Novel, Customizable Scoring Functions, Parameterized Using N-PLS, for Structure-Based Drug Discovery

Cornel Catana*,† and Pieter F. W. Stouten‡

Computer Assisted Drug Discovery, Pfizer Global Research and Development, 2800 Plymouth Road, Ann Arbor, Michigan 48105, and Molecular Modelling & Design, Nerviano Medical Sciences, Viale Pasteur 10, 20014 Nerviano (MI), Italy

Received August 16, 2006

The ability to accurately predict biological affinity on the basis of in silico docking to a protein target remains a challenging goal in the CADD arena. Typically, "standard" scoring functions have been employed that use the calculated docking result and a set of empirical parameters to calculate a predicted binding affinity. To improve on this, we are exploring novel strategies for rapidly developing and tuning "customized" scoring functions tailored to a specific need. In the present work, three such customized scoring functions were developed using a set of 129 high-resolution protein−ligand crystal structures with measured Ki values. The functions were parametrized using N-PLS (*N*-way partial least squares), a multivariate technique well-known in the 3D quantitative structure−activity relationship field. A modest correlation between observed and calculated pKi values using a standard scoring function ($r^2 = 0.5$) could be improved to 0.8 when a customized scoring function was applied. To mimic a more realistic scenario, a second scoring function was developed, not based on crystal structures but exclusively on several binding poses generated with the Flo+ docking program. Finally, a validation study was conducted by generating a third scoring function with 99 randomly selected complexes from the 129 as a training set and predicting pKi values for a test set that comprised the remaining 30 complexes. Training and test set r^2 values were 0.77 and 0.78, respectively. These results indicate that, even without direct structural information, predictive customized scoring functions can be developed using N-PLS, and this approach holds significant potential as a general procedure for predicting binding affinity on the basis of in silico docking.

INTRODUCTION

With the increasing number of therapeutic targets, the need for a rapid search for small molecules that may bind to these targets is of crucial importance in the drug discovery process. One way of achieving this is the in silico or virtual screening of large compound collections to identify a subset of compounds that contain relatively many hits against the target, compared to a randomly selected subset. With the increasing availability of biologically relevant protein structures,[1] structure-based virtual screening will play an ever more important role in finding new leads. Finding such leads typically involves two steps: docking and scoring. Docking involves the computational generation of many orientations and conformations of a small molecule in the active site of a target protein structure. A variety of methods such as genetic algorithms, Monte Carlo simulations, simulated annealing, and distance geometry can be used to carry out these orientational and conformational searches. Scoring refers to the calculation of the interaction energy with the target for all binding modes ("poses") that are generated during this search process, with the ultimate goal of identifying the energetically most favorable pose. This docking and scoring process is repeated for all molecules of interest. Molecules can subsequently be rank-ordered by their scores, and top-scoring compounds can then be selected for purchase, synthesis, or biological evaluation. The result of a virtual screening campaign depends heavily on the quality of the docking program used. A critical assessment of the performance of several such programs was recently published.[2]

To assess the binding affinity of a receptor−ligand complex on the basis of in silico docking, one uses a so-called scoring function, which basically is a mathematical equation that approximates the thermodynamics of binding.[3−5] Scoring functions can serve a dual purpose. They can be used to differentiate between poses of a single ligand in the target binding site, and they can be used to estimate relative binding affinities of different receptor−ligand complexes. Owing to the crucial role of scoring, a large number of functions have been developed. These methods can be classified into three broad categories: force-field-based (e.g., AutoDock,[6] G-Score,[7] and D-Score[8]), knowledge-based (e.g., PMF[9] and DrugScore[10]), and empirical (e.g., LigScore,[11] PLP,[12] LUDI,[5] F-Score,[13] Chem-Score,[14,15] and X-Score[16]). In addition, consensus scoring[17] has emerged as a technique whereby multiple scoring functions are combined to reduce statistical noise and to improve hit rates yielded by virtual screening.

The first category is based on molecular mechanics force fields. The issue here is that classical force fields have not normally been developed to describe intermolecular interactions. Therefore, truncating the noncovalent part of a force

* Corresponding author phone: (734) 622-4479; fax: (734) 622-2782; e-mail: cornel.catana@pfizer.com.
† Pfizer Global Research and Development.
‡ Nerviano Medical Sciences.

