Journal of Computer-Aided Molecular Design **17:** 597–605, 2003.

A novel scoring function for molecular docking

A.E. Muryshev*, D.N. Tarasov, A.V. Butygin, O.Yu. Butygina, A.B. Aleksandrov & S.M. Nikitin
Algodign LLC, Bolshaya Sadovaya street 8, Moscow 123379, Russia

Received 2 June 2003; accepted in revised form 27 August 2003

Key words: binding affinity, scoring function, molecular docking, structure-based drug design

Summary

We present a novel scoring function for docking of small molecules to protein binding sites. The scoring function is based on a combination of two main approaches used in the field, the empirical and knowledge-based approaches. To calibrate the scoring function we used an iterative procedure in which a ligand's position and its score were determined self-consistently at each iteration. The scoring function demonstrated superiority in prediction of ligand positions in docking tests against the commonly used Dock, FlexX and Gold docking programs. It also demonstrated good accuracy of binding affinity prediction for the docked ligands.

Introduction

Enormous advances in genomics have moved drug design efforts to new heights. Decoded genome sequences open the door to a large number of new potential therapeutic targets. This growth in potential targets has increased the demand for powerful and reliable technologies that can identify high-quality lead drug candidates. However, lead discovery techniques are still unable to keep up with the new pace of target discovery.

In recent years, the main approaches to increasing productivity of lead discovery were experimental and included the development of combinatorial chemistry and high-throughput assays that provide automated screening of hundred of thousands or even millions of potential compounds. Despite these impressive technical advances, this methodology has failed to provide many new drugs [1].

As some assert, the pharmaceutical industry is now starting to go beyond the high-throughput screening era [2]. The new hopes are related to structure-based drug design, an approach in which new leads are constructed on the basis of the 3D structure of target proteins. Such 3D structures of proteins of therapeutic interest are becoming increasingly numerous, as new technology and industrial techniques are applied to X-ray crystallography. Having been introduced in the early 1990s, the structure-based design approach is now increasingly used [3]. The main computational trends in modern structure-based drug design are an increase of the number of molecules tested in *in silico* binding affinity assays and their integration into high-throughput experimental techniques (see, e.g., [4, 5]).

A very basic concept of computational structure-based drug design is that of a scoring function. A scoring function estimates the free energy of binding of a ligand as a function of its position in a protein binding site. The global minimum of the scoring function for a given protein and ligand should estimate the actual binding energy. A procedure for determining the ligand position that provides the global minimum of the scoring function is known as molecular docking [6, 7].

Since the pioneering work of Bohm [8], a variety of scoring functions have been developed (see [9] for a review). The accuracy of binding energy estimation still remains insufficient and leads to a substantial portion of false positives in the output of modern *in silico* lead generation techniques [10, 11]. At the same time,

*To whom correspondence should be addressed. E-mail: andrey.muryshev@algodign.com. Phone +7 095 2091419, fax +7 095 2091415.

the need to increase the number of compounds tested *in silico* against a given target demands higher speed for score evaluation. Thus, a continuous effort is being made to improve both accuracy and efficiency of scoring schemes [9, 11].

There are two main approaches to derive a scoring function for protein–ligand interaction: the empirical approach and the knowledge-based approach [9]. In the empirical approach, the binding energy of a ligand atom in a protein binding site is expressed as a function of its position via an expression containing a set of adjustable parameters. These parameters have to be fitted so that the constructed scoring function approximates the values of binding energies found in experiments. A problem with this method is that there are rather few protein–ligand complexes for which both the 3D structures and binding energies are known [12]. This makes it impossible to derive the spatial dependencies of protein–ligand interatomic interactions that are obviously important for proper positioning of a ligand.

In the knowledge-based approach, a scoring function is obtained from statistical counts derived from known 3D structures: the so-called Boltzmann hypothesis converts frequencies of finding atom A at a distance r from atom B into an effective interaction energy between A and B as a function of r (see, e.g., [13]). Since the corresponding amount of distance data is much larger than the binding energy data used for the empirical method, the knowledge-based approach provides smooth radial interatomic interaction potentials. However, there are three serious problems with this method, which are indicated below:

- the Boltzmann hypothesis has a very weak justification;
- The ideology of the method originates from the statistics of a spatially uniform liquid. Its adaptation to a two-component non-uniform medium such as a protein–ligand complex is a hard and non-trivial problem;
- Knowledge-based potentials are typically pairwise, while the probability to find atoms A and B at a distance r is non-pairwise, i.e., depends also on surrounding atoms C, D etc. because of strong multi-particle correlations in structures of interacting molecules.

