ELSEVIER

Journal of Molecular Graphics and Modelling 20 (2002) 281–295

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Consensus scoring for ligand/protein interactions

Robert D. Clark [a,*], Alexander Strizhev [a], Joseph M. Leonard [a], James F. Blake [b], James B. Matthew [b]

[a] *Tripos Inc., 1699 S. Hanley Road, St. Louis, MO 63144, USA*
[b] *Pfizer Central Research Laboratories, Eastern Point Road, Groton, CT 06340, USA*

Received 2 January 2001; received in revised form 6 April 2001; accepted 6 April 2001

Abstract

Several different functions have been put forward for evaluating the energetics of ligand binding to proteins. Those employed in the DOCK, GOLD and FlexX docking programs have been especially widely used, particularly in connection with virtual high-throughput screening (vHTS) projects. Until recently, such evaluation functions were usually considered only in conjunction with the docking programs that relied on them. In such studies, the evaluation function in question actually fills two distinct roles: it serves as the objective function being optimized (fitness function), but is also the scoring function used to compare the candidate docking configurations generated by the program. We have used descriptions available in the open literature to create free-standing scoring functions based on those used in DOCK and GOLD, and have implemented the more recently formulated PMF [J. Med. Chem. 42 (1999) 791] scoring function as well. The performance of these functions was examined individually for each of several data sets for which both crystal structures and affinities are available, as was the performance of the FlexX scoring function. Various ways of combining individual scores into a consensus score (CScore) were also considered. The individual and consensus scores were also used to try to pick out configurations most similar to those found in crystal structures from among a set of candidate configurations produced by FlexX docking runs. We find that the reliability and interpretability of results can be improved by combining results from all four functions into a CScore.

Keywords: Docking; Scoring function; Fitness function; Ligand/protein complexes; Ligand binding; Ligand/protein interaction; Consensus scoring; Genetic algorithms

1. Introduction

Computer programs dedicated to docking small molecules into protein binding pockets in silico are currently a focus of attention for many research groups. The commercially distributed programs DOCK,[1] GOLD[2] and FlexX[3] are of particularly widespread interest at present. The approach used is somewhat different for each program, as is the degree of protein and ligand flexibility accommodated. Each program has the same two goals, however: to rapidly determine the most likely binding mode of each ligand, and to return some measure of its estimated binding affinity for the target protein. Moreover, each goal needs to be met as rapidly as possible.

The need for speed is especially critical when one wishes to screen large databases to identify important potential ligands, since each extra second of CPU time required per molecule translates into an extra day of screening time for a 100,000 compound database. Such virtual high-throughput screening (vHTS) must necessarily involve some computational trade-offs and compromises.

Vagaries of active site solvation and the uncertainties of local dielectrics make it difficult to calculate enthalpies of interaction both precisely and quickly enough for database screening. Accurate prediction of affinities, moreover, entails estimation of free energies of interaction, which necessitates inclusion of problematic entropic terms for both protein and ligand. A number of general reviews of the difficult problems involved have been compiled [2–9].

As a result of the approximations involved in estimating free energies of interaction, the evaluation function used is generally empirically parameterized even when the form of that function is based on energetic first principles. Each program's scoring function has its own form, relies

* Corresponding author. Tel.: +1-314-647-1099; fax: +1-314-647-9241.
E-mail address: bclark@tripos.com (R.D. Clark).

[1] DOCK is available from the Department of Pharmaceutical Chemistry at the University of California, San Francisco; www.cmpharm.ucsf.edu/kuntz.

[2] GOLD is distributed by the Cambridge Crystallographic Data Centre, Cambridge, UK; www.ccdc.ac.uk/prods/gold.html.

[3] FlexX was developed at the German National Research Center for Information Technology (GMD), and is distributed by Tripos Inc., St. Louis MO; www.tripos.com/software/flexx.html.

1093-3263/02/$ – see front matter
PII: S1093-3263(01)00125-5

on a distinct atom typing scheme, uses different atomic partial charge calculation methods, and has been trained on different ligand/protein data sets. Moreover, the scoring functions used in DOCK, GOLD and FlexX were developed to guide distinct optimization routines as well as to score the candidate binding configurations obtained as end results.

Given these differences, it is not surprising that each program returns a somewhat different estimate of relative binding affinity. One way of reacting to such differences is to search for a single "correct" function to use in both docking and scoring. Another approach is to use a different function to score docked configurations (poses) from that used as an objective function during docking [10–12]. The alternative pursued in the work documented here is to look for a productive way to combine estimates from a variety of scoring functions into a single consensus score (CScore). This approach complements work by others [13] and has provided the basis for a widely disseminated commercial software product, CScore^{TM},[4] now being used for vHTS [14]. Here we document the original work [15] underpinning this methodology, illustrating its use for successfully ranking binding configurations of ligands in crystal structures and for re-ranking poses obtained from FlexX docking runs. Ranking with respect to literature binding energies are taken as the performance criterion in the former analyses; heavy atom root mean square deviations (RMSD) are used in the latter.

2. Methodology

2.1. Structure preparation

Crystal structures obtained from the Brookhaven Protein Databank were converted from .pdb to .mol2 format using standard functionality from the Biopolymer module of SYBYL 6.6.[5] This included automatic typing of each ligand atom based on the geometry found in the crystal structure and addition of hydrogens as necessary. All crystallographic waters were removed. Atom typing of protein residues and addition of hydrogens to them was carried out by reference to standard Biopolymer libraries keyed to residue type. Terminal rotors such as methyl and hydroxyl groups were then relaxed to avoid distortion of scores by spurious steric clashes with the added hydrogen atoms. Charges were assigned to both protein and ligand atoms as described by Gasteiger and Marsili [16,17]. Binding sites were defined for all scoring functions as including all atoms in protein residues where at least one atom was within 6.5 Å of an atom in the ligand as found in the parent crystal structure. FlexX docking runs utilized the default parameters provided in SYBYL 6.6.

[4] CScore^{TM} is available from Tripos Inc., St. Louis, MO 63144, USA.
[5] SYBYL® is available from Tripos Inc., St. Louis, MO 63144, USA.

2.2. Individual scoring functions

DOCK uses an evaluation function composed of both steric and electrostatic terms [18] based on the AMBER force-field [19]. The GOLD evaluation function is a sum of hydrogen bonding stabilization energy calculated from donor/acceptor pair atom types and geometries; internal van der Waals energy for the ligand conformer in question; and the strength of steric interactions between ligand and protein [20,21]. Both scoring functions were implemented in C within the SYBYL environment using the published descriptions. For the former, pairwise interaction energies between all ligand and binding site atoms were calculated directly, rather than from a rectilinear grid approximation; for the latter, scores were re-scaled so as to be roughly commensurate with the nominal free energies produced by most other scoring functions. For both, contributions from unfavorable steric interactions were capped at 0.5 kcal/mol for each pair of atoms considered.

Because of these differences, scores obtained from our implementations are generally not identical to those obtained from the functions used in the corresponding docking programs. The output from our programs will henceforth be referred to as D-SCORE and G-SCORE, respectively, to emphasize this point. Our implementations gave results in reasonable qualitative agreement with published values for docking results obtained for literature complexes. More direct validation of our versions of these scoring functions was not possible, because published applications include minimizations within the respective docking programs, which are not available to us (it is the lack of such relaxation which necessitates the cap on steric repulsion energies mentioned above). Regardless, particular deficiencies specific to our implementations clearly should not be construed as reflecting negatively on the originals. Our intent here is to show how different kinds of scoring function behave alone and when used in combination with one another, not to assess the relative merits of the particular scoring functions considered.

The FlexX scoring function [22] was used as implemented in the commercially available software (see footnote 3). It is derived directly from that introduced by Böhm [23] and is considerably more complex than are either the DOCK or GOLD evaluation functions. It considers the number of rotatable bonds in the ligand, hydrogen bonds (including atom types and geometry of interaction), ion pairing, aromatic interactions, and the lipophilic contact energy.

The PMF scoring function developed by Muegge and Martin [1] is not itself part of any docking program, though it has been used to guide DOCK [24]. It is much simpler in form, being simply a summation over all pairwise interaction terms. The magnitude and sign of each interaction potential is based on the atom types of the interacting pair and the intervening distance. Potential curves for each atom type pairing are derived from a survey of crystal structures drawn from the Brookhaven Protein Data Bank. Minor ambiguities in the published methodology were clarified by direct

communication with the authors, particularly as regards atom typing definitions (I. Muegge, private communication). [6]

ChemScore [25] has shown considerable promise as well, and has been incorporated into recent commercial releases of CScore, but is not included in the experiments described here.

2.3. Consensus scoring functions

The simplest way to combine different scoring functions is to simply add up the individual scores (or equivalently to average them). Although the values generated by various scoring functions are all nominally energies of some kind, they are not in fact commensurate and a more robust approach is called for. Therefore, a rank transform was applied to each set of scores, such that the configurations predicted to be most stable were given the lowest rank. The ranks obtained from each of the scoring functions considered (here, four) were then added to give a rank-sum. The advantage of a sum over an average is that the contribution from the rank for each individual score can more easily be split out for illustrative purposes in the former instance.

A potentially more robust, non-linear combination score is obtained by dropping the largest (worst) rank(s) for each complex or candidate configuration, and sorting the remainder to identify the worst-best rank. If a given candidate configuration has ranks 3, 5, 6 and 12 for four different scoring functions, for example, the worst-best rank is 6 (assuming lower ranks correspond to lower, and therefore, more favorable, energies). Note that, if five or six scoring functions are being use, it may be appropriate to drop the two worst ranks from consideration. The deprecated rank-sum, which is the sum of those ranks not dropped (here the sum of the three best ranks — 14 in this example), may also be a useful criterion. Like the rank-sum, it is easily split up graphically so as to show the contributions of the worst-best, second best and best ranks.