10.1021/ci600357t CCC: $37.00
Published on Web 12/06/2006

field equation and then applying it to protein−ligand binding (as is done in D-Score) is not expected to give robust docking results, although this was a standard practice in the early years. Reparametrization of force-field-based scoring functions definitely improves their ability to predict crystallographically observed protein−ligand conformations, as evidenced by results from AutoDock and G-Score analyses. However, these functions still do not adequately account for hydrophobic and entropic effects. A detailed analysis of these types of scoring functions has recently been published.[18]

Knowledge-based scoring functions (also known as potentials of mean force)[19] are derived from statistical analyses of experimentally determined protein−ligand crystal structures. The underlying assumption is that specific interatomic contacts that occur more frequently than average (based on the distribution of interactions between specific types of protein and ligand atoms extracted from many crystal structures) are energetically favorable. Apart from donor−acceptor or lipophilic−lipophilic contact terms that exhibit the expected distance dependence, the other terms are difficult to interpret. DrugScore combines a potential of mean force with molecular surface-based terms and fits the resulting function to experimental data.[10] Such molecular surface-based terms have been introduced to more effectively capture the hydrophobic effect, a practice commonly employed in the empirical scoring functions described below.

Empirical scoring functions approximate the free energy of binding as a weighted sum of terms, where each term describes a different type of ligand/receptor interaction, such as lipophilic contacts, electrostatic interactions, hydrogen bonds, and so forth. They are calibrated with a set of protein−ligand complexes with experimentally determined structures and binding affinities via multivariate regression analysis. Empirical scoring functions have several appealing features. First, because they are usually calibrated with a diverse set of protein−ligand complexes, their application is not limited to a congeneric series of ligands or a particular protein class. Second, the individual terms have a clear physical meaning. Thus, the weight assigned to each term can contribute to an understanding of the receptor−ligand binding process. Third, an accuracy level (approximately 2 kcal/mol) in predicting binding affinity can be achieved that is acceptable for most structure-based virtual screening processes. However, despite their appeal and the encouraging results obtained to date, there remains ample opportunity for improvement in prediction accuracy and robustness. Empirical scoring functions are only as accurate as the training set used to build them. While they seem to work on ligands not present in the training set, one cannot expect results of docking experiments to extrapolate well beyond the classes of proteins used to derive the function in the first place.

In this paper, we describe the advantages of training a scoring function (i.e., adjusting the coefficients) on a specific set of protein−ligand complexes and correlating the predictions with observed binding affinity. The resulting "customized (affinity) scoring function" can be thought of as a 3D quantitative structure−activity relationship (QSAR) model with the descriptors being supplied by the docking program as protein−ligand interaction terms.

Prerequisites for developing such customized scoring functions are a docking tool that is capable of generating ligand conformations close to the crystal structure and access to the individual protein−ligand interaction terms (either directly from the docking program or indirectly by rescoring with a different program). We chose QXP_Flo+[20] because, in our hands on our projects, it was able to consistently reproduce observed protein−ligand crystal structure geometries, and because it allows the output of individual protein−ligand interaction terms. A recent study[2] also confirmed its capabilities to generate the correct ligand conformations observed in crystal structures.

