DrugScore^{CSD}—Knowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of Near-Native Ligand Poses and Better Affinity Prediction

Hans F. G. Velec,[†] Holger Gohlke,[‡] and Gerhard Klebe*,[†]

Institute of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, D-35032 Marburg, Germany, and Department of Biology and Computer Science, J. W. Goethe-University, Marie-Curie-Straße 9, D-60439 Frankfurt/Main, Germany

Received May 9, 2005

Following the formalism used for the development of the knowledge-based scoring function DrugScore, new distance-dependent pair potentials are obtained from nonbonded interactions in small organic molecule crystal packings. Compared to potentials derived from protein−ligand complexes, the better resolved small molecule structures provide relevant contact data in a more balanced distribution of atom types and produce potentials of superior statistical significance and more detailed shape. Applied to recognizing binding geometries of ligands docked into proteins, this new scoring function (DrugScore^{CSD}) ranks the crystal structures of 100 protein−ligand complexes best among up to 100 generated decoy geometries in 77% of all cases. Accepting root-mean-square deviations (rmsd) of up to 2 Å from the native pose as well-docked solutions, a correct binding mode is found in 87% of the cases. This translates into an improvement of the new scoring function of 57% with respect to the retrieval of the crystal structure and 20% with respect to the identification of a well-docked ligand pose compared to the original Protein Data Bank-based DrugScore. In the analysis of decoy geometries of cross-docking studies, DrugScore^{CSD} shows equivalent or increased performance compared to the original PDB-based DrugScore. Furthermore, DrugScore^{CSD} predicts binding affinities convincingly. Reducing the set of docking solutions to examples that deviate increasingly from the native pose results in a loss of performance of DrugScore^{CSD}. This indicates that a necessary prerequisite to successfully resolving the scoring problem with a more discriminative scoring function is the generation of highly accurate ligand poses, which approximate the native pose to below 1 Å rmsd, in a docking run.

Introduction

The generation and recognition of "relevant" binding geometries is an essential prerequisite for successful virtual screening. While the former (the "docking" problem) is considered to be solved in the case of rigid receptors,[1,2] the latter (the "scoring" problem) still provides a major challenge. In consequence, several new scoring functions have been developed over the last two years[3−9] and established ones have achieved improvement.[10−12]

The fastest growing is obviously the class of regression-based scoring functions,[3−5] in particular, because their predictive power with respect to binding affinities appears most promising. Nevertheless, the general applicability of these functions strongly depends on the training set. In the field of knowledge-based scoring functions, besides the new function described here, PLASS[6] was introduced recently. The group of empirical functions has been enhanced by a new specific free energy function for carbohydrate−protein interactions[7] and by a method combining empirical and knowledge-based approaches.[8] Finally, a scoring function based on fast quantum mechanical calculations was introduced.[9] As the number of scoring functions grows steadily, many researchers focus more and more on improved evaluation protocols to assess scoring results or to perform consensus scoring,[13−16] and an increasing number of elaborate comparative evaluations of scoring functions have been reported that are very supportive when selecting the appropriate function for a given problem.[17−21]

In a comparative study of 11 scoring functions,[21] Wang et al. demonstrated that most of them do not recognize the experimentally determined ligand geometry as most favorable but succeed to retrieve a deviating pose as optimum. This indicates that scoring function minima do not necessarily correspond to those found in crystal structures or that they are not tolerant enough to cope with widely accepted geometrical deviations from native poses. Accordingly, the native geometry could be scored significantly worse.

Crystal coordinates are accurate to about 0.5 Å in position, and in docking, overall rms deviations of up to 2 Å are generally accepted as near-native. However, with respect to drug design, it is highly desirable to detect a ligand pose that matches the native geometry as closely as possible (clearly less than 2 Å rmsd). In consequence, it is questionable whether pronounced positional deviations should still be tolerated, in particular in view of the detrimental influence they may have on predicting binding affinities.

When developing a knowledge-based function to score protein−ligand complexes, the most obvious approach appears to extract structural information from experi-

* Corresponding author. Phone: +49/6421/2821313. Fax: +49/6421/2828994. E-mail: klebe@mailer.uni-marburg.de.
[†] University of Marburg.
[‡] J. W. Goethe-University.

10.1021/jm050436v CCC: $30.25
Published on Web 09/07/2005

mentally determined protein−ligand complexes. This strategy has been pursued in DrugScore.[22,23] The information retrieved from crystallographically determined protein−ligand complexes as stored in the Protein Data Bank (PDB)[24] has been converted into distance-dependent pair preferences as well as into singlet potentials scaled by the solvent accessible surface area (SASA). Hereby, the statistical significance of each compiled pair potential strongly depends on the frequency of occurrence of the individual atom−atom pair contacts in the database. In consequence, some interaction potentials are less significant. In that respect, assuming that nonbonded interactions in small molecule crystal packings are governed by the same physical principles as the interactions between proteins and ligands, the Cambridge Structural Database (CSD) provides a rich source to learn about interaction geometries that are not or only insufficiently available from crystallographically determined protein−ligand complexes. The idea of like forces driving the formation of protein−ligand complexes and small molecule crystalline assemblies is supported by the fact that the same packing density is found for both systems. Furthermore, several studies demonstrate that the local interaction geometries in small molecule crystals and protein−ligand complexes agree convincingly well.[25−28]

In the present contribution, we present distance-dependent pair potentials derived from small molecule crystal data for scoring protein−ligand interactions (DrugScore^{CSD}). We show the substantial improvement of the statistical significance of many pair potentials and the superior performance of the CSD-based potentials compared to that of the original DrugScore (DrugScore^{PDB}) in the recognition of near-native ligand binding modes and the mutual ranking with respect to experimentally determined binding affinities.