A third approach relies on relatively coarse quantiles or positions in range for values generated by each scoring function rather than ranks. We have found that positions in range are generally somewhat better behaved. In this case, each scoring function casts one vote in favor of a particular crystal structure or candidate configuration if its score falls into the best half (or best third) of the range of values obtained for that scoring function across the data set of interest. The CScore is then the total number of votes received. The intersection of the top quantiles across all scoring functions under consideration, which has been employed by others [13], is a special case of this criterion.

Calculating CScore values using a single criterion (best half or best third) may produce a problematically large number of tied scores, which complicates rank correlation analysis unnecessarily. Such ties under one criterion are best broken by reference to the other. In the present work, CScore ranks were obtained by applying the more stringent best-third criterion first, and breaking ties based on the more relaxed best-half criterion.

[6] The definitions used here are fully specified in the PMF. Ligand and PMF. Protein files found in the sybylbase/tables/ascii directory distributed with SYBYL.

2.4. Statistics

Results from different scoring functions can be compared in terms of direct correlations between scores and binding energies or RMSD. For vHTS applications, however, it is the monotonicity of the relationship between scores and binding energy or RMSD which is important, not linearity. Hence, we have chosen to use the rank correlations of scores and their various combinations with the relevant experimental values. Spearman's rank correlation coefficient (r_S) is well suited to this purpose. In the absence of ties, r_S is given by

$$r_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{1}$$

where d_i is the rank difference for the ith observation under two different criteria [26], e.g. the difference between the rank with respect to binding energy and that obtained from a scoring function. For cases where there are ties, the statistic was calculated as [26]:

$$r_S = \frac{T_x^2 + T_y^2 - \sum d_i^2}{2\sqrt{T_x^2 \times T_y^2}} \tag{2}$$

where

$$T_q^2 = \left(\frac{1}{12}\right)\left(n(n^2-1) - \sum t_{jq}(t_{jq}^2 - 1)\right) \tag{3}$$

and t_{jq} is the number of ties in classification q (i.e. x or y) at rank j, and the summation is across all ranks. The variables x and y are the experimental values and calculated scores, respectively.

When alternative ligand configurations are evaluated, many are often quite similar to each other and to that found in the crystal structure. In such cases, it is important that some reasonably accurate solution be among the top ranked, but less so that all good solutions be identified. Using this "first acceptable configuration" statistic puts a high value on giving good scores to good configurations and entails a high cost for giving good scores to bad configurations (false positives). It is much more forgiving of false negatives than is rank correlation, however. For the FlexX docking runs examined here, we have attempted to quantitate this criterion by noting how far into the ranked list of configurations one must go to include a configuration within some critical RMSD — 2 or 3 Å, depending on the system — of the crystal structure used. When ties are present, the lowest acceptable rank cited is the expected value assuming a random distribution of ranks across the tied scores.

Table 1
Binding free energy scoring for HIV-1 protease complexes

Ligand	PDB	Stability (kJ/mol)[a]	FlexX	Scores[b]		
				D-SCORE	G-SCORE	PMF
A-78791	1hvj	−60	−59	−53	−58	−59
A-76928	1hvk	−58	−60	−61	−63	−66
A-77003	1hvi	−57	−64	−55	−58	−62
JG-365	7hvp	−55	−46	−60	−57	−64
XK263	1hvr	−54	−60	−51	−52	−27
VX-478	1hpv	−53	−40	−36	−37	−33
A-76889	1hvl	−51	−53	−56	−59	−63
AAFΨGVVOMe[c]	1aaq	−48	−41	−44	−35	−49
GR126045	1htf	−46	−44	−40	−41	−34
Ac-pepstatin	5hvp	−44	−48	−47	−44	−51
SB203238	1hbv	−36	−43	−41	−35	−34
MVT-101	4hvp	−35	−39	−54	−57	−56

[a] Free energy of binding taken from Eldridge et al. [25].
[b] Each score has been normalized to the experimental data to give a common mean and range.
[c] Ψ indicates replacement of the peptide bond by its hydroxyethylene isostere.

3. Results

3.1. Estimating relative affinities from crystal structures

FlexX, G-SCORE, D-SCORE and PMF were each applied to a series of HIV protease and thermolysin complexes for which crystal structures and free energies of binding were both available [25], with the results shown in Tables 1 and 2. No particular pattern is evident across these two target proteins in the degree of correlation between experimentally determined values and the various scores.

Plots of the rank-sums obtained across the various HIV protease and thermolysin complexes are shown in Fig. 1A and B, respectively. The complexes are shown in order of decreasing stability, with the lowest ranks assigned to complexes with the ligands of highest affinity. Fig. 1C and D show the analogous results obtained for complexes of two sets of proprietary ligands with squalene synthase [27] (‘a’ series) and FKBP-12 (‘b’ series), respectively. These latter data sets are of particular interest because the affinities and complexes were all obtained in the same laboratory and so are more directly comparable. Ligand affinities for squalene synthase (Fig. 1C) fall in the 1 μM range, which is particularly relevant for identifying leads in virtual screening programs. The FKBP-12 data set includes results from replicate X-ray crystal structure determinations for two complexes — b01a–c share a common ligand, as do b02a and b02b.

Parallel results using the worst-best criterion and deprecated rank-sums are shown in Fig. 2, where the worst (highest) rank for each complex has been dropped and ranks from the remaining three scores have been stacked to show the deprecated rank-sum as well. The contribution from the highest rank of those remaining (i.e. the worst-best) is shown as the bottom section of each bar, with the contribution of second best atop it and the best rank specifying the top section. Note that, the top of the middle section of each bar in these plots is the rank-sum for the two middle scores — i.e. twice the median. The behavior of the median falls between

Table 2
Binding free energy scoring for thermolysin complexes.

Ligand	PDB	Stability (kJ/mol)[a]	FlexX	Scores[b]		
				D-SCORE	G-SCORE	PMF
ZFP(O)LA	4tmn	−58	−35	−60	−55	−59
Thorphan	5tmn	−46	−27	−49	−43	−45
Phosphoramidon	1tlp	−43	−50	−52	−45	−45
Peptidomimetic	1tmn	−42	−64	−46	−43	−44
Nitroanilide	5tln	−36	−33	−11	−38	−38
VW	3tmn	−34	−34	−28	−25	−25
P-Leu-NH_2	2tmn	−34	−35	−28	−27	−27
ZGP(O)LL	6tmn	−29	−30	−47	−42	−42
Leu-NHOH	4tln	−21	−34	−21	−19	−1

[a] Free energy of binding taken from Eldridge et al. [25].
[b] Each score has been normalized to the experimental data to give a common mean and range.

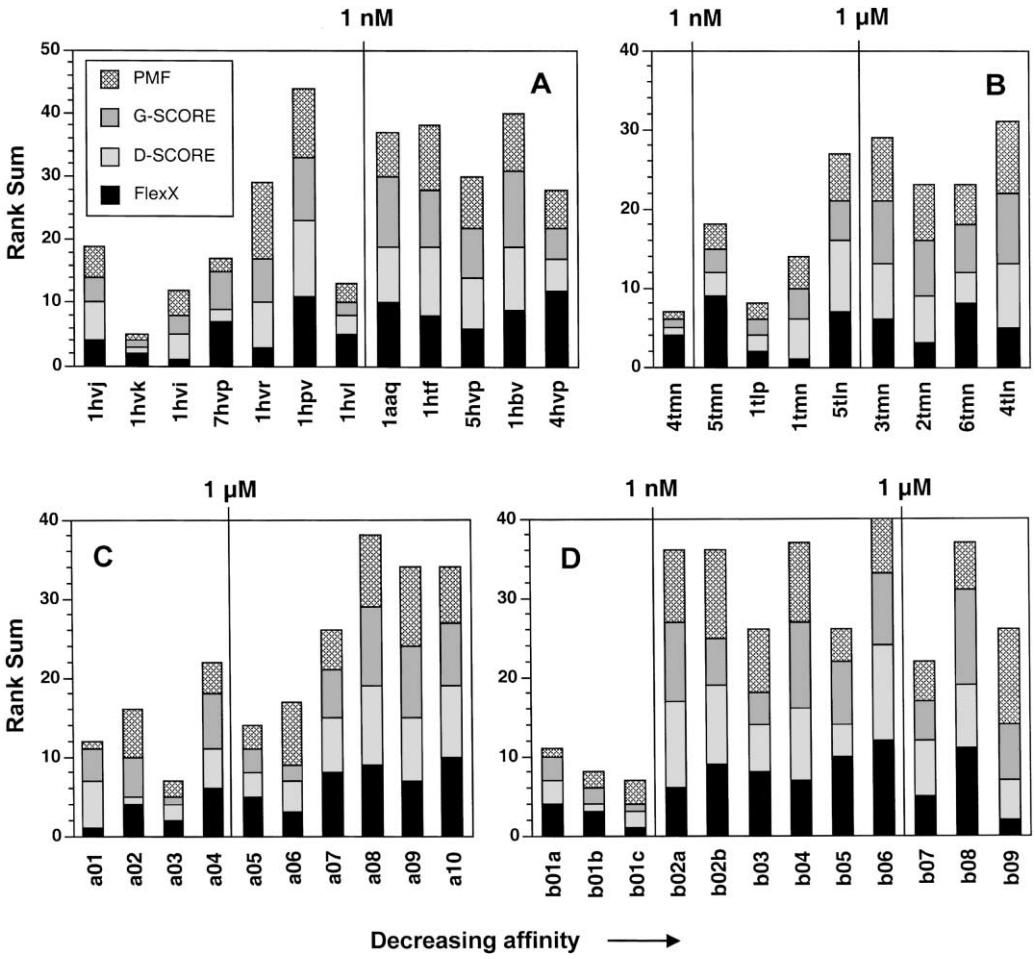

Fig. 1. Scoring HIV-1 protease complexes by simple rank-sums (1 = best score). Sections of each bar represent the separate contributions from the respective scoring function to each sum. Complexes have been sorted in order of decreasing stability. (A) HIV-1 protease complexes; (B) thermolysin complexes; (C) squalene synthase complexes; (D) FKBP-12 complexes.

that of the worst-best criterion and the deprecated rank-sum, and so is not considered separately here.

