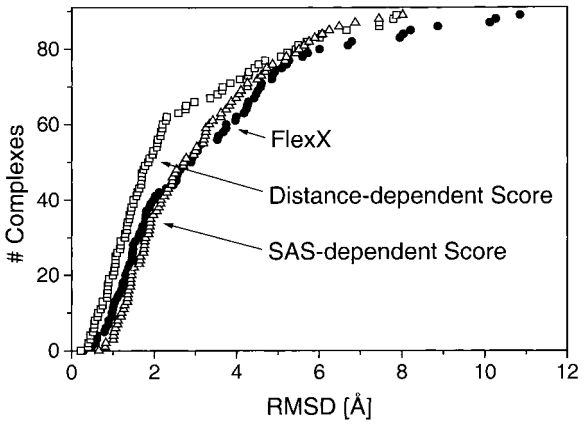

Figure 14. The performance of both statistically derived preference schemes applied to the first test set of 91 protein-ligand complexes to recognize ''well-docked'' solutions in comparison to the original FlexX scoring.

† The new scoring function is still under development. It is planned to incorporate this function into ReLiBase (http://www2.ebi.ac.uk:8081/home.html), a database for the handling of protein-ligand complexes, and to make it available in the framework of this data analysis tool.

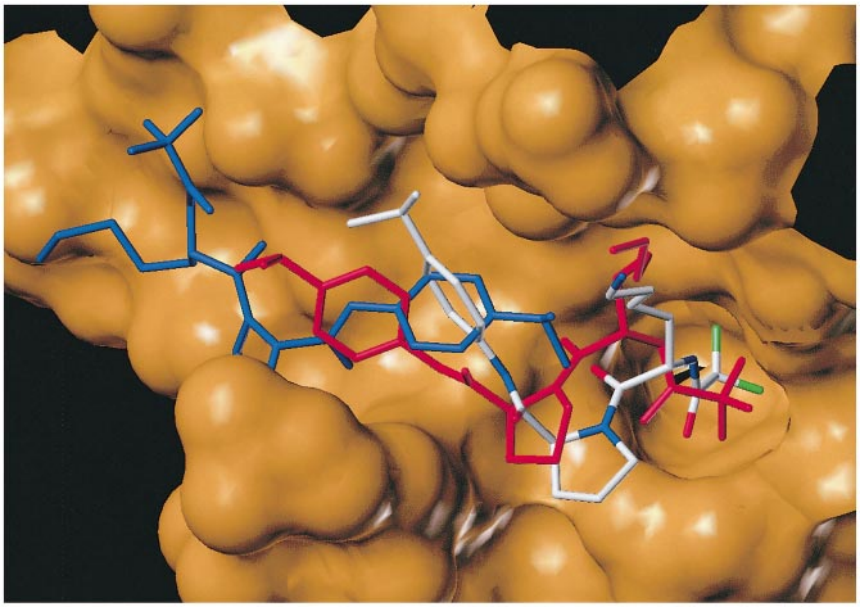

Figure 15. The binding pocket of elastase (PDB code: 1ela) is displayed. The crystal structure of the ligand is color-coded by atom type, the best ranked solution (rmsd = 11.9 Å) applying only the distance-dependent pair preference scheme is colored in blue, the solution (rmsd = 2.5 Å) found on rank 1 applying the pair and the SAS-dependent singlet preferences is colored in red. The deep subpocket on the right is composed of the side-chain of Val224 and two methyl groups of Thr221 and Thr236.

those derived by multiple linear regression, the contribution of each individual term in these former equations is difficult to understand due to the fact that implicitly ''new'' orthogonal descriptors are calculated as linear combinations in terms of the original ones.

Knowledge-based approaches are assumed to be more general, since they implicitly incorporate even those effects that are yet not fully understood. The conversion of structural database information of experimentally determined geometries into statistical preferences regards cooperative effects as well as mainly entropy-driven effects arising from the solvation. They are expected to be taken into account as a result of the mean-field character of the approach. Moreover, less frequently populated states are considered with lower statistical preferences, thus implicitly penalizing computer-generated artifacts.

Since no explicit training set is used in contrast to the derivation of regression-based scoring functions, our scoring function should be generally applicable. This was demonstrated for exemplary cases where all proteins with a homology >30 % to the test case protein were excluded from the database used to derive the preferences. Nevertheless, several parameters had to be adjusted in the context of our approach to reveal optimal results: e.g. the interatomic distance cut-off, the bin size for data sampling and the scaling of the pair preferences to the singlet preferences.