Table 1. Protein−Ligand Complexes Used and Their Experimental pKi (mol/L) Values

complex	pKi	complex	pKi	complex	pKi
181L	3.76	1DWB	2.90	1TLP	7.55
182L	3.95	1DWC	7.41	1TMN	7.30
183L	3.71	1DWD	8.18	1TNG	2.93
184L	4.71	1ELA	6.36	1TNH	3.37
185L	3.54	1ELB	7.15	1TNI	1.70
186L	4.84	1ELC	6.66	1TNJ	1.96
187L	3.37	1ETR	7.41	1TNK	1.49
188L	3.33	1ETS	8.22	1TNL	1.88
1AAQ	8.40	1ETT	6.19	2CGR	7.27
1ABE-1	7.01	1FKF	9.70	2CTC	3.89
1ABE-2	7.01	1HBV	6.37	2GBP	7.60
1ABF-1	5.42	1HIH	8.53	2OLB	5.54
1ABF-2	5.42	1HOS	8.55	2TMN	5.88
1APB	5.82	1HPS	9.22	2UPJ	10.39
1APB	5.82	1HPS	9.22	3PTB	4.50
1APT	9.39	1HPV	9.22	3TMN	5.90
1B05	7.11	1HPX	8.22	4DFR	9.70
1B0H	6.69	1HSG	9.42	4HVP	6.11
1B1H	7.06	1HTE	7.00	4PHV	9.15
1B2H	4.54	1HTF	8.09	4PHV	9.15
1B32	7.1	1HTF	8.09	4TLN	3.72
1B3F	6.88	1HTG	9.68	4TMN	10.19
1B3G	6.71	1HTG	9.68	5ABP	6.63
1B3H	6.20	1HVI	10.07	5ABP	6.63
1B3L	5.89	1HVJ	10.45	5TLN	6.37
1B40	7.27	1HVK	10.11	5TMN	8.04
1B46	5.27	1HVL	9.00	6ABP-1	6.35
1B4H	5.45	1HVR	9.51	6ABP-2	6.35
1B4Z	5.22	1HVS	10.08	6CPA	11.52
1B51	7.36	1JET	7.20	6TMN	5.05
1B58	6.57	1JEU	6.81	7ABP-1	6.45
1B5H	6.01	1JEV	6.87	7ABP-2	6.45
1B5I	7.04	1L83	3.76	7CPA	13.96
1B5J	7.43	1MNC	9.00	7EST	7.60
1B6H	7.83	1NHB	4.17	7HVP	9.62
1B7H	8.08	1PPC	6.16	7TLN	2.47
1B9J	5.94	1PPH	6.22	8ABP−1	8.00
1BAP-1	6.85	1PRO	11.30	8ABP−2	8.00
1BAP-2	6.85	1QKA	5.92	8CPA	9.15
1BRA	1.82	1QKB	7.34	8HVP	9.00
1CBX	6.30	1RBP	6.72	9ABP-1	8.00
1CPS	6.66	1SBG	7.74	9ABP-2	8.00
1DIF	10.66	1SBP	6.92	9HVP	8.35

METHODS

Data Set. A total of 129 well-resolved (<2.5 Å) protein−ligand complexes (see Table 1) were obtained from the Protein Data Bank (PDB).[21] The set spans seven protein families with diverse active-site topographies (lysozyme, OppA, sugar-binding, HIV protease, trypsin-like, carboxypeptidase, and thermolysin). All ligands bind to the receptor noncovalently, have a molecular weight < 874, and have less than 37 rotatable bonds. There are no cofactors present. If the complex had two ligands, it was split in two separate protein−ligand complexes. An example is 1ABE, which consists of α-l-arabinose (1ABE-1) and β-l-arabinose (1ABE-

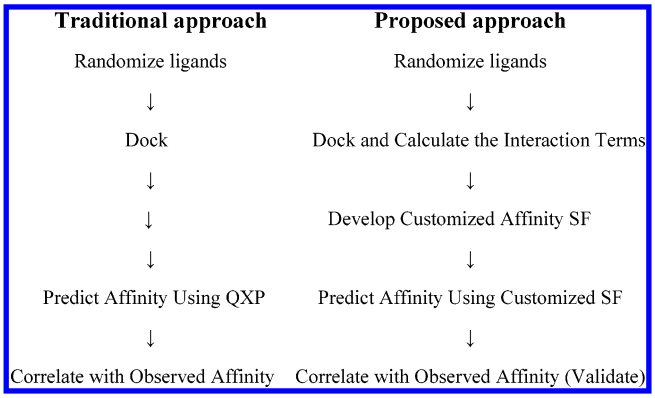

Figure 1. Traditional and proposed schemes for estimating binding affinities on the basis of in silico docking; SF stands for scoring function.

2). The associated binding affinities are the ones also used by Accelrys to develop their LigScore scoring function.[11,31] pKi (−log Ki) values range from 1.49 to 13.96. From the outset, it was our intention to work on a diverse set of complexes and not on a specific chemical series or specific protein target class, in order to assess the overall power of the proposed approach.

Binding-Site Preparation. Coordinates for each complex were obtained from the PDB.[21] Water molecules were removed, and each complex was subjected to the "pdb2mod" script in Flo+ for automatic conversion to a binding-site model. The binding-site models were checked carefully for ligand atom types and connectivity, ligand bond orders, ligand protonation, and protonation states of acidic and basic protein residues that line the binding site. Imidazole rings were rotated by 180° as needed to form the most appropriate hydrogen-bonding pattern. Flo+ uses only polar hydrogens on the ligand and protein, and they are optimized during the conversion process. Ligands were minimized inside the protein active site to optimize the structures for the Flo+ force field. We observed only minor differences between the crystallographically observed and the minimized versions of the ligands.