Methods

Most of the methods for the derivation of pair potentials applied here have been adapted from Gohlke et al.[22]

Distance-Dependent Pair Potentials. Following Sippl's approach,[29] specific interactions $\Delta W_{i,j}$ between atoms of type i and j, located at a distance r, can be obtained from the normalized radial pair distribution function $g_{i,j}(r)$ and the normalized mean radial pair distribution function $g(r)$.

$$\Delta W_{i,j}(r) = -\ln \frac{g_{i,j}(r)}{g(r)}; g(r) = \frac{\sum_i \sum_j g_{i,j}(r)}{ij}$$

Here, $g_{i,j}(r)$ was compiled from occurrence frequencies N of atom pairs with types i and j

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

where distances $1.0 \le r \le 6.0$ Å were considered. All other parameters were chosen as defined in ref 22.

Scoring of Protein−Ligand Complexes. The specific interaction ("binding score") between two molecules I and J (e.g., a protein and a ligand) is calculated by DrugScore as the sum of all occurring atom−atom interactions.

$$\Delta W_{I,J} = \sum_{i \in I} \sum_{j \in J} \Delta W_{i,j}(r)$$

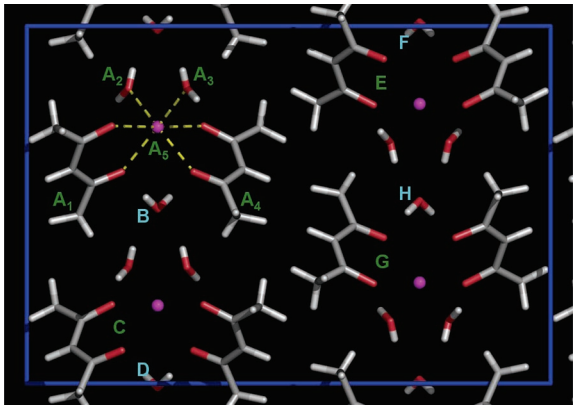

Figure 1. Exemplary crystal packing for the derivation of pair potentials. Shown is a calcium−acetylacetone complex (refcode BOLTIF). Molecules in the unit cell are denoted by letters. To derive the pair potentials, only complex portions A_1−A_4 and calcium ion A_5 are considered to be central molecules. Molecules C, E, and G as well as all other molecules in the crystal packing form the interacting environment. Water molecules (B, D, F, and H) were not taken into account. The dimensions of the unit cell are indicated by blue lines. This visualization was produced using PyMOL.[32]

Small Molecule Data To Derive Pair Potentials. Pair potentials for nonbonded interactions are derived from small molecule crystal packing data as stored in the Cambridge Structural Database[30] by applying ConQuest[31] as a query engine. Only structures that have been determined by crystal structure analysis and refined to an R-factor of less than 0.05 and are comprised of at least one of the elements C, H, O, S, P, N, F, Cl, Br, I, Ca, Fe, and Zn are considered in the following analysis. Structures that have been flagged as erroneous were excluded, as were all structures with incomplete atomic coordinates. Furthermore, structures containing bonding information incompatible with our assignment, e.g., hydrogen bonds coinciding with a symmetry element or functional groups involved in hypervalent complexes, were discarded. Crystal packings of the appropriate size were generated for each of the 28 642 CSD entries (for a comprehensive list, see the Supporting Information) so that in all cases one central molecule is fully embedded into a complete contact environment of neighboring molecules. Each molecule present in the crystallographic asymmetric unit was used once to determine nonbonded contact distances to all neighboring molecules in the crystal packing. Accordingly, in case of multiple molecular entries in the asymmetric unit (Figure 1), all other molecules in the packing were included. Complexed metal ions were treated as independent "molecules". Uncomplexed water molecules, nonmetal ions, and molecules that contained less than six heavy atoms were excluded from the sampling.

Potentials were derived for all DrugScore standard atom types (similar to the Sybyl atom-type notation[33]): 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 amide 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 (tetrahedral sulfur), P.3 (tetrahedral phosphorus), F (fluorine), Cl (chlorine), and Br (bromine). Calcium, zinc, and iron were subsumed in the atom type Met. Additionally, the atom type I (iodine) was introduced. To guarantee full compatibility to DrugScore^{PDB}, the following atom types were grouped together: S.2 (bivalent sulfur) and S.3 as well as N.4 (positively charged nitrogen) and N.3.

Validation. We have assessed the performance of the derived potentials by ranking docked ligand poses as well as by comparing predicted to experimentally determined binding affinities. We used the data set of protein−ligand complexes published by Wang et al.[21,34] We selected this data set because it served as a reference panel for the most comprehensive

assessment of presently available scoring functions described in the literature. Thus, it allows a very exhaustive comparison with our results achieved on this data sample. The Wang data set consists of 100 crystallographically determined complexes for which additional ligand geometries have been generated using AUTODOCK.[35] Therefore, docking parameters were adjusted to cover a broad range of geometries, such that also poses strongly deviating from the experimentally determined structures were obtained. Following this route, Wang et al. produced data sets of 100 complex geometries considering arrangements of up to 20 Å rmsd from the crystal structure in addition to the experimentally determined ligand poses. Taking the measured affinity data as reference, they subjected the entire collection of 100 complexes to 11 scoring functions. In the present contribution, we follow a similar procedure to validate DrugScore^{CSD}. Another common benchmark to assess the robustness of scoring functions is their application to score ligand poses in a cross-docking run. Recently, Ferrara et al. tested DrugScore^{PDB} among other scoring functions[18] on a set of six trypsin (1tni, 1tng, 1tpp, 1ppc, 1pph, and 3ptb) and seven HIV-1 protease complexes (1ajv, 1gno, 1hih, 1hps, 1htf, 1hvi, and 2upj). Up to 100 decoys for the parent complex and 100 cross-decoys using the geometry of the other protein complexes with rms deviations between 0 and 20 Å were generated by molecular dynamics and subsequently minimized for all 85 possible protein−ligand combinations. In the present contribution, we performed a very similar validation using the complex geometries generated by Ferrara et al.[18] and applied DrugScore^{CSD} to recognize near-native geometries in this cross-docking test.

