Journal of Computer-Aided Molecular Design, 10 (1996) 427–440
ESCOM

J-CAMD 351

Scoring noncovalent protein–ligand interactions: A continuous differentiable function tuned to compute binding affinities

Ajay N. Jain

Arris Pharmaceutical Corporation, 385 Oyster Point Boulevard, San Francisco, CA 94080, U.S A.

Received 7 February 1996
Accepted 28 April 1996

Keywords: Molecular docking; Protein–ligand interaction; Scoring function

Summary

Exploitation of protein structures for potential drug leads by molecular docking is critically dependent on methods for scoring putative protein–ligand interactions. An ideal function for scoring must exhibit predictive accuracy and high computational speed, and must be tolerant of variations in the relative protein–ligand molecular alignment and conformation. This paper describes the development of an empirically derived scoring function, based on the binding affinities of protein–ligand complexes coupled with their crystallographically determined structures. The function's primary terms involve hydrophobic and polar complementarity, with additional terms for entropic and solvation effects. The issue of alignment/conformation dependence was solved by constructing a continuous differentiable nonlinear function with the requirement that maxima in ligand conformation/alignment space corresponded closely to crystallographically determined structures. The expected error in the predicted affinity based on cross-validation was 1.0 log unit. The function is sufficiently fast and accurate to serve as the objective function of a molecular-docking search engine. The function is particularly well suited to the docking problem, since it has spatially narrow maxima that are broadly accessible via gradient descent.

Introduction

As the number of proteins whose structure is known to atomic resolution expands, approaches to 3D database searching for small-molecule leads are becoming increasingly important. There have been numerous reports of leads discovered through database screening [1–4]. One key area for improvement in docking systems is in the methods for scoring putative protein–ligand interactions. Scoring is central to the docking problem. The goal of a molecular docker, given a protein and a ligand, is to extremize the value of a scoring function by varying the alignment and conformation of the ligand relative to the protein (a conformation/alignment of a molecule will be referred to as a *pose* in the remainder of the paper). This paper presents an empirically tuned scoring function for predicting the binding affinities of protein–ligand interactions, with the specific goal of use in a flexible molecular docking system.

There are three key properties in an ideal scoring function: (i) accuracy in predicting binding affinities; (ii) speed; and (iii) tolerance to inaccurate ligand pose. The first two are obvious, but the last requires some discussion. A geometric search engine attempting to generate suitable poses for a ligand given a protein binding site will generate the 'correct' pose within some tolerance with some probability within some amount of time. As the tolerance increases, so does the probability of generating a good enough pose quickly. So, the degree to which a scoring function is able to compute an accurate binding affinity given an inaccurate pose will have a significant impact on its utility in such a system.

One way to achieve the goal of pose insensitivity is to construct a spatially coarse function – one where small perturbations in pose will not cause significant changes in the predicted affinity. However, this will necessarily lead to a loss of specificity in the function. The broader the 'sweet spots' in a function, the more ligands will appear (incorrectly) to score well. The problem of false positives in docking is much more serious than false negatives. Suppose that in a database of 100 000 compounds just 20 will bind above some threshold affinity to the target in question. Even if the false negative rate is high (say 50%), one would likely be happy with finding 10 true hits. How-

ever, depending on how many compounds may be practically obtained (through purchase or synthesis) and assayed, the false positive rate has a significant impact on the likelihood of finding the true hits. If the false positive rate is just 1%, there will be 1000 false positive 'hits' in this screen. If there is capacity to obtain and assay only 50 of the 1010 hits, with no principled way of distinguishing true positives from false positives, there is a 60% chance of not finding a single true hit. In order to achieve a 90% chance of finding at least one true hit, the absolute false positive rate must be 0.2%.

While the foregoing analysis is speculative, qualitatively it depends only on the observation that the number of true specific binding ligands in a large database of compounds is very much smaller than the number of non-binders, which is certainly true. It is clearly important to produce as *specific* a function as possible. The approach taken here is to construct a function with well-defined narrow maxima in order to minimize the false positive rate. However, this makes it difficult for the docker to find ligand poses corresponding to maxima. An additional constraint on the function addresses this problem. The function is continuous and differentiable so that adocker need only find a pose *close* to a functional maximum, and gradient-based search is relied on to find the maximum. So, inaccurate initial ligand poses are corrected by the scoring function instead of relying on the docker to generate perfect poses. To accommodate this, the fit of the function to its calibration data is defined as the maximum value of the function under gradient-based variation of ligand pose. The desired behavior of the function is that the maximum score for a protein–ligand interaction is close to the correct binding affinity and the corresponding pose is close to the true structure.

The functional form used here follows that of the Compass technique using a combination of nonlinear

TABLE 1
COMPLEXES USED IN THE CALIBRATION SET

Number	Protein	Ligand	Affinity[a]	Source
1	Carboxypeptidase	ZFV^{P}(O)F	14.0	7CPA
2	Streptavidin	Biotin	13.4	1STP
3	Carboxypeptidase	A-ZAA^{P}(O)F	11.52	6CPA
4	Thermolysin	ZF^{P}LA	10.19	4TMN
5	DHFR	Methotrexate	9.70	4DFR
6	HIV Protease	L700,417	9.15	4PHV
7	Thrombin	NAPAP	8.52	1DWD
8	Thermolysin	ZG^{P}LL	8.04	5TMN
9	Galactose BP	Galactose	7.60	2GBP
10	Thermolysin	Phosphoramidon	7.55	1TLP
11	Thrombin	MQPA	7.40	1ETR
12	Thermolysin	CLT	7.30	1TMN
13	Retinol BP	Retinol	6.72	1RBP
14	Trypsin	NAPAP	6.46	1PPC
15	Thermolysin	HONH-BAGN	6.37	5TLN
16	Trypsin	3-TAPAP	6.22	1PPH
17	Thrombin	TAPAP	6.19	1ETT
18	Cytochrome P450	4-Phe-imidazole	6.07	1PHF
19	DHFR	2,4-Diaminopteridine	6.00	4DFR[b]
20	Cytochrome P450	Adamantone	5.88	5CPP
21	Xylose isomerase	Xylose	5.82	1XIS
22	Fatty acid BP	C_{15}COOH	5.43	2IFB
23	PNP	Guanine	5.30	1ULB
24	TIM	Phosphoglycilic acid	4.82	2YPI
25	Trypsin	Benzamidine	4.74	3PTB
26	PHBH	*p*-Hydroxybenzoate	4.68	2PHH
27	Thermolysin	PLN	4.67	2TMN
28	Trypsin	Phenylguanidine	4.14	3PTB[c]
29	Thrombin	Amidinopiperidine	3.82	1DWD[c]
30	Thermolysin	Leu-NHOH	3.72	4TLN
31	Trypsin	Benzylamine	3.42	3PTB[c]
32	Chymotrypsin	Indole	3.10	4CHA[c]
33	Thrombin	Benzamidine	2.92	1DWB
34	Trypsin	Butylamine	2.82	3PTB[c]

[a] Units are $-\log_{10}(K_d)$.
[b] The protein structure from 5 was used by deleting all but the pteridine fragment of methotrexate.
[c] Ligands were docked into the binding sites using Hammerhead [11].

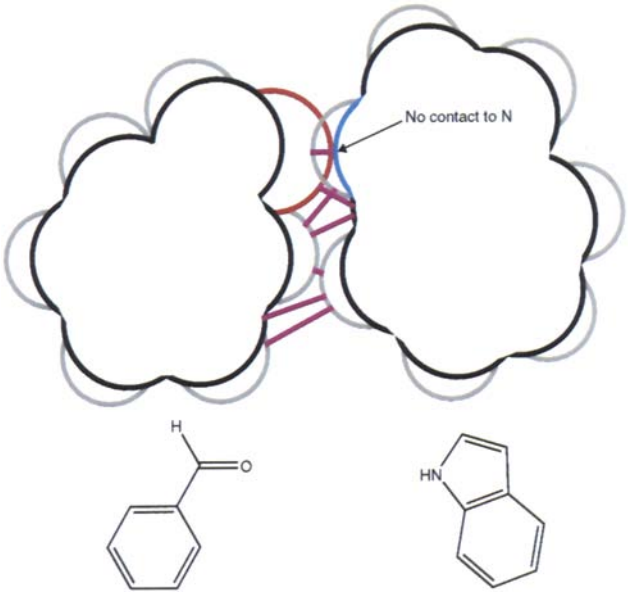

Fig. 1. Protein–ligand atom pairs involved in the computed interaction. Purple lines indicate interatomic contacts that are not occluded and thus form the basis of the interaction energy sum.

basis functions with tunable parameters estimated by iterative gradient descent [5–7]. Apart from the Compass functional approach, this work borrows most directly from Böhm [8] and from Bohacek and McMartin [9]. Böhm showed that it is possible to construct a simple and accurate empirically tuned scoring function by accounting for hydrophobic surface contact area, favorable polar contacts, and an entropic correction. Earlier, Bohacek had shown that a weighted linear sum of the number of complementary hydrophobic and polar protein–ligand contacts fits the K_i's of a series of thermolysin inhibitors extremely well. The primary terms in the Compass-like scoring function reported here are measures of hydrophobic and polar complementarity, with additional terms for entropy and solvation effects.