We believe that the two approaches described above can complement each other. Thus, we have developed a novel method to derive a protein–ligand scoring function based on a combination of the two approaches. In our approach we start with the knowledge-based potentials as an initial approximation to protein–ligand interaction potentials. We then correct the potentials by introducing adjustable parameters calibrated with a set of protein–ligand complexes with known binding energies and 3D structures.

Our calibration has an important difference from the one typically used in the empirical approach. Usually scores calculated for the ligand positions as they were determined by X-ray crystallography are fit to experimental values of binding energies. Here, we first optimize the ligand positions with our scoring function and then fit the scores to experiment. We re-optimize ligand positions iteratively until the ligand positions and calibrated parameters have finally converged. Our work demonstrates the importance of this self-consistent procedure for accuracy of the derived scoring function.

We present tests of our scoring function applied to docking of rigid ligands to protein active sites and demonstrate both good accuracy and high speed of docking.

Methods and results

In our scoring function we distinguish the following seven atom types:

- Carbon
- Oxygen in OH group (alcohol oxygen)
- Oxygen in C-O-C group (ether oxygen)
- Oxygen in C=O-OH group (carboxyl oxygen)
- Oxygen in C=O group (sp^2-oxygen)
- Nitrogen in aromatic rings
- Nitrogen hydrogen bond donor (other than aromatic)

In this initial study we focus on protein active sites containing no metal atoms and exclude metal atoms from the scoring function. Also, we accelerate score evaluation by not considering hydrogen atoms explicitly. Any other atom type (such as sulfur, phosphorus, and types of nitrogen not listed above) is treated as a carbon atom.

A general form of our scoring function is

$$\Delta G = \sum_{i,j} F_{A,B}(r_{i,j}), \tag{1}$$

where i and j enumerate ligand and protein atoms in a given protein–ligand complex, A and B denote the types of atoms i and j, correspondingly, and $r_{i,j}$ is the distance between atoms i and j. Here ΔG is the binding free energy of a protein–ligand complex and

$F_{A,B}(r_{i,j})$ approximates the free energy of interaction between atoms i and j.

We constructed our scoring function in three steps. At the first step for the functions $F_{A,B}(r)$ we took knowledge-based potentials. A method of derivation of these potentials is well described in the literature (see, e.g., [13]). The basic assumption in this method is known as the Boltzmann hypothesis. Informally speaking, it states that the probability $P_{A,B}(r)$ to find a ligand atom A at the distance r from the protein atom B is proportional to

$$P_{A,B}(r) \propto \exp\left(\frac{-F_{A,B}(r)}{T}\right),$$

where T is some effective temperature usually supposed to be $T = 300$ K. From this we arrive at the formula used to derive knowledge-based potentials

$$F_{A,B}(r) = -T\log(P_{A,B}(r)) + C.$$

Here the coefficient C is determined by the condition that $F_{A,B}(r) \to 0$ at $r \to \infty$. In practice this means that $F_{A,B}(r) = 0$ for $r \geq R$ where R is sufficiently large (in our calculations we used $R = 7$ Å). The probability $P_{A,B}(r)$ is found from statistics on protein–ligand structures available from the Protein Data Bank (PDB) with the formula

$$P_{A,B}(r) = \frac{n_{A,B}(r)}{r^2}.$$

Here $n_{A,B}(r) = \Delta N_{A,B}(r)/\Delta r$ and $\Delta N_{A,B}$ is the number of pairs of protein atoms of type A and ligand atoms of type B separated by a distance $d \in [r, r+\Delta r]$ which was observed in the PDB. We counted $\Delta N_{A,B}$ for $\Delta r = 0.1$ Å in more than two thousand PDB files with resolution finer than 2 Å and no metal atoms in the protein active sites.

In Figures 1 and 2 we present typical knowledge-based potentials we obtained (solid lines).

At the second step we approximated the derived knowledge-based potentials for each pair of atom types A, B with the following simple smooth functions:

$$F(r) = \begin{cases} -\epsilon + \kappa(r-\rho)^2; & r < \rho \\ -\epsilon + \kappa(r-\rho)^2 - 2\kappa^{3/2}(r-\rho)^3/3\sqrt{3\epsilon}; & \rho < r < \rho + \sqrt{3\epsilon/\kappa} \\ 0; & \rho + \sqrt{3\epsilon/\kappa} < r \end{cases} \quad (2)$$

with values of adjustable parameters $\epsilon_{A,B}$, $\rho_{A,B}$, and $\kappa_{A,B}$ depending on the types A, B of atoms i, j. The values of the parameters were determined by a multivariate regression fitting the function 2 to the corresponding knowledge-based potentials. In this way, our scoring function took a simple analytical form approximating knowledge-based potentials (see Figures 1 and 2, dashed lines).

