

Articles

Evaluation of Docking Functions for Protein–Ligand Docking

Carlos Pérez and Angel R. Ortiz*

Department of Physiology & Biophysics, Mount Sinai School of Medicine, New York University, One Gustave Levy Plaza, Box 1218, New York, New York 10029

Received March 29, 2001

Docking functions are believed to be the essential component of docking algorithms. Both physically and statistically based functions have been proposed, but there is no consensus about their relative performances. Here, we propose an evaluation approach based on exhaustive enumeration of all possible docking solutions obtained with a discretized description of a rigid docking process. We apply the approach to study both molecular mechanics and statistical potentials. It is found that the statistical potential evaluated is less effective than the AMBER molecular mechanics function to provide an accurate description of the docking process when the exact experimental coordinates are used. However, when coordinates of crystal structures obtained with analogous ligands are used, similar performances are obtained in both cases. Possible reasons for the successes and failures of both docking schemes have been uncovered using linear discriminant analysis, on the basis of a set of physicochemical descriptors capturing the main physical effects at play during protein–ligand docking. In both types of potentials steric effects appear critical to obtain a successful docking. Our results also indicate that neglecting desolvation effects and the explicit treatment of hydrogen bonds are the main source of the failures observed with the molecular mechanics potential. On the other hand, detailed consideration of steric interactions, with a careful treatment of dispersive forces, seems to be needed when using statistical potentials derived from a structural database. The possibility of filtering combinatorial libraries in order to maximize the probability of correct docking is discussed.

Introduction

The growing interest and widespread application of structure-based virtual screening in Medicinal Chemistry¹ has evidenced the need for the development of better, more accurate docking functions.² Two different types of scoring schemes are being used at present: first principles molecular mechanics (MM) force fields and statistical potentials (either in the form of empirical scoring functions or knowledge-based methods). In general, physically motivated force fields, such as those used in MM calculations, have yielded limited results in protein–ligand docking, which has stimulated the search for alternative docking force fields. There have been recent advances with incorporation of improved electrostatics models and approximate descriptions of hydrophobic effects.^{3–5} Also recently, knowledge-based or potential of mean force (PMF) statistical potentials have shown some encouraging results in the related problem of affinity ranking,^{6–10} and for this reason they have been strongly advocated in the docking problem as well,^{11,12} but so far there is only limited experience with them. Despite all this progress, no docking function seems to be still effective enough. The situation is summarized by the results of the docking section of the CASP2 meeting, where a variety of methods were used

to score the docking solutions. In his analysis of CASP2 results, Dixon¹³ concluded that while conformational sampling was reasonable, docking functions were not reliable enough to separate nativelike answers from alternative solutions.

The docking function is therefore, at this juncture, the component of the docking algorithm that needs better understanding. To advance, we need a way to single out which elements in the docking function need to be introduced or modified, coupled to a method of assessing improvements in the function. When addressing the question of which of the various alternative scoring schemes is more adequate in protein–ligand docking, it becomes essential to have an accurate assay to measure success, since only with the availability of a good test at hand it will be possible to discover those elements of the function which are still lacking or in need of refinement. Usually, tests of docking force fields focus on the deviation from the experimental structure of the minimum energy solution found by a given docking algorithm.^{13,14} Recent experience¹⁵ suggests that this approach is insufficient. First, by using standard docking programs, the issue of the ability of the function to discriminate the experimental solution from alternative decoys or *poses* gets convoluted with the ability of the search scheme to surmount kinetic barriers created by the potential.¹⁶ In other words, there is no guarantee that the global minimum is found, and for

* To whom correspondence should be addressed. Phone: (212) 241-6533. Fax: (212) 860-3369. E-mail: ortiz@inka.mssm.edu.

this reason, while ultimately the docking function needs to be implemented together with a given search algorithm, it is convenient to decouple both problems. Second and more important, looking at the minimum energy solution only, even if that solution contains the global minimum, can be misleading. The global energy minimum of the effective energy function (even if it contains entropy contributions coming from degrees of freedom averaged out, such as in the PMF case) can be entropically unfavorable (for the configurational and conformational solute entropies making up the width of the docking funnel) and hence thermodynamically unstable. In addition, large energy barriers can also surround it, making it kinetically inaccessible.

A recent paper by Verkhivker et al.¹⁵ illustrates clearly this concept. These authors have studied docking failures in three ligand–protein systems using a combined thermodynamic and kinetic analysis of their binding energy landscapes. In their study, Verkhivker and co-workers found that docking algorithms can be mistaken when the native binding mode, even if having the lowest stabilization energy, corresponds to a narrow and isolated region on the binding energy landscape with a small energy gap with alternative binding modes. In these cases, neither the determination of a single structure with the lowest energy nor finding the largest cluster of structurally similar conformations allows the docking algorithm to detect with confidence the native-like binding mode. In these cases both thermodynamic stability and kinetic accessibility are compromised, and the docking algorithm is prone to failure. The conclusion to draw from this study is that it is necessary to consider the optimization of the complete shape of the energy landscape, and not only the position of its global minimum, when developing and testing docking functions.

Here, we present a new approach, the coverage-error plots (CEPs),¹⁷ to evaluate quantitatively docking force fields using the complete pool of solutions obtained after exhaustive enumeration of all rigid docking solutions at a given resolution level. The exhaustive enumeration also allows an approximate reconstruction of the shape of the energy landscape. Using a diverse dataset of protein–ligand complexes, we evaluate two different types of docking functions: the well-known AMBER MM force field¹⁸ and the statistical potential developed by Muegge and Martin.⁹ These two potentials are taken only as representatives of the two types of functions, and the comparison between them is established only in order to obtain insight about their respective strengths and deficiencies as force field classes. First, we show the close correspondence between CEPs and docking energy landscapes. Then, we use CEPs to study the docking functions under two different situations: using the experimental coordinates for ligand and receptor and also using the coordinates of the receptor obtained with a close analogue of the ligand. The comparison allows us to evaluate the effects that coordinate shifts in the receptor have in the performance of the potentials. Finally, we apply the multivariate technique of linear discriminant analysis¹⁹ to the quantitative results provided by CEPs in order to better understand the physical origin of the successes and failures of the potentials. To this end, each complex is encoded as a

set of physicochemical descriptors to capture the main interactions taking place during the docking process. The models obtained are interpreted and used to suggest ways to improve the current use and future development of docking functions.

Methods

Dataset. A test set of 34 crystallographic non-covalent protein–ligand complexes, selected to have diversity in their ligand atom types and binding site shapes, has been used (Chart 1). It comprises a set of 17 pairs of complexes of the same protein bound to two different, usually related, ligands. For each complex, protein hydrogen atoms were first positioned using the AMBER 5.0 package,²⁰ while for the ligand hydrogen atoms were added using SYBYL.²¹ No further preparation of the complexes was carried out. Specifically, complexes were not energy minimized with the scoring functions. In the case of the MM force field, AMBER charges were used for the protein, while Gasteiger–Marsilli²² atomic charges as implemented in SYBYL²¹ were used for the ligands.