The calibration data set consisted of 34 protein–ligand complexes with known affinities and geometries. The mean error of predicted binding affinity under leave-one-out cross-validation was 1.0 log unit. The maxima of the scoring function subject to ligand pose variation corresponded well to the experimentally determined structures. The maxima are very narrow: a divergence of 0.5 Å rms from the maximum pose can produce a drop of 3 log units in computed binding affinity. However, the gradients of the function are sufficiently well-behaved that displacements of up to 2.0 Å do not result in poses that cannot be optimized through gradient descent. The function is quite fast to compute (0.03 s per evaluation on the trypsin–benzamidine interaction), which makes it suitable for molecular docking. The technique forms the basis for a method of automatically identifying and characterizing protein binding sites [10]. It is also being used in a flexible molecular docking system [11].

Methods

Calibration data set

The complexes used to estimate the parameters of the function were a subset of the complexes presented by Böhm [8]. Complexes that failed to meet the following

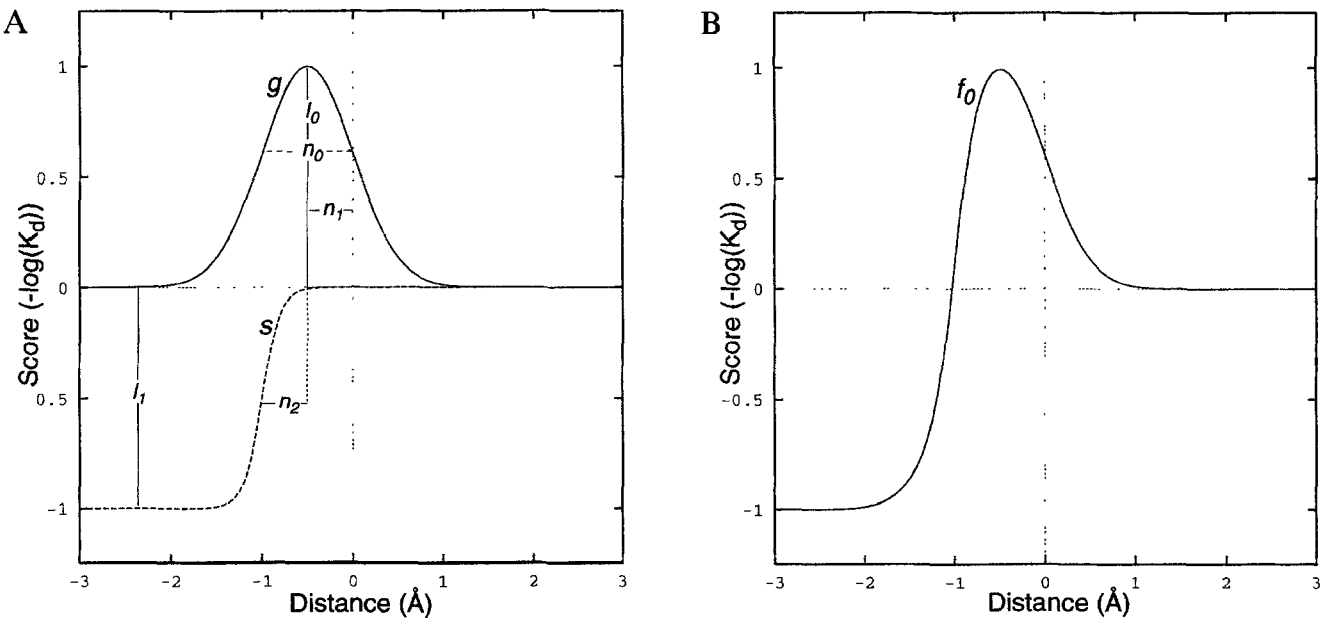

Fig. 2. Plot of basis functions. (A) Gaussian-like and sigmoidal basis functions have tunable parameters. (B) Combined, they form a function of interatomic distance with a singular maximum, with large positive distances resulting in an asymptotically decreasing contribution, and with large interpenetrations resulting in a negative contribution.

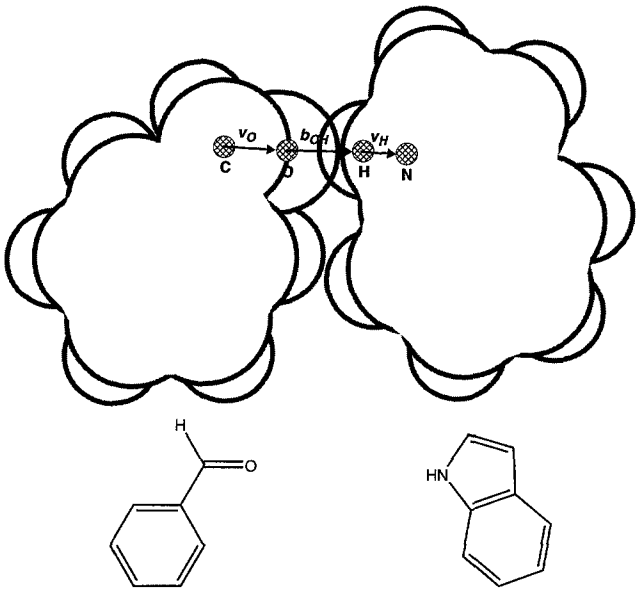

Fig. 3. Directionality in polar interactions. The vectors defining the direction of the C=O bond, the O···H hydrogen bond, and the H-N bonds are used in computing the directional contribution of the hydrogen bond.

criteria were eliminated: available or derivable structure, a dissociation constant of less than 10^{-3} M, and no significant protein–ligand steric overlap after protonation. Hydroxyls and thiols on the proteins and ligands were systematically repositioned to yield minimal unfavorable interpenetration prior to the assessment of protein–ligand overlap. This resulted in 34 complexes with K_d ranging from 10^{-3} to 10^{-14} M. Table 1 lists the complexes, affinities, and sources of data. The binding interactions in all cases are believed to be noncovalent. Note that a small number of the complexes were derived by docking small, relatively rigid ligands into the active sites of the proteins. In these cases, complexes with other very similar molecules or fragments thereof had been solved experimentally.

Functional form

The scoring function is a weighted sum of nonlinear functions of exposed protein–ligand atomic van der Waals surface distances. It accounts for hydrophobic contacts, polar contacts and directionality, solvation effects, and entropic effects. It is piecewise continuous and differentiable. Binding affinities are expressed in $-\log_{10}(K_d)$ units throughout the paper.

Figure 1 depicts a 2D cartoon of the atom pairs that are used in computing the interaction energy. For each pair of atoms, the nearest surface distance is computed (interpenetrations are considered as negative distances). If the distance is greater than 2.0 Å, or if the path interpenetrates another atom, the pair is ignored. While it is possible to compute the nearest nonoccluded distance, this involves a more complicated computation. The assumption is that in cases where an occlusion prevents an atom-pair interaction, the interaction involving the atoms causing the occlusion will tend to dominate. A protein–ligand atom pair that passes these tests will be referred to as a pair.