Results and Discussion

Information about nonbonded interactions derived from small molecule crystal data is widely used for analyzing and predicting protein−ligand interactions. For example, propensity maps generated by the program SuperStar[36] allow prediction of "hot-spots" of binding in protein pockets favorable for occupation with certain ligand atom types. By comparing scatter plots of interactions extracted from both PDB and CSD, Boer et al.[37] demonstrated the geometrical equivalence of nonbonded interactions retrieved from both sources. Similar findings have been discussed by Taylor.[38,39] However, differences in the occurrence frequencies of hydrophobic contacts between CSD- and PDB-derived interaction data have been noticed.[40] Here, we examine the similarity of CSD-derived knowledge-based potentials with respect to the original DrugScore^{PDB} potentials and compare their predictive power with respect to the scoring of protein−ligand interactions.

Properties of CSD-Based Pair Distributions. The statistical significance of knowledge-based potentials increases with the amount of data evaluated to assemble the various atom−atom contact histograms. In our case, about 500 pair interactions per histogram were considered to yield sufficiently smooth potential curves, which means that each 0.1 Å width bin is populated on average by 10 contacts. Due to the large number of CSD entries that have been enumerated to derive the 364 potentials, 79% (including iodine; 84% excluding I) of all potentials are based on pair distributions that contain more than 500 contacts (Figure 2). In contrast to the occurrence frequencies obtained from the PDB, which are inherently biased toward interactions formed by atom types present in amino acids (in particular C.3, N.am, and O.2), the contact pair frequencies extracted from CSD are based on a better balanced distribution across all atom types found in druglike molecules. Also a sufficient number of contacts is available for atom types that rarely occur in protein−ligand complexes but increasingly matter in drug research and, accordingly, require reliable scoring, such as fluorine,[41] chlorine, bromine, or sulfur in different coordination states.

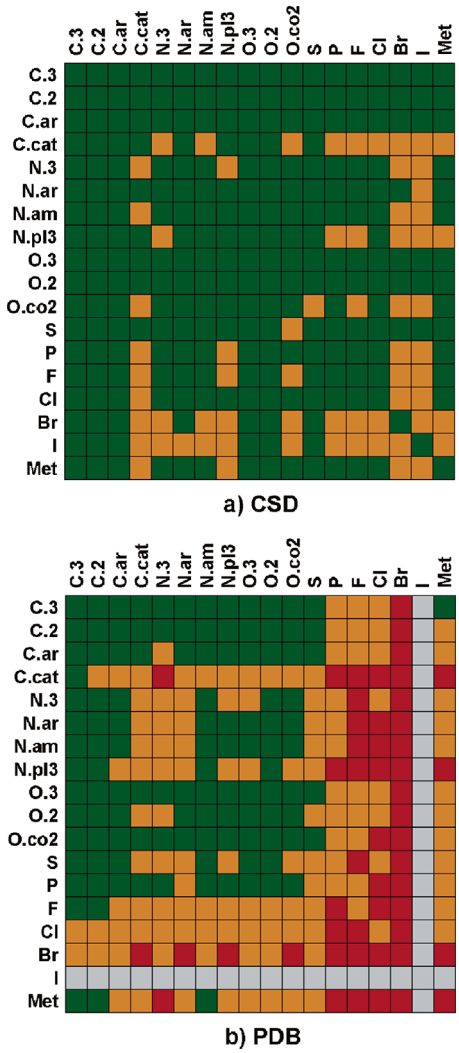

Figure 2. Occurrence frequencies of contacts between two atoms of denoted types for (a) CSD-derived contact data and (b) PDB-derived contact data (iodine was not examined here). Sufficient contact numbers (≥500) are marked green, orange areas mark low contact numbers (<500), and red areas denote contacts which do not occur in the database.

Characteristics of Derived Pair Potentials. Figure 3 shows examples of pair potentials for interactions between lipophilic atoms (C.3 and C.3), the charged interaction between two atoms of types O.co2 and N.pl3, and the interaction between chlorine and carbon (Cl and C.3). Potentials that are derived from a large population of nonbonded contacts in both of the corresponding databases show similar overall shapes and relative positions of the minima, as is apparent in Figure 3 for C.3−C.3 and O.co2−N.pl3 pair preferences. Since the resolution of the individual crystal structures stored in the CSD is remarkably higher and, accordingly, the uncertainties in atomic coordinates are lower, the resulting pair potentials show steeper potential wells compared to those of the PDB-based data. Furthermore, second-order minima corresponding to interactions with the shell of second-nearest neighbors are apparently better defined, whereas the PDB-based potentials appear more blurred and seem to spread favorable inter-