Docking Functions. (a) Molecular Mechanics Based Function. The nonbonded interaction energies of the AMBER force field¹⁸ using an all atom model were employed. The equation used has the form

$$E_{\text{MM}} = \sum_i^{\text{prot}} \sum_j^{\text{lig}} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + 332 \frac{q_i q_j}{\epsilon r_{ij}} \right] \quad (1)$$

In eq 1, A_{ij} and B_{ij} represent the van der Waals parameters of the atom types to which atoms i and j belong, q_i and q_j are the partial charges of atoms i and j , respectively, and r_{ij} is the distance between them. We refer to the AMBER potential for details about parameters.¹⁸ A constant dielectric of $\epsilon = 4$ was used to scale down the electrostatic term. We note that, while more complex MM approaches have been proposed for the docking problem,^{23,24} the use of eq 1 or close variants is still a fairly common approach in many docking programs.^{25–27}

(b) Statistical Potential Based Function. PMF potentials were generously provided by Dr. Muegge.⁹ Since values are tabulated at 0.2 Å intervals from 0.2 to 12 Å, a linear interpolation scheme is required in order to compute the potential for each atom–grid point interaction. The function being computed for each kl protein–ligand atom pair of type ij is

$$E_{\text{PMF}} = \sum_{r^{kl} < r_{\text{cutoff}}^{kl}} -k_{\text{B}} T \left[\frac{f_{\text{vol corr}}^j(r) \rho_{\text{seg}}^{ij}(r)}{\rho_{\text{bulk}}^{ij}} \right] \quad (2)$$

In eq 2, k_{B} is the Boltzmann factor, T is the absolute temperature, $f_{\text{vol corr}}^j(r)$ is the ligand volume correction factor, $\rho_{\text{seg}}^{ij}(r)$ is the number density of atom pair ij occurrences at a certain distance, and ρ_{bulk}^{ij} is the reference number density for that pair ij . For details about derivation of these parameters, we refer the reader to the original paper by Muegge and Martin.⁹

The correct implementation of this potential in our docking scheme has been validated by reproducing the reported linear correlation with observed activities for the set of 16 serine protease complexes analyzed in Figure 4 of Muegge and Martin.⁹ Our implementation gives a squared correlation coefficient $r^2 = 0.88$, while Muegge and Martin report a similar value, of 0.87. Likewise, PMF scores produced with our implementation have a squared correlation coefficient $r^2 = 0.96$ with the scores reported by Muegge and Martin for this set of complexes. The slight discrepancies observed can be attributed to differences in our automated procedure for atom type assignment compared to that of Muegge and Martin.

Additionally, a Lennard-Jones repulsion term ($1/r^{12}$ term in eq 1) using the AMBER parameters for each atom pair is added to provide the steric effects lacking in the PMF potentials at short distances. Our implementation is consistent with the

Chart 1. Set of 34 Crystallographic Protein–Ligand Complexes Used in This Work^a

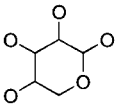
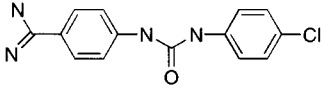
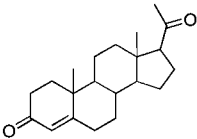
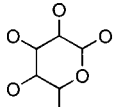
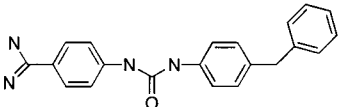
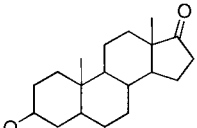
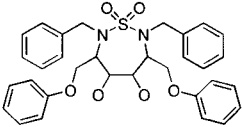
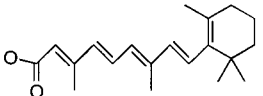
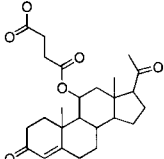
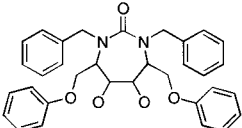
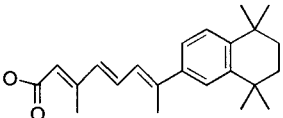
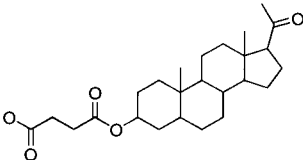
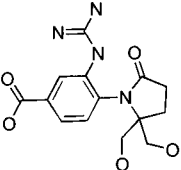
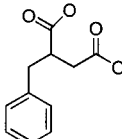
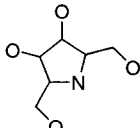
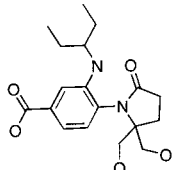
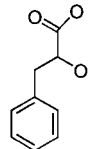
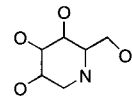
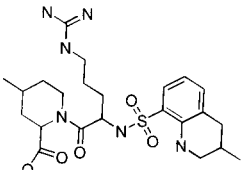
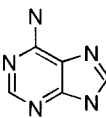
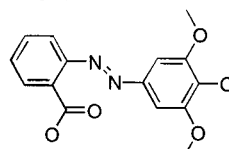
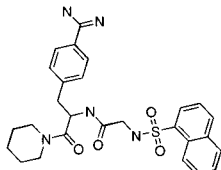
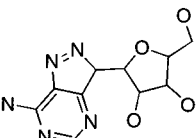
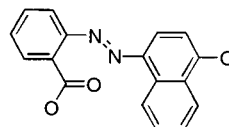
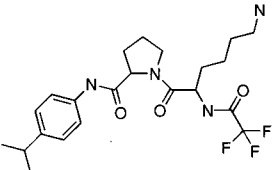
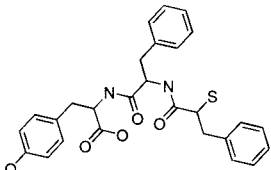
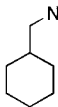
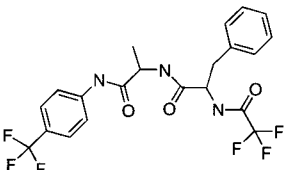
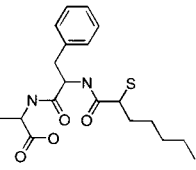
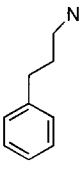
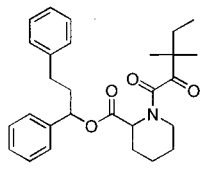
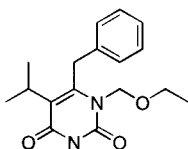
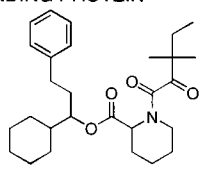
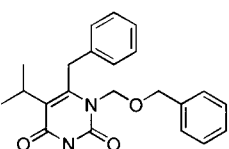
<p>1abe 1.7 L-ARABINOSE-BINDING PROTEIN</p> 	<p>1bju 1.8 BETA-TRYPSIN</p> 	<p>1dbb 2.7 DB3 ANTI-STEROID MON. ANTIBODY</p> 
<p>1abf 1.9 L-ARABINOSE-BINDING PROTEIN</p> 	<p>1bjv 1.8 BETA-TRYPSIN</p> 	<p>1dbj 2.7 DB3 ANTI-STEROID MON. ANTIBODY</p> 
<p>1ajv 2.0 HIV-1 PROTEASE</p> 	<p>1cbs 1.8 RETINOIC-ACID-BINDING PROTEIN</p> 	<p>1dbm 2.7 DB3 ANTI-STEROID MON. ANTIBODY</p> 
<p>1ajx 2.0 HIV-1 PROTEASE</p> 	<p>2cbs 2.1 RETINOIC-ACID-BINDING PROTEIN</p> 	<p>2dbl 2.9 DB3 ANTI-STEROID MON. ANTIBODY</p> 
<p>1b9t 2.4 NEURAMINIDASE</p> 	<p>1cbx 2.0 CARBOXYPEPTIDASE A</p> 	<p>1did 2.5 D-XYLOSE ISOMERASE</p> 
<p>1b9v 2.4 NEURAMINIDASE</p> 	<p>2ctc 1.4 CARBOXYPEPTIDASE A</p> 	<p>1die 2.5 D-XYLOSE ISOMERASE</p> 
<p>1dwc 3.0 ALPHA-THROMBIN</p> 	<p>1mrg 1.8 RIBOSOME INACTIVATING PROTEIN</p> 	<p>1srh 2.2 STREPTAVIDIN</p> 
<p>1dwd 3.0 ALPHA-THROMBIN</p> 	<p>1mrk 1.6 RIBOSOME INACTIVATING PROTEIN</p> 	<p>1srj 1.8 STREPTAVIDIN</p> 