Each atom on the protein and ligand is labeled as being nonpolar (e.g. the H of a CH_3) or polar (e.g. the H of an N-H or the O of a C=O), and polar atoms are also assigned a charge. A discussion of the form of the scoring function will be broken down in rough order of significance: hydrophobic complementarity, polar complementarity, solvation terms, and entropic terms. The full scoring function is the sum of each of these terms. The remainder of this section may be skipped by readers more interested in the practical aspects of the function's utility than the details of its implementation.

Hydrophobic complementarity

The hydrophobic effect is captured by the weighted sum of a Gaussian-like function g and a sigmoid function s of pairwise surface distances as in Compass [5–7]:

$$g(x,\mu,\sigma) = e^{-(x+\mu)^2/\sigma} \quad (1)$$

$$s(x,\mu) = 1/(1+e^{10(x+\mu)}) \quad (2)$$

The function g captures the positive portion of atomic contacts, while s captures the portion due to excess interpenetration. In the equations that follow, tunable linear parameters will be denoted by l_i, and tunable parameters with a non-linear relationship to binding affinity will be denoted by n_i:

$$f_0(x) = l_0\, g(x,n_0,n_1) + l_1\, s(x,n_2+n_1) \quad (3)$$

The function f_0 defines the hydrophobic contribution to the binding affinity. Figure 2 shows a plot of f_0 with $l_0 = 1.0$, $l_1 = -1.0$, $n_0 = 0.5$, $n_1 = 0.5$, and $n_2 = 0.5$ (the parameters are for illustrative purposes only). Panel A shows the g

TABLE 2
SUMMARY OF EQUATIONS DEFINING F

Hydrophobic term	$\Sigma_{i,j}\, f_0(d(i,j))$
Polar term	$\Sigma_{i,j}\, f_1(d(i,j),i,j)$
Repulsive term	$\Sigma_{i,j}\, f_2(d(i,j),i,j)$
Solvation term	$(l_5 \cdot \text{phbe}) + (l_6 \cdot \text{lhbe})$
Entropic term	$(l_7 \cdot \text{n\_rot}) + (l_8 \log(\text{mol weight}))$
$f_0(x)$	$l_0\, g(x,n_0,n_1) + l_1\, s(x,n_2+n_1)$
$f_1(x,i,j)$	$f_{1a}(x)\, f_{1b}(i,j)\, (1+n_6 c_i)\, (1+n_6 c_j)$
$f_2(x)$	$l_5\, g(x,n_7,n_8)\, f_{1a}(x)\, f_{1b}(i,j)\, (1+n_6 c_i)\, (1+n_6 c_j)$
$f_{1a}(x)$	$l_2\, g(x,n_3,n_4) + l_3\, s(x,n_2,n_4)$
$f_{1b}(i,j)$	$s(-(b_{ij} \cdot v_i)\, (b_{ij} \cdot v_j), -n_5$
$g(x,\mu,\sigma)$	$e^{-(x+\mu)^2/\sigma}$
$s(x,\mu)$	$1/(1+e^{10(x+\mu)})$
$d(i,j)$	$((x_i-x_j)^2+(y_i-y_j)^2+(z_i-z_j)^2)^{1/2} - r_i - r_j$

See text for further details.

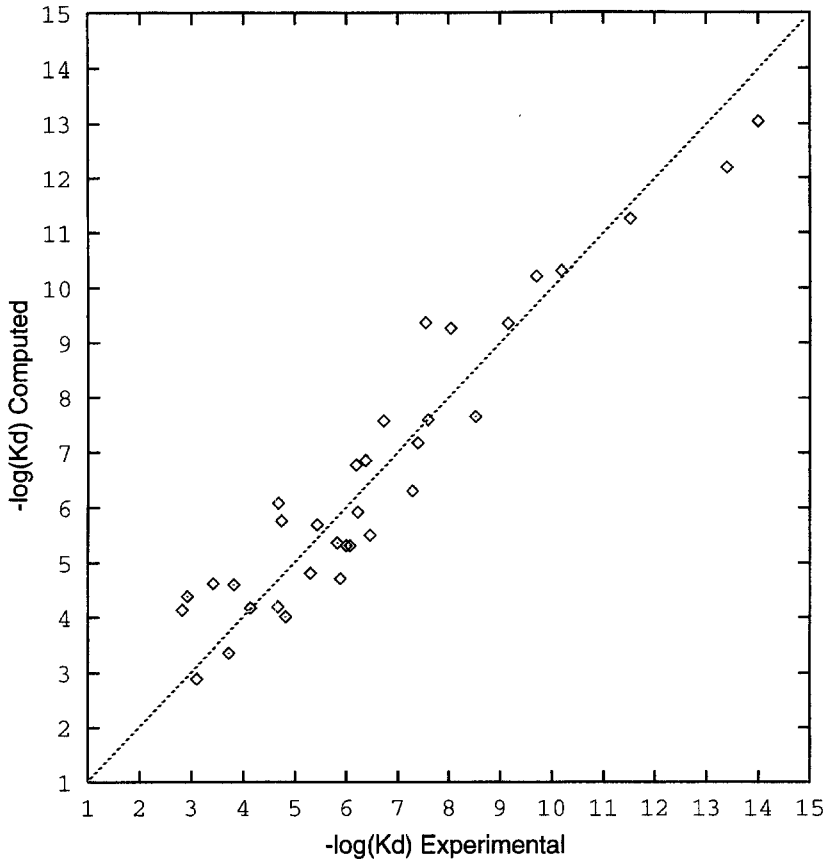

Fig. 4. Plot of computed versus experimental affinities.

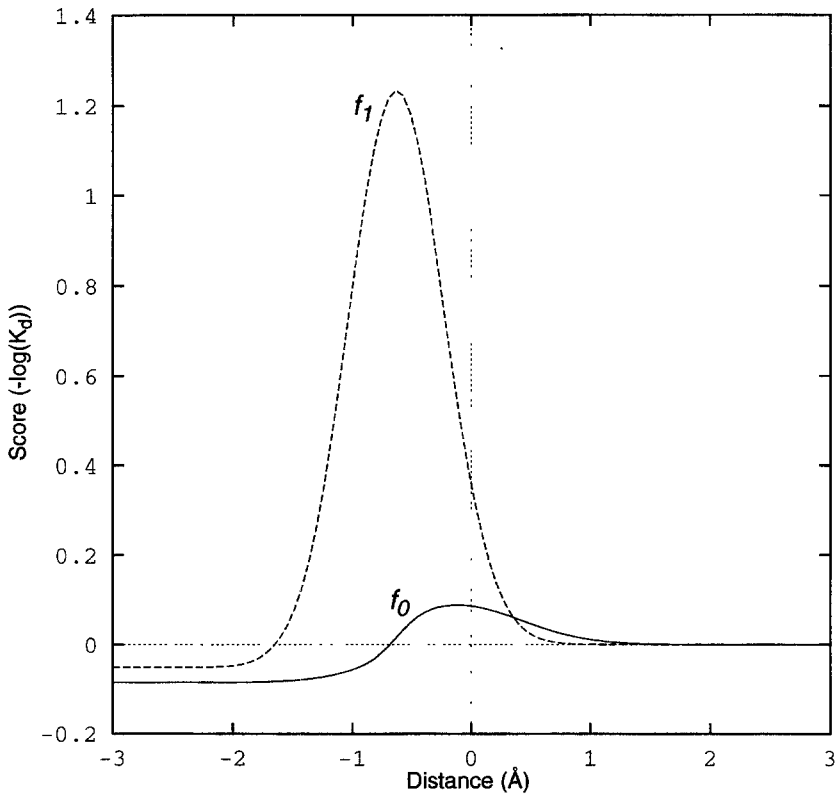

Fig. 5. Plot of f_0 and f_1 for F.

and s contributions separately, and panel B shows the combined function. It has a positive maximum at a particular surface distance; the function drops off to a negative minimum as the distance decreases.

The linear parameters control the overall contribution to the binding affinity, while the nonlinear parameters control the underlying shape of the basis function. The parameter l_0 represents the positive hydrophobic contribution of a single ideal nonpolar contact and l_1 is the maximum cost of excess steric overlap. The parameter n_0 corresponds to the narrowness of the hydrophobic interaction, n_1 is the degree of interpenetration required to achieve maximal interaction, and n_2 is the amount of excess interpenetration required to achieve half the maximum interpenetration penalty.