Applying a cutoff worst-best rank threshold of 5 to the data for HIV protease successfully picks out four of the seven highest affinity complexes among the 12 examined, but misses JG-365 (7hvp; NCI [28]), Xk263 (1hvr; DuPont Merck [29]), and VX-478 (1hpv; Vertex Pharmaceuticals [30]) (Fig. 2A; see Fig. 3 for ligand structures). Raising the cutoff to 7 captures JG-365 and Xk263 but does so at the cost of picking up MVT-101 (4hvp; NCI [31]) as a false positive.

The worst-best criterion also fares well for the thermolysin data, where there is a smooth monotonic increase with decreasing stability and a cutoff worst-best rank of 4 identifies the four best compounds of the nine examined (Fig. 2B). Note, however, that the affinity range is considerably broader in this case than for HIV protease, and that CbzGP(O)LL (6tmn [32], with an affinity of 9 μM [33]), which would constitute a false positive in this context, falls only two ranks above the cutoff.

For the proprietary data sets, the worst-best criterion picks up five of the six best squalene synthase complexes at a threshold of 5, missing only a04 (Fig. 2C), and cleanly picks out the three best FKBP-12 complexes (Fig. 2D), which are all the same ligand in this case. Increasing the threshold to 7 picks out all of the seven best squalene synthase complexes, whereas the only incremental effect in the FKBP-12 series is to pick out b07 and b09, which are (nominally) false positives.

It is worth noting that the worst-best criterion is much more reproducible for individual complexes (b01a–c and b02a–b) than is any single scoring function.

Results for CScore are illustrated for the HIV protease, thermolysin and proprietary complex data sets in Fig. 4. Complexes whose scores fall in the top third (or top half) of the range for any scoring function earn one “yes” vote for that criterion. The best possible CScore (here, 4) is then obtained when all scoring functions considered agree that the complex is a “good” one [13]. Note that, here, in contrast to the rank scoring used above, larger values are more favorable.

The results obtained with CScore are broadly similar to those obtained by applying the more relaxed worst-best

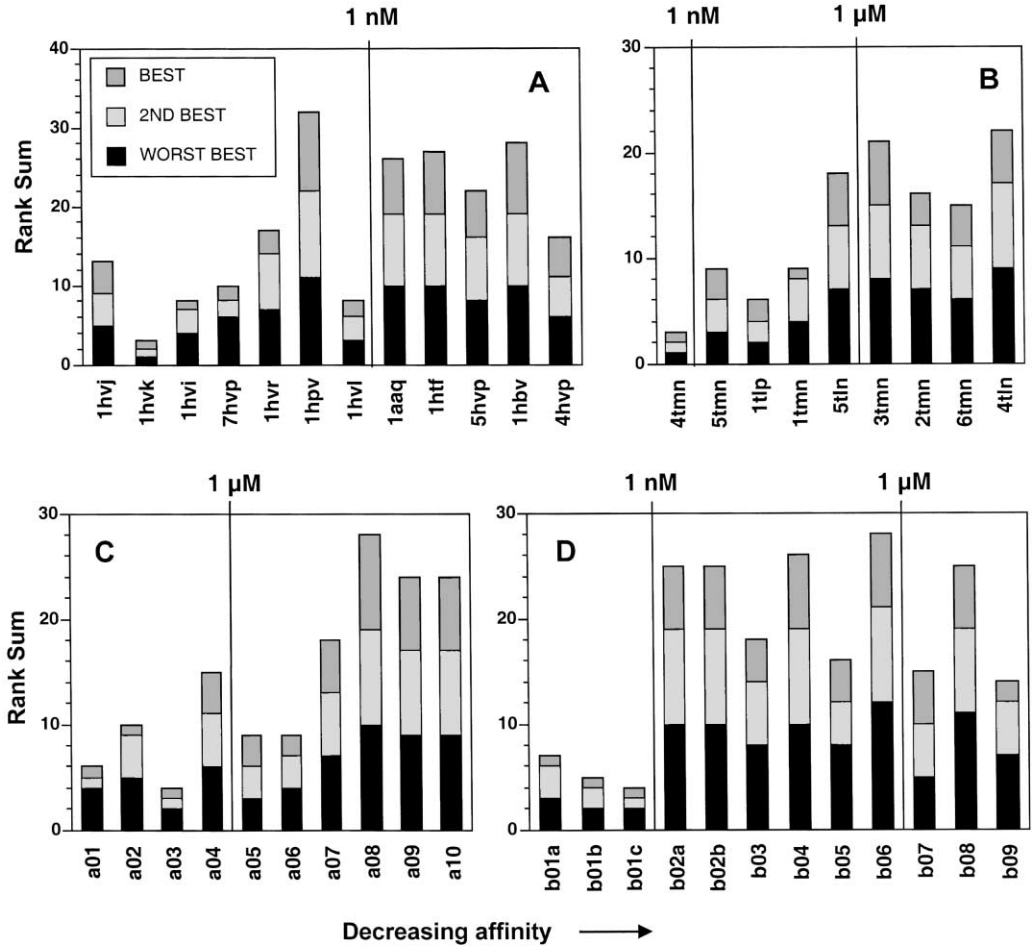

Fig. 2. Scoring by the worst-best criteria. The worst (highest) rank for each complex was discarded, and the remaining ranks are plotted with the worst assigned to the bottom section of each bar. The top of the middle section of each bar is the sum of the middle ranks and corresponds to twice the median rank. The full bar represents the deprecated rank-sum. A complex satisfies this criterion if the worst-best (bottom) rank falls at or below the selected cutoff. Complexes have been sorted in order of decreasing stability. (A) HIV-1 protease complexes; (B) thermolysin complexes; (C) squalene synthase complexes; (D) FKBP-12 complexes.

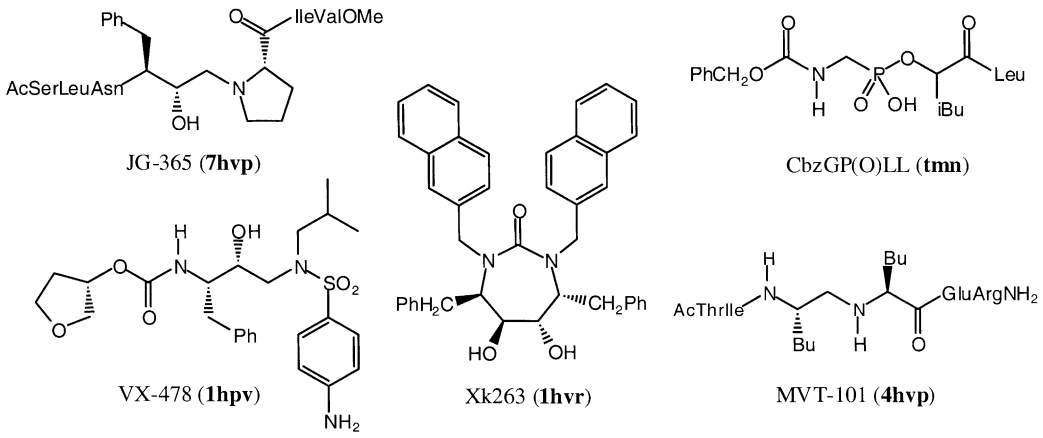

Fig. 3. Structures for ligands discussed in the text in connection with the data shown in Figs. 1, 2 and 4. PDB access codes of the corresponding crystal structures are indicated in parentheses.