Customized Scoring Functions Based on the Crystallographic Ligand Poses. The Flo+ docking software has two components: a graphic interface called LZM (lazy mouse) and a command-line option known as QXP (quick explore). Flo+ has its own scoring function, developed to simultaneously predict binding modes as well as relative affinities within SAR series. As a benchmark, the compounds were docked with QXP_Flo+ (after randomizing their poses) and the affinity was calculated with the built-in Flo+ scoring function; that is, the "traditional approach" in Figure 1 was followed. The correlation between the predicted affinity using Flo+ and experimental pKi was $r^2 = 0.50$. r^2 is clearly very data-set-dependent as Flo+ achieves $r^2 = 0.70$ when applied to its own 200-complex training set.[32]

In parallel, the "proposed approach" in Figure 1 was pursued in two ways: using crystallographic poses (this section) and using poses generated by QXP (see below). For each of the 129 minimized complex crystal structures, the protein−ligand interaction terms calculated by Flo+ were output using the "bscore12" option. There are 15 such terms calculated by the program: five for contact energy, three for hydrogen bonds, two for binding-site entropy, three for polar desolvation, one for ligand strain, and one for steric clashes. They can be grouped into six main categories. From an energetic point of view, there are two favorable categories: contact (distance-dependent ligand-binding-site contacts) and hydrogen bond (distance-dependent hydrogen-bonded interactions). The four unfavorable categories are entropy (protein binding-site entropy), polar desolvation (contact energy between polar and nonpolar atoms; this is not incorporated in the contact category), ligand internal energy (ligand strain energy, entropy, and internal contacts), and bump energy (a weakly repulsive term accounting for steric clashes between the ligand and protein). These 15 interaction terms make up the **X** matrix, and the experimental pKi values represent the **Y** vector. The **X** matrix was autoscaled (variance one) and the **Y** vector centered, and partial least squares (PLS)[22] was applied using the leave-one-out cross-validation (XV) technique. The number of optimum components (latent variables) found using the root-mean-square (RMS) error of cross-validation was 7. The following statistics were obtained:

$$r^2 = 0.80,\ \text{XV}\ r^2 = 0.76,\ \text{and RMS} = 1.00\ \text{pKi units}$$

where

$$\text{RMS} = \sqrt{\sum(\text{pKi, pred} - \text{pKi, obs})^2/(n-1)}$$

and n = the number of complexes.

Thus, a new "customized" scoring function that was trained on a specific set of complexes was obtained. We did not validate this function by applying it to an independent test set because the main purpose of this study is to evaluate the performance of our approach when no crystallographic information on ligand poses is used, and also because the gap between r^2 and XV r^2 was fairly small. The results from this first study showed that it is possible to develop a customized scoring function with improved affinity prediction capabilities ($r^2 = 0.80$ versus 0.50) using known crystallographic ligand poses.

Customized Scoring Functions That Do Not Employ Crystallographic Ligand Poses. In real life situations, often few (or no) protein−ligand crystal structures exist for the protein being studied. The possibility of developing customized scoring functions in such cases was assessed next. For this study, all ligands were removed from their respective protein binding sites, randomized (transformed into 2D structures), and reinserted into the proteins using the "complex" script in Flo+. Each starting complex was subjected to 700 steps of MCDock+ (Monte Carlo-based docking program within Flo+) to generate 25 ligand binding poses. MCDock+ combines Autodock[6] (Monte Carlo Metropolis-type grid box search) and MacroModel[23] (Monte Carlo with energy minimization and a candidate pool of "best solutions to date") strategies and gradually reduces the size of the Monte Carlo steps.