The total hydrophobic component of the interaction energy is:

$$\Sigma_{i,j}\, f_0(d(i,j)) \qquad (4)$$

where

$$d(i,j) = ((x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2)^{1/2} - r_i - r_j \qquad (5)$$

The atomic coordinates of atom i are denoted by x_i, y_i, and z_i. The van der Waals radius of atom i is denoted by r_i. The sum is over all pairs that are not both polar atoms.

Polar complementarity

The effects of hydrogen bonds and salt bridges are computed using a sum similar to that of the hydrophobic term. The polar basis function is summed over all pairs of *complementary* polar atoms. It has the same form as f_0, but with different parameters:

$$f_{1a}(x) = l_2\, g(x,n_3,n_4) + l_3\, s(x,n_2,n_4) \qquad (6)$$

The function f_{1a} defines the interaction contribution for a neutral hydrogen bond of optimal directionality. Since hydrogen bonding is influenced by factors in addition to pure distance, a directionality term is also required:

$$f_{1b}(i,j) = s(-(b_{ij}\cdot v_i)\,(b_{ij}\cdot v_j), -n_5 \qquad (7)$$

The vector b_{ij} is the normalized vector from atom i to atom j, v_i is the 'out' direction of atom i, and v_j is the 'in' direction of atom j. Figure 3 illustrates this (O is atom i, H is atom j). Note that the direction factor is symmetric: switching atoms i and j results in the same value. The function reaches a maximum of 1.0 as the vectors become coincident and drops to 0.0 as the directions diverge. The point at which the strength is half-maximal is determined by n_6. The function essentially defines a cone of favorability around each of the polar contact participants relative to their mutual contact direction. The in/out direction for

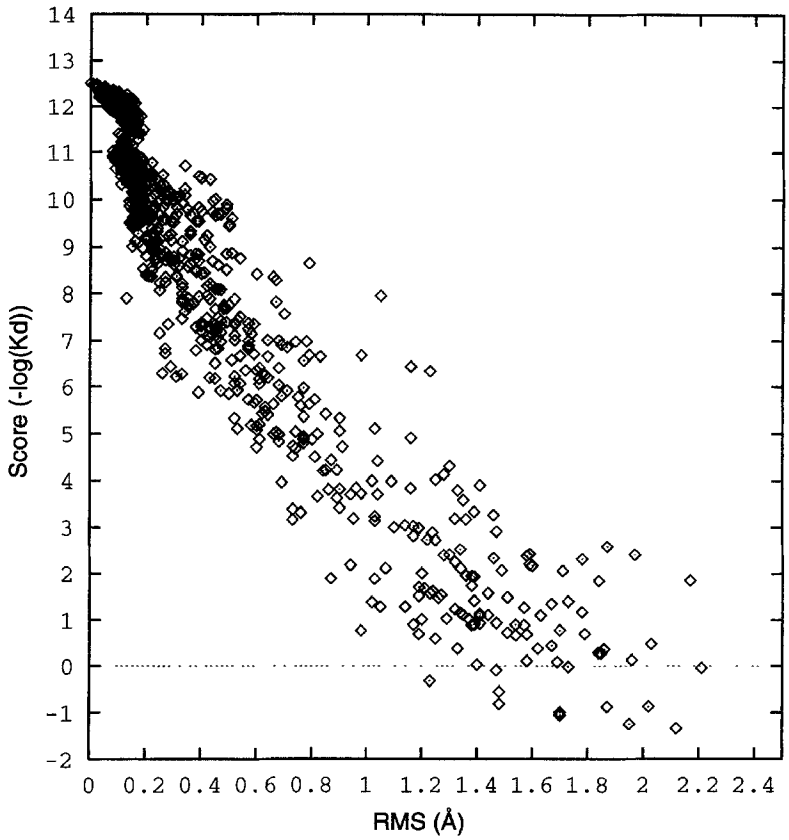

Fig. 6. Plot of rms deviation from maximal pose versus computed affinity.

each polar atom is computed using the centroid of the atoms connected to it along with the polar atom position. For carbonyls, two alternative vectors are also generated, emanating from the oxygen and parallel to the vectors from the sp^2 carbon to its other substituents. The strength of the interaction arising from the most favorable set of vectors is used in the computation. This approximates the geometric distribution of hydrogen-bonding geometries observed in crystallographically determined structures [12].

One last addition is required to account for the effects of formally charged interactions. The final equation defining the strength of a single polar contact is

$$f_1(x,i,j) = f_{1a}(x)\, f_{1b}(i,j)\, (1 + n_6 c_i)\, (1 + n_6 c_j) \qquad (8)$$

The formal charge of atom i is denoted by c_i. So, as the formal charge of either atom increases, the strength of the interaction increases linearly, as in Coulomb's law. Partial charges are not used.

To account for unfavorable polar contacts, the following function is summed for pairs of polar atoms of the same sign:

$$f_2(x) = l_5\, g(x,n_7,n_8)\, f_{1a}(x)\, f_{1b}(i,j)\, (1 + n_6 c_i)\, (1 + n_6 c_j) \qquad (9)$$

This function has a single peak and falls off as distance increases.

Solvation effects

One can view solvation effects as the costs associated with breaking favorable interactions with water and failing to remake those interactions with ligand or protein. So, packing a lipophilic portion of a ligand against polar moieties on a protein or vice versa should cost something. Solvation effects on the protein and ligand are treated symmetrically. The object is to compute the total possible 'hydrogen-bond equivalents' for the protein and ligand compared with the actual number achieved in the interaction. So, a protein carbonyl pointed directly at the ligand and proximal to it will receive one hydrogen-bond equivalent. A polar protein atom that is oriented away from the ligand or which is too far away will receive a value close to zero. In both cases, it is assumed that such atoms will be able to interact with solvent molecules or other protein atoms.

For the protein, each pair involving a polar atom on the protein is identified. For each of these, the strength (directionality and charge effect) of the best potential polar interaction arising from any ligand atom is computed (each ligand atom is considered to be a neutral hydrogen-bond participant of the right polarity). This value is attenuated by the sigmoidal distance function $s(x,n_7)$. The parameter n_7 determines the point at which the attenuation function reaches its half-maximal value.

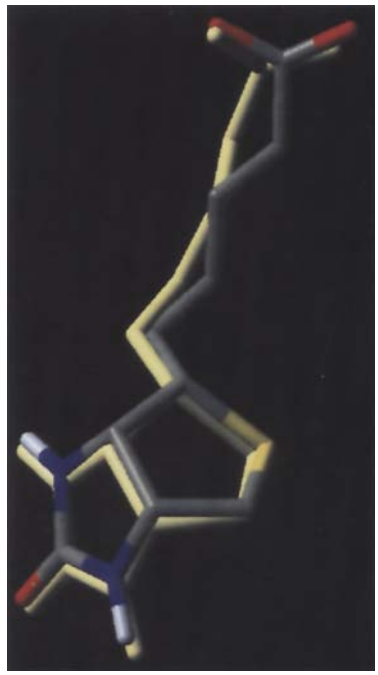

Fig. 7. Gradient-optimized pose of biotin (grey) versus crystallographically determined structure.

The difference between the total number of potential protein hydrogen-bond equivalents and the actual polar interaction amount (phbe) is multiplied by l_5. Similarly, the difference between the total number of ligand hydrogen-bond equivalents and the actual polar interaction amount (lhbe) is multiplied by l_6.

Entropic costs

Entropic costs due to fixation of rotatable ligand bonds and loss of rotational and translational entropy form the remaining terms of the function. These do not depend on the protein and are captured by two simple terms: the first is simply the number of freely rotatable bonds in the ligand (n_rot) multiplied by a scale factor l_7, and the second is the log of the molecular weight of the ligand multiplied by a scale factor l_8 (motivated by the rough mass dependence of the translational/rotational entropy [13]). Fixation costs are not computed for the protein, since the target application of the scoring function is in a molecular docking system where the protein is treated as rigid. The variations in the fixation costs of the protein from ligand to ligand for a particular site are assumed to be small.

All the equations presented in this section are summarized in Table 2. The full scoring function is the sum of the hydrophobic, polar, repulsive, and entropic terms over the appropriate atom pairs.