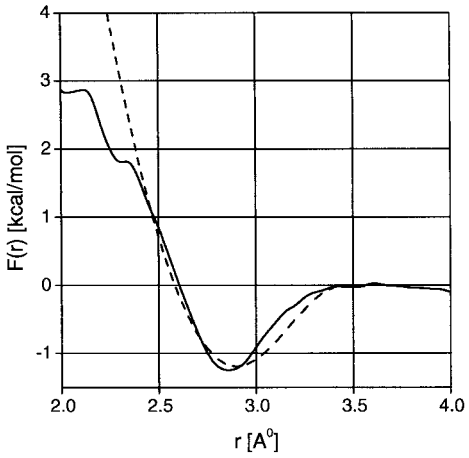

Figure 1. Knowledge-based potential for interaction between sp^2-oxygen and nitrogen-donor (solid) and the function (2) fitted to it (dashed).

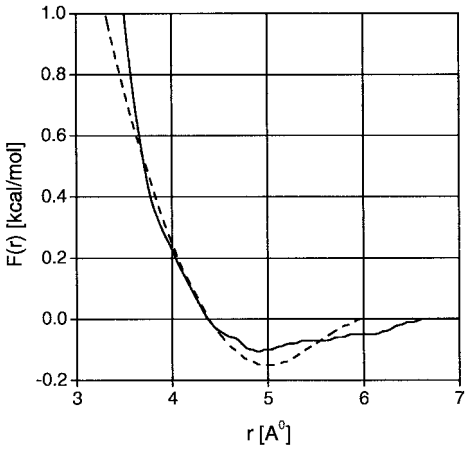

Figure 2. Knowledge-based potential for interaction between two carbon atoms (solid) and the function (2) fitted to it (dashed).

At the third step of construction of our scoring function we corrected some of the adjustable parameters introduced at the previous step by calibrating them against known binding energies. We proceeded with the calibration along the following lines. First we separated all the potentials into three groups: hydrophobic interaction, hydrogen bonds, and pure van der Waals repulsion. In our atom type classification, the only hydrophobic interaction is that between two carbon atoms. To the hydrogen bonds we attributed potentials with a clear minimum at the interatomic distance of 2.5–3 Å (as shown in Figure 1). All other potentials

were treated as pure repulsion. Next we constrained the parameters of the scoring function by the following conditions that are to be kept during the calibration procedure:

- For pure van der Waals repulsion group we put $\epsilon_{A,B} = 0$ and $\rho_{A,B} = \rho_{vdw}$. For the initial value of ρ_{vdw} we took the average of values of $\rho_{A,B}$ found for potentials of this group at the previous step of the scoring function construction;
- For hydrogen bonds we put $\epsilon_{A,B} = \epsilon_{hb}$. For the initial value of ϵ_{hb} we took the average of $\epsilon_{A,B}$ previously found for all hydrogen bonds;
- The parameters $\kappa_{A,B}$ for all interactions and $\rho_{A,B}$ for hydrogen bonds we fixed to the previously found values.

This leaves just four adjustable parameters which should be calibrated against known binding energies:

- Energy of hydrophobic interaction ϵ_{cc};
- Radius of hydrophobic interaction ρ_{cc};
- Energy of hydrogen binding ϵ_{hb};
- Radius of pure van der Waals repulsion ρ_{vdw}.

The four adjustable parameters were calibrated using multivariate regression to fit values of binding energy predicted with our scoring function to experimentally known values for a training set of 164 protein–ligand complexes with known PDB structures. All the ligands in the training set were flexible while the rigid ligands were put into the testing set (reasons for this are explained below). In addition, our training set contained only complexes with small ligands and no metal atoms in the protein active sites.

To calibrate the adjustable parameters we used a self-consistent procedure in which we first optimized the ligand positions with our scoring function and then adjusted the parameters to fit the obtained scores to experiment. We re-optimized ligand positions iteratively until the ligand positions and calibrated parameters finally converged. In Figure 3 we present predicted versus known values of binding energy for our training set. The mean-squares deviation for the binding energy was 2.1 kcal/mol. The average of mean-squares of shifts of ligand atoms from their PDB positions was 0.49 Å.