To mimic a scenario where a crystal structure is unavailable, the most diverse set of four MCDock+ poses were chosen from the 25 for subsequent statistical analyses. The philosophy behind this procedure is that this way we should generate a pose close to the crystallographically observed one plus three very different "decoy" poses. The diverse set of four ligand poses was obtained by applying the CombiChem I Library Analysis module from Cerius2[11] (distance-

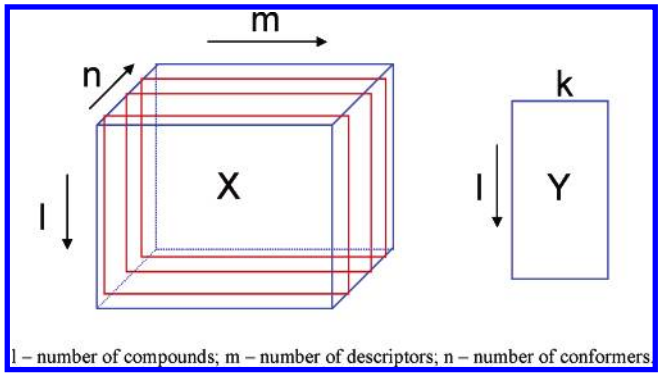

Figure 2. Schematic representation of N-PLS (a 3D data matrix is shown).

based, max−min-spantree, Euclidian, 100 000 Monte Carlo steps, terminate after 1000 steps). As descriptors, we used rmslig (RMS distance of each pose to the first one), in addition to the 15 interaction terms calculated by Flo+. Thus, for each ligand, we retained four poses, and for each pose, we calculated 15 interaction terms. The resulting three-dimensional matrix (129 complexes × 15 descriptors × 4 poses) was used as input for the N-PLS[22] regression procedure. The **Y** vector with experimental pKi values was left unchanged.

To generalize the PLS approach to multiway data, a general multilinear PLS (N-PLS)[24,25] regression model was developed. Multiway data are matrices with more than two dimensions. They are used quite frequently in *N*-dimensional QSAR studies. For example, a set of compounds with measured pKi values, a given number of conformers, and a set of descriptors calculated for each conformer will generate a 3D data matrix (see Figure 2).

Another example of the application of N-PLS would be the analysis of the interactions between a set of compounds and three different probes at defined grid points. Such an approach would generate a matrix with five dimensions: the compounds, three grid directions, and the probes.

If in the first example we substitute the conformer dimension with a pose dimension, we will exactly be in the context of docking studies where, for each compound, we could have multiple poses without knowing a priori which is the one that occurs in real life and that correlates best with the experimentally determined binding affinity. Applying N-PLS in such studies leads to the identification of the pose that correlates best with the observed binding affinity, which may well be a pose that is also close to the crystallographically observed pose.

The final step in this study was to validate the true predictive ability of N-PLS by randomly splitting the set of 129 complexes into 99 and 30 complex training and test sets, respectively. An N-PLS model developed on the training set was used to predict pKi values of the test set. The procedure outlined here is the "proposed approach" described in Figure 1.

RESULTS AND DISCUSSION

The statistics obtained for our customized scoring functions and described in the literature for several other scoring functions are detailed in Table 2.

When the r^2 and q^2 (cross-validated r^2) values are considered, only Head's model[27] performs better than our crystallographic model (PLS with crystal structures in Table 2), but his is based on a smaller set of 51 complexes. The statistics show that our proposed approach [see Table 2, N-PLS with diverse poses (training set)], which only uses docking poses generated with QXP and no crystallographic poses, generates comparable predictive statistics. Additional support for the validity of our approach by prediction of an external test set was recently provided by Giordanetto et al.[33]

The small gap between $r^2 = 0.69$ and $q^2 = 0.76$ in Table 2 for "N-PLS with diverse poses" indicates that this is not a chance correlation. To further test the possibility of chance correlation, all pKi values were scrambled (i.e., randomly assigned to the protein−ligand complexes) 50 times and models were developed for each scrambled set. The resulting r^2 values ranged from 0.18 to 0.29, providing clear proof that the correlation reported in Table 2 is real.

A retrospective analysis of the diverse four poses selected for analysis in the second study revealed that at least one pose was very close to the ligand conformation from the crystal structure. The above statistics show that one can develop adequate generic scoring functions in the absence of complex structures, but the real power of the approach lies in the opportunity to specifically develop scoring functions for a single target or target family, without the need for a large number of crystal structures to calibrate the functions.

How well this approach works if the training set contains many complexes for which (nearly) no correct pose is identified remains to be established. The present work suggests that, in order to develop a successful "customized scoring function", a docking tool capable of reproducing the crystal structure is mandatory. Additionally, a ligand pose close to the observed crystal structure must be part of the 3D matrix that is used as input for N-PLS. In real life situations, one usually has at least one crystal structure to compare with. In these cases, we recommend calculating the "rmslig" to the ligand from the crystal structure and including in the 3D data matrix the pose with the smallest value.