Experimental results

Recall that the scoring function is to be used by a

molecular docker whose goal is to maximize the score of a ligand by varying its pose. Therefore, the score of the ligand at its maximal pose must be considered its effective score. The desired behavior of a scoring function is that it has its global maximum close to the crystallographically determined pose. An accurate scoring function that depends on knowing the crystallographically determined pose is of little utility. This issue is best illustrated by considering a function constructed using static poses and then looking at the effects of pose variation.

A function was constructed using the 34 native crystal structures. The tunable parameters of the function were computed by gradient descent from a random initial starting point with small values for the linear parameters. Table 3 shows the fit to the data of this function (called F_1). All error values are reported in terms of $-\log_{10}(K_d)$ units. As a nonparametric measure of the computed rank of the complexes, the pair-ranking correlation coefficient (PRCC) is also given [7]. The PRCC is an estimate of the likelihood of ranking two predictions correctly. The performance appears reasonable: 1.2 logs $\overline{err}$ (mean error) and 1.4 rmse (root-mean-squared error). However, if the poses of the ligands are optimized by gradient descent, many of the scores increase significantly. The second set of values reported in Table 3 for F_1 results from pose optimization beginning with the native poses. While no ligand diverges a great deal from its native pose in terms of rms, the fit in terms of computed affinities is significantly worse: 1.5 logs $\overline{err}$ (2.1 rmse). Clearly it is important to account for this effect during the parameter optimization process, else the on-line performance of the function will suffer.

This issue was approached in the same manner that the Compass algorithm approaches the problem [5–7]. The training algorithm iterates parameter estimation and ligand pose optimization. The initial random parameter values are optimized given the static poses, then the poses are optimized with the current function, and the process is repeated. During parameter variation, it is required that the *maximum* value of the scoring function on a complex be close to the experimental value. Convergence to a stable set of parameters occurs within five iterations.

This approach raises an additional issue in how to deal with unfavorable contacts between ligand and protein. It is somewhat risky to begin with a function that 'knows nothing' and to iteratively develop a tuned function. There is nothing in the training set that implies that steric overlap is unfavorable, since the data set only contains examples of ligands that do not exhibit significant protein penetration. So, it is possible that hard physical constraints (such as steric overlap) might be ignored during intermediate pose optimization, with the resulting poses being treated as perfectly legal. To avoid this problem, during pose optimization, a steric overlap term is added to the function (this was also done in the pose variation experiment above with F_1). The steric overlap term for atom pair ij is $-10.0(d_{ij} + \delta)(d_{ij} + \delta)$, where δ is 0.7 for complementary polar contacts and 0.1 for others (these values were not chosen in a systematic fashion). So, a steric overlap of 0.5 Å generates a penalty of 1.6 log units. In order to maximize the scoring function, the steric overlap term must be respected. An ideal system would learn appropriately hard negative penalties for steric overlap by explicitly modeling negative data. However, lacking negative data, imposition of the additional penalty generates a tighter function, since the function is more highly constrained. The additional term coerces the system to learn a set of parameters subject to the constraint that steric overlap be avoided.

The steric overlap term is only used during pose optimization; it does not enter into the final score. One might expect that there should be a significant penalty for even minor interpenetrations in the final score. However, even in high-resolution structures of tight-binding ligands, one often finds some steric overlap. There is little correlation between such overlaps and binding affinities. Clearly, very significant overlaps cannot be tolerated, but setting the threshold on how much can be tolerated is an empirical issue relating to the screening aspect of the docking problem.

F_2 was constructed using online pose optimization during the parameter estimation process beginning from the native poses. Its performance was significantly better than F_1 under the pose optimization condition, with 1.0 logs $\overline{err}$ and 1.2 rmse (versus 1.5 and 2.1 for F_1 respectively). Pose optimization is required by the end-application, but it should also help ameliorate problems with

TABLE 3
GOODNESS OF FIT OF THE DIFFERENT FUNCTIONS

Function	Mean error	Rmse[a]	PRCC[b]	Conditions of test
F_1	1.15	1.39	0.90	Static native poses
F_1	1.51	2.09	0.90	Native poses, pose optimization
F_2	0.97	1.17	0.92	Native poses, pose optimization
F	0.72	0.85	0.95	Minimized ligands, pose optimization

[a] Root-mean-squared error.
[b] Pair rank correlation coefficient (1.0 is a perfect ranking, 0.5 is random, 0.0 is reversed) [7].

TABLE 4
PARAMETER VALUES OF F

Parameter	Value	Parameter	Value
l_0	0.0898	n_0	0.6213
l_1	−0.0841	n_1	0.1339
l_2	1.2338	n_2	0.4880
l_3	−0.1796	n_3	0.3234
l_4	−0.0500	n_4	0.6313
l_5	−0.1539	n_5	0.6139
l_6	0.0000	n_6	0.5000
l_7	−0.2137	n_7	0.5010
l_8	−1.0406		

experimental uncertainty in the detailed atomic positions in the complexes. This is crucial since the scoring function is completely dependent on interatomic distances. A small variation in hydrogen-bond distances and angles from structure to structure within the resolution of the X-ray data is essentially just noise. Such variation might have significant impact on the estimated parameters. By allowing the function to 'regularize' such geometries, one can derive a function with tighter maxima.

One last issue is that of conformational strain on the ligand and on the protein. In many of the complexes, the ligands (and less often the proteins) exhibited poor bond angles and/or dihedrals. This energy is hidden from the scoring function, and complexes with significant conformational strain will tend to have stronger computed affinities than the experimentally determined ones. With a ligand exhibiting some degree of flexibility, it may only be necessary to account for the ligand's strain, since both the ligand and protein will share strain (like two springs pushing against each other). The final function F was constructed to help account for this conformational strain. For each of the complexes, the ligand was removed from the binding site and minimized in vacuo before being returned to the binding site. In most cases, the minimized ligand structures deviated only slightly from the originals, but in some cases the conformational strain was significant. F and F_2 differ only in that the ligands used for F were preminimized.

The fit to the data for the final scoring function F was significantly better than for any of the intermediate functions. The mean error of computed affinity was 0.72 log units (0.85 rmse), with a near-perfect affinity ranking of the complexes (0.95 pair-ranking correlation coefficient [7]). The result is quite striking, since the minimization process applied to the ligand poses perturbed them in a manner independent of the complexes and of the scoring function. Figure 4 shows a plot of the computed versus experimental binding affinities for F_3. The parameter values for F are given in Table 4. The rms deviations of the final gradient optimized poses of the ligands compared with the original crystal structures were small (mean 0.70 Å, standard deviation 0.37).

Discussion

It is interesting to consider what the relative contributions of the individual terms are to the binding affinities. After computing all of the affinities using F, the absolute values of each of the terms were computed as a percentage of the total. The breakdown of scores for the 34 complexes across the different terms of F was: 44% hydrophobic, 26% polar, 25% entropic, and 5% solvation. Of the terms that are complex-specific (the entropy terms do not depend on the protein), 93% come from the hydrophobic and polar terms f_0 and f_1. Figure 5 shows a plot of these terms of F. Note that the sigmoidal portions of each function do not carry much weight. The reason is twofold. First, the complexes exhibit very little excess interpenetration since all of the ligands fit in the protein binding sites. Second, the steric overlap term used during pose optimization prevents such interpenetrations from occurring during the training process. This is problematic from the perspective of screening, since the majority of ligands need to be rejected based on not fitting into the target site. Ideally, the computed score would have penetration effects appropriately weighted. In practice, a threshold is employed that allows for only a small amount of interpenetration to account for some protein mobility.

The peak for the hydrophobic component is centered at a slight interpenetration of van der Waals surfaces (the van der Waals radii used are slightly larger than those deduced from crystallographic studies [5]). A single ideal hydrophobic contact yields 0.09 log units of binding affinity. While this may seem insignificant, since the preponderance of contacts are of this variety, this term is the most significant. There are particular motifs that yield significant hydrophobic scores using small numbers of atoms. For example, a single nonpolar proton packing against the face of a six-membered ring generates more than 0.5 log units of affinity. This will be illustrated further in the streptavidin–biotin interaction, where key tryptophan side chains offer ideal surfaces for hydrophobic packing.