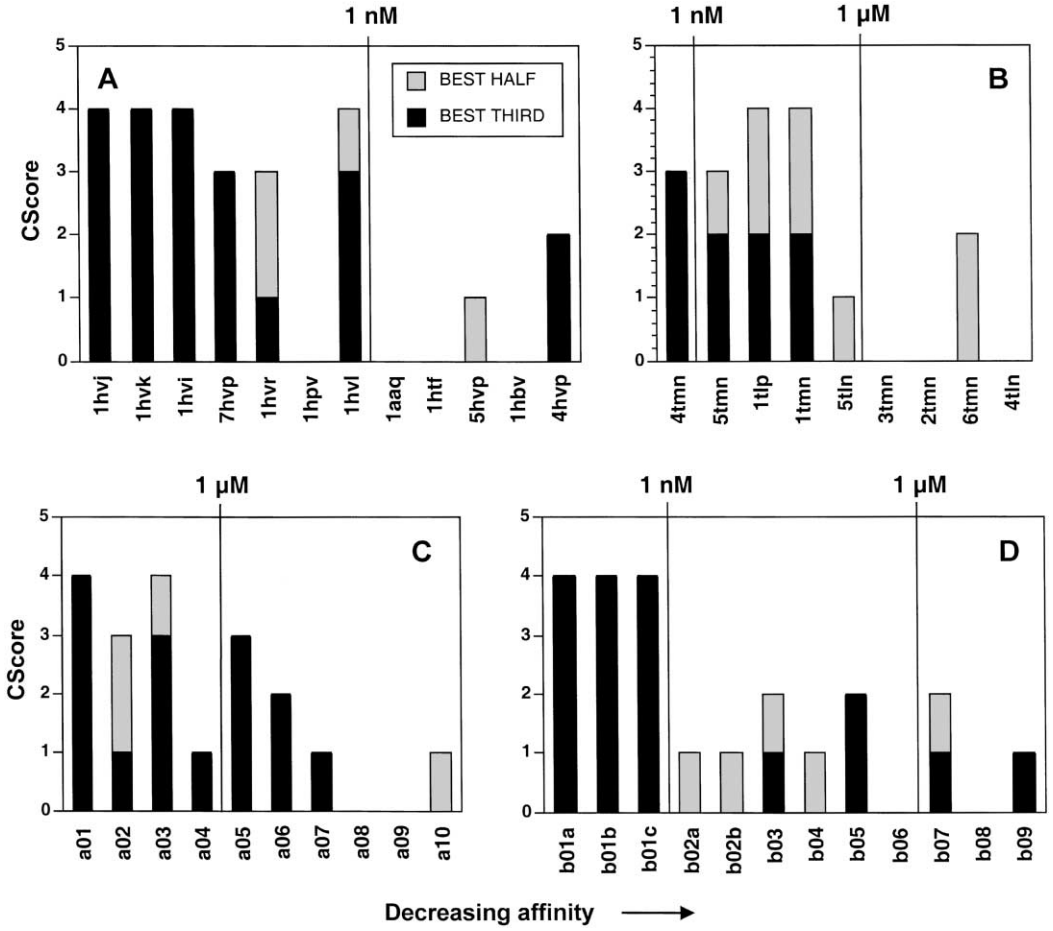

Fig. 4. Applying CScore. All complexes were scored, and each score falling in the best third or half of the range for a scoring function earned one "vote" from that scoring function. The CScore shown is then the sum of the votes obtained for each complex. The lower section of each bar is the CScore obtained using the best third of each scoring range; the upper section indicates the number of additional votes obtained by relaxing the threshold for each individual function to the best half of each range. Complexes have been sorted in order of decreasing affinity. (A) HIV-1 protease complexes; (B) thermolysin complexes; (C) squalene synthase complexes; (D) FKBP-12 complexes.

criterion described above in that the same six of the seven most active ligands are identified as worthy of follow-up: VX-478 (1hpv) is still missed. Note, however, that the much less tight-binding MVT-101 (4hvp) falls short of a three-vote CScore cutoff, even when the relaxed best-half criterion is used for each score. Moreover, the discrimination between "good" and "bad" ligands is clearer for CScore than for the worst-best criterion (compare Figs. 2 and 4).

For the 'a' and 'b' series of complexes, CScore is clearly the best way to combine scores, in that the highest affinity ligands (a01 and b01) are clearly identified. Furthermore, CScore gives a clear, albeit imperfect, ordering of expected affinities for the seven best ligands in the 'a' series: a01 > a03 > a05 > a02, a06 > a04, a07.

Table 3 shows the rank correlations obtained for the individual scoring functions and their various combinations, with the highest correlations for each target protein system highlighted in boldface type and the lowest correlations italicized. In each case, all four consensus scores performed better than two or three of the four individual scoring functions, and each individual scoring function was least effective for at least one series of complexes. Among the consensus scoring functions, CScore performed best, followed closely by rank-sums. Indeed, for two of the four targets, CScore outperformed all other scoring methods examined here.

3.2. Scoring candidate ligand configurations

FlexX was used to dock five different ligands (Fig. 5) into their respective protein targets: 2-MQPA or NAPAP into thrombin (1etr [34] and 1dwd [35]; Figs. 6 and 7, respectively), L-3-phenyllactic acid into carboxypeptidase A (2ctc [36]; Fig. 8), 1-deoxynojirimycin into glucoamylase (1dog [37]; Fig. 9), and DANA into neuraminidase (1nsd [38]; Fig. 10). These targets do not represent a rigorously random sample, but are intended to span the range of challenges which might be encountered in a virtual screening program. In each case, the 30 configurations having the most favorable FlexX scores were selected for further examination and sorted in order of increasing heavy atom root mean square

Table 3
Spearman's rank correlation (r_S) between complex stabilities and scores[a]

Method	Protein			
	HIV protease	Thermolysin	Squalene synthase	FKBP-12
FlexX	0.720	*0.133*	0.855	*0.399*
D-SCORE	0.490	0.683	0.745	*0.399*
G-SCORE	0.573	**0.917**	*0.636*	**0.628**
PMF	*0.448*	0.867	0.758	0.480
Rank-sum	0.594	0.812	0.851	0.535
Deprecated RS[b]	0.581	0.778	0.841	0.412
Worst-best	0.546	0.787	0.762	0.534
CScore	**0.746**	0.805	**0.860**	0.597

[a] The least effective scoring method(s) for each set of complexes is (are) italicized. The most effective scoring method for each set of complexes is highlighted in boldface.
[b] Deprecated sum, i.e. rank-sum obtained after dropping the worst rank.

deviation (RMSD) from the configuration found in the parent complex itself. Note that, the x-axis in these figures is ordinal in RMSD, not linear. Candidate configurations are often quite similar to one another, which tends to produce clumps of solutions with clear RMSD "gaps" between them. The vertical separators in Figs. 6–10 have been placed at such gap points, with the range of RMSD values found within each block indicated below the bottom panel of each figure.

The results obtained by applying simple ranked-sums, worst-best criterion and CScore are shown for each complex. All of the consensus methods (rank-sums; worst-best and deprecated rank-sums; and CScore, panels A–C, respectively) clearly identify binding configuration 2 as optimal for MQPA (Fig. 6), which falls well within the resolution of the crystal structure (2.2 Å).

The rank-sums discrimination is poorer for NAPAP (Fig. 7A), whereas the worst-best criterion picks out one solution close to the crystal structure (4) as well as a pair (13 and 14) of very similar configurations which are quite different from that found in 1dwd. CScore picks out 1, 2, 6 and 7 (all of which are within the crystal structure resolution of 3.0 Å) as best, with 13 and 14 showing up in the second tier.

The complex of L-phenyl lactate with carboxypeptidase A is a higher resolution structure (1.4 Å), and several FlexX results come within less than 1 Å RMSD of the 2ctc configuration (Fig. 8), but these are not the best scoring, either by individual functions or by consensus. Rather, pose 12 scores optimally by rank-sums and by the worst-best criterion (Fig. 8A and B). CScore does put configurations 8, 9 and 11, which are somewhat closer to the crystal structure, on a par with 12, but all four of these configurations are very similar to each other, and all are within 2 Å of the configuration found in 2ctc. The worst-best criterion and CScore both identify L-phenyl lactate configurations 24 and 25 as good solutions; this group of solutions does not come out of the rank-sums analysis because the configurations involved are given relatively poor scores by FlexX (Fig. 8A). Note that, in this case no FlexX solution earned a CScore of 4 (unanimous agreement) at any stringency.

Note that, the nominal loss in discrimination in going from the worst-best criterion to CScore is compensated for, at least in part, by the clear indication that the five configurations which diverge from the crystal structure by more

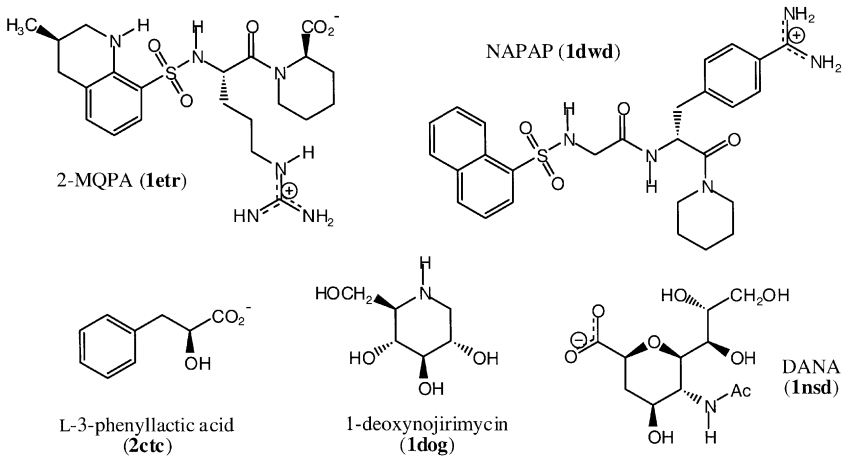

Fig. 5. Structures of ligands used in the analyses shown in Figs. 6–10. PDB access codes of the corresponding crystal structures are indicated in parentheses.