As mentioned already, the choice of the reference state is crucial and could lead to the underestimation of several contributions (Godzik *et al*., 1995). To account for solvent effects, not sufficiently considered by pair preferences collected up to 6 Å (Bahar & Jernigan, 1997), we introduced a SAS-dependent singlet potential.

Only non-hydrogen atom types are used within our scoring function. On the one hand, most complexes in PDB either lack or only contain force-field generated polar hydrogen atoms. Accordingly, most of these atoms would have to be model-built prior to the derivation of the preferences. On the other hand, in particular hydrogen-atom positions strongly depend on the influences of their molecular environment. The local electrostatic field in a protein can change upon ligand binding and might result in substantial pK_a shifts of ionizable groups. In consequence, defining protonation states *a priori,* e.g. during a docking experiment, is by no means straightforward. Although at a first glance, the neglect of H-atom positions appears to imply the loss of information about directionality of polar interactions, the simultaneous consideration of many-fold pair-preferences in a compact molecular environment recovers these features (Figure 5).

Using the data in the Cambridge structural database (CSD; Allen *et al*., 1991) instead of the PDB to derive statistical preferences of intermolecular interactions could cope with these shortcomings. Furthermore, additional data and an enhanced scope of atom types could be studied. However, 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 here the influence of the hydrophobic effect is expected to be smaller in the data derived from the CSD compared to the PDB. This influence was first recognized by Verdonk *et al*. (1999) during the development of SuperStar. Furthermore, since the CSD data are of higher accuracy compared to those in the PDB, the derived statistical preferences are expected to exhibit less scatter.

The recent studies of Verkhivker *et al*. (1995) and Muegge & Martin (1999) use the same formalism to derive potentials; however, they follow a quite

different objective: the application of a knowledge-based scoring function to predict ΔG values of experimentally determined protein-ligand complexes. Our goal is to render prominent the ligand geometry most closely resembling the native structure. The rationale behind this intention is the observation, that scoring functions such as SCORE2 (Bohm, 1998) are already reliable enough to correctly predict ΔG within 1.3 log units if applied to native-like geometries. Virtual screening tools presently generate a whole set of artifical ligand binding modes that have to be detected as such. Moreover, both studies follow a different procedure to derive the potentials or preferences. While Verkhivker and co-workers compile ligand-protein interactions up to 6-7 Å from a small data set of 30 HIV and SIV proteases and relate these pair-interaction potentials to other explicit entropic and solvation terms, Muegge & Martin derive potentials of mean force from 697 complexes stored in the PDB using a cut-off at 12 Å together with a correction term regarding the volume occupied by the ligand in order to incorporate solvent effects into their pair potentials.

Several improvements and enhancements can be imagined. So far our scoring function is applied to post-process solutions from FlexX and DOCK. It is encouraging that DrugScore performs comparably well, independent of the methods applied to generate ligand poses. This suggests its general applicability. Nevertheless, it is not totally obvious whether solely using our approach would reveal comparable results or whether only the combination of two subsequently applied scoring functions (Bohm's SCORE in FlexX or ''chemical scoring''/''energy scoring'' in DOCK and our scoring function) achieves the reported success rate. We are currently implementing DrugScore into FlexX to study these effects. Furthermore, geometrical complementation of the binding sites due to crystal contacts are not yet taken into account during the derivation and the application of our approach. Even if only a minor influence on the shape of the preferences as well as the predictive power is expected, this assumption has to be investigated. The same holds for water molecules often being a mediator for ligand-protein interactions. Cofactors are considered as part of the protein during the compilation of atom/atom preferences, yet they are neglected during ligand scoring. For the latter two situations, a correlation with the actually evaluated binding mode is obvious. These effects have to be taken into account.

Note Added in Proof

During revision of this paper, Mitchell *et al.* (1999b) published the development of an atomic level potential of mean force using high-resolution X-ray structures from the PDB. In total 820 possible atom-atom pairs are evaluated. The performance to identify low-energy binding modes from decoy conformations is tested for one case, the heparin binding to bFGF (PDB code: 1bfc) (Mitchell *et al.*, 1999a). While the crystal structure was ranked lowest, the best scored geometry generated by FTdock deviates largely from the experimental one.