To test the accuracy of our scoring function we have performed docking of a set of ligands for which positions in the corresponding active sites are known from the PDB. We used for docking an in-house developed docking program AlgoDock, which will be described in detail elsewhere. For flexible ligands the docking accuracy depends on the algorithm used to introduce flexibility, so we decided to choose only rigid

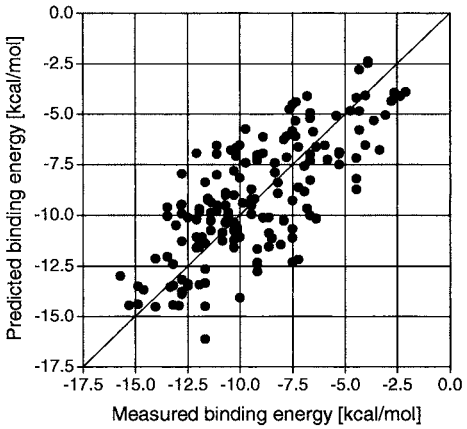

Figure 3. Predicted versus measured values of binding energy for the training set. Predicted values are calculated for optimized ligand positions.

ligands for the docking tests. We treated as rigid the ligands which have no single-valence bond other than in rings or between two sp^2 hybridized atoms.

For docking tests we used 36 protein–ligand complexes, none of which was included in the training set. Before docking we optimized the 3D structures of the ligands in vacuum with Amber94. At the same time we left 3D structures of the protein active sites as they are given in the PDB, making the only change by removal of water molecules.

To make docking faster we used a grid approach for the score evaluation. In each active site we built a grid of size 20 × 20 × 20 Å with a spatial step of 0.125 Å. The grid was centered at the center of mass of a studied ligand in its native PDB position. At each point of the grid, the scores for individual atoms placed at this point were calculated for all atom types. Then at each step of calculation the score for each atom was set equal to the pre-calculated value at the nearest grid point.

To test the accuracy of our scoring functions against scoring functions implemented in well known docking programs we collected all rigid ligands contained in the testing set of the FlexX docking program [14] (19 ligands in total). For most of these ligands docking results are also available for the Gold and Dock docking programs [15–17]. We docked these ligands with AlgoDock. The results of docking are presented in Table 1 and Figure 4. For the docked ligands, AlgoDock demonstrated both higher accuracy and higher speed of docking relative to the above-mentioned programs. Figure 4 shows that FlexX and Gold produce less accurate RMS deviation values than

Table 1. Results of docking with AlgoDock against those of FlexX, Gold, and Dock. Data for FlexX and Gold were taken from their web pages [14, 15] and data for Dock are from [17]. Root-mean-square deviations are given in Angstroms and times are in seconds.

Code		RMSD				Time		
PDB	Ligand	AlgoDock	FlexX	Gold	Dock	AlgoDock	FlexX	Gold
1abe	ARA	0.31	1.16	0.86	0.35	0.12	31.98	464
1acj	THA	0.35	0.49	4.00	0.33	0.15	7.56	468
1aha	ADE	0.39	0.56	0.51	0.36	0.05	1.39	320
1cbs	REA	0.73	1.68			0.07	29.69	
1coy	AND	1.14	1.06	0.86	0.65	0.11	3.44	766
1dbj	AE2	0.31	1.22	0.72	3.22	0.21	16.32	619
1dbk	ANO	0.64	0.76			0.18	1.67	
1dwb	BEN	0.56	0.54			0.05	4.91	
1epb	REA	0.52	2.77	2.08	1.70	0.07	37.06	763
1fen	AZE	0.82	1.39			0.17	7.63	
1ldm	OXM	0.51	0.74	1.00	2.55	0.03	3.27	256
1pbd	PAB	0.36	0.33	0.57	0.44	0.05	5.96	339
1ulb	GUN	0.62	3.37	0.32	0.43	0.08	8.71	374
2cpp	CAM	0.45	2.94			0.31	2.98	
2phh	PHB	0.56	0.43	0.72	3.50	0.01	8.01	331
3aah	PQQ	0.37	5.93	0.42	0.29	0.23	23.55	710
3ptb	BEN	0.31	0.55	0.96	0.37	0.05	5.12	308
4fab	FDS	4.76	4.95	5.69	1.43	0.04	145.73	579
6abp	ARA	0.29	1.12	1.08	0.42	0.11	21.07	443
Average		0.74	1.68	1.41	1.15	0.11	19.27	481

AlgoDock for most ligands, while Dock is as accurate as AlgoDock for many ligands, but produces much less accurate results for several of them. In particular, FlexX, Gold and Dock each have several ligands for which the RMS deviation is substantially greater than 1.0, while Algodock has only one such ligand in this set. Comparing the docking times one should keep in mind that AlgoDock was tested on Pentium IV 1.7GHz, FlexX was tested on UltraSparc III 750 MHz [14] (≈two times slower), and Gold was tested on SGI R4400 200 MHz [18] (≈10 times slower). Average mean-squares shift for the docked ligands from their PDB positions was 0.74 Å for AlgoDock, 1.68 Å for FlexX, 1.41 Å for Gold, and 1.15 Å for Dock. The average time to dock a ligand was 0.11 s for AlgoDock (note that the time to fill the grid, which only needs to be performed once, was not included in the docking time).