Chart I (Continued)

1ela 1.8 ELASTASE 	1qf0 2.2 THERMOLYSIN 	1tng 1.8 TRYPSIN 
1eld 1.8 ELASTASE 	1qf1 2.0 THERMOLYSIN 	1tnk 1.8 TRYPSIN 
1fkg 2.0 FK506 BINDING PROTEIN 	1rt1 2.6 HIV-1 REVERSE TRANSCRIPTASE 	
1fkh 2.0 FK506 BINDING PROTEIN 	1rt2 2.6 HIV-1 REVERSE TRANSCRIPTASE 	

^a The PDB code,⁵⁹ structure resolution (Å), protein name (from the PDB header), and the ligand 2D structure are displayed.

implementation of Muegge and collaborators of the same PMF parameters for docking.²⁸ They added the complete 6–12 potential²⁹ for distance pairs shorter than the longest unoccupied distance bin for the respective atom types in the database, and overwrite the PMF potential if the van der Waals (VDW) interaction is larger than 4 kcal/mol. Thus, at the short distances at which the VDW potential becomes operative, the VDW interaction is heavily dominated by the repulsion term in their approach, which thus becomes effectively the same as ours.

Finally, we note that the original PMF potential encoded in eq 2 was initially formulated with the aim of solving the rank affinity problem, not the docking problem. However, both problems are related, and in fact Muegge and collaborators have recently tested successfully the same PMF potential as a docking function.¹¹ Moreover, Golhke et al.¹² have recently proposed a very similar formalism explicitly for the docking problem.

Docking Method. (a) Grid Description of the Binding Space and Energy Precomputation. The protein–ligand intermolecular energy is precomputed using an underlying three-dimensional grid. A gridded box is created by adding a 3.0 Å cushion to the maximum dimensions of the ligand complexed with the protein and using a grid spacing of 0.3 Å. The atom–atom interaction energy is precomputed in the grid and then used to calculate for each ligand atom the protein–ligand interaction energy from the nearest eight surrounding grid points using a trilinear interpolation³⁰ method:

$$E_i = \sum_{n=1}^8 (1 - a/s)(1 - b/s)(1 - c/s)V(k, l, m) \quad (3)$$

where a – c are distances along the x -, y -, and z -coordinates from atom i to the point (k, l, m) ; s is the grid spacing; and V is

the precalculated potential on each grid point. The test charge and atom type method is used for the computation of the potentials.

The error accumulated from the energy interpolation and atom type simplification was evaluated. A squared correlation factor (R^2) of 0.97 on the linear fitting was found when a grid spacing of 0.3 Å was used (Figure 1). This level of error is similar to that reported in other implementations of the same approach.³¹

(b) Exhaustive Search Algorithm. In our docking search, we carried out a complete enumeration of all possible orientations of the rigid ligand in the active site of the rigid protein. The six relevant degrees of freedom, three translational and three rotational, are first discretized. The molecule is translated and rotated in the docking region using the ligand's center of mass, which is moved consecutively to every grid point in the box using a grid spacing of 0.3 Å. At each grid point, a complete sampling of the rotational space is achieved by computing all nondegenerate sets of Euler angles³² obtained with a resolution of 27° arc. Initial tests indicated that it is necessary to relax the lattice conformation off-lattice in order to avoid artifacts arising from the discrete sampling. For this reason, at each rotational and translational point the ligand was subjected to a rigid body off-lattice energy minimization using the SIMPLEX algorithm from Nelder and Mead³³ as described in *Numerical Recipes*.³⁴ Structures having steric clashes after minimization ($E_{VDW} > 10$ kcal/mol) were discarded for further analysis.

Analysis Methods. (a) Docking Energy Landscapes. Energy landscapes are studied by computing docking energies for each point in conformational space, considered as a function of the displacement from the experimental orientation, and represented by the six relevant degrees of freedom. To visualize the seven-dimensional energy landscape, for every stored

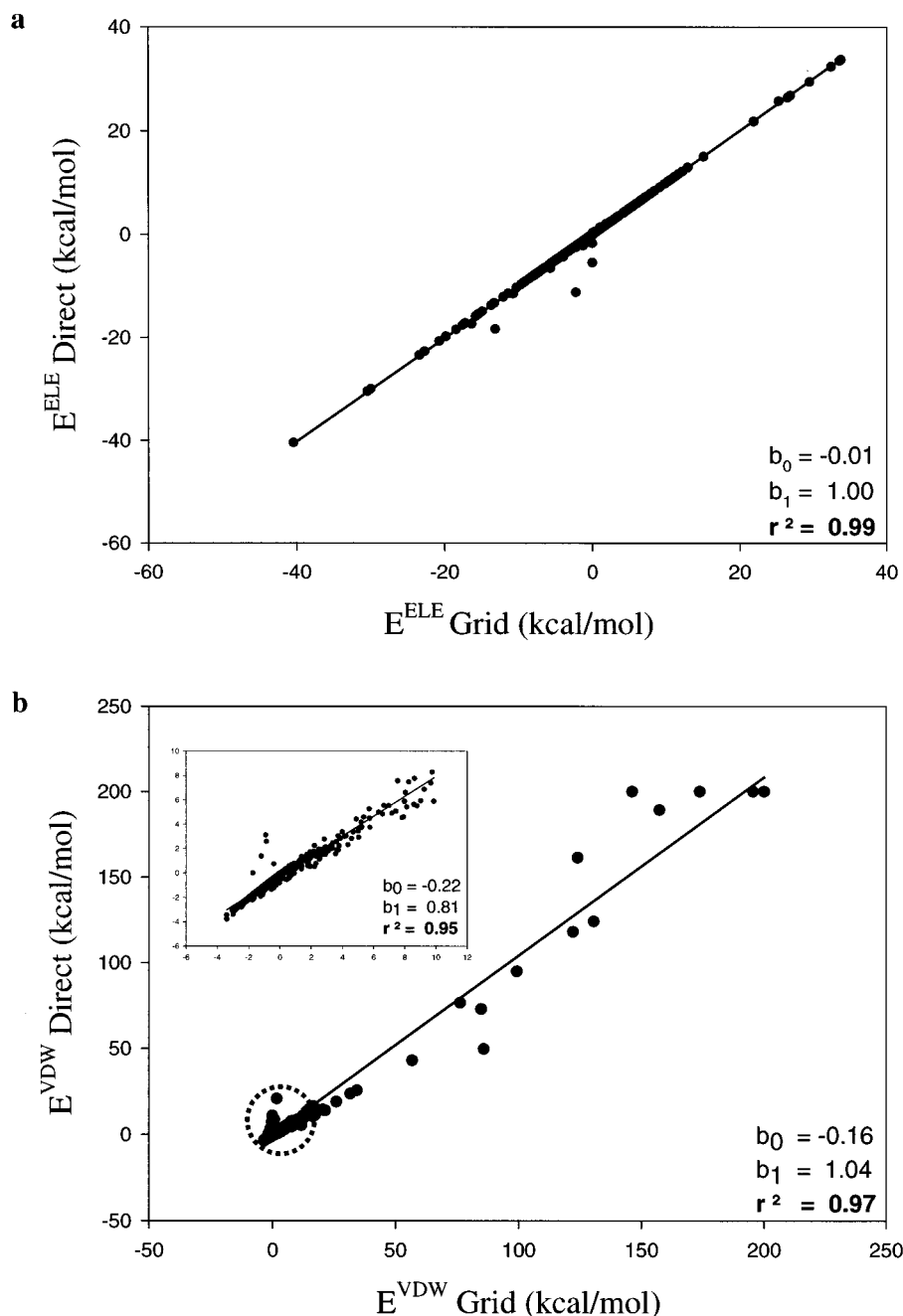


Figure 1. Correlation between *per-atom* energies interpolated from the grid approximation vs their counterparts obtained by direct calculation: (a) electrostatic energy; (b) van der Waals energy. The inset shows the van der Waals energies for those atoms with values below 10 kcal/mol. Regression coefficients are also shown.