The peak for the polar component is centered at an interpenetration corresponding to an H···O hydrogen-bond distance of 2.0 Å. The angular displacement from the H···O vector at which the polar strength drops by half is 52°. An ideal neutral hydrogen bond yields 1.2 log units of binding affinity. The effect of charge ($n_6 = 0.5$) is best illustrated by an example. An amidino proton (formal charge +0.25) interacting with a carboxylate oxygen (formal charge −0.5) yields 1.7 log units of affinity versus 1.2 for a neutral hydrogen bond. The repulsive polar term receives little weight for essentially the same reason as the lack of weight on the interpenetration terms.

The solvation terms are not particularly significant overall, but make a difference of up to 2 log units in some cases (e.g. complex 12, thermolysin/CLT). Despite

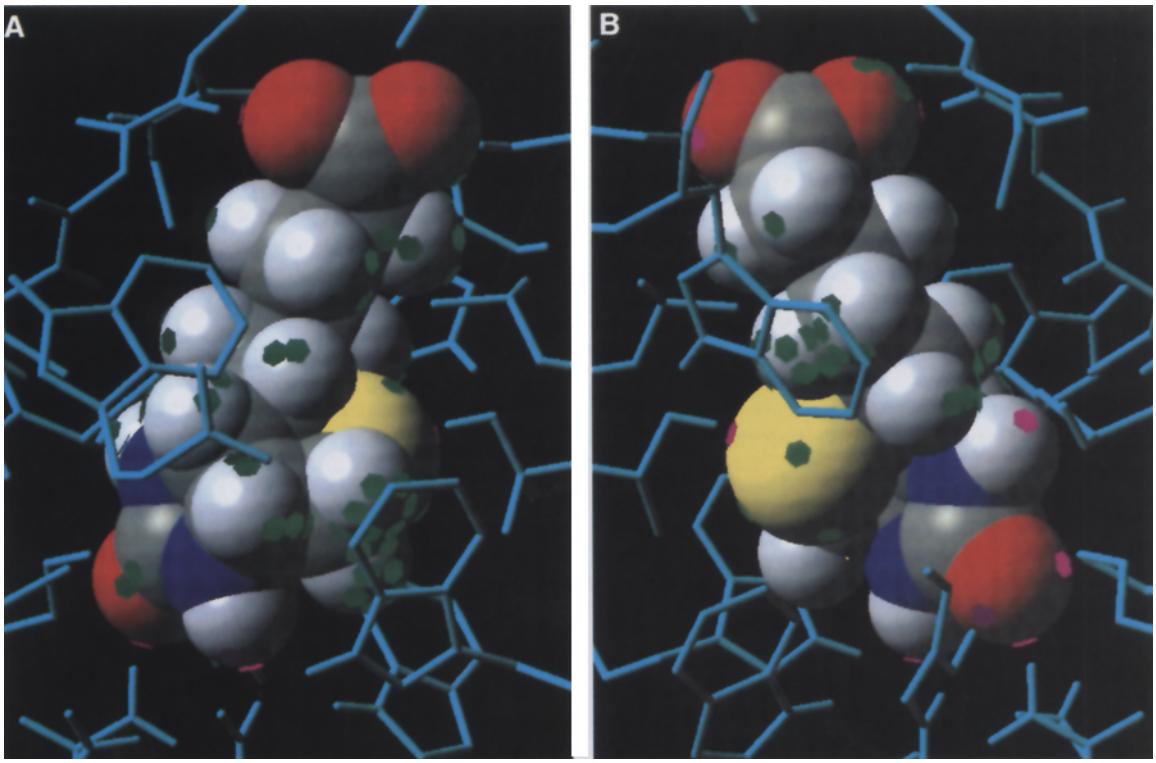

Fig. 8. Interactions of biotin bound to streptavidin (two views). Purple patches indicate favorable polar interactions; green patches indicate favorable hydrophobic interactions.

being defined symmetrically, only the protein solvation component received significant weight. The reason for this appears to be that the dynamic range of the protein solvation effect was quite large, whereas that of the ligand

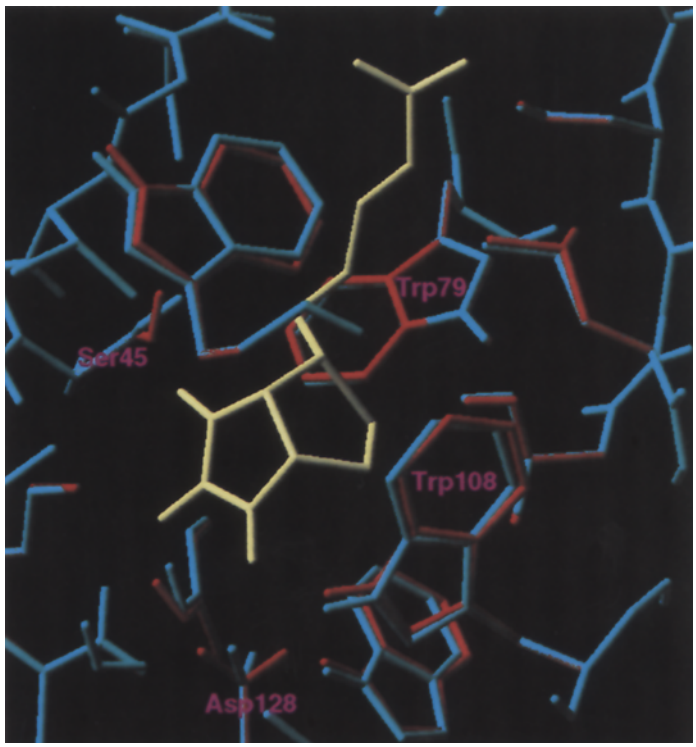

Fig. 9. Variation of protein side chains using F (two views). The original structure of streptavidin is in cyan; the modified structure is in red; biotin is shown in yellow.

solvation effect was much smaller owing to the fact that the ligands had fewer polar moieties. To force symmetry in the solvation computation, l_5 and l_6 could be made into a single parameter, but this would not make a significant difference in the function's performance.

Several factors known to influence binding are not treated explicitly by the function. Examples include long-range polar effects and polar effects of aromatic rings. While terms for each could be added to the function, the data are well explained using the current set. This is due in part to the limited size of the data set. As additional data become available, and as examples of poor computed affinity arise, the addition of such terms may be necessitated.

Comparison to Böhm's function

The function is very similar to that derived by Böhm. Since the term for loss of translational/rotational entropy is different for the two functions, the dynamic range of the other linear parameters is slightly different. A direct comparison is easiest for a particular example, and the trypsin–benzamidine case can be analyzed since the raw data are available for Böhm's function. Böhm's computed affinity for this case is 5.49, broken down as follows: 3.52 hydrophobic, 3.16 polar, –0.25 rotatable bond entropy, and –0.95 constant rigid body fixation entropy. For F, the computed affinity is 5.69, broken down as follows: 5.21 hydrophobic, 3.50 polar, –0.64 protein solvation, –0.21 rotatable bond entropy, and –2.17 constant rigid body fixation entropy. To compare the numbers directly, it is helpful to rescale the values so that both use the same constant: 4.38, 2.94, –0.54, –0.18, and –0.95. One last change is to combine F's hydrophobic and protein solvation terms to match Böhm's single lipophilic term: 3.84 (3.52) hydrophobic, 2.94 (3.16) polar, –0.18 (–0.25) entropy, and –0.95 (–0.95) fixation. The breakdown from both functions is very close. The match is striking considering that the underlying functional approaches are quite different.

Böhm [8] discusses the relationship of these derived values with experimental estimates at some length, and the discussion is not repeated here for those values on which the functions agree. One significant departure in F from Böhm's function is that the fixation cost due to loss of translational and rotational entropy is higher. The cost for F ranges from –1.95 to –2.9 (11.2–16.5 kJ) depending on log(molecular weight) versus Böhm's constant –0.95 (5.4 kJ). The values for F are closer to the estimate by Searle and Williams [13] of 'anywhere in the range 9 to 45 kJ/mol'.

Internal validation tests

The tight fit to the calibration data is not terribly surprising given the expressiveness of the nonlinear functions being used. However, it is important to note that the number of tunable parameters is significantly exceeded by the number of complexes in the calibration set. Further, the 13 parameters defining the hydrophobic, polar, and entropic terms account for 95% of the total scores. Also, pose variation increases the effective size of the data set, since the function is constrained to fit the maximum of many possible poses for each ligand.