A standard test for the accuracy of a scoring function is a comparison of its binding energy predictions to the experimental values for protein–ligand complexes with known PDB structures. Typically, in such a test the experimentally observed positions

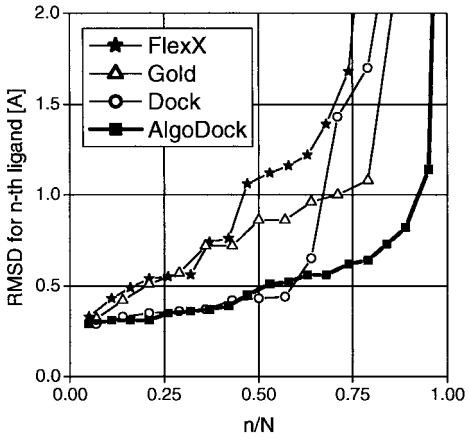

Figure 4. Atomic root-mean-square deviations from the X-ray conformations of the docked ligands in Table 1. The X-axis gives the rank number of the ligand divided by the total number of docked ligands. Ligands are ordered separately on the X-axis for each program according to increasing RMSD for that program. The RMSD is indexed against n/N rather than n because the total number of available ligands N differs for the various programs.

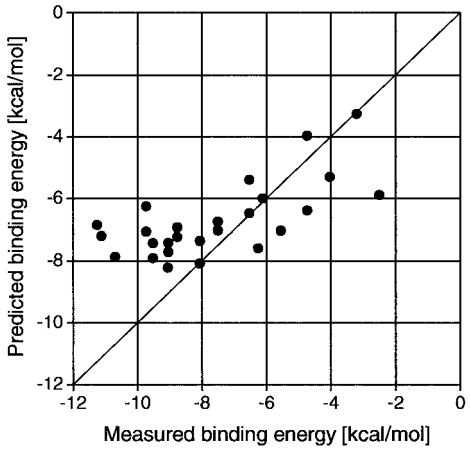

Figure 5. Predicted values of binding energy are shown with the measured values for the ligands from Table 2. The predicted values are calculated for positions of ligands determined by docking.

are used for scoring. However, for docked ligands their positions may significantly differ from the native PDB positions. Instead of scoring native positions, we performed docking of rigid ligands for which both the PDB structure and binding energy is known (25 protein–ligand complexes in total) and compared scores for the docked positions to measured binding energies. These ligands were not a part of the training set used to calibrate our scoring function. The results of docking are presented in Table 1 and Figure 5. For the docked ligands, the RMS deviation for binding energy prediction was 2.0 kcal/mol and the average RMS deviation for prediction of ligand positions was 0.75 Å.

Table 2. Binding energies predicted with our scoring function for the docked ligands. Root-mean-square deviations are given in Angstroms, times are in seconds, and energies are in kilocalories per mole.

PDB code	Ligand code	AlgoDock RMSD	Predicted energy	Measured energy	Energy error
1abe	ARA	0.31	−6.26	−9.73	3.47
1abe	ARB	0.39	−7.07	−9.73	2.66
1abf	FCA	0.31	−7.03	−7.51	0.48
1abf	FCB	0.22	−6.75	−7.51	0.76
1apb	FCA	0.28	−7.37	−8.06	0.69
1apb	FCB	0.50	−8.09	−8.06	−0.03
1bap	ARA	0.42	−7.92	−9.52	1.60
1bap	ARB	0.19	−7.44	−9.52	2.08
1bit	BEN	0.69	−7.04	−5.56	−1.48
1bra	BEN	1.74	−5.89	−2.50	−3.39
1dbj	AE2	0.31	−7.88	−10.70	2.82
1dbk	ANO	0.64	−6.86	−11.26	4.40
1dwb	BEN	0.56	−5.30	−4.03	−1.27
1l83	BNZ	0.69	−3.97	−4.73	0.76
1ulb	GUN	0.62	−6.00	−6.12	0.12
2dri	RIP	0.26	−8.22	−9.06	0.84
2phh	PHB	0.56	−6.39	−4.73	−1.66
3ptb	BEN	0.31	−7.60	−6.26	1.34
4fab	FDS	4.76	−7.21	−11.12	3.91
5tim	DTT	3.89	−3.26	−3.20	−0.06
6abp	ARA	0.29	−6.93	−8.76	1.83
6abp	ARB	0.28	−7.24	−8.76	1.52
7abp	FCA	0.27	−7.43	−9.04	1.61
7abp	FCB	0.28	−7.72	−9.04	1.32
9ldt	OXM	0.31	−5.40	−6.53	1.13
Average		0.76			
RMS					2.02