conformation from the exhaustive search (from 4000 to 1000 000 depending on the complex), the six ligand degrees of freedom are projected onto two orthogonal axes: translational and rotational. Translational projections are computed as Euclidean distances between the centers of mass at the grid point i and the experimentally observed docking position; while rotational projections are calculated as the root mean square deviation (RMSD) of the rotated unit vector local coordinate frame with that of the docked conformation.

(b) Coverage-Error Plots. Coverage-error plots (CEPs)¹⁷ are used to quantify, at different levels of precision (i.e., RMSD deviation from experiment), the behavior of the docking energy function. Three levels of precision (0.5, 1.0, and 2.0 Å) were evaluated to assess the maximal resolution of every function. However, and for the sake of clarity, functions are discussed here using only the 2.0 Å cutoff. The amount of error that the docking function accumulates in order to recover a given percentage of the pool of successful solutions is computed as

follows: (i) We fix a given RMSD threshold (defining good conformations), and generate a “sorted by RMSD” list of poses. The size of the list is determined by the number of conformations found during the simulation to have an RMSD below the threshold. (ii) We create another list, with an identical number of entries, this time using a “sorted by energy” convention, containing the lowest energy conformations found during the simulation. (iii) We count down the “sorted by energy” list and check whether each entry is found also in the “sorted by RMSD” list. (iv) We keep track of the fraction of the conformations in the “sorted by RMSD” list assigned (coverage), and at each point we register the cumulative error (error per query) obtained in the “sorted by energy” list. (v) At the end of the count (coverage is one), we obtain the total error accumulated by the function (a number between 0 and 1). Figure 2 summarizes the overall procedure. We note that in practice we use two additional auxiliary lists: an RMSD auxiliary list for the energy and an energy auxiliary list for the RMSD.

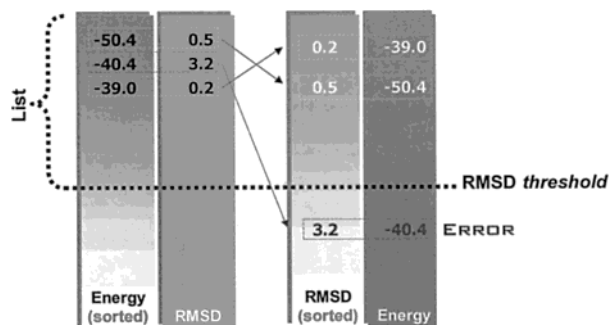


Figure 2. Description of a coverage–error (CEP) calculation. See Methods for a detailed explanation of the algorithm.

When we check the existence of a particular conformation in step iv, we actually check the matching of both RMSD and energy using the auxiliary lists. We used this approximation in order to avoid the storage of the translational vector and rotational matrices defining each conformation. Tests demonstrated that the error introduced does not have impact in the results.

(c) Discriminant Analysis. Linear discriminant analysis (LDA)¹⁹ is a pattern recognition technique normally used to determine which variables discriminate between two or more occurring groups. Specifically, LDA provides answer to the question of whether two or more groups of objects are significantly different from each other with respect to the mean of a particular set of descriptors used to characterize them. LDA seeks to find a low-dimensional space in which groups are projected while group discrimination is optimal. The linear discriminant functions can be used to identify outcome variable constructs (or latent variables) that underlie the group differences. It is known that the method is Bayes-optimal in the case where the groups under study follow a normal distribution and have the same covariance structure. We will describe the quality of these discriminant functions with four indices:

(1) Wilks' Lambda (Λ^*). It compares the matrix of pooled within-group variances and covariances on one hand with the matrix of total variances and covariances on the other. It is defined as

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} = \frac{\sum_{i=1}^g \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \langle \mathbf{x} \rangle) (\mathbf{x}_{ij} - \langle \mathbf{x} \rangle)'}{\sum_{i=1}^g \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \langle \mathbf{x} \rangle) (\mathbf{x}_{ij} - \langle \mathbf{x} \rangle)'} \quad (4)$$

where $\mathbf{B} + \mathbf{W}$ is the total variance-covariance matrix and \mathbf{W} is the within groups variance-covariance matrix, g is the number of groups, and n_i is the number of objects within each group. Vectors $\langle \mathbf{x} \rangle$ and $\langle \mathbf{x} \rangle_i$ are the total and group averages, respectively. Matrices are built after projection onto the low-dimensional solution found by LDA. The Wilks' Λ varies from 0 to 1, and the smaller the Λ , the larger the differences, and the better the discrimination.

(2) *F*-Test. It provides a statistic related to the Wilks' Λ corresponding to the tested discriminant function, but focused on the separation of the group centroids:

$$F = \frac{\frac{1}{g-1} \sum_{i=1}^g n_i (\langle \mathbf{x} \rangle_i - \langle \mathbf{x} \rangle)^2}{\frac{1}{\sum_{i=1}^g n_i - g} \sum_{i=1}^g \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \langle \mathbf{x} \rangle)^2} \quad (5)$$

(3) Significance Level (*P*-Value) of the *F*-Test. This tests the probability that the Fischer *F* computed above could have

been obtained by chance. The test rejects the null hypothesis at the α confidence level if

$$F > F_{\alpha}(g-1, \sum_{i=1}^g n_i - g) \quad (6)$$

The level of α for which $F = F_{\alpha}$ is selected as the *P*-value.

(4) Discrimination Power. This is the classification accuracy achieved by the discriminant function as compared with the expected error in classification. This expectation is computed by leave-one-out cross-validation of each of the members of the sample using the Lachenbruch's hold-out procedure.¹⁹ This procedure provides a nearly unbiased estimate of the expected actual error rate ($\hat{\theta}$). The discrimination power (*D*) is then $D = 100(1 - \hat{\theta})$. We considered models with at least 70% discrimination.

Each complex was characterized by a set of physicochemical descriptors representing the main effects at play during the docking process (steric effects; hydrogen bond effects; hydrophobic effects; electrostatic and polar effects), encoded in the following variables.

SA: Ligand van der Waals surface area (\AA^2), calculated using the grid method defined by Bodor et al.³⁵

VO: Ligand van der Waals volume area (\AA^3), calculated using the grid method.³⁵

SL: Ligand electrostatic component of solvation energy (kcal/mol), obtained by multiplication of the solvent reaction field by the charges of the ligand:

$$\Delta G_{\text{sol,elec}} = \frac{1}{2} \sum_i^Q q_i [\phi_i^{\text{sol}}(\bar{r}_i) - \phi_i^{\text{vac}}(\bar{r}_i)] \quad (7)$$

where Q represents the charge set of the ligand. Potentials generated by the ligand charges under vacuum and solvent (ϕ_i^{vac} and ϕ_i^{sol} , respectively) are evaluated using the finite difference method to solve the Poisson equation. An internal dielectric of 2 was assigned to the ligand molecular interior, defined as the region inaccessible to a 1.8 \AA probe sphere. A solvent dielectric constant of 80 was used to represent water and 1 for vacuum. A grid box with 75% occupancy by the ligand was defined, containing 8 grid points/ \AA on each axis.