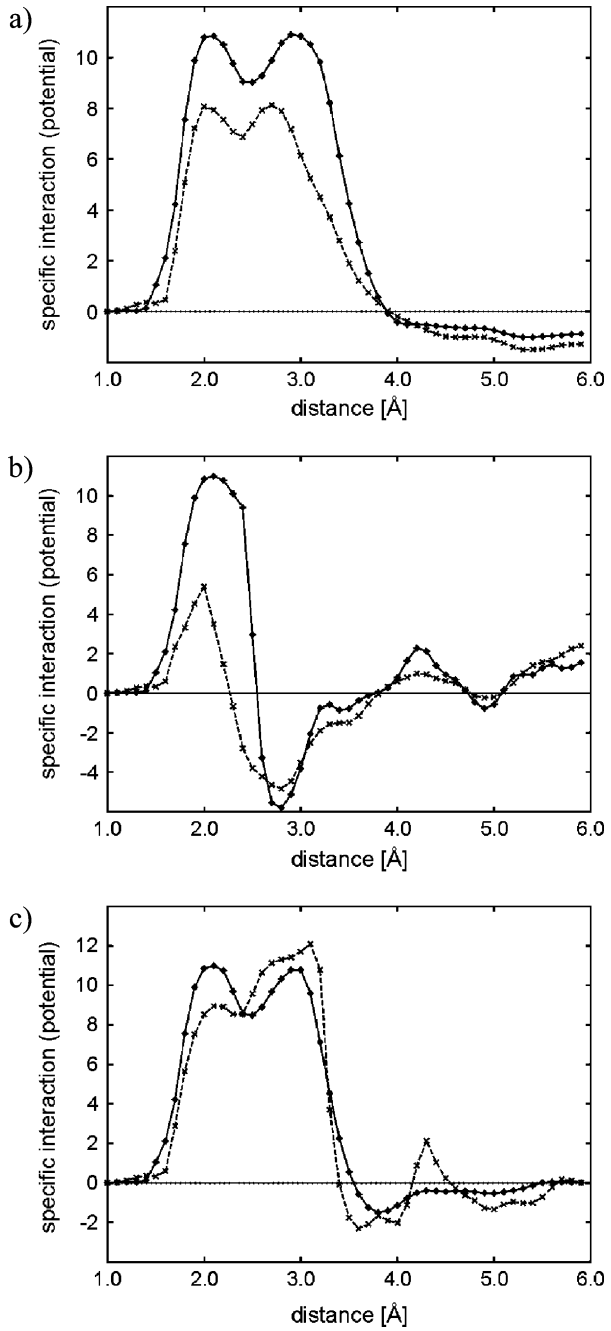

Figure 3. Distance-dependent pair potentials derived from the CSD (◆) and PDB (×) for (a) the hydrophobic contact between atoms of types C.3 and C.3, (b) the polar/charged interaction between atoms of types O.co2 and N.pl3, and (c) the interaction between atoms of types Cl and C.3, which was derived from a high frequency of contacts (49 705) in the CSD case and from a rather limited number of contacts (317) for the PDB data (see also Figure 2).

actions across a wider range of distances. In total, CSD-based potentials show more pronounced extrema compared to those of the PDB-based analogues. Not unexpectedly, differences in the general shapes of the potentials originating from both databases are experienced in those cases in which at least one of the potentials is based on a low occurrence frequency of contacts. For example, the potential given in Figure 3c (Cl−C.3) was derived from about 50 000 contacts observed in 2312 crystal packings in the CSD. In contrast, only 317 contacts were available from 11 database entries in the case of the PDB data.

To quantify the similarity of the different contact potentials derived from either CSD or PDB data, a similarity index introduced by Hodgkin et al.[42] and comparably applied by Nissink et al.[43] was computed. All mutual similarity indices were compiled in a similarity matrix. This matrix shows high correspondence for potentials that were derived from highly populated distributions in either the PDB or the CSD, whereas deviating potentials are mainly observed if the amount of data extracted from at least one of the databases was insufficient.

With respect to the hydrophobic effect mentioned above, the CSD-based pair potentials reveal lower probabilities for contacts between any of the C.3-, C.2-, or C.ar-type atoms in comparison to the case of PDB-based potentials. For example, considering a hydrophobic (C.3−C.3) contact at a distance of 4.1 Å and a hydrophilic (O.3−O.3) contact at 2.7 Å, which correspond to the first maxima in the distribution frequencies normalized with respect to $4\pi r^2$, the hydrophobic interaction as detected by the CSD potentials contributes only 18% to the total score, whereas, in the case of the PDB potential, the hydrophobic contribution amounts to 27%. This trend is also found throughout most hydrophilic/hydrophobic interaction pairs, which means that hydrophobic contacts contribute less to the total score for the CSD-based potentials in comparison to the PDB-based ones. An actual explanation for the observed deviations is difficult to give; however, the phenomenon has already been described by others.[37] Possibly, the difference in crystallization conditions might provide some reasoning, as protein complexes are usually crystallized from aqueous solutions, whereas in the case of small organic molecules frequently nonaqueous solvents are applied.

Reranking of Ligand Poses. Using DrugScore based on CSD-derived pair potentials (DrugScore^{CSD}), for 77% of the 100 evaluated protein−ligand complexes of the Wang data set, the experimentally determined ligand geometries are ranked best out of the total set including all 100 decoys. For the remaining examples, the crystal structure is found in 11% and 2% of all cases in the second or third rank, respectively. Thus, for 90% of all cases, the crystallographically determined geometries are found in the first three ranks. Allowing the successful retrieval of docking solutions considered to be "well-docked" (rmsd ≤ 2.0 Å) for 87% of the 100 test cases, either the crystal structure or one of these well-docked solutions is found in rank 1. This means that the application of CSD-based potentials results in an improvement of 57% with respect to the retrieval of the crystal structure and 20% with respect to the identification of a well-docked ligand pose compared to the case of the original PDB-based DrugScore^{PDB}. A summary of the scoring results in terms of different rmsd thresholds is given in Table 1. We also tested DrugScore^{CSD} in a more "real-life" scenario where the crystallographically determined binding modes of the ligands have been omitted. In this case, only those complexes were included for which at least one docking solution was present that showed an rmsd ≤0.5 or 2 Å (16 or 91 out of 100, respectively). In addition to these solutions, all other decoys were taken into account. DrugScore^{CSD} finds in 88% or 66%, respectively, of these examples a