To further validate the function's accuracy, a leave-one-out cross-validation of F was carried out. Thirty-four models were constructed using the same conditions as for F, each time leaving one complex out of the calibration set. The cross-validation yielded a mean error predicted affinity of 1.00 log units (1.26 rmse), a cross-validated r^2 of 0.79, and a PRCC of 0.91. This is, of course, somewhat less accurate than the *fit* to the data, but the pattern of false positives and false negatives in the cross-validation experiment mirrors that in the calibration experiment. This suggests that the function parameters are overdetermined by the data. The most significant deviations in the cross-validation experiment correspond to the most significant deviations in the fitting experiments from above, but the magnitudes of the deviations are greater. Formal experiments were not conducted on a 'blind set' due to the lack of a sufficient number of appropriate complexes. Experiments using F in flexible docking to several protein sites are ongoing and will provide definitive data regarding its usefulness and predictive accuracy.

Insensitivity to precise pose

The utility of F in molecular docking depends on how well it is able to recover from inaccurate initial poses generated by a geometric search engine. To assess this property, the minimized pose of biotin was varied randomly by perturbing both its torsion angles and the alignment parameters. Beginning from the perturbed poses, gradient descent was applied in order to maximize the value of the scoring function. Figure 6 shows a plot of rms deviation from the maximal pose of biotin versus score for all of the poses considered during the multiple gradient descents. Two features are important. First, the width of the maximum in the function with respect to rms is narrow. As mentioned earlier, narrow maxima in pose space should reduce the false positive rate observed in screening. Within 0.2 Å of the maximal pose exists a plateau of very close scores, but this drops off very rapidly. At 0.5 Å rms, the score has dropped off about 3 logs, and at 1.0 rms, the score has dropped off about 8 logs. Second, the gradient of the function with respect to pose is well enough behaved that all of the perturbed poses were pulled into an acceptable final pose (near maximal score and small rms). Figure 7 shows a superposition of the maximal pose of biotin in grey (score 12.5) with the crystal

structure pose in yellow (score 9.7). The poses are essentially indistinguishable, with an rms difference of 0.45 Å.

Speed

Speed is of paramount importance in a function intended to be evaluated repeatedly in the inner loop of docking tens of thousands of molecules. One of the more time-consuming parts of the computation is identifying the protein–ligand atom pairs that are involved in the interaction. Rather than computing all pairwise distances, a 3D grid of nearby protein atoms is precomputed. Given the position of a ligand atom, it is possible to do a constant-time lookup in the grid for all protein atoms that might be near enough to affect the function. With this simple optimization, the average time to compute the affinity of benzamidine to trypsin is 0.03 s on an SGI R4400 150 MHz machine. This is fast enough that evaluation of the scoring function is not the bottleneck in the flexible molecular docking system that utilizes F [11].

Training set bias

Bias from the complexes used in the calibration set is a source of concern. A relatively minor concern is overrepresentation of certain types of complexes (e.g. proteases). The mean error of predicted affinity for proteins represented by one exemplar was 1.0 log unit, just as with complexes with multiple exemplars. Variability in the ligands for each protein appears to be sufficient that each of the complexes is quite different. A more significant concern is that all the data are derived essentially from *positive* examples. All of the ligands are known to bind the proteins. Further, they are all sufficiently well behaved that they can be crystallized. The lack of negative data manifests itself in small weights applied to repulsive interactions such as steric overlap. It is possible to set conservative thresholds on allowable overlap in order to avoid false positives from this effect. However, in the case of solvation effects, which may well be quite a bit more important than they are weighted, there is no simple way to eliminate false positives.

One may be tempted to take negative results from screening and use them as negative data by generating putative poses for nonbinders and assigning very low scores to them. However, a piece of data saying that a ligand does not bind to a protein better than some threshold may mean many things. The desired interpretation is that the ligand does not bind to the identified binding site. Another explanation is that it binds to another site better. So, the ligand may fit into the docked site, but the assay would not detect it. There are other explanations relating to false negatives from assays that may also explain negative binding results. A relatively safe source of 'negative' data is molecules that are predicted to bind better than they actually do. Following a prediction that a molecule binds at 1 μM in a particular pose with an assay result of 100 μM, it is safe to retune the function to predict a *maximum* of 4.0 ($-\log(K_d)$) in *any* pose.

Streptavidin–biotin complex

The streptavidin–biotin complex is a well-studied system, owing to its remarkable binding affinity ($K_d = 10^{-13.4}$ M) [14]. Figure 8 shows the maximal pose of biotin bound to streptavidin, with the hydrophobic and polar interactions painted on the surface of biotin. The purple patches indicate polar interactions, and the green patches indicate hydrophobic interactions. Biotin makes eight strong specific polar interactions with the protein: three from the ureido oxygen and one from each of the remaining polar atoms. The pattern of hydrophobic packing is worth noting. The methylene marked on the lower right on panel A makes 13 contacts with the indole of Trp^{108}. The methylene in the middle of panel B makes 10 contacts with the indole of Trp^{79}. The breakdown of the score (12.5) is: hydrophobic 7.61, polar 8.99, protein solvation −0.53, rotatable bond entropy −1.07, and rotational/translational entropy −2.48.

It is difficult to relate these values to other studies of the complex. Thermodynamic studies such as that by Weber et al. have partitioned the free energy into ΔH and ΔS terms [14]. However, the key hydrophobic and polar terms of F each capture both enthalpic and entropic effects. The hydrophobic term lumps together the enthalpic dispersion–attraction forces, the entropy of freeing waters, and the enthalpy of waters making improved hydrogen bonds outside the largely hydrophobic protein cavity. The polar term lumps together the Coulombic attraction and the mixed effects of displacing fixed waters. Further complicating matters, the experimental ΔH and ΔS may be significantly influenced by variation in the protein that is unaccounted for here. For example, a flexible loop of streptavidin becomes fixed on binding biotin [14].

However, whereas previous work on this complex has attempted to explain the high affinity in part by assuming partial polarization of the ureido head of biotin [8,14], the computations reported here do not make any special assumptions about that moiety. The polar atoms on the head are treated as neutral hydrogen bonders. This is another point where this work diverges from that of Böhm. Despite assuming partial polarization, the computed biotin–streptavidin affinity was still 2.7 log units low. The difference may lie in the precise computation of the hydrophobic component of binding affinity. Böhm's approach linearly relates lipophilic surface contact area to binding affinity. The case of the methylenes interacting with Trp^{108} and Trp^{79} tests the implicit assumption that every $Å^2$ of lipophilic contact is equal to every other. The density of atom-pair contacts in these areas is very high.

Indeed, the density of nonpolar atom-pair contacts throughout the binding interface is quite high. If contact density has a significant effect on binding energy per $Å^2$, the streptavidin/biotin complex presents a difficult case for functions assuming a linear relationship between surface contact area and binding affinity.

Miyamoto and Kollman [15] have done extensive free-energy perturbation (FEP) calculations on the streptavidin–biotin complex. They compute the free energy of association by mutating biotin into a 'dummy molecule' in two environments: inside streptavidin and in solvent. The difference between these energy deltas is the energy of association. While the computed polar interaction energies are quite large, they observe that biotin makes similar interactions in solvent and conclude that the hydrophobic component dominates the interaction. They suggest that Trp^{108} and Trp^{79} may be particularly important. The importance of the hydrophobic component of the interaction is in agreement with this work in explaining the very high affinity, but there is significant disagreement on the net importance of the polar interactions. In the FEP computation, the energetic contributions of the polar interactions present in the complex are largely canceled out by the solvation effects.

While it seems likely that F somewhat underestimates solvation effects, it is hard to reconcile Miyamoto and Kollman's FEP computation with the microcalorimetry data [14]. Weber et al. [14] showed that the free energy of binding of biotin to streptavidin is strongly enthalpically driven ($\Delta H = -32.0$) and is entropically disfavored ($T\Delta S = -13.7$). Miyamoto and Kollman explain that the sign and magnitude of enthalpy and entropy are poor predictors of the hydrophobic effect. The computation presented here reconciles the binding affinity of the complex somewhere between that of Weber et al., with its emphasis on the polar component, and that of Miyamoto and Kollman, with its emphasis on the hydrophobic component.