Discussion

Smooth radial profiles

The basic goal of our study was to construct a scoring function for docking of small molecules to protein active sites, i.e., a function able to distinguish the ligand's true binding position from all other possible positions. Such a function should approximate the free energy of protein–ligand interaction since physically it is this quantity that distinguishes the true binding position from others. At the same time the scoring function should properly describe radial dependencies of protein–ligand atom-atom interactions, which are obviously important for proper positioning of a ligand.

There are two basic types of scoring functions used for docking of small molecules to protein active sites [9]. Empirical scoring functions express binding free energy via a set of adjustable parameters which are calibrated against experimental data on 3D structures and binding affinities of protein–ligand complexes. Unfortunately, publicly available data of ligand–protein complexes allow reliable calibration of about 5 parameters for a scoring function [12]. This is too few to derive distance dependencies of atom-atom interaction potentials. Therefore these potentials are usually described by a step function, which is the simplest possible form.

Knowledge-based potentials form another class of scoring functions describing protein–ligand interactions. Knowledge-based potentials are constructed with the aid of statistics on 3D structures of protein–

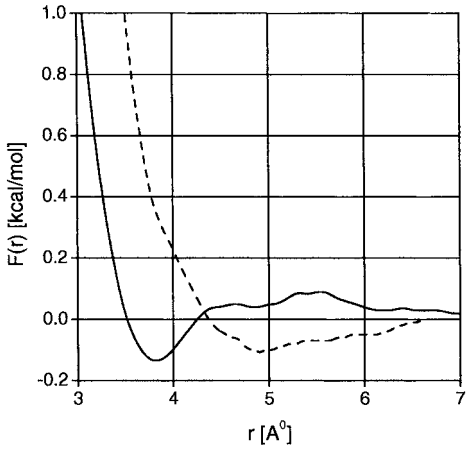

Figure 6. Knowledge-based potentials for interaction between sp^2-oxygen and carbon (solid) and between two carbon atoms (dashed).

ligand complexes. They typically have smooth radial dependencies (see, e.g., Figures **??** and **??**). However, they have a set of serious drawbacks mentioned in the Introduction. The most important one comes from multi-particle correlations which are typically disregarded but can be very strong. We illustrate this point in Figure 6, where we present a knowledge-based potential we obtained for the interaction of sp^2-oxygen and carbon atoms in comparison with a carbon-carbon interaction. Molecules or fragments consisting mainly of carbon atoms in water experience an effective attraction known as hydrophobic interaction – an effect analogous to surface tension. Hydrophobic interaction is a collective phenomenon but it is often approximated by some effective pair-wise attraction between carbon atoms. In Figure 6, the effective hydrophobic attraction is seen as an energy minimum at the dashed curve. On the other hand, there is no physical reason for an effective attraction between carbon and oxygen atoms. However, the corresponding curve in Figure 6 has a sharp minimum which is even deeper than that for two carbon atoms. The explanation is that a carbon atom is often chemically bonded to an oxygen or nitrogen atom; when these adjacent atoms make a hydrogen bond to another oxygen atom, the carbon atom seems attracted to the oxygen atom. This demonstrates that direct use of knowledge-based potentials for description of interactions between protein and ligand atoms may lead to qualitatively wrong results.

We argue that both approaches need additional input to describe spatial dependencies of interaction potentials properly. Our idea is to combine these two approaches. In our unified approach, we started with the knowledge-based potentials as an initial approximation to protein–ligand interaction potentials. We then corrected the potentials by introducing adjustable parameters to the potentials. The adjustable parameters were calibrated with a set of protein–ligand complexes with known binding energies and 3D structures along the lines of the empirical approach. Importantly, we used only four adjustable parameters, which should provide transferability of our scoring function to protein–ligand complexes dissimilar to those included in the training set (see [12] for a discussion).

Optimization of ligand positions

Another important concept employed here is the use of a regression procedure to calibrate our scoring function, which differs in an important way from the schemes usually used for calibration of empirical scoring functions. Typically, scores calculated for the ligand positions as they are in PDB data are fit to experimental values. Here we first optimized the ligand positions with our scoring function and then fitted the obtained scores to experiment. We re-optimized ligand positions iteratively until the ligand positions and calibrated parameters finally converged.