Table 1. Retrieval Rate of Ligand Poses in the First Scoring Rank[a]

	success rate				
scoring function	rmsd 0.0 Å	rmsd ≤0.5 Å[b]	rmsd ≤1.0 Å	rmsd ≤1.5 Å	rmsd ≤2.0 Å[b]
DrugScore^{CSD}	**77**	**82 (88)**	**83**	**85**	**87 (66)**
Cerius2/PLP	52	58 (75)	63	69	76 (70)
SYBYL/F-Score	38	47 (63)	56	66	74 (68)
Cerius2/LigScore	48	58 (88)	64	68	74 (60)
DrugScore^{PDB}	49	58 (81)	63	68	72 (65)
Cerius2/LUDI	23	33 (63)	43	55	67 (64)
X-Score	25	33 (50)	40[c]	54	65 (64)[c]
AutoDock	8	19 (69)	34	52	62 (66)
Cerius2/PMF	32	35 (38)	40	46	52 (48)
SYBYL/G-Score	13	15 (25)	24	32	42 (43)
SYBYL/ChemScore	7	8 (6)	12	26	35 (34)
SYBYL/D-Score	3	3 (0)	8	16	26 (29)
Lennard Jones 12−6[d]	57	61 (38)	65	66	68 (40)
N_{cplx}[e]	100	100 (16)	100	100	100 (91)

[a] The table corresponds to Table 2 from ref 21. The success rate is given as a percent with respect to all complexes analyzed, allowing rmsd deviations as indicated. Scoring functions are ranked by their success rates at rmsd values ≤2.0 Å. [b] Percentages in parentheses denote results obtained when excluding the crystal structure geometry. [c] Values differ from the results given in ref 21. [d] Scoring function is a standard Lennard-Jones 12−6 potential. [e] Number of complexes that comprise docking solutions below the given thresholds.

"well-docked" ligand geometry in the best rank. This recognition rate is slightly better compared to that of DrugScore^{PDB} (81%) if only minor deviations (rmsd ≤0.5 Å) are allowed. It is equivalent if deviations up to 2 Å are tolerated (66/65%). Considering the 66% achieved using a threshold of 2 Å, it is remarkable that, for the remaining 34% of the test examples, no docking solutions are available in the data set that show rmsd values ≤0.5 Å. This result indicates that superior performance of DrugScore^{CSD} with respect to the PDB-based version is particularly achieved if more accurate ligand poses are present in the data set.

To further validate the influence of the quality of the ligand poses on the recognition rate of well-docked ligand geometries, we also removed all poses with an rmsd ≤1 Å. Accordingly, only cases with at least one docking solution between 1 and 2 Å rmsd have been considered (90 out of 100 complexes). Under these conditions, the recognition rate decreases to 54%. This indicates that, for the 11 protein−ligand complexes of the test sample, the recognition of a correct ligand pose depends on the presence of a geometry with an rmsd ≤1.0 Å. Taken together with the above finding, generating highly accurate docking solutions, thus, is a necessary prerequisite for reliably recognizing well-docked poses.

Two recent studies report on the influence of steric complementarity between the receptor and its natural ligand for recognizing near-native structures among a set of decoys,[18,44] although it is noted that this effect may be overemphasized by the way the validation data sets of these studies were prepared.[18] Hence, to test whether evaluating steric complementarity alone is sufficient to identify the crystal structure or one of the well-docked solutions in rank 1, all ligand poses were also rescored with a Lennard-Jones 12−6 potential (parameters were taken from AUTODOCK[35]). Interestingly, in the case of identifying the native ligand pose, using this approach yields the second best recognition

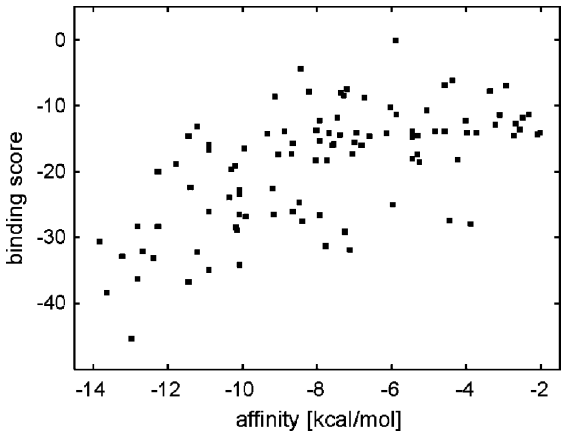

Figure 4. Correlation between experimentally determined binding free energies (kcal/mol) and binding scores of 100 protein−ligand complexes (R_S = 0.624).

rate (57%) of all scoring functions investigated (Table 1), which again confirms the role of steric complementarity in these experiments. Convincingly, however, the recognition rate obtained by DrugScore^{CSD} is higher by 20%. This indicates that the success of the new scoring function can be attributed to not only a better representation of steric interactions due to the increased steepness of the potentials compared to the case of DrugScore^{PDB} but also the better description of chemical complementarity provided by these potentials. In the case of recognizing well-docked solutions, both DrugScore^{PDB} and DrugScore^{CSD} outperform the Lennard-Jones potentials, a fact that becomes particularly pronounced if the crystal structure geometries are excluded from the decoy sets. In the latter case where ligands sterically fit into the receptor pocket in a less-than-ideal manner, the knowledge-based scoring functions, thus, appear to be more robust.