In our experience, optimum correlation between observed and calculated pKi values resulted when between four and six poses per ligand were retained. Reducing the number of poses increases the probability that the correct pose will be removed by chance, and increasing the number retained may result in the introduction of significant noise. The number of poses retained for N-PLS should therefore be guided by the quality of the customized scoring function statistic (r^2, q^2, and RMS).

If one has confidence that the highest-scoring pose from the docking program will reproduce the crystal structure (this usually works for simple proteins without metal ions involved in the interaction with the ligand), it is not necessary to use N-PLS. Applying data analysis methods like support vector machines[34] or random forests[35] on interaction terms of the highest-scoring pose may result in superior models (better correlation in the test phase, because these methods have a higher power to generalize) compared to those using N-PLS (data not shown).

On the other hand, we can look at the crystal structure as a snapshot of the interactions between the ligand and binding site. In solution, matters are more complicated, multiple conformations can coexist and N-PLS may be a good

Table 2. Comparison of Various Scoring Functions, Including Customized Ones

scoring function	no. ligands	r^2	q^2(LOO)[a]	RMS	#LV[b]
PLS with crystal structures	129	0.80	0.76	1.00	7
N-PLS with diverse poses	129	0.76	0.69	1.24	5
N-PLS with diverse poses (training set)	99	0.77	0.70	1.22	5
N-PLS with diverse poses (test set)	30	0.78		1.15	-
Böhm - Ludi[26]	82	0.79	N/A	1.28	
Head - Validate[27]	51	0.85	0.78	1.02	
Eldridge[28]	82	0.71	0.66	1.40	
Gschwend[29]	103	0.75	0.70	1.26	
Wang[30]	170	0.78	0.74	1.16	
Accelrys - Ligscore2[31]	118	0.75	0.73	1.04	
QXP[32]	200	0.70	0.65	1.17	

[a] q^2(LOO) = cross validated r^2 using the "leave-one-out" technique. [b] The number of components (latent variables).

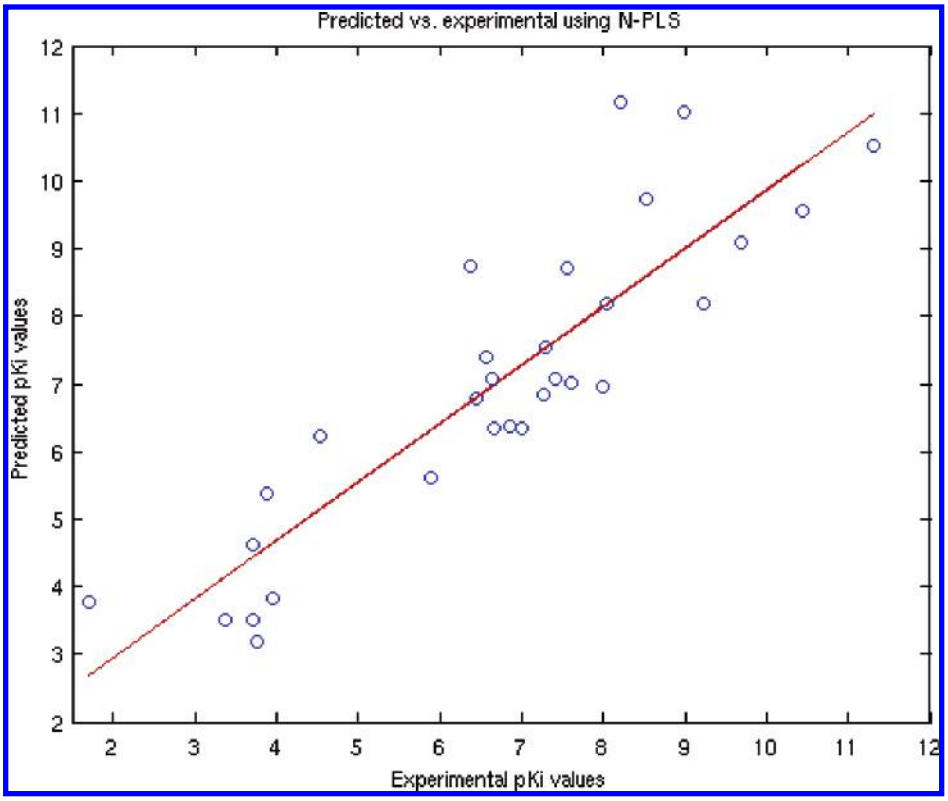

Figure 3. Correlation between experimental and predicted pKi for the test set.

substitute for Boltzmann-weighted poses. The small difference in r^2 values we observed when we used the information from the crystal structure and when we did not seems to support this argument, at least for this set of complexes.