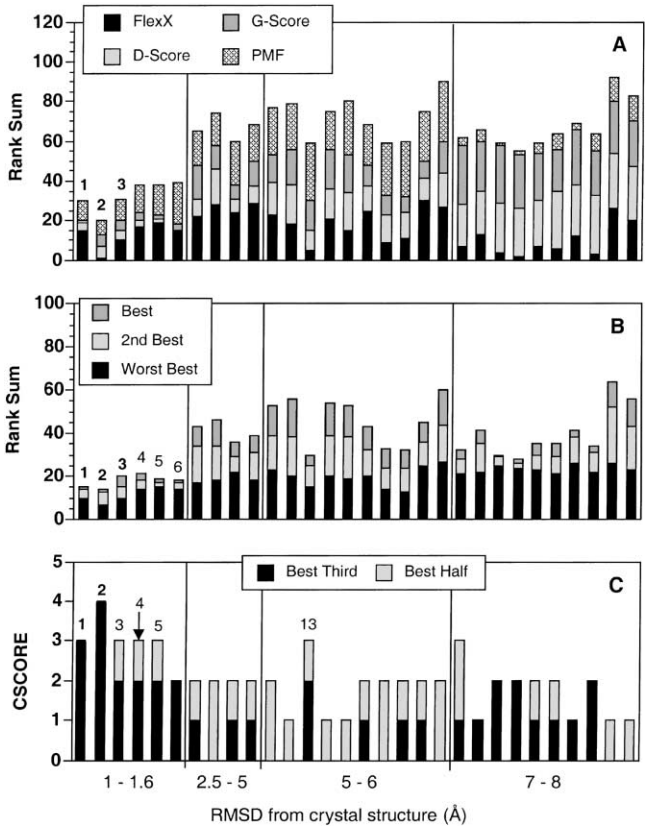

Fig. 6. Scoring across binding configurations for MQPA docked into thrombin using FlexX. The 30 configurations obtained which scored best in terms of the FlexX scoring function were also evaluated using G-SCORE, D-SCORE and PMF functions and sorted from best (pose 1) to worst (pose 30) in terms of RMSD with respect to the ligand pose found in 1etr. (A) Simple rank-sums, with each segment representing the rank for the corresponding scoring function; (B) worst-best and deprecated rank-sum criteria, with the worst (highest) rank dropped and the remaining scores sorted from worst (bottom segment in each bar) to best (top). Poses for which the worst-best rank is 10 or less are labelled; (C) CScore across FlexX output configurations. The bottom segment in each bar indicates the value obtained when only the best third of each score's range earns a "vote", whereas the top segment indicates the increase in CScore obtained when the criterion is relaxed to include the top half of each range.

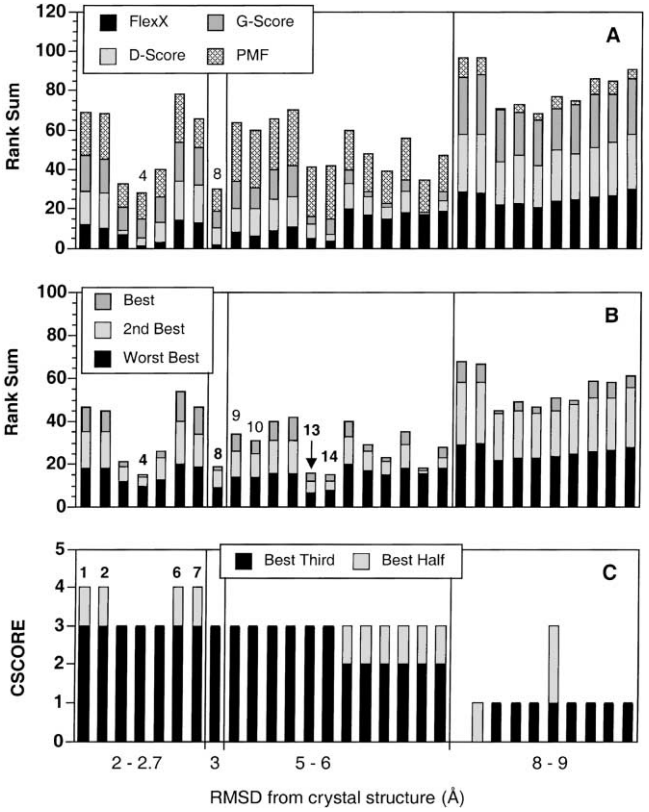

Fig. 7. Scoring across binding configurations for NAPAP docked into thrombin using FlexX. Details are the same as for Fig. 6, except that the protein and reference ligand structures were drawn from 1dwd.

than 3.0 Å (26–30) are all "bad" dockings. And those five are very bad indeed — their RMS deviations range from 6.1 to 6.9 Å!

At first glance, FlexX appears to have performed poorly with regard to docking 1-deoxynojirimycin into glucoamylase, in that the solution nearest to the crystal structure has an RMSD of 2.7 Å with respect to that found in 1dog (Fig. 9). Furthermore, none of the three consensus scoring methods shows good discrimination among the 21 configurations with heavy-atom RMSDs of 5 Å or below. This result suggests that the enzyme has a fairly broad binding domain; in fact, 1dog includes two ligand molecules in the binding site, which presumably correspond to the terminal and penultimate residues of the natural substrate [37]. The configurations clustered around 9, which CScore picks out as the optimal configuration, straddle the two adjacent binding sites found in the crystal structure.

Analysis of the DANA complex with neuraminidase identified very similar configurations close to the crystal structure as optimal — 4 for rank-sums (Fig. 10A); 1, 3, 4 and 5 for the worst-best criterion (Fig. 10B); and 4, 5 and 7 for CScore (Fig. 10C). Of these, only 7 falls beyond the resolution of the crystal structure (1.8 Å). It is of more concern that CScore also highlights configuration 28 as a good one (as does the worst-best criterion, albeit less definitively), despite the fact that it falls at a considerable distance from the binding site found in 1nsd.

Rank correlations obtained for the various individual and consensus scoring methods as applied to docking runs are shown in Table 4. Again, no individual scoring function was able to give consistently accurate rankings for all five systems. Indeed, results were not even qualitatively consistent between 1etr and 1dwd, despite the fact that both involve the same target protein — thrombin. Results for the four consensus functions considered, on the other hand, are qualitatively similar and comparable in quality across all five complexes. Here the worst-best criterion is most effective, with CScore edging out rank-sums for second place because of its excellent performance on 1dwd.

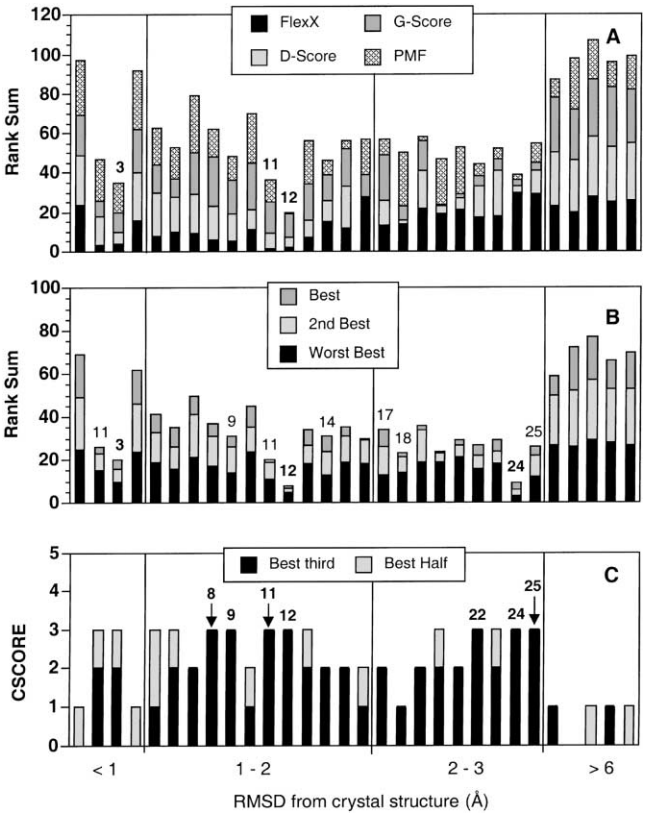

Fig. 8. Scoring across binding configurations for L-3-phenyllactic acid docked into carboxypeptidase A using FlexX. Details are the same as for Fig. 6, except that the protein and reference ligand structures were drawn from 2ctc.

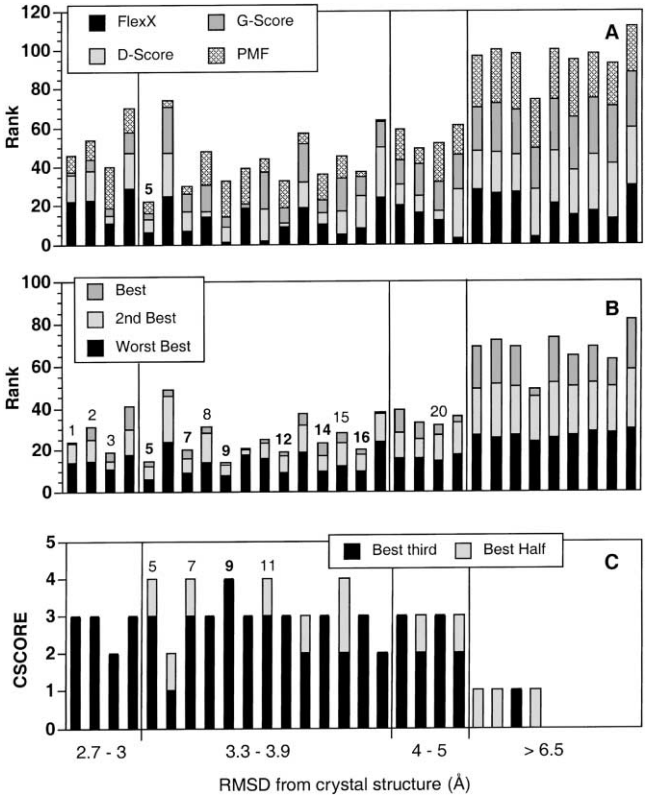

Fig. 9. Scoring across binding configurations for 1-deoxynojirimycin docked into glucoamylase using FlexX. Details are the same as for Fig. 6, except that the protein and reference ligand structures were drawn from 1dog.