We stress that such position optimization, first introduced by Jain [19], is important for two reasons. First, in drug design applications such as high-throughput docking one usually has no crystallographic data on the position of the ligand. Therefore the scoring function should predict both the position and the binding energy for this position. It is this prediction that must be trained against the experiment. Second, in our experience a proper account of van der Waals repulsion is crucial for successful docking. However, PDB structures often have atomic clashes which may impose huge energetic penalties on ligands' scores if van der Waals repulsion is included in the scoring function. These penalties may corrupt the calibration results.

To illustrate our statements, in Figure 7 we present predicted versus measured energies for our training set with non-optimized ligand positions. First, it is seen that all the predicted values are biased to the region of positive energies by a value of $\sim$2 kcal/mol. This is quite natural since the optimization of position by definition decreases the energy of a ligand. From this it becomes clear that if one trains a scoring function against non-optimized positions, one gets in docking experiments a systematic error of about 2 kcal/mol.

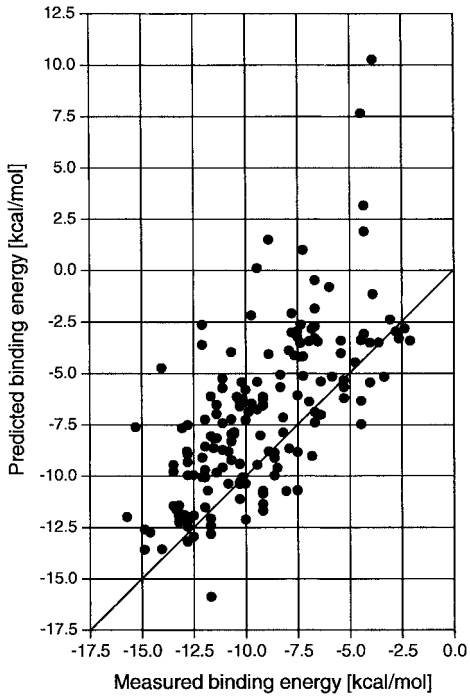

Figure 7. Predicted versus measured values of binding energy for the training set. Predicted values are calculated for non-optimized ligand positions (compare to Figure 3).

Second, some points are in the positive energy region. This is evidence of van der Waals clashes.

The comparison of Figures 7 and 3 demonstrates the sensitivity of our scoring function to ligand positions. Moreover, it is seen that our scoring function cannot be used for prediction of binding energy for non-optimized PDB structures. This is a natural consequence of the way the scoring function was constructed.

Efficiency

Drug design applications such as high-throughput docking absolutely require scanning of a huge variety of ligand variants: their ability to find new leads is directly related to the rapidity of score evaluation. Among the scoring algorithms used in structure-based drug design, the fastest ones utilize gridding of a protein [20, 21]. In such an algorithm a grid is built inside a protein active site. At each point of the grid the scores for individual atoms placed at that point are calculated for all atom types. Then at each step of the calculation the score for each atom is set equal to a pre-calculated value at the nearest grid point. Moreover, interpolation between grid points is not necessary because the amount of memory available in contemporary computers allows the construction of a grid of almost any desired fineness. This makes the score evaluation much faster.

However, not any scoring function is suitable for a grid algorithm. For non-pairwise scoring functions containing angular dependencies or surface matching, the total score of a molecule is not additive, i.e., cannot be expanded as a sum of independent scores of individual atoms. Therefore such a score cannot be calculated on grid. The scoring function we have constructed is grid-oriented and has a simple additive form (1) allowing very rapid calculation. This becomes especially important when flexible ligands, for which many configurations must be considered, are screened.

Another feature of our scoring function contributing to its computational efficiency is implicit consideration of hydrogen atoms. We do not explicitly consider hydrogen atoms for several reasons. First of all, addition of hydrogen atoms to molecules increases the total number of particles by a factor of 2, making the calculation proportionally slower. Also, the small size of a hydrogen atom adds smaller scale features to a potential energy landscape, making it more rugged, which in its turn makes the overall calculation harder. Finally and most importantly, explicit introduction of hydrogen atoms greatly increases the number of degrees of torsional freedom, and the time of calculation depends on this number exponentially.

The latter effect is the most important one and therefore we illustrate it with the 1ABE protein–ligand complex as an example. In this complex the ligand α-arabinose has 4 COH groups in which hydrogen atoms may rotate along the CO axis. Thus, the number of rotatable bonds is 4. If one considers, say, 6 angular conformations per each rotatable bond, one has to dock $6^4 = 1296$ conformers. However, if one disregards positions of hydrogen atoms the number of rotatable bonds is zero and docking of only 1 conformer is needed. For large proteins a binding site is often rigid, so that its heavy atoms may be considered as immobile. However, this is not the case with hydrogen atoms which are much more mobile and in many cases may move due to torsions. Therefore implicit account of hydrogen atoms makes the rigidity approximation for a binding site more realistic.