LP: log *P*, using atomic parameters derived by Ghose et al.³⁶

MR: Molar refractivity,³⁶ using parameters derived by Ghose et al.³⁶ Molar refractivity is given by the Lorentz–Lorentz equation:

$$MR = \frac{(\mu^2 - 1) M}{(\mu^2 + 1) \delta} \quad (8)$$

where *M* is the molecular weight, δ is the density of a crystal of the molecule, and μ is the refractive index.

MP: Molecular polarizability, following the atomic parameters derived by Miller et al.³⁷

HB: Total number of possible hydrogen bonds (donor + acceptor) in the ligand.

HBZ: Ligand hydrogen bonds formed in the ligand binding site.

L/BS: Percentage of the free binding site volume occupied by the ligand volume (VO/BSV \times 100). The grid used to store the precomputed potentials is employed to approximate the binding site volume (BSV) by counting the volume of the cubes defined by all the grid points that are not in contact with protein atoms.

L/HB: Surface area of the ligand (\AA^2) per hydrogen bond (SA/HB).

SA, VO, LP, MR, and MP were calculated using HYPER-CHEM 6.³⁸ SL was calculated using DELPHI.³⁹ HBZ was calculated using LIGPLOT.⁴⁰

Results and Discussion

Coverage-Error Plots and Energy Landscapes. CEP analysis provides information about the error that the

docking function needs to accumulate in order to recover a given percentage of the pool of the user-defined set of *good* solutions. In Figure 3 we show that CEP plots bear considerable correspondence with docking energy landscapes. For the sake of illustration, we divide the CEP at 100% coverage in three different regions (I–III in Figure 3). Region III comprises those cases for which nativelylike solutions in the CEP plot are well separated in energy (i.e., recovered first) from all (or most of) alternative poses. Consequently, an energy gap exists between nativelylike and nonnativelylike poses. This is reflection of a funnellike shape of the binding energy landscape for the complex, as observed for instance in the case of 1abe in Figure 3a, where a clear funnel toward the nativelylike basin region is observed. Region II, on the other hand, contains cases for which nativelylike solutions are mixed in their energy levels with alternative poses. For example, the binding energy surface of 1ela in Figure 3a shows two minima of similar energy separated by an energy barrier, one nativelylike and the other corresponding to a family of misdocked solutions. Finally, in region I coverage is complete and yet no nativelylike poses have been detected. Therefore, and as shown for the case of 1dbm, the global minimum is in a region of conformational space separated from the nativelylike basin. A similar situation can be observed for the PMF potential in Figure 3b.

In Table 1a, we summarize the evaluation of the docking results and the function performance as measured using CEPs, and compare it with the traditional criterion of analyzing only the global energy minimum provided by the algorithm. In general, there is a good correspondence between both sets. Thus, in cases where the function cannot find a correct docking solution (in red in Table 1a), both evaluation methods agree. However, some differences also emerge. In some cases a good solution—as considered by the usual criterion—is found by the program, but the CEP results indicate these good solutions are close in their energy levels to wrong alternative states, which is reflected by a medium to high error in the CEP plot (yellow background). As discussed in the Introduction, this mixing creates difficulties to stochastic algorithms in finding correct solutions, and therefore should be avoided. The colored blue background highlights the resolution of the function, that is, the RMSD threshold required to obtain a successful result in each case. The higher the resolution, the more reliable the function is.

Function Performance with Exact Coordinates. When experimental coordinates for both ligand and receptor are used by the rigid docking algorithm (*exact* coordinates, hereafter), we find that the MM function slightly, but significantly, outperforms the PMF function. By judging our results using the standard approach of measuring the RMSD of the lowest energy conformation, our rigid body approach succeeded in 27 out of 34 cases (i.e., 79% success rate) for the MM force field and in 22 out of 34 (59%) for the PMF potential (Table 1a, using 1.5 Å cutoff to define success). The success rate is comparable to those obtained with other docking programs (rigid or flexible) for the MM potential, but inferior for the PMF one. We find that overall there is a close correspondence between PRO_LEADS results and our own results with the MM potential. In direct

docking, PRO_LEADS success rates range from 76⁴² to 86%,⁴¹ depending upon the dataset chosen to test the program in the publication. Percentages are slightly different if a 2 Å cutoff is used, since PRO_LEADS authors report a number of borderline cases which change the percentage in going from the 1.5 to the 2 Å cutoff, while our percentages remain essentially constant. Similarly, prediction rates reported for GOLD in direct docking using the 1.5 Å RMSD cutoff are 55%.²⁶ Finally, David et al.²⁵ have recently reported a molecular mechanics based docking algorithm that, upon testing on a dataset of 27 complexes, gave a success rate of 63%, again using the 1.5 Å RMSD cutoff. Since both MM and PMF share the modeling of the excluded volume penalty and the same rigid body approach is employed in both cases, it is unlikely that the underperformance in the PMF case is due to an artifact in the setup of the docking experiments. Thus, although results of rigid docking should always be interpreted and analyzed with care, it seems that the use of a rigid body approach together with the 12–6 VDW potential is not introducing serious artifacts in our results.

The same qualitative difference between MM and PMF can be observed using the CEP (see Table 1a as well as Figure 4). In Table 2 we quantify the differences between MM and PMF and establish their significance. For that purpose, the distribution of errors obtained at all three resolutions for all different complexes is compared with the distribution of errors obtained when a random number replaces the interaction energy. A Student's *t*-test is then used to determine whether the observed differences are significant. Our results indicate that error distributions obtained by MM and PMF are significantly different, and also very different from what should be expected by using a random number generator as the energy function, supporting the significance of the differences.

Given the superiority of MM in this case, we also decided to explore in more detail the MM function and went on to study the relative weights of steric and electrostatic interactions in the performance of the docking score. We found that omitting the electrostatic term has a relatively small impact in the performance of the function. As expected, the largest differences are observed for small molecules with strong charge–charge interactions (Figure 5). Overall, however, the results suggest that the Coulombic term is playing a relatively minor role compared with the steric term, as it is only responsible for fine-tuning the orientation of the ligand, whereas the steric term controls the overall positioning.

Function Performance with Coordinate Shifts. While the experiments reported so far give information about the behavior of the function when exact coordinates are used, the most significant case from the point of view of virtual screening applications is the study of the behavior of the function in cases where the coordinates of the receptor are taken from a cocrystal obtained with a ligand analogue (*cross-docking*, hereafter). Due to computational and experimental constraints, virtual screening experiments normally use a rigid receptor derived from a cocrystal with a structurally related ligand. Thus, cross-docking experiments provide insight into the effect that these inaccuracies can have in the performance of the docking functions.