Despite the sometimes low rank correlations between heavy atom RMSD and score, most of the scoring methods place at least one configuration similar to that found in the crystal at or near the top of the respective scores list — i.e. most of the lowest acceptable ranks cited parenthetically in Table 4 have a value of 1. By this standard, the deprecated rank-sum and CScore performed best among the consensus scores considered, all of which proved themselves more completely reliable than the individual scoring functions.

Again, it bears noting that these data cannot fairly be used to compare the performance of individual scoring functions to one another. Indeed, the results shown in Table 4 are inherently biased in favor of FlexX, since the configurations considered have already been relaxed with respect to its scoring function but not to the other individual scoring functions.

Table 4
Spearman's rank correlation (r_S) for FlexX docking analyses with respect to heavy atom RMSD[a]

Method	Complex ID				
	1etr[b]	1dwd[c]	2ctc[b]	1dog[c]	1nsd[b]
FlexX	0.181(1)	0.842(1)	**0.705**(1)	*0.134*(11)	*0.416*(2)
D-SCORE	**0.900**(1)	0.543(1)	0.199(4)	0.637(4)	**0.716**(1)
G-SCORE	0.816(1)	0.429(9)	0.095(7)	**0.810**(1)	0.591(1)
PMF	*−0.289*(7)	*−0.636*(9)	*−0.297*(1)	0.624(8)	0.676(1)
Rank-sum	0.492(1)	0.547(1)	0.216(1)	0.694(8)	0.655(1)
Deprecated RS[d]	0.441(1)	0.515(2)	0.126(1)	0.728(2.5)	0.694(1)
Worst-best	0.767(1)	0.696(3)	0.288(2)	0.723(6.0)	0.704(1)
Cscore	0.481(1)	**0.910**(1)	0.222(1.9)	0.716(9.6)	0.527(1)

[a] The most effective scoring method for each complex is highlighted in boldface type. The least effective scoring method for each complex is italicized.
[b] The expected value for the lowest rank with RMSD < 2 Å (i.e. the first acceptable configuration) is shown in parentheses.
[c] The expected value of the lowest rank with RMSD < 3 Å is shown in parentheses.
[d] Deprecated rank-sum, i.e. rank-sum obtained after dropping the worst rank.

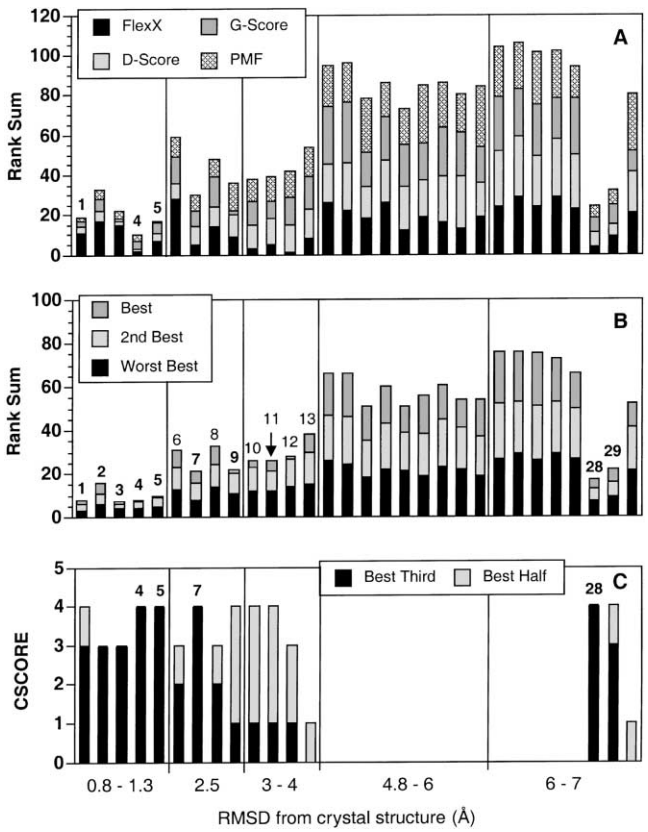

Fig. 10. Scoring across binding configurations for DANA docked into neuramidinase using FlexX. Details are the same as for Fig. 6, except that the protein and reference ligand structures were drawn from 1nsd.

4. Discussion

Two different kinds of validation are important in evaluating scoring functions for use in virtual screening programs. On one hand, it is important to distinguish more stable docking configurations from less stable ones (Figs. 1, 2 and 4). On the other hand, a scoring function which works well for ligands already positioned in an energetically optimal orientation is not necessarily well-suited for differentiating well-docked ligands from poorly docked ones (Figs. 6–10). Being able to unambiguously identify ligands for which no energetically favorable docking configuration exists involves satisfying both of these needs, but is not as readily amenable to direct detailed validation.

If different scoring functions all estimate a property independently and that property relates smoothly to the free energy of binding (or RMSD) for all configurations, then averaging values from several scoring functions should usually be a better predictor of stability than any individual score is. Examination shows that the various scores have quite different scales, but it is nonetheless reasonable to expect that averaging or summing the ranks produced would generally improve performance.

Applying rank-sums does, however, entail an implicit assumption that the individual scoring functions contribute equally for each configuration being scored, which may not always be the case. Each scoring function brings something distinctive to the ensemble, but each is also liable to miss something in some cases. It follows that any one score may sometimes be inaccurate, but it is unlikely that all four scoring functions will be in error at the same time. This rationale has been used by others when combining data from diverse compound selection [39] and QSAR [40] methods, and in applying data fusion techniques [41]. This leads naturally to the idea that a useful way to combine the scores might be to simply disregard the worst rankings for a configuration, de-emphasize the best, and focus in on the intermediate values. The worst-best, deprecated rank-sums and CScore criteria represent different ways to accomplish this.

4.1. False positives

In several cases, configurations that deviate substantially from the crystal structure are identified as “good” by consensus. In some cases, such false positives may occur because uncertainties in the crystal structure with respect to protein side chains distort scores in favor of a secondary or transitional binding mode that is in fact slightly higher in energy. An examination of the configuration found in 2ctc reveals several steric clashes between ligand and protein atoms, for example (Fig. 11A). These may reflect incidental uncertainty in the crystal structure (the reported resolution of 2ctc is 1.4 Å) or cumulative uncertainties in the position of individual atoms (temperature factors for proximal heavy atoms range from 11 to 24 $Å^2$, corresponding to isotropic vibrational amplitudes [42] between 0.37 and 0.55 Å). Configuration 12 relieves these bad contacts without changing hydrogen bonding or ionic interactions while maintaining a snug fit into the binding pocket (Fig. 11B): the ligand carboxylate still interacts with the distal guanidinium NH_2 of Arg127 and its α-hydroxy group hydrogen bonds to the carbonyl of Glu270.

The identification of 24 as a “good” configuration is, at first glance, more disturbing. Its high RMSD (2.65 Å) comes from a substantial rotation of the ligand within the active site with respect to the crystal structure (Fig. 11C). This produces a scrambling of hydrogen bonds but no net loss of interactions: the carboxylate interacts with the proximal amino group of Arg127 and is positioned directly over the catalytic zinc atom complexed to His196, whereas the α-hydroxy group participates in hydrogen bonding interactions with Tyr248 and with Arg127. Good steric complementarity is maintained with few bad contacts. It can be argued, then, that configuration 24 in Fig. 8 is a reasonable hypothesis as an alternative mode for phenyl lactate binding in the carboxypeptidase A active site. It is likely higher in energy than the pose found in the crystal structure itself, yet it may well inspire equally good ideas for follow-up chemistry.

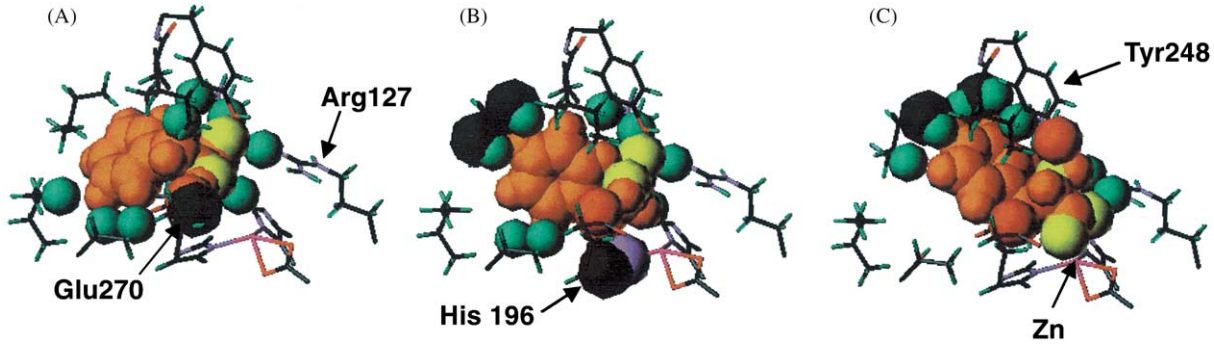

Fig. 11. Close contacts for L-phenyllactate dockings into carboxypeptidase A. Protein atoms which fall within 2.5 Å of any ligand atom are shown as space-fill by atom type (black for carbon, cyan for hydrogen, blue for nitrogen, and red for oxygen). The bound Zn is shown in magenta. Carbon and hydrogen atoms in the ligands are highlighted in orange, whereas ligand oxygens are shown in yellow. (A) The crystal structure 2ctc; (B) configuration 12 from Fig. 8; (C) configuration 24 from Fig. 8.