Together grid score evaluation and implicit account of hydrogens are the two basic factors that provide us with a high efficiency scoring function: we make 10^6 score evaluations per second on a Pentium IV processor.

Accuracy

It is common to test accuracy of binding energy prediction using crystallographic 3D structures of protein–ligand complexes for which data on binding affinity are available. However, the scroes for ligand positions found by docking may differ significantly from scores for X-ray positions. It is thus much more important to test energy predictions for docked ligands. That is why we made our choice in favor of this type of test.

Besides accuracy we would also like to demonstrate efficiency of our scoring function. Unfortunately, in docking both speed and accuracy depend not only on the scoring function, but on the docking algorithm, too. For example, with flexible ligands both accuracy and speed substantially depend on the way in which the flexibility is taken into account. Thus, in order to eliminate the influence of the docking algorithm as much as possible we decided to choose only rigid ligands for the docking tests done here.

We chose 25 protein–ligand complexes with both known PDB structure and binding energy for the tests. None of these complexes was included in the training set of our scoring function. We used for docking an in-house developed program, AlgoDock, which will be described in detail elsewhere. The results of docking presented in Figure 5 and Table 2 demonstrate both good accuracy of our scoring function and high speed of score evaluation.

We wanted to compare our scoring function with scoring functions implemented in most popular docking programs. For this purpose we extracted all rigid ligands from the testing set of the FlexX docking program [14] (19 ligands in total). For most of these ligands docking results are also available for the Gold and Dock docking programs [15, 17]. Results of docking for these 19 ligands are given in Table 1 and Figure 4. For these ligands, AlgoDock is generally superior in both speed and accuracy of docking.

Conclusions

We have developed a new approach to derive a scoring function for docking of small molecules to protein binding sites. Our approach is based on a combination of two main approaches used in the field, i.e., the empirical and knowledge-based approaches. Two basic features of our approach are smooth radial profiles of atom-atom potentials and a procedure of calibration of adjustable parameters in which positions and scores of ligands are determined self-consistently. For a representative set of 25 rigid ligands, the RMS deviation of the binding energy of ligands docked using our scoring function is 2 kcal/mol relative to the experimental values, which is equivalent to a 30-fold difference in binding affinity. We believe that the demonstrated improvement in accuracy makes our scoring function valuable for structure-based drug design.

Acknowledgements

We acknowledge fruitful discussions with M. Levitt, A. Finkelstein, M. Roytberg, V. Ozrin, M. Subbotin, and I. Bodrenko, and thank C. Queen for careful review of the manuscript.

References

1. Crossley, R., Modern Drug Discov., 12 (2002) 18.
2. Augen, J., Drug Discov. Today, 7 (2002) 315.
3. Henry, C.M., Chem. Eng. News, 79 (2001) 23, 69.
4. Bajorath, J., Nature Rev., Drug Discov., 1 (2002) 882.
5. Li, J., Drug Discov. Today, 3 (1998) 105.
6. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., J. Mol. Biol., 161 (1982) 269.
7. Taylor, R.D., Jewsbury, P.J. and Essex, J.W., J. Comput.-Aided. Mol. Des., 16 (2002) 151.
8. Bohm, H.J., J. Comput.-Aided Mol. Des., 8 (1994) 243.
9. Gohlke, H. and Klebe, G., Curr. Opin. Struct. Biol., 11 (2001) 231.
10. Lyne, P. D., Drug Discov Today, 7 (2002) 1047.
11. Bohm, H. J., Drug Discov. Today, 7 (2002) 64.
12. Wang, R., Lai, L. and Wang, S., J. Comput.-Aided Mol. Des., 16 (2002) 11.
13. Muegge, I. and Martin, Y.C., J. Med. Chem., 42 (1999) 791.
14. http://cartan.gmd.de/flexx/
15. http://www.ccdc.cam.ac.uk/prods/gold/
16. http://www.cmpharm.ucsf.edu/kuntz/
17. Paul, N. and Rognan, D., Proteins, 47 (2002) 521.
18. Jones, G., Willet, P., Glen, R. C., Leach, A.R. and Taylor, R., J. Mol. Biol., 267 (1997) 727.
19. Jain, A. N., J. Comput.-Aided Mol. Des., 10 (1996) 427.
20. Pattabiraman, N., Levitt, M., Ferrin, T. E. and Langridge, R., J. Comput. Chem., 6 (1985) 432.
21. Meng, E.C., Shoichet, B.K. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 505.