In addition, the N-PLS models provide insight into the most important interaction terms for the targets considered. It should also increase the probability of correctly identifying the ligand pose most similar to the crystal structure and could identify alternative poses that may contribute significantly to the observed binding affinity.

Table 2 shows that the customized scoring function developed on 80% of the initial set of complexes generated statistics similar than those when the entire set was analyzed. Cross validation was done for both models using the leave-one-out technique. We identified as the most important descriptors hydrogen bonds, contact energy, and polar desolvation; pose numbers 1 and 3 make strong contributions to the model, or in other words, these conformations are closest to the ligand conformation from the crystal structures (a retrospective analysis in fact proved this fact). If you do not have any crystal structure, these poses help you in hypothesizing a binding mode close to the real one. The model applied to the test set (see Figure 3) provided the following statistics:

$$r^2(\text{experimental, predicted}) = 0.78, \text{RMS} = 1.15 \text{ pKi, and } n = 30$$

The test set statistics are a bit better than those of the training one (usually, the statistics for the training set are superior to those of the test set), and a possible explanation could be the fact that the training set covers uniformly the chemical space of interactions terms.

CONCLUSIONS

The present study demonstrates that the dependency of empirical scoring functions on the training set used in their development can be overcome by developing customized

scoring functions. The correlation coefficient increased from $r^2 = 0.5$, obtained when the Flo+ affinity scoring function was applied to 129 complexes, to $r^2 = 0.8$, when a customized scoring function that took into consideration the pose of the ligand in the crystal structure was used. This correlation decreased slightly to $r^2 = 0.78$ when the customized scoring function used ligand poses generated by the QXP docking program rather than crystallographically observed poses. Thus, the potential exists to use docked structures by themselves to develop a scoring function capable of predicting observed affinities.

To apply this approach, one must have a robust docking program capable of generating the crystallographic binding mode, a way to calculate and output the individual protein−ligand interaction terms, a well-defined binding site for each protein, and experimental pKi values for complexes that will be part of the training set. At this moment, no general recipe exists for choosing which and how many docking poses should be used to generate parameters for the 3D data matrix used as input to the N-PLS analysis. Thus, this process should be guided by the statistics of the customized scoring function.

Customized scoring functions should be able to generate more precise affinity predictions for a set of ligands that are assumed to bind in a mutually comparable way to a single protein (or class of proteins) than generic scoring functions, particularly if the protein of interest was not part of the training set on which the original empirical scoring function was developed. The approach described here can be applied in numerous practical situations. For example, one could develop a precise customized scoring function on a congeneric series bound to a certain protein, or to different proteins, or on a diverse set of ligands bound to a certain protein, and so forth.

On the basis of the present work, there appears to be significant promise in developing a "next generation" empirical scoring function that incorporates the ability to adjust its interaction terms to achieve the highest correlation for a given training set.

ACKNOWLEDGMENT

The authors thank Dan Ortwine for his constructive remarks and comments.

REFERENCES AND NOTES

(1) National Institute of General Medical Sciences - Protein Structure Initiative (NIGMS-PSI). http://www.nigms.nih.gov/Initiatives/PSI (accessed Sep 29, 2006).

(2) Warren, G. L.; Andrews, W. C.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, *49*, 5912−5931.

(3) Murcko, M. A; Mark, A. Computational Methods to Predict Binding Free Energy in Ligand−Receptor Complexes, *J. Med. Chem*. **1995**, *38*, 4953−4967.

(4) Tame, J. R. H. Scoring Functions: A View from the Bench. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 99−108.

(5) Böhm, H. J. 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.* **1998**, *12*, 309−323.

(6) (a) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem*. **1998**, *19*, 1639−1662. (b) *Autodock*, version 2.04. http://www.scripps.edu/mb/olson/doc/autodock (accessed Sep 29, 2006).