Acknowledgments

This work was performed as part of the RELIMO-Project (grant number 0311619) funded by the German Federal Ministry for Education, Science, Research, and Technology (BMBF). We also acknowledge the help of A. Kulke and H. Velec (University of Marburg) in preparing the DOCK input files. Parts of the work were presented at the 12th European QSAR Symposium, Copenhagen/Denmark, September 1998, and the International Workshop on Virtual Screening, Rauischholzhausen/Germany, March 1999.

References

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. & Watson, D. G. (1991). The development of version-3 and version-4 of the Cambridge Structural Database system. *J. Chem. Inf. Comput. Sci.* **31**, 187-204.

Bahar, I. & Jernigan, R. L. (1997). Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.* **266**, **(1)**, 195-214.

Ben-Naim, A. (1987). *Solvation Thermodynamics*, Plenum Press, New York.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, **(3)**, 535-542.

Beveridge, D. L. & DiCapua, F. M. (1989). Free energy-*via* molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431-492.

Blokzijl, W. & Engberts, J. B. F. N. (1993). Hydrophobe effekte-ansichten und tatsachen. *Angew. Chem.* **105**, 1610-1648.

Bohm, H. J. (1992). The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des.* **6**, **(1)**, 61-78.

Bohm, H. J. (1994). The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* **8**, **(3)**, 243-256.

Bohm, H. J. (1998). Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput. Aided Mol. Des.* **12**, **(4)**, 309-323.

Bohm, H.-J. & Klebe, G. (1996). What can we learn from molecular recognition in protein-ligand complexes for the design of new drugs? *Angew. Chem. Int. Ed. Engl.* **35**, **(22)**, 2566-2587.

Bostrom, J., Norrby, P. O. & Liljefors, T. (1998). Conformational energy penalties of protein-bound ligands. *J. Comput. Aided Mol. Des.* **12**, **(4)**, 383-396.

Bowie, J. U., Luthy, R. & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science,* **253**, **(5016)**, 164-170.

Bruno, I. J., Cole, J. C., Lommerse, J. P., Rowland, R. S., Taylor, R. & Verdonk, M. L. (1997). IsoStar: a library of information about nonbonded interactions. *J. Comput. Aided Mol. Des.* **11**, **(6)**, 525-537.

Burgi, H. B. & Dunitz, J. D. (1988). Can statistical analysis of structural parameters from different crystal environments lead to quantitative energy relationships. *Acta Crystallog. sect. B,* **44**, 445-448.

Burley, S. K. & Petsko, G. A. (1985). Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science,* **229**, **(4708)**, 23-28.

Clark, M., Cramer, R. D., III & Van Opdenbosch, N. (1989). Validation of the general purpose Tripos 5.2 force field. *J. Comp. Chem.* **10**, 982-1012.

Davis, A. M. & Teague, S. J. (1999). Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Angew. Chem. Int. Ed. Engl.* **38**, **(6)**, 736-749.

DeWitte, R. S. & Shaknovich, E. I. (1996). SMoG: *de novo* design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. *J. Am. Chem. Soc.* **118**, 11733-11744.

Dill, K. A. (1997). Additivity principles in biochemistry. *J. Biol. Chem.* **272**, **(2)**, 701-704.

Dixon, J. S. (1997). Evaluation of the CASP2 docking section. *Proteins: Struct. Funct. Genet.* **Suppl. 1**, 198-204.

Eads, J., Sacchettini, J. C., Kromminga, A. & Gordon, J. I. (1993). *Escherichia coli*-derived rat intestinal fatty acid binding protein with bound myristate at 1.5 Å resolution and I-FABPArg106 → Gln with bound oleate at 1.74 Å resolution. *J. Biol. Chem.* **268**, **(35)**, 26375-26385.

Ewing, T. (1997). Editor of *DOCK Version 4.0 Manual*, Regents of the University of California, San Francisco, CA.

Ewing, T. J. A. & Kuntz, I. D. (1997). Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **18**, **(9)**, 1175-1189.

Finkelstein, A. V., Gutin, A. M. & Badretdinov, A. Y. (1995). Perfect temperature for protein structure prediction and folding. *Proteins: Struct. Funct. Genet.* **23**, **(2)**, 151-162.

Godzik, A. (1996). Knowledge-based potentials for protein folding: what can we learn from known protein structures? *Structure,* **4**, **(4)**, 363-366.

Godzik, A., Kolinski, A. & Skolnick, J. (1995). Are proteins ideal mixtures of amino acids? Analysis of energy parameter sets. *Protein Sci.* **4**, **(10)**, 2107-2117.