Cross-Decoy Analysis. The obvious importance of steric complementarity on the recognition rate of the crystal structure or near-native geometries questions the sole use of self-docking experiments to assess the quality of a scoring function in this context. In addition, the reduced recognition rate of near-native geometries compared to those of the crystallographically determined ones raises the question of whether the enhanced steepness of the CSD-derived potentials parallels in a loss of scoring power once less precise geometries are used. To analyze such influences, the test data generated by Ferrara et al.[18] on six trypsin and seven HIV protease complexes has been scored by DrugScore^{CSD}. It shows equivalent or increased performance compared to the original DrugScore^{PDB} with respect to successfully scoring either trypsin or HIV protease decoys (diagonal in Figure 5) or cross-decoys (off-diagonal data in Figure 5). The overall rmsd of the retrieved geometries with respect to the corresponding crystal structures is better for DrugScore^{CSD} in both cases. This underlines the fact that DrugScore^{CSD} recognizes more reliably near-native geometries, provided they are present in the sample. For trypsin, the more rigid reference enzyme in this test, DrugScore^{CSD} performs in a more balanced way compared to DrugScore^{PDB} and succeeds in 80% of the cross-decoys. No geometry deviating by more than 2.9 Å rmsd was placed in the best-scored rank throughout all combinations, which is a remarkable result and was not

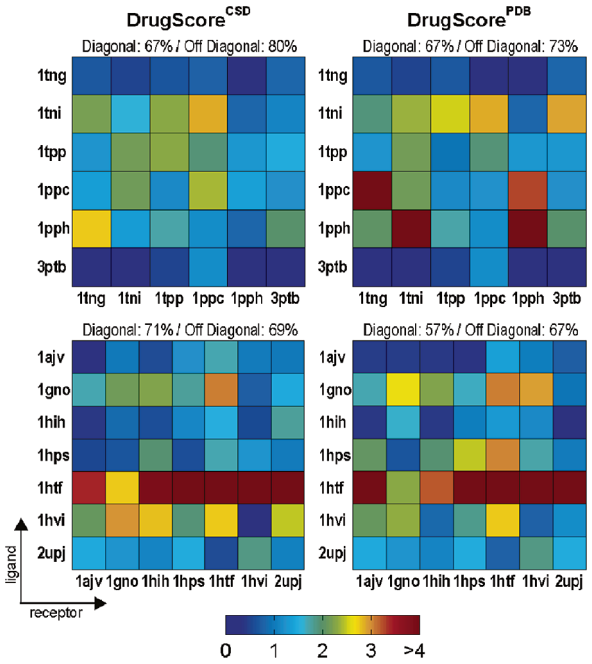

Figure 5. Rmsd (in Å) of the lowest energy configuration from the native structure for the all-pairs decoys of trypsin (top) and HIV-1 protease (bottom) as found by DrugScore^{CSD} (left) and DrugScore^{PDB} (right). The abscissa and ordinate represent the different receptors and ligands, respectively. The average recognition rates for the decoys (diagonal elements) and cross-decoys (off-diagonal elements), expressed in terms of a percentage of complexes for which the lowest energy decoy deviates ≤2 Å rmsd from the crystal structure, are indicated above the plots. This figure corresponds to Figure 6 from ref 18.

achieved by CHARMm or Chemscore.[18] For both of the unsatisfactorily performing HIV complexes with respect to the recognition of the decoys (1gno and 1htf), it must be noted that a considerable portion of the ligands interact with neighboring protein molecules in the crystal packing. These additional contacts have not been considered in the analysis performed by Ferrara et al. As a consequence of the disregarding of these additional contacts, both DrugScore^{CSD} and DrugScore^{PDB} select better-buried decoys in the first ranks.

Assessment of Binding Affinities. Correctly ranking different ligands with respect to their binding affinities toward a given target still provides one of the major challenges in structure-based drug design. To demonstrate the predictive power of DrugScore^{CSD} with respect to affinity prediction, a Spearman's rank order correlation coefficient R_S was calculated for the correlation between experimentally determined binding affinities and computed scores for all 100 test examples. R_S is the nonparametric analogue of Pearson's correlation coefficient, and nonparametric correlation is considered to be more robust than linear correlation.[45] An R_S value of 1 indicates a perfect correlation, and an R_S value of −1 a perfect anticorrelation. DrugScore^{CSD} correlates binding affinities and binding scores well (Figure 4) and yields an R_S of 0.62, which is slightly better than the result obtained by DrugScore^{PDB} (R_S = 0.59). A comparative summary of R_S values for different scoring functions is given in Table 2. It is interesting to note, however, that already the correlation between molecular weight and experimentally determined affinities yields

Table 2. Correlations between Binding Scores and Experimentally Determined Binding Affinities[a]

	Spearman correlation coefficient (R_S) based on	
scoring funtion	the experimentally observed conformations	the best-scored conformations
X-Score	0.660	0.698
DrugScore^{CSD}	**0.624**	**0.622**
Cerius2/PLP	0.593	0.609
DrugScore^{PDB}	0.589	0.603
SYBYL/G-Score	0.570	*0.534*
SYBYL/D-Score	*0.476*	*0.491*
SYBYL/ChemScore	*0.432*	*0.437*
Cerius2/LUDI	*0.431*	*0.458*
Cerius2/PMF	*0.370*	*0.368*
Cerius2/LigScore	*0.368*	*0.421*
SYBYL/F-Score	*0.287*	*0.257*
AutoDock	*0.118*	*0.425*

[a] This table corresponds to Table 6 from ref 21 and was recalculated using the data provided in the Supporting Information. Scoring functions are ranked with respect to correlation coefficients that are calculated by using the experimentally observed conformation of each ligand. The italicized values show correlations worse than a pure correlation with the molecular weight of the ligands (R_S = 0.56).

an R_S of 0.56 (actually an anticorrelation of −0.56 due to the positive values of the molecular weights in contrast to the negative values of the affinities). Hence, for the given data set, affinity predictions purely based on the ligand's molecular weight give in fact better results than many scoring functions listed in Table 2, which is a striking and, however, also puzzling observation.

Summary and Conclusions

Using the concept and formalism of DrugScore,[22,23] a new variant, named DrugScore^{CSD}, has been developed. It is based on contact data found in the crystal packing of small organic molecules. Apparently, the better resolved small molecule structures provide relevant contact data in a more balanced distribution and produce potentials of superior statistical significance and more detailed shape.