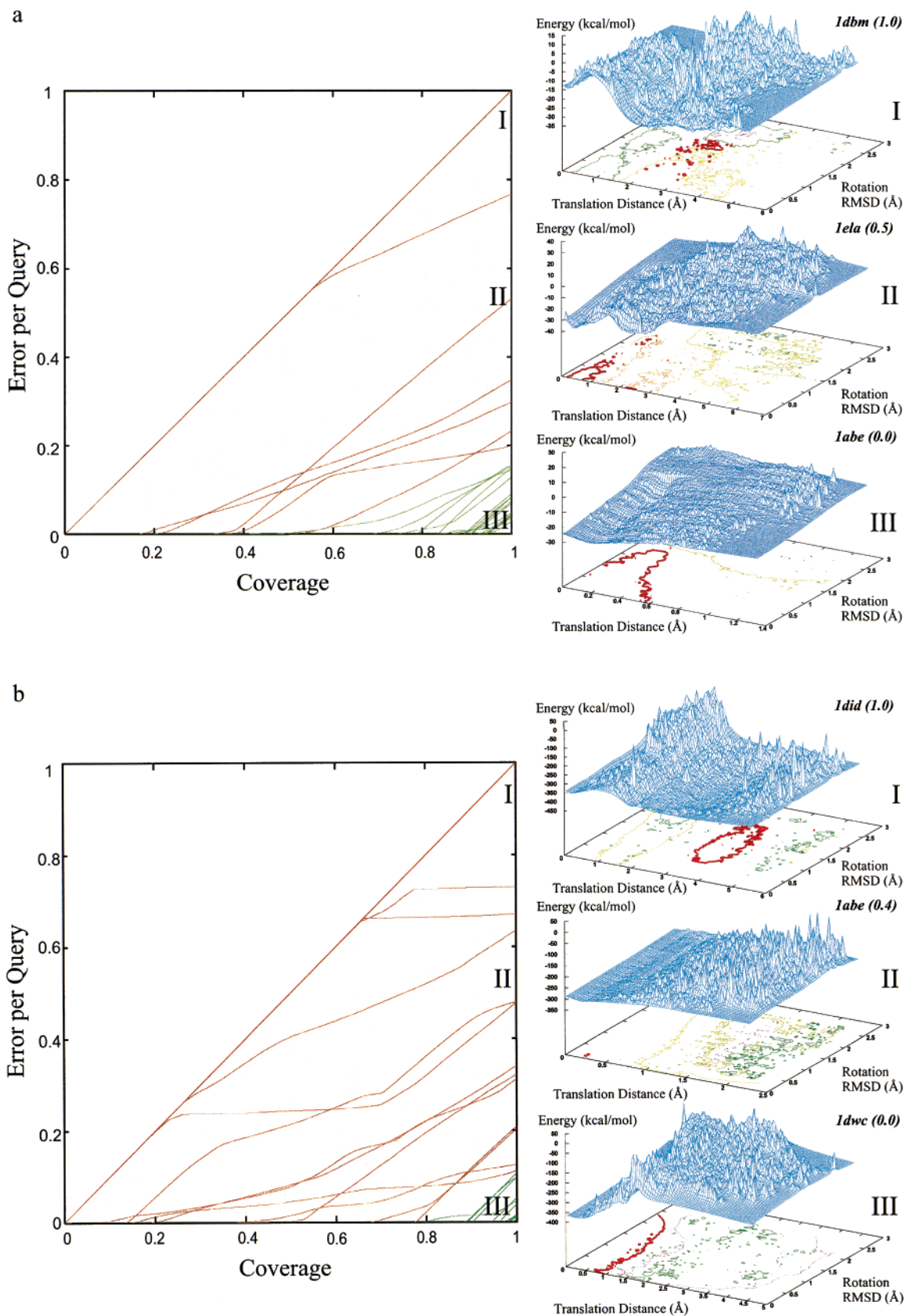


Figure 3. Coverage–error plots (CEPs) at 2.0 Å resolution and the corresponding energy landscapes. Examples at three different error levels are shown. Background colors in the CEP plots represent the docking success region (green, region III) or the docking failure region (red, regions I and II). Each line in the CEP plot represents a different complex studied. Simultaneously, docking energy landscapes are also shown. Examples of energy surfaces (pdb id of the corresponding complex is indicated) in cases of low (green in the CEP plot, cases of 1abe (a) and 1dwc (b)), medium and high (both in red in the CEP plot, cases of 1ela and 1dbm (a), and 1abe and 1did (b)) function error are given to illustrate the correspondence with the numerical results obtained with the CEP plots. The different isocontour lines are in different colors in the figure. The isocontour containing the global energy minimum is shown in red. (a) MM; (b) PMF.

Table 1. Differences in the Evaluation of Docking Results Following the Traditional Criterion and the CEPs Based: (a) Direct Docking; (b) Cross-Docking

	MM						PMF					
	Energy		Error per Query		Resolution		Energy		Error per Query		Resolution	
	Lowest	RMSD	0.5	1	2	Lowest	RMSD	0.5	1	2		
labe	-28.96	0.3	0.03	0.07	0.08	-303.68	0.4	0.48	0.44	0.48		
labf	-28.83	0.6	0.79	0.03	0.15	-368.28	0.6	1.00	0.07	0.11		
lajv	-50.82	0.2	0.02	0.01	0.00	-544.87	0.2	0.04	0.01	0.00		
lajx	-65.77	0.3	0.03	0.08	0.09	-562.67	0.2	0.01	0.04	0.05		
laju	-30.64	0.3	0.29	0.07	0.15	-167.96	1.2	1.00	0.40	0.20		
lajv	-29.77	0.7	1.00	0.21	0.23	-198.87	8.0	1.00	1.00	1.00		
lcbv	-51.22	0.3	0.59	0.54	0.07	-456.63	9.8	1.00	0.97	0.67		
lcbv	-61.54	0.7	0.86	0.03	0.04	-475.61	0.5	0.46	0.01	0.01		
lbbv	-40.36	0.8	0.53	0.34	0.05	-395.15	6.9	1.00	1.00	0.64		
ldbv	-43.15	0.3	0.11	0.02	0.04	-438.97	0.4	0.06	0.01	0.04		
ldvc	-53.46	0.3	0.02	0.02	0.04	-369.08	0.2	0.06	0.01	0.02		
ldvd	-76.18	0.6	1.00	0.07	0.16	-336.95	0.9	1.00	0.09	0.05		
lela	-37.39	0.7	0.59	0.32	0.53	-257.37	0.5	0.81	0.16	0.32		
leld	-27.82	5.8	1.00	1.00	1.00	-175.81	6.1	1.00	1.00	1.00		
lfgk	-43.59	0.3	0.11	0.05	0.03	-540.79	0.4	0.13	0.01	0.01		
lflh	-42.15	0.8	0.76	0.42	0.04	-546.49	0.4	0.03	0.03	0.01		
lrtl	-37.42	0.4	0.02	0.00	0.00	-409.42	0.3	0.02	0.01	0.01		
lrr2	-48.83	0.2	0.01	0.00	0.00	-399.71	0.3	0.01	0.00	0.00		
lsrh	#	#	#	#	#	-180.24	7.2	1.00	1.00	1.00		
lsrj	-33.61	7.4	1.00	1.00	1.00	-316.41	7.7	1.00	1.00	1.00		
lqf0	-32.68	0.5	0.36	0.02	0.04	-488.33	0.4	0.08	0.02	0.01		
lqf1	-23.52	0.3	0.06	0.01	0.00	-433.55	0.4	0.15	0.02	0.00		
lby1	-66.34	0.8	1.00	0.16	0.13	-429.61	0.4	0.04	0.01	0.10		
lbyv	-75.76	0.4	0.12	0.04	0.08	-523.03	0.3	0.01	0.01	0.10		
lcbx	-145.60	0.7	1.00	0.02	0.00	-328.40	0.6	1.00	0.06	0.21		
lctc	-52.10	0.3	0.32	0.09	0.00	-237.40	3.6	1.00	0.30	0.31		
ldbm	-31.81	3.0	1.00	1.00	1.00	-427.45	2.5	1.00	0.97	0.73		
ldbl	-37.87	0.9	1.00	0.50	0.30	-439.80	1.7	1.00	1.00	0.11		
ldid	-29.32	4.7	1.00	0.81	0.77	-439.33	4.2	1.00	1.00	1.00		
ldie	-29.00	3.6	1.00	1.00	1.00	-440.59	3.3	1.00	1.00	1.00		
lmrg	-20.56	0.7	1.00	0.32	0.35	-153.33	2.9	1.00	1.00	1.00		
lmrk	-39.76	0.7	1.00	0.24	0.07	-300.46	0.8	1.00	0.19	0.13		
ltng	-12.50	0.4	0.10	0.09	0.14	-135.40	2.0	1.00	0.47	0.48		
ltnk	-14.12	1.6	1.00	1.00	0.20	-132.44	1.7	1.00	1.00	0.34		