Head, R. D., Smythe, M. L., Oprea, T. I., Waller, C. L., Green, S. M. & Marshall, G. R. (1996). VALIDATE: a new method for the receptor-based prediction of binding affinities of novel ligands. *J. Am. Chem. Soc.* **118**, 3959-3969.

Hendlich, M. (1998). Databases for protein-ligand complexes. *Acta Crystallog. sect. D,* **54**, 1178-1182.

Hendlich, M., Lackner, P., Weitckus, S., Floeckner, H., Froschauer, R., Gottsbacher, K., Casari, G. & Sippl, M. J. (1990). Identification of native protein folds amongst a large number of incorrect models. The calculation of low energy conformations from potentials of mean force. *J. Mol. Biol.* **216**, **(1)**, 167-180.

Honig, B. & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science,* **268**, **(5214)**, 1144-1149.

Hossain, M. A. & Schneider, H.-J. (1999). Flexibility, association constant, and salt effects in organic ion pairs: how single bonds affect molecular recognition. *Chem. Eur. J.* **5**, **(4)**, 1284-1290.

Jain, A. N. (1996). Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities. *J. Comput. Aided. Mol. Des.* **10**, **(5)**, 427-440.

Jernigan, R. L. & Bahar, I. (1996). Structure-derived potentials and protein simulations. *Curr. Opin. Struct. Biol.* **6**, **(2)**, 195-209.

Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). A new approach to protein fold recognition. *Nature,* **358**, **(6381)**, 86-89.

Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, **(3)**, 727-748.

Knegtel, R. M., Bayada, D. M., Engh, R. A., von der Saal, W., van Geerestein, V. J. & Grootenhuis, P. D. (1999). Comparison of two implementations of the incremental construction algorithm in flexible docking of thrombin inhibitors. *J. Comput. Aided Mol. Des.* **13**, **(2)**, 167-183.

Kocher, J. P., Rooman, M. J. & Wodak, S. J. (1994). Factors influencing the ability of knowledge-based potentials to identify native sequence-structure matches. *J. Mol. Biol.* **235**, **(5)**, 1598-1613.

Koehl, P. & Delarue, M. (1994). Polar and nonpolar atomic environments in the protein core: implications for folding and binding. *Proteins: Struct. Funct. Genet.* **20**, **(3)**, 264-278.

Kollman, P. (1993). Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **93**, 2395-2417.

Kollman, P. A. (1996). Advances and continuing challenges in achieving reaistic and predictive simulations of the properties of organic and biological molecules. *Acc. Chem. Res.* **29**, **(10)**, 461-469.

Koppensteiner, W. A. & Sippl, M. J. (1998). Knowledge-based potentials-back to the roots. *Biochemistry (Moscow),* **63**, **(3)**, 247-252.

Kossiakoff, A. A., Randal, M., Guenot, J. & Eigenbrot, C. (1992). Variability of conformations at crystal contacts in BPTI represent true low-energy structures: correspondence among lattice packing and molecular dynamics structures. *Proteins: Struct. Funct. Genet.* **14**, **(1)**, 65-74.

Kubinyi, H. (1998). Structure-based design of enzyme inhibitors and receptor ligands. *Curr. Opin. Drug Discov. Devel.* **1**, **(1)**, 4-15.

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **161**, 269-288.

Kuntz, I. D., Meng, E. C. & Shoichet, B. K. (1994). Structure-based molecular design. *Acc. Chem. Res.* **27**, **(5)**, 117-123.

Lengauer, T. & Rarey, M. (1996). Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* **6**, **(3)**, 402-406.

Li, A.-J. & Nussinov, R. (1998). A set of van der Waals and Coulombic radii of protein atoms for molecular

and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins: Struct. Funct. Genet.* **32**, 111-127.

Makino, S. & Kuntz, I. D. (1997). Automated flexible ligand docking method and its application for database search. *J. Comput. Chem.* **18**, **(14)**, 1812-1825.

Mitchell, J. B. O., Laskowski, R. A., Alex, A., Forster, M. J. & Thornton, J. M. (1999a). BLEEP-potential of mean force describing protein-ligand interactions: II. Calculation of binding energies and comparison with experimental data. *J. Comput. Chem.* **20**, **(11)**, 1177-1185.

Mitchell, J. B. O., Laskowski, R. A., Alex, A. & Thornton, J. M. (1999b). BLEEP-potential of mean force describing protein-ligand interactions: I. Generating potential. *J. Comput. Chem.* **20**, **(11)**, 1165-1176.