Application of these potentials to a data set of 100 protein−ligand complexes (each represented by the experimentally found geometry and up to 100 decoy docking poses maximally deviating up to 20 Å rmsd from the crystal reference) demonstrates the considerable improvement of the new function. As such, DrugScore^{CSD} is capable of retrieving from the test sets the crystal structure as the best solution in the majority of all cases (77%). The superior performance of DrugScore^{CSD} diminishes once the crystal geometry and an increasing fraction of the near-native solutions are excluded from the test sets. This indicates the existence of the requirement to generate accurate docking geometries to achieve good recognition rates of relevant binding modes. Accordingly, in contrast to the professed opinion that in docking the geometry problem has been resolved to a sufficient extent and the scoring problem remains as an open question,[1,2] the present study shows that both are intimately related. The scoring problem can only be increasingly alleviated if better and more relevant binding poses are produced by docking programs. Accordingly, the usually accepted accuracy limit of a 2 Å deviation for the recognition of a near-native

docking solution seems by far too large to perform subsequently a reliable scoring on such crude geometries, at least if scoring functions of the present fine-grain type are used. As such, the enhanced recognition rate of near-native poses most likely results from the rather steep, well-structured and better-discriminating CSD potentials. Accordingly, it appears most advisable to drive docking solutions as close as possible to the native geometries by minimizing them with respect to a scoring function such as DrugScore^{CSD}. Nevertheless, it has to be stressed that DrugScore^{CSD} shows by far the best retrieval rate yet reported for any scoring function in recognizing either the native crystal geometry or any pose deviating by less than 0.5 Å among a sample of decoy poses. Considering the fact that crystal structures are affected by coordinate errors of about 0.3−0.5 Å of deviation, the latter slightly deviating poses can be assumed to be equivalent to the crystal structure. The analysis of decoy geometries in our cross-docking study on trypsin and HIV-1 protease again indicates that the CSD potentials are comparable to or even better in performance than the original PDB ones, once less accurate geometries affected by larger positional uncertainties of atomic coordinates are considered.

In addition, DrugScore^{CSD} shows convincing predictive power with respect to binding affinity predictions. In contrast to DrugScore^{PDB}, the DrugScore^{CSD} function does not evaluate singlet potentials. We examined a formalism equivalent to the one applied in the original DrugScore^{PDB}. However, no significant improvement in the predictive power was observed. As a consequence, we neglect such potentials. This results in a significant reduction of the computational requirements (factor of 15). Across the considered test data set of 100 complexes, DrugScore^{CSD} shows good performance (only superseded by X-Score) and displays improvement with respect to DrugScore^{PDB} in binding affinity predictions. However, such considerations highly depend on the composition of the evaluated data set. Taking only the molecular weight of the ligands as a descriptor to estimate the binding affinity, a fair correlation can already be obtained for this data set. Interestingly enough, the molecular weight performs better than many of the applied scoring functions listed in Table 2. Accordingly, future developments of scoring functions must consider this fact (and include test sets showing anticorrelation with respect to molecular weight) in order to avoid a simple size dependence of scoring functions. Otherwise, in a virtual screening application, such scoring functions will tend to retrieve preferentially larger ligands (which possibly turn out to be weak binders) and fail to capture small but actually strong-binding ligands.

Acknowledgment. We acknowledge the Cambridge Crystallographic Data Centre (CCDC) for making the CSD data available to us and Charles L. Brooks, III, for providing us with the structural data of the trypsin and HIV-1 protease decoys. H.F.G.V. is grateful for financial support by CCDC and the Scoring Function Consortium (SFC) assembled by the following drug companies: Agouron, AstraZeneca, Aventis, BASF, Boehringer Ingelheim, GlaxoSmithKline, Novo Nordisk, Pfizer, Roche, and Schering.

Supporting Information Available: The refcode list of molecules used for the derivation of the pair potentials is available free of charge via the Internet at http://pubs.acs.org. For individual trials, an online version of DrugScore^{CSD} is available at http://www.agklebe.de/drugscore.