	MM						PMF					
	Energy		Error per Query		Resolution		Energy		Error per Query		Resolution	
	Lowest	RMSD	0.5	1	2	Lowest	RMSD	0.5	1	2		
labe	-29.20	0.3	0.08	0.17	0.33	-325.38	3.5	1.00	1.00	0.88		
labf	-14.80	2.8	1.00	1.00	1.00	-319.77	3.3	1.00	1.00	0.94		
lajv	-31.97	9.5	1.00	1.00	1.00	-508.23	9.5	0.56	0.29	0.28		
lajx	-26.61	0.3	0.01	0.00	0.00	-438.00	0.4	0.07	0.02	0.02		
laju	-27.16	0.4	0.37	0.12	0.27	-155.75	1.4	1.00	0.74	0.42		
lajv	-33.38	0.6	0.88	0.16	0.14	-208.26	8.0	1.00	1.00	0.95		
lcbv	-56.66	0.4	0.18	0.02	0.11	-454.62	0.5	0.09	0.01	0.07		
lcbv	-53.94	1.0	1.00	0.39	0.02	-459.59	9.9	1.00	0.58	0.40		
lbbv	-42.07	0.2	0.10	0.04	0.11	-430.16	0.7	1.00	0.04	0.18		
ldbv	-33.39	5.6	1.00	1.00	1.00	-387.95	5.6	1.00	1.00	1.00		
ldvc	-44.16	0.7	1.00	0.07	0.18	-358.13	1.1	1.00	0.61	0.07		
ldvd	-62.95	1.0	1.00	0.43	0.11	-250.64	0.8	1.00	0.70	0.10		
lela	-34.82	0.7	0.76	0.62	0.63	-235.96	7.2	0.47	0.37	0.42		
leld	-29.00	5.9	1.00	1.00	1.00	-183.18	6.0	1.00	1.00	1.00		
lfgk	-40.74	0.4	0.47	0.10	0.03	-535.08	0.5	0.35	0.01	0.01		
lflh	-42.38	0.4	0.43	0.18	0.06	-546.29	0.3	0.07	0.00	0.02		
lrtl	-36.92	0.6	1.00	0.01	0.00	-375.90	0.6	0.82	0.01	0.00		
lrr2	-1.99	0.6	1.00	0.00	0.00	-323.99	0.8	1.00	0.04	0.00		
lsrh	-28.20	2.7	1.00	1.00	1.00	-300.87	7.9	1.00	1.00	1.00		
lsrj	-30.94	8.0	1.00	1.00	1.00	-297.30	7.8	1.00	1.00	1.00		
lqf0	11.73	0.6	0.84	0.08	0.19	-457.83	0.5	0.79	0.04	0.02		
lqf1	13.10	0.4	0.36	0.01	0.00	-454.29	0.4	0.06	0.01	0.00		
lby1	-66.12	0.9	1.00	0.48	0.07	-435.02	0.5	1.00	0.08	0.09		
lbyv	-61.78	4.1	1.00	1.00	1.00	-373.27	0.5	1.00	0.26	0.34		
lcbx	-38.55	5.5	1.00	1.00	1.00	-284.18	5.5	1.00	0.18	0.21		
lctc	-81.40	2.1	1.00	0.17	0.07	-297.93	3.6	1.00	0.82	0.27		
ldbm	-33.04	2.9	1.00	1.00	0.99	-420.31	2.1	1.00	1.00	0.84		
ldbl	-32.92	0.9	1.00	0.27	0.29	-437.15	0.8	1.00	0.24	0.07		
ldid	-27.56	4.9	1.00	1.00	1.00	-434.35	4.3	1.00	1.00	1.00		
ldie	-29.95	3.1	1.00	1.00	1.00	-445.84	2.2	1.00	1.00	1.00		
lmrg	-25.47	3.5	1.00	1.00	1.00	-169.42	3.5	1.00	1.00	1.00		
lmrk	-23.70	3.9	1.00	1.00	1.00	-216.98	5.7	1.00	1.00	1.00		
ltng	-10.98	0.4	0.47	0.08	0.15	-130.27	2.0	1.00	0.80	0.61		
ltnk	-14.52	1.6	1.00	1.00	0.33	-132.16	2.4	1.00	1.00	0.68		

^a Yellow background shows singularity type effects. Blue background shows the minimum resolution required in order to recover solutions without error. Red foreground shows the cases where no docking solutions are found. (#) indicates that all the explored docking solutions have a VDW energy larger than 10 kcal/mol and therefore are not considered in the CEP analysis.³⁰

Table 2. Statistical Significance of Docking Results at Different Resolutions (CEP Results Used To Quantify Docking Success)

potential	resolution (Å)		
	0.5	1.0	2.0
(a) Direct Docking			
MM	0.35	0.05	0.06
MM (VDW only)	0.82	0.07	0.17
PMF	0.74	0.17	0.14
(b) Cross-Docking			
MM	0.78	0.02	0.05
PMF	0.25	0.16	0.08

^a *P*-value for a students' *t*-test,³⁴ comparing the distribution of errors obtained at three resolutions for the different complexes with the distribution of errors obtained when a random number replaces the interaction energy value.³¹

In cross-docking the performance of both functions is, as expected, poorer than with exact coordinates. As summarized in Table 1b, both functions fail to predict correct orientations in about 50% of the cases. To be more specific, using RMS as a figure of merit, and a 1.5 Å cutoff to define success, MM has a success rate of 56% in cross-docking while in the case of PMF the success rate is 41%. There are very few docking studies addressing cross-docking performance. The most exhaustive study to date is from Murray et al.,⁴² using PRO_LEADS, who obtained a 49% success rate in cross-docking all pairs of complexes using a database of complexes for three enzymes (thrombin, thermolysin, and neuraminidase). Thus, a similar deterioration in performance is obtained, suggesting again that deterioration is not due to the rigid docking protocol or implementation of the VDW function. Figure 6 shows the error changes in switching from the exact coordinates case to the cross-docking one. MM seems to be more sensitive to the changes in coordinates than PMF, where even some improvements in docking quality are observed in cross-docking compared with the direct docking, although these are probably the results of chance effects. As can be observed in Figure 4 by looking at the distribution of errors, cross-docking performance for both functions is about the same.

We compared the differences in docking performance with the amount of induced fit observed in protein by measuring the RMSD between pairs of proteins complexed with analogous ligands. These differences were evaluated separately for backbone and side chains. In agreement with Murray et al.,⁴² we observed that backbone RMSD is not negligible in cross-docking (Figure 7). In addition, we did not find any common pattern able to explain the observed differences in docking results (data not shown). Similar degradation in performance with cross-docking has also been reported by Murray et al.,⁴² who found the changes in structural features that lead to misdocking to be highly case dependent.