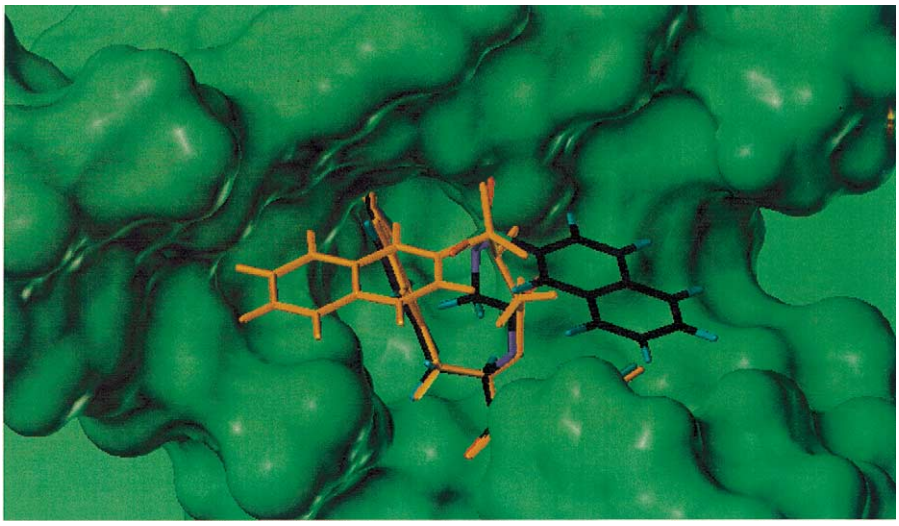

Fig. 12. Alternative docking mode for NAPAP in thrombin identified by FlexX and CScore. The alternative configuration is colored by atom type. The naphthyl ring as found in the crystal structure (1dwd) is colored orange; the rest of the inhibitor structure, which is positioned essentially identically to that for the FlexX solution, is omitted for clarity.

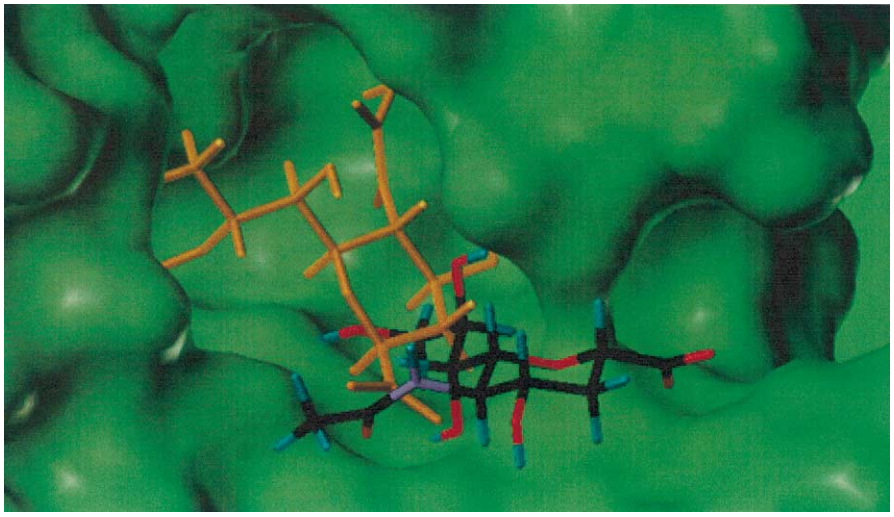

Fig. 13. Alternative docking mode for DANA in neuraminidase identified by FlexX and CScore. The alternative configuration is colored by atom type, whereas the binding mode found in the crystal structure.

Such potentially useful ambiguity is not uncommon, as is illustrated in Fig. 12 for the binding of NAPAP to thrombin (configuration 13 from Fig. 7). Here the naphthalene ring from the crystal structure (1dwd) is highlighted in red. The alternative binding mode which is also found by FlexX and picked out by CScore is colored by atom type. Related work in which output configurations from FlexX were relaxed and re-scored using a more rigorous and time-consuming approach identified the same alternative docking mode [10].

The mode of DANA binding to neuraminidase corresponding to configuration 28 from Fig. 10 and to the configuration of DANA in the crystal (1nsd) are shown in Fig. 13. The inhibitor structures are colored by atom type and orange, respectively, with the former probably representing a secondary binding site occupied by the penultimate glycosidic residue in the natural substrate, similar to the situation for 1dog. This suggests that the affinity of DANA for neuraminidase might well be enhanced by creating a hybrid of the two configurations which bridges the two binding sites identified by docking. Hence, the alternative kind of protein/ligand interaction corresponding to configuration 28 for DANA points up a potentially profitable direction for structural elaboration of the known ligand, even though it is probably not itself useful as a template for drug design.

4.2. Caveats

The work presented here is limited to applications of scoring functions to candidate docking configurations, and should not be interpreted as extending to their use as target functions for docking, i.e. as fitness functions. This distinction between fitness and scoring functions has also been noted in related work by other groups [10,43].

Nor have we tried to address the question of which individual functions work best on their own, but have focused instead on ways in which results from diverse scoring functions might profitably be combined so as to arrive at a CScore which outperforms the individual component scores. At least for the data sets considered here, CScore and rank-sums are more robust and reliable than are individual scoring functions. A similar conclusion was reached for vHTS using CScore with the original DOCK and GOLD docking and scoring functions [14], and by the Vertex group [13]. The latter workers used the rigid docking variation of DOCK and their in-house rigid docking program GAMBLER; they surveyed a range of scoring functions and concluded that the intersection of the best-ranking sets from three of them — piecewise linear potential, ChemScore and the DOCK energy — improved the ratio of valid hits to false positives substantially. This is equivalent to using the most stringent possible CScore criterion, i.e. a unanimous vote, which will sharply reduce the number of false positives obtained from virtual screening program. With the panel of scoring functions used here, however, such a conservative approach would fail to identify any good ligand for thermolysin (Fig. 4B) and would only find a good binding pose for one of the FlexX dockings (Figs. 5–10). In most cases, the modest false positive rate expected for a three-vote CScore minimum is likely to be acceptable provided that it results in a reduced false negative rate, i.e. in fewer missed leads.

In other related work, two groups have reported that applying alternative scoring functions to ensembles of candidate docking configurations produced by FlexX significantly increases the chances that the best-scoring configuration will be very close to that found in the crystal structure [10,43].

5. Conclusions

The version of CScore presented here effectively reduces errors in properly ordering protein complexes of ligands with affinities in the nM to μM range (Fig. 4A and C), including errors due to uncertainties involved in crystallization and X-ray structure determination (compare b01a–c and b02a–b in Fig. 1D with the corresponding entries in Fig. 4D). Furthermore, it allows clear-cut identification of good binding poses among ensembles of candidate ligand configurations produced by FlexX. Several other distinctive consensus scoring functions of potential interest have come to our attention since the worst-best criterion and CScore were originally presented [15], most notably ChemScore [25] and the knowledge-based function developed by the group at Marburg [12]. We are actively working to determine how including these functions or others in our consensus scoring panel will affect performance.

As scoring functions to sort the output from docking programs, these and other consensus scoring schemes clearly have a lot to offer. Researchers want to be able to identify ligand configurations which are reasonably close (if not necessarily identical) to binding modes found in crystal structures. We have shown here that consensus scoring functions are an effective way to accomplish that goal, at least when used in conjunction with FlexX. Work recently published by other groups supports the expectation that output from other docking programs will behave similarly [13,14].

We have also found consensus scoring useful for identifying potentially useful alternative binding mode hypotheses such as those shown in Figs. 11C, 12 and 13. They are clearly not “correct”, yet they may well be thermodynamically accessible and so could serve as distinctive jumping-off points for molecular design. In particular, applying CScore with moderately tight thresholds and relaxed constraints (e.g. the best third/three-vote minimum criterion from Fig. 6C) seems to be an excellent way to pick out good configurations.

As noted above, we are looking to expand the range of scoring functions provided for consensus scoring and want to determine what characterizes good consensus panels. The use of CScore in vHTS to identify stable complexes has already been reported [14]. That application was

limited to rescoring the single configuration identified as best by the particular docking program being considered. We hope to extend that work by using CScore to select the best configurations as well. We would also like to examine how sensitive individual scoring functions are to protein changes within the binding pocket of the target protein and whether consensus scoring might ameliorate problems created by such sensitivity. The cross-docking experiments recently carried out by Murray et al. [44] represent a particularly interesting approach to this task.

Acknowledgements

The authors wish to thank Susan Froshauer and J. Pandit for making the structures of FKBP-12 available to us, as well as for helpful suggestions along the way. Dr. Yvonne Martin (Abbott Laboratories) was kind enough to provide us with an updated list of PMF parameters. Drs. Denise Beusen and Dennis Sprous provided useful input on the organization of the manuscript, as did Steven Burkett and Peter Fox and the reviewers of the original manuscript.