(7) Jones G.; Willet, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol*. **1997**, *267*, 727−748.

(8) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. Dock 4.0: Search Strategies for Automated Molecular Docking of Flexible Molecule Databases. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411−428.

(9) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Interaction Approach. *J. Med. Chem*. **1999**, *42*, 791−804.

(10) Gohlke, H; Hendlich, M; Klebe, G. Knowledge-Based Scoring Function to Predict Protein−Ligand Interactions. *J. Mol. Biol.* **2000**, *295*, 337−356.

(11) *Cerius2*, version 4.6; Accelrys: San Diego, CA. http://www.accelrys.com/products/cerius2/cerius2products/listing.html (accessed Sep 29, 2006).

(12) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular Recognition of the Inhibitor AG-1343 by HIV-1 Protease: Conformationally Flexible Docking by Evolutionary Programming. *Chem. Biol*. **1995**, *2*, 317−324.

(13) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol*. **1996**, *261*, 470−489.

(14) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical Scoring Functions: The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.

(15) Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Flexible Docking Using Tabu Search and an Empirical Estimate of Binding Affinity. *Proteins* **1998**, *33*, 367−382.

(16) Wang, R.; Lai, L.; Wang, S. Further Development and Validation of Empirical Scoring Functions for Structure-Based Binding Affinity Prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11−26.

(17) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins *J. Med. Chem.* **1999**, *42*, 5100−5109.

(18) Wang, R.; Lu, Yipin; Wang, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J. Med. Chem*. **2003**, *46*, 2287−2303.

(19) Gohlke, H.; Klebe, G. Statistical Potentials and Scoring Functions Applied to Protein−Ligand Binding. *Curr. Opin. Struct. Biol*. **2001**, *11*, 231−235.

(20) (a) McMartin, C.; Bohacek, R. S. QXP: Powerful, Rapid Computer Algorithms for Structure-Based Design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 333−344. (b) *QXP_Flo+*, version August 2002; Thistlesoft: Colebrook, CT.

(21) Research Collaboratory for Structural Bioinformatics − Protein Data Bank. http://www.rcsb.org/pdb/ (accessed Sep 29, 2006).

(22) *PLS-Toolbox for Use with MATLAB*, version 3.0; Eigenvector Research: Manson, WA, 2003.

(23) *MacroModel*; Schrödinger: Portland, OR. http://www.schrodinger-.com/ProductDescription.php?mID=6&sID=8 (accessed Sep 29, 2006).

(24) Bro, R. Multiway Calibration - Multilinear PLS. *J. Chemom.* **1996**, *10*, 47−61.

(25) Bro, R.; Smilde, A. K.; de Jong, S. On the Difference between Low Rank and Subspace Approximation: Improved Model for Multi-Linear PLS Regression. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 3−13.

(26) Böhm, H. J. 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.* **1998**, *12*, 309−323.

(27) Head, R. D.; Smythe, M. L; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R. VALIDATE: A New Method for the Receptor-Based Prediction of Binding Affinities of Novel Ligands. *J. Am. Chem. Soc.* **1996**, *118*, 3959−3969.

(28) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.

(29) Smith, R.; Hubbard, R. E.; Gschwend, D. A.; Leach, A. R.; Good, A. C. Analysis and Optimization of Structure-Based Virtual Screening Protocols. (3). New Methods and Old Problems in Scoring Function Design. *J. Mol. Graphics Modell.* **2003**, *22*, 41−53.

(30) Wang, R.; Gao, Y.; Lai L. SCORE: A New Empirical Method for Estimating the Binding Affinity of a Protein−Ligand Complex. *J. Mol. Model.* **1998**, *4*, 379−394.
(31) Krammer, A.; Kirchhoff, P. D.; Jiang, X.; Venkatachalam, C. M.; Waldman, M. LigScore: A Novel Scoring Function for Predicting Binding Affinities. *J. Mol. Graphics Modell.* **2005**, *23*, 395−407.
(32) McMartin, C. QXP_Flo+ 200 Complex Training Set. Personal communication.
(33) Giordanetto, F.; Cotesta, S.; Catana, C.; Trosset, J. Y.; 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.
(34) Burges, C. J. C. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowledge Discovery* **1998**, *2*, 121−167.
(35) Breiman, L. Random Forest. *Machine Learning* **2001**, *45*, 5−32.

CI600357T