Miyazawa, S. & Jernigan, R. L. (1996). Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **256**, **(3)**, 623-644.

Miyazawa, S. & Jernigan, R. L. (1999). Self-consistent estimation of inter-residue protein contact energies based on an equilibrium mixture approximation of residues. *Proteins: Struct. Funct. Genet.* **34**, **(1)**, 49-68.

Muegge, I. & Martin, Y. C. (1999). A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **42**, **(5)**, 791-804.

Muegge, I., Martin, Y. C., Hajduk, P. J. & Fesik, S. W. (1999). Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein. *J. Med. Chem.* **42**, **(14)**, 2498-2503.

Muller, K. (1995). De novo design. In *Perspectives in Drug Discovery and Design* (Anderson, P. S., Kenyon, G. L. & Marshall, G. R., eds), vol. 3, ESCOM, Leiden.

Murray, C. W., Auton, T. R. & Eldridge, M. D. (1998). Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model. *J. Comput. Aided Mol. Des.* **12**, **(5)**, 503-519.

Park, B. & Levitt, M. (1996). Energy functions that discriminate X-ray and near native folds from well-constructed decoys. *J. Mol. Biol.* **258**, **(2)**, 367-392.

Pickett, S. D. & Sternberg, M. J. (1993). Empirical scale of side-chain conformational entropy in protein folding. *J. Mol. Biol.* **231**, **(3)**, 825-839.

Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **261**, **(3)**, 470-489.

Rose, P. W. (1997). *Scoring Methods in Ligand Design*, 2nd UCSF Course in Computer-Aided Molecular Design, San Francisco.

Sharp, K. A., Nicholls, A., Friedman, R. & Honig, B. (1991). Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry,* **30**, **(40)**, 9686-9697.

Sippl, M. J. (1990). Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.* **213**, **(4)**, 859-883.

Sippl, M. J. (1993). Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Comput. Aided Mol. Des.* **7**, **(4)**, 473-501.

Sippl, M. J. (1995). Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.* **5**, **(2)**, 229-235.

Skolnick, J., Jaroszewski, L., Kolinski, A. & Godzik, A. (1997). Derivation and testing of pair potentials for protein folding. When is the quasichemical approximation correct? *Protein Sci.* **6**, **(3)**, 676-688.

Stahl, M. & Bohm, H.-J. (1998). Development of filter functions for protein-ligand docking-fast, fully automated docking of flexible ligands to protein binding sites. *J. Mol. Graph. Model.* **16**, **(3)**, 121-132.

Testa, B., Carrupt, P. A., Gaillard, P., Billois, F. & Weber, P. (1996). Lipophilicity in molecular modeling. *Pharm. Res.* **13**, **(3)**, 335-343.

Torda, A. E. (1997). Perspectives in protein-fold recognition. *Curr. Opin. Struct. Biol.* **7**, **(2)**, 200-205.

Vajda, S., Sippl, M. & Novotny, J. (1997). Empirical potentials and functions for protein folding and binding. *Curr. Opin. Struct. Biol.* **7**, **(2)**, 222-228.

Van Drie, J. H. & Lajiness, M. S. (1998). Approaches to virtual library design. *Drug Discov. Today,* **3**, **(6)**, 274-283.

Verdonk, M. L., Cole, J. C. & Taylor, R. (1999). SuperStar: a knowledge-based approach for identifying interaction sites in proteins. *J. Mol. Biol.* **289**, **(4)**, 1093-1108.

Verkhivker, G., Appelt, K., Freer, S. T. & Villafranca, J. E. (1995). Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Eng.* **8**, **(7)**, 677-691.

Wallqvist, A. & Covell, D. G. (1996). Docking enzyme-inhibitor complexes using a preference-based free-energy surface. *Proteins: Struct. Funct. Genet.* **25**, **(4)**, 403-419.

Wallqvist, A., Jernigan, R. L. & Covell, D. G. (1995). A preference-based free-energy parameterization of enzyme-inhibitor binding. Applications to HIV-1-protease inhibitor design. *Protein Sci.* **4**, **(9)**, 1881-1903.

Walters, W. P., Stahl, M. T. & Murcko, M. A. (1998). Virtual screening - an overview. *Drug Discov. Today,* **3**, **(4)**, 160-178.

Warshel, A. & Aqvist, J. (1991). Electrostatic energy and macromolecular function. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 267-298.

Edited by R. Huber

(Received 2 August 1999; received in revised form 8 November 1999; accepted 8 November 1999)