References

(1) Dixon, J. S. Evaluation of the CASP2 Docking Section. *Proteins* **1997**, Suppl., 198−204.
(2) Owens, J. Chris Lipinski discusses life and chemistry after the Rule of Five. *Drug Discovery Today* **2003**, *8*, 12−16.
(3) Deng, W.; Brenemann, C.; Embrechts, M. J. Predicting Protein−Ligand Binding Affinities Using Novel Geometrical Descriptors and Machine-Learning Methods. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 699−703.
(4) Giordanetto, F.; Cotesta, S.; Catana, C.; Trosset, J.; Vulpetti, A.; Stouten, P. F. W.; Kroemer, R. T. Novel Scoring Functions Comprising QXP, SASA, and Protein Side-Chain Entropy Terms. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 882−893.
(5) Bock, J. R.; Gough, D. A. A New Method to Estimate Ligand Receptor Energetics. *Mol. Cell. Proteomics* **2002**, *1*, 904−910.
(6) Ozrin, V. D.; Subbotin, M. V.; Nikitin, S. M. PLASS: Protein−ligand affinity statistical score − a knowledge-based force-field model of interaction derived from the PDB. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 261−270.
(7) Laederach, A.; Reilly, P. J. Specific Empirical Free Energy Function for Automated Docking of Carbohydrates to Proteins. *J. Comput. Chem*. **2003**, *24*, 1748−1757.
(8) Muryshev, A. E.; Tarasov, D. N.; Butygin, A. V.; Butygina, O. Y.; Aleksandrov, A. B.; Nikitin, S. M. A novel scoring function for molecular docking. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 597−605.
(9) Raha, K.; Merz, K. M. A Quantum Mechanics-Based Scoring Function: Study of Zinc Ion-Mediated Ligand Binding. *J. Am. Chem. Soc.* **2004**, *126*, 1020−1021.
(10) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(11) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750−1759.
(12) Ishchenko, A. V.; Shakhnovich, E. I. Small Molecule Growth 2001 (SMoG2001): An Improved Knowledge-Based Scoring Function for Protein−Ligand Interactions. *J. Med. Chem.* **2002**, *45*, 2770−2780.
(13) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, T. M.; Murray, C. W.; Taylor, R. D.; Watson, P. Virtual Screening Using Protein−Ligand Docking: Avoiding Artificial Enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793−806.
(14) Klon, A. E.; Glick, M.; Thoma, M.; Acklin, P.; Davies, J. W. Finding More Needles in the Haystack: A Simple and Efficient Method for Improving High-Throughput Docking Results. *J. Med. Chem.* **2004**, *47*, 2743−2749.
(15) Jacobsson, M.; Liden, P.; Stjernschantz, E.; Boström, H.; Norinder, U. Improving Structure-Based Virtual Screening by Multivariate Analysis of Scoring Data. *J. Med. Chem.* **2003**, *46*, 5781−5789.
(16) Krovat, E. M.; Langer, T. Impact of Scoring Functions on Enrichment in Docking-Based Virtual Screening: An Application Study on Renin Inhibitors. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1123−1129.
(17) Marsden, P. M.; Puvanendrampillai, D.; Mitchell, J. B. O.; Glen, R. C. Predicting protein−ligand binding affinities: a low scoring game? *Org. Biomol. Chem*. **2004**, *2*, 3267−3273.
(18) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing Scoring Functions for Protein−Ligand Interactions. *J. Med. Chem.* **2004**, *47*, 3032−3047.
(19) Perola, E.; Walters, W. P.; Charifson, P. S. A Detailed Comparison of Current Docking and Scoring Methods on Systems of Pharmaceutical Relevance. *Proteins* **2004**, *56*, 235−249.
(20) Wang, R.; Lu, Y.; Fang, X., Wang, S. An Extensive Test of 14 Scoring Functions Using the PDBbind Refined Set of 800 Protein−Ligand Complexes. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2114−2125.
(21) Wang, R.; Lu, Y.; Wang, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J. Med. Chem.* **2003**, *46*, 2287−2303.
(22) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring-function to predict protein−ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337−356.
(23) Gohlke, H.; Hendlich, M.; Klebe, G. Predicting binding modes, binding affinities and 'hot spots' for protein−ligand complexes using a knowledge-based scoring function. *Perspect. Drug Discovery Des.* **2000**, *20*, 115−144.

(24) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(25) Murray-Rust, P.; Glusker, J. P. Directional Hydrogen Bonding to sp2- and sp3-Hybridized Oxygen Atoms and Its Relevance to Ligand-Macromolecule Interactions. *J. Am. Chem. Soc.* **1984**, *106*, 1018−1025.
(26) Klebe, G. The use of composite crystal-field environments in molecular recognition and the *de-novo* design of protein ligands. *J. Mol. Biol.* **1994**, *237*, 215−235.
(27) Bruno, I. J.; Cole, J. C.; Lommerse, J. P. M.; Rowland, R. S.; Taylor, R.; Verdonk, M. L. IsoStar: a library of information about nonbonded interactions. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 525−537.
(28) Meyer, E. A.; Castellano, R. K.; Diederich, F. Interactions with Aromatic Rings in Chemical and Biological Recognition. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210−1250.
(29) Sippl, M. J. Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.* **1995**, *5*, 229−235.
(30) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr.* **2002**, *B58*, 380−388.
(31) Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. New Software for searching the Cambridge Structural Database and visualizing crystal structures. *Acta Crystallogr.* **2002**, *B58*, 389−397.
(32) DeLano, W. L. The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA 2002.
(33) *SYBYL*, 7.0; Tripos Inc.: 1699 South Hanley Rd., St. Louis, Missouri, 63144.
(34) The complete test results of ref 21 as well as the test set can be found at http://sw16.im.med.umich.edu/software/xtool/ (accessed July 2003).
(35) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998,** *19*, 1639−1662.
(36) Verdonk, M. L.; Cole, J. C.; Taylor, R. SuperStar: a knowledge-based approach for identifying interaction sites in proteins. *J. Mol. Biol.* **1999**, *289*, 1093−1108.
(37) Boer, D. R.; Kroon, J.; Cole, J. C.; Smith, B.; Verdonk, M. L. SuperStar: comparison of CSD and PDB-based interaction fields as a basis for the prediction of protein−ligand interactions. *J. Mol. Biol.* **2001**, *312*, 275−287.
(38) Taylor, R. Life-science applications of the Cambridge Structural Database. *Acta Crystallogr., Sect. D* **2002**, *D58*, 879−888.
(39) Nissink, J. W. M.; Taylor, R. Combined use of physicochemical data and small-molecule crystallographic contact propensities to predict interactions in protein binding sites. *Org. Biomol. Chem.* **2004**, *2*, 3238−3249.
(40) Tanford, C. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 2nd ed.; Wiley: New York, 1980.
(41) Olsen, J. A.; Banner, D. W.; Seiler, P.; Wagner, B.; Tschopp, T.; Obst-Sander, U.; Kansy, M.; Müller, K.; Diederich, F. Fluorine interactions at the thrombin active site: Protein backbone fragments H−C#-C═O comprise favourable C−F environment and interactions of C−F with electrophiles. *ChemBioChem* **2004**, *5*, 666−675.
(42) Hodgkin, E. E.; Richards, W. G. Molecular similarity based on electrostatic potential and electric field. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1987**, *14*, 105−110.
(43) Nissink, J. W. M.; Verdonk, M. L.; Klebe, G. Simple knowledge-based descriptors to predict protein−ligand interactions. Methodology and validation. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 787−803.
(44) Pérez, C., Ortiz, A. R. Evaluation of Docking Functions for Protein−Ligand Docking. *J. Med. Chem.* **2001**, *44*, 3768−3785.
(45) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipies in C: The art of scientific computing*, 2nd ed.; Cambridge University Press: New York, 1992; pp 639−642.

JM050436V