References

[1] I. Muegge, Y.C. Martin, A general and fast scoring function for protein–ligand interactions: a simplified potential approach, J. Med. Chem. 42 (1999) 791–804.
[2] T. Liljefors, Progress in force-field calculations of molecular interaction fields and intermolecular interactions, in: H. Kubinyi, G. Folkers, Y.C. Martin (Eds.), 3D QSAR in Drug Design, Vol. 2, Kluwer Academic Publishers, Dordrecht, 1998, pp. 3–17.
[3] R.C. Wade, A.R. Ortiz, F. Gago, Comparative binding energy analysis, in: H. Kubinyi, G. Folkers, Y.C. Martin (Eds.), 3D QSAR in Drug Design, Vol. 2, Kluwer Academic Publishers, Dordrecht, 1998, pp. 19–34.
[4] T.I. Oprea, G.R. Marshall, Receptor-based prediction of binding affinities, in: H. Kubinyi, G. Folkers, Y.C. Martin (Eds.), 3D QSAR in Drug Design, Vol. 2, Kluwer Academic Publishers, Dordrecht, 1998, pp. 35–61.
[5] M.K. Holloway, A priori prediction of ligand affinity by energy minimization, in: H. Kubinyi, G. Folkers, Y.C. Martin (Eds.), 3D QSAR in Drug Design, Vol. 2, Kluwer Academic Publishers, Dordrecht, 1998, pp. 63–84.
[6] M.R. Reddy, V.N. Viswanadhan, M.D. Erion, Rapid estimation of relative binding affinities of enzyme inhibitors, in: H. Kubinyi, G. Folkers, Y.C. Martin (Eds.), 3D QSAR in Drug Design, Vol. 2, Kluwer Academic Publishers, Dordrecht, 1998, pp. 85–98.
[7] R.M.A. Knegtel, P.D.J. Grootenhuis, Binding affinities and non-bonded interaction energies, in: H. Kubinyi, G. Folkers, Y.C. Martin (Eds.), 3D QSAR in Drug Design, Vol. 2, Kluwer Academic Publishers, Dordrecht, 1998, pp. 99–114.
[8] I.T. Weber, R.W. Harrison, Molecular mechanics calculations on protein–ligand complexes, in: H. Kubinyi, G. Folkers, Y.C. Martin (Eds.), 3D QSAR in Drug Design, Vol. 2, Kluwer Academic Publishers, Dordrecht, 1998, pp. 115–127.
[9] References 2–8 also appear in: Perspect. Drug Discovery Design 1998, 9/10/11.
[10] D. Hoffmann, B. Kramer, T. Washio, T. Steinmetzer, M. Rarey, T. Lengauer, Two-stage method for protein–ligand docking, J. Med. Chem. 42 (1999) 4422–4433.
[11] M. Stahl, Modifications of the scoring function in FlexX for virtual screening applications, Perspect. Drug Discovery Design 20 (2000) 83–98.
[12] H. Gohlke, M. Hendlich, G. Klebe, Predicting binding modes, binding affinities and 'hot spots' for protein–ligand complexes using a knowledge-based scoring function, Perspect. Drug Discovery Design 20 (2000) 115–144.
[13] P.S. Charifson, J.J. Corkery, M.A. Murcko, W.P. Walters, Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins, J. Med. Chem. 42 (1999) 5100–5109.
[14] C. Bissantz, G. Golkers, D. Rognan, Protein-based virtual screening of chemical databases. Part 1. Evaluation of different docking/scoring combinations, J. Med. Chem. 43 (2000) 4759–4767.
[15] R. Clark, A. Strizhev, A. Nayeem, Working toward a consensus-scoring function. in: Proceedings of the 218th ACS National Meeting, 1999, COMP 15 (abstract).
[16] J. Gasteiger, M. Marsili, Iterative partial equalization of orbital electronegativity — a rapid access to atomic charges, Tetrahedron 36 (1980) 3219–3228.
[17] J. Gasteiger, M. Marsili, Org. Magn. Reson. 15 (1981) 353–360.
[18] E.C. Meng, B.K. Shoichet, I.D. Kuntz, Automated docking with grid-based energy evaluation, J. Comp. Chem. 13 (1992) 505–524.
[19] S.J. Weiner, P.A. Kollman, D.T. Nguyen, D.A. Case, A new force field for molecular mechanical simulation of nucleic acids and proteins, J. Comp. Chem. 7 (1986) 230–252.
[20] G. Jones, P. Willett, R.C. Glen, Molecular recognition of receptor sites using a genetic algorithm with a description of solvation, J. Mol. Biol. 245 (1995) 43–53.
[21] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748.
[22] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, A fast flexible docking method using an incremental construction algorithm, J. Mol. Biol. 251 (1996) 470–489.
[23] H.-J. Böhm, 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. Design 8 (1994) 243–256.
[24] I. Muegge, Y.C. Martin, P.J. Hajduk, S.W. Fesik, Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein, J. Med. Chem. 42 (1999) 2498–2503.
[25] M.D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, R.P. Mee, Empirical scoring functions. Part I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes, J. Comput.-Aided Mol. Design 11 (1997) 425–445.
[26] W.W. Daniel, Applied Nonparametric Statistics, Houghton Mifflin Company, Boston, 1978.
[27] J. Pandit, D.E. Danley, G.K. Schulte, S. Mazzalupo, T.A. Pauly, C.M. Hayward, E.S. Hamanaka, J.F. Thompson, H.J. Harwood, Crystal structure of human squalene synthase. A key enzyme in cholesterol biosynthesis, J. Biol. Chem. 275 (2000) 30610–30617.
[28] A.L. Swain, M.M. Miller, J. Green, D.H. Rich, J. Schneider, S.B. Kent, A. Wlodawer, X-ray crystallographic structure of a complex between a synthetic protease of human immunodeficiency virus 1 and a substrate-based hydroxyethylamine inhibitor, Proc. Natl. Acad. Sci. U.S.A. 87 (1990) 8805–8809.
[29] P.Y. Lam, P.K. Jadhav, C.J. Eyermann, C.N. Hodge, Y. Ru, L.T. Bachler, J.L. Meek, M.J. Otto, M.M. Rayner, Y.N. Wong, C.-H. Chang, P.C. Weber, D.A. Jackson, T.R. Sharpe, S. Erickson-Viitanen, Rational design of potent, bioavailable nonpepide cyclic ureas as HIV protease inhibitors, Science 263 (1994) 380–384.
[30] E.E. Kim, C.T. Baker, M.D. Dwyer, M.A. Murcko, B.G. Rao, R.D. Tung, M.A. Navia, Crystal structure of HIV-1 protease in complex with Vx-478, a potent and orally bioavailable inhibitor of the enzyme, J. Am. Chem. Soc. 117 (1995) 1181–1182.

[31] M. Miller, J. Schneider, B.K. Sathyanarayana, M.V. Toth, G.R. Marshall, L. Clawson, L. Selk, A. Wlodawer, Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution, Science 246 (1989) 1149–1152.
[32] D.E. Tronrud, H.M. Holden, B.W. Matthews, Structures of two thermolysin-inhibitor complexes that differ by a single hydrogen bond, Science 235 (1987) 571–574.
[33] P.A. Bartlett, C.K. Marlowe, Evaluation of intrinsic binding energy from a hydrogen bonding group in an enzyme inhibitor, Science 235 (1987) 569–571.
[34] H. Brandstetter, D. Turk, H.W. Hoeffken, D. Grosse, J. Sturzebecher, P.D. Martin, B.F. Edwards, W. Bode, Refined 2.3 Å X-ray crystal structure of bovine thrombin complexes formed with the benzamidine and arginine-based thrombin inhibitors NAPAP, 4-TAPAP and MQPA. Astarting point for improved antithrombotics, J. Mol. Biol. 226 (1992) 1085–1099.
[35] D.W. Banner, P. Hadvary, Crystallographic analysis at 3.0 Å resolution of the binding to human thrombin of four active site-directed inhibitors, J. Biol. Chem. 266 (1991) 20085–20093.
[36] A. Teplyakov, K.S. Wilson, P. Orioli, S. Mangani, The high resolution crystal structure of the complex between carboxypeptidase A and L-phenyl lactate, Acta Crystallogr D. Biol. Crystallogr 49 (1993) 534–540.
[37] E.M. Harris, A.E. Aleshin, L.M. Firsov, R.B. Honzatko, Refined structure for the complex of 1-deoxynojirimycin with glucoamylase from *Aspergillus awamori* var. 100× to 2.4 Å resolution, Biochemistry 32 (1993) 1618–1626.
[38] W.P. Burmeister, B. Henrissat, C. Bosso, S. Cusack, R.W. Ruigrok, Influenza B virus neuraminidase can synthesize its own inhibitor, Structure 1 (1993) 19–26.
[39] S.K. Kearsley, S. Sallamack, E.M. Fluder, J.D. Andose, R.T. Mosley, R.P. Sheridan, Chemical similarity using physicochemical property descriptors, J. Chem. Inf. Comput. Sci. 36 (1996) 118–127.
[40] S.-S. So, M. Karplus, A comparative study of ligand-receptor complex binding affinity methods based on glycogen phosphorylase inhibitors, J. Comput.-Aided Mol. Design 13 (1999) 243–258.
[41] C.M.R. Ginn, P. Willett, J. Bradshaw, Combination of molecular similarity measures using data fusion, Perspect. Drug Discovery Design 20 (2000) 1–16.
[42] G. Rhodes, Crystallography Made Crystal Clear, Academic Press, San Diego, 1993, pp. 162–164.
[43] H. Gohlke, M. Hendlich, G. Klebe, Knowledge-based scoring function to predict protein–ligand interactions, J. Mol. Biol. 295 (2000) 337–356.
[44] C.W. Murray, C.A. Baxter, A.D. Frenkel, The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase, J. Comput.-Aided Mol. Design 13 (1999) 547–562.