The case of HABA binding to streptavidin provides another example. The binding affinity is 10^{-4} M, being driven by entropic effects [14]. The computed affinity for HABA using F from the native structure (PDB ref. 1SRE) was $10^{-5.9}$ M. The hydrophobic term dominated the polar term (7.8 and 1.7 log units respectively). This is in qualitative agreement with the calorimetry data. The substantial decrease in polar contribution relative to biotin simultaneously decreases the enthalpy of binding and the fixation of both the protein and the ligand. The crystallographic observation that the flexible loop of streptavidin in the complex is not completely immobilized supports this interpretation [14].

Effects of protein conformation

Nearly all of the complexes used in the calibration were taken from crystal structures in which the ligand was present. Thus, the conformation of the protein was optimized for the particular ligand in question. Clearly, this situation does not mirror the docking problem where the protein is held rigid. Thousands of putative ligands are to be docked into a fixed protein binding site. To help assess the possible magnitude of the error arising from this effect on a small molecule with good affinity, biotin was docked into an uncomplexed structure of streptavidin [16]. The computed affinity was 9.9 (following ligand pose optimization), which is 3.5 log units lower than the correct affinity and 2.6 log units lower than that computed from the co-crystal structure. This may be more indicative of the degree of error to be expected in a true docking application than the result from the cross-validation. Fortunately, it seems likely that most such errors should be on the negative side (low computed affinity), assuming a reasonably low-energy conformation of the protein. This should tend to increase the false negative rate, but as was discussed earlier, this is not as serious as adversely affecting the false positive rate. In this particular case, one would still find biotin as a true positive hit beginning from the uncomplexed structure, since it scored in the subnanomolar range.

Correcting for this effect involves explicit protein conformation variation. This was done by varying the conformations of the protein's side chains using F while holding the remainder of the protein as well as the ligand constant. In this experiment, the side chains could not distinguish between protein and ligand, so it was possible that the resulting protein conformation would be optimized for intraprotein interactions at the expense of protein–ligand interactions. This would have resulted in a computed decrease in protein/ligand binding affinity relative to the native uncomplexed structure. The experiment was done iteratively, beginning from an approximate docking of biotin to the uncomplexed streptavidin structure (generated by superposition of the native structure with the minimized biotin used in training). First the pose of biotin was adjusted, then each side chain in contact with biotin was optimized by varying only the side-chain torsions, and the process was iterated until convergence. The protein side chains were able to increase their 'affinity' to the combination of protein and ligand by 4.0 log units. Most of the gains that the side chains made came from intraprotein contacts. The final computed affinity for biotin to the modified streptavidin was 11.4, an increase of 1.5 log units over the original uncomplexed structure (and just 1 log unit from the native structure's computed affinity).

Figure 9 shows the differences in the conformation of the modified structure to the original structure. The side chains moved relatively little, with the largest atomic displacement being 0.76 Å (the overall side-chain rms was 0.27 Å). The largest movement corresponded to Asp^{128} (at the bottom), which reoriented to make two hydrogen

bonds with other protein atoms while maintaining a hydrogen bond with an N-H of biotin. Trp^{108} and Trp^{79} both reorient during the optimization. The most significant remaining difference between the interactions made here and those in the native structure is the hydrogen bond from the other N-H of biotin to Ser^{45}. The distance in the former case is 2.34 Å, but in the native structure it is 2.06 Å. Side-chain motion alone improves the distance from 2.56 Å after the initial biotin pose optimization in the original structure, but it cannot bring the interaction into the optimal distance. Additional work is required in order to speed the side-chain variation computation enough to be practical in docking searches. Main-chain protein variation requires substantially more work.

Screening

The only way to fully assess the utility of F is to screen large databases for novel small-molecule ligands using F as the objective function. These experiments are underway utilizing F in conjunction with a method for automatic binding-site identification and characterization and a flexible search engine [10,11]. Preliminary results on the streptavidin system are encouraging. In particular, of 80 000 compounds docked into the streptavidin binding site, F predicts that biotin will be the highest affinity ligand by 2 log units. This is an encouraging result with respect to the issue of false positives. Assays of predicted tight-binding ligands are underway and will be reported in a subsequent paper. Other systems being screened include cytokine receptors and serine proteases.

Conclusions

The three critical requirements on a scoring function for a molecular docking system are accuracy, speed, and tolerance to inaccurate poses of putative ligands in protein binding sites. The function F defined here satisfies these requirements. The expected mean error of predicted affinity, estimated by cross-validation across a diverse set of binding sites and ligands, is 1.0 log unit. Using a simple optimization to speed up the identification of protein atoms near the ligand, the time to compute the affinity of benzamidine to trypsin is 0.03 s, which is fast enough for use in a docking search engine. F is a continuous and differentiable function whose maxima correspond closely to crystallographically determined structures. So, imprecise putative ligand poses can be efficiently optimized by gradient descent.

In the database screening problem, it is important that the false positive rate of a scoring function be minimized, since of thousands of compounds only tens will actually bind to a particular target. The parameter estimation regime for F yielded a function with spatially narrow maxima. Divergence from the maximal pose of a ligand by 0.5 Å rms can lead to a decrease of 3 log units in computed affinity. The high spatial specificity of F is likely to decrease the false positive rate observed in screening compared with other approaches involving functions that were designed to be spatially coarse.

The function accounts for hydrophobic interactions, polar interactions (including directionality and charge), entropic effects, and solvation effects. Without adding any special-purpose terms, parameters, or local polarization effects, the breakdown of computed affinities into these components is in qualitative agreement with experimental observations of the enthalpy and entropy of binding of ligands to streptavidin. Pilot experiments using F to vary protein conformation hold promise for a more general treatment of flexibility in molecular docking.

Acknowledgements

I thank Will Welch, Jim Ruppert, and Teri Klein for their collaborative efforts and for comments on the manuscript, and Mike Ross and Mike Venuti for support.

References

1 Rutenber, E., Fauman, E.B., Keenan, R.J., Fong, S., Furth, P.S., Ortiz de Montellano, P.R., Meng, E., Kuntz, I.D., DeCamp, D.L., Salto, R., Rose, J.R., Craik, C.S. and Stroud, R.M., J. Biol. Chem., 268 (1993) 15343.
2 Shoicet, B.K., Stroud, R.M., Santi, D.V., Kuntz, I.D. and Perry, K.M., Science, 259 (1993) 1445.
3 Bodian, D.L., Yamasaki, R.B., Buswell, R.L., Stearns, J.F., White, J.M. and Kuntz, I.D., Biochemistry, 32 (1993) 2967.
4 Ring, C.S., Sun, E., McKerrow, J.H., Lee, G.K., Rosenthal, P.J., Kuntz, I.D. and Cohen, F.E., Proc. Natl. Acad. Sci. USA, 90 (1993) 3583.
5 Jain, A.N., Harris, N.L. and Park, J.Y., J. Med. Chem., 38 (1995) 1295.
6 Jain, A.N., Dietterich, T.G., Lathrop, R.L., Chapman, D., Critchlow, R.E., Bauer, B.E., Webster, T.A. and Lozano-Perez, T., J. Comput.-Aided Mol. Design, 8 (1994) 635.
7 Jain, A.N., Koile, K. and Chapman, D., J. Med. Chem., 37 (1994) 2315.
8 Böhm, H.-J., J. Comput.-Aided Mol. Design, 8 (1994) 243.
9 Bohacek, R.S. and McMartin, C., J. Med. Chem., 35 (1992) 1671.
10 Ruppert, J., Welch, W. and Jain, A.N., manuscript submitted for publication.
11 Welch, W., Ruppert, J. and Jain, A.N., Chem. Biol., 3 (1996) 449.
12 Murray-Rust, P. and Glusker, J.P., J. Am. Chem. Soc., 106 (1984) 1018.
13 Searle, M.S. and Williams, D.H., J. Am. Chem. Soc., 114 (1992) 10690.
14 Weber, P.C., Wendoloski, J.J., Pantoliano, M.W. and Salemme, F.R., J. Am. Chem. Soc., 114 (1992) 3197.
15 Miyamoto, S. and Kollman, P.A., Proteins, 16 (1993) 226.
16 Katz, B.A., Biochemistry, 34 (1995) 15421.