Linear Discriminant Analysis. LDA calculations were carried out for both functions in order to find physico-chemical descriptors able to classify the results of docking with exact coordinates as "successes" or "failures", as defined by CEP analysis. An error percentage in the function of 15% for MM and 10% for PMF was selected as boundary for group classification, based on the observed distribution of group errors (Figure 4). The stability of the results (i.e., robustness in the models)

was tested by systematically changing the boundary from 10 to 60% of error. Shifts in the boundary definition affected slightly to the quality of the discrimination models, but barely had any influence on the set of descriptors selected (Table 3).

Initial LDA models obtained from the MM function indicated that molecular refractivity (MR) was a major descriptor for classifying docking results. The best model had 53% of discrimination power, a Wilks' Λ of 0.66, and a *P*-value of 0.06. Two different physical effects come into play in the calculation of the MR value: molecular bulk and polarizability (reflecting in turn dispersion interactions). Since we obtained almost equally significant models replacing MR by molecular volume as a descriptor, we suspected that in this case MR was reflecting mainly a volume effect. This was supported by the bimodal distribution of the volume of our ligands in the set (not shown). Thus, in an attempt to decouple both effects, we tested discriminant models including only molecules above 200 Å³, on the basis of our observation of volume distribution. This eliminated only nine molecules from the set. When the volume filter is applied, the best discriminant equation includes SL, LP, L/BS, and L/HB as descriptors (Table 4) and has a discrimination power of 84%, with a Wilks' Λ of 0.40, *F*-test of 7.4, and a *P*-value of 0.0008 (Tables 5 and 6). The significance was also established by randomly scrambling the dependent variable 20 times, giving averaged values of 39% for discrimination power, Wilks' Λ of 0.78, *F*-test of 1.23, and a *P*-value of 0.48 (Table 6). Thus, the models obtained are significant and robust, and, since previously observed volume effects disappear, this suggests that the initial models were suffering from artifacts of the volume distribution in our set.

The discriminant equation (Figure 8a) is dominated by SL and LP, with a *F*-ratio three times larger than that corresponding to the other two variables (Table 4). Analysis of the means and standard deviations of each group (Table 4) led us to conclude that molecules predominantly polar (SL, L/NHB), but with concomitant hydrophobic patches (LP), and not entirely filling the binding site (L/BS), are more prone to failure with the MM function. It seems therefore that polar effects, particularly in the form of hydrogen bonds and charge desolvation, are not fully considered by the MM function, but have significant effects in the likelihood of a successful docking. The effect of ligand desolvation is not surprising, since our simplified MM potential does not include it, but the results stress its important contribution to the physics of the docking process and underline the need for its systematic incorporation in docking force fields. Neglecting electrostatic desolvation contributions was deemed necessary in the past, due to the computational difficulties posed by the numerical solution of the Poisson–Boltzmann equation in the docking problem (solving a system of partial differential equations is a computationally costly procedure), and the lack of satisfactory simplified approaches. Recently, encouraging developments have been presented that allow for less expensive estimations of desolvation penalties.⁴³ On the basis of our results, we expect that introduction of these techniques will considerably improve current docking functions. On the other hand, the effect of hydrogen bonds might not have been expected

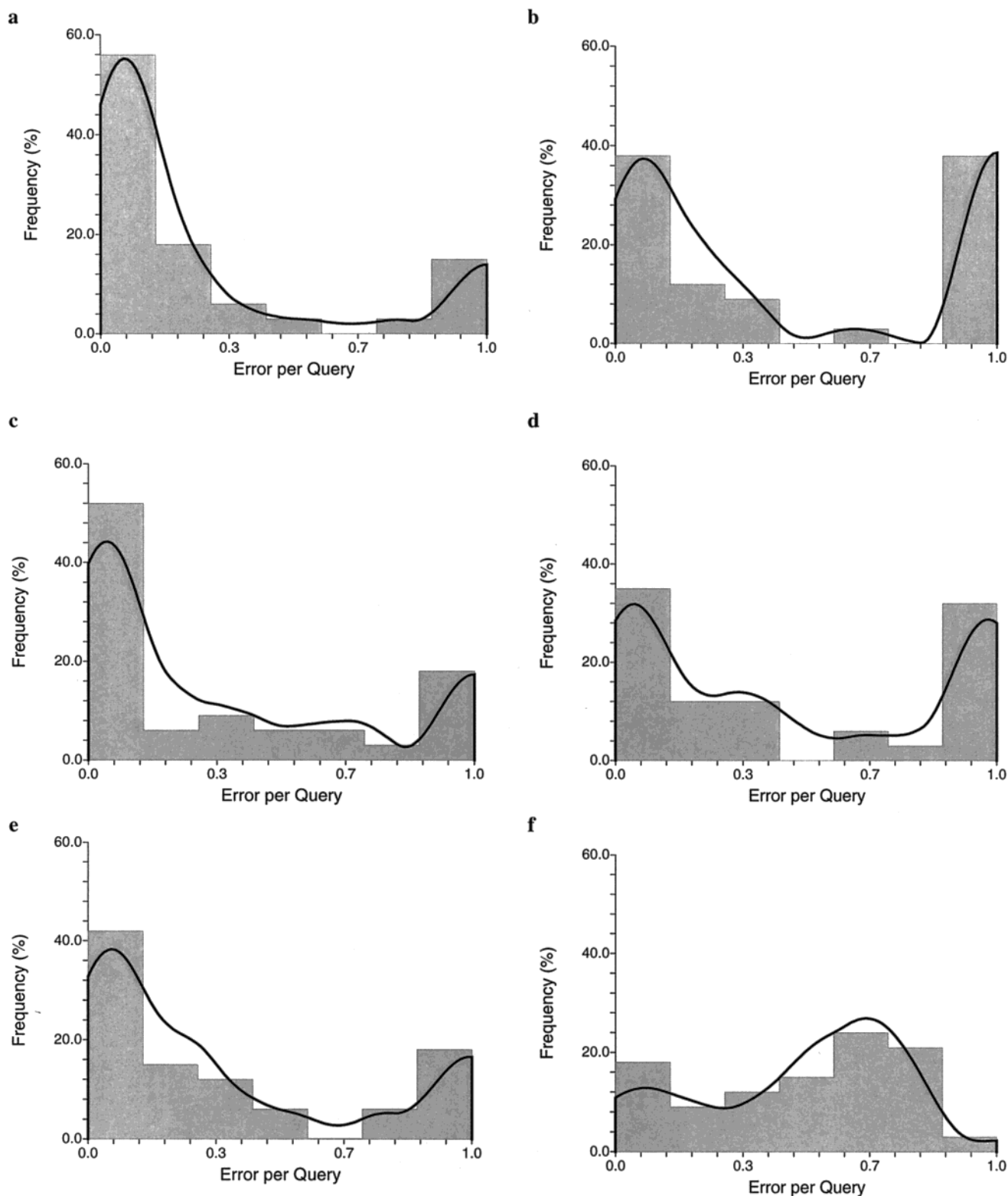


Figure 4. Histograms showing the distribution of errors in our set of complexes in the following cases: (a) MM with exact coordinates; (b) MM with cross-docking; (c) PMF with exact coordinates; (d) PMF with cross-docking; (e) MM without electrostatic term; (f) random energy function.

a priori, since the AMBER force field, like most MM potentials, accounts for hydrogen bond effects through their van der Waals and electrostatic components. However, hydrogen bond interactions are far from being completely understood. Experimentally, it has been shown that hydrogen bonds are highly directional, and in many cases their interaction energies are found to

be nonadditive and nontransferable.⁴⁴ Theoretically, while the main stabilization energy appears to be electrostatic, some other effects are found nonnegligible.⁴⁵ Thus, many-body effects such as polarization are estimated to contribute up to 30% of the total interaction energy in small water clusters,⁴⁶ and about 10% in other clusters.⁴⁷ Charge transfer also appears to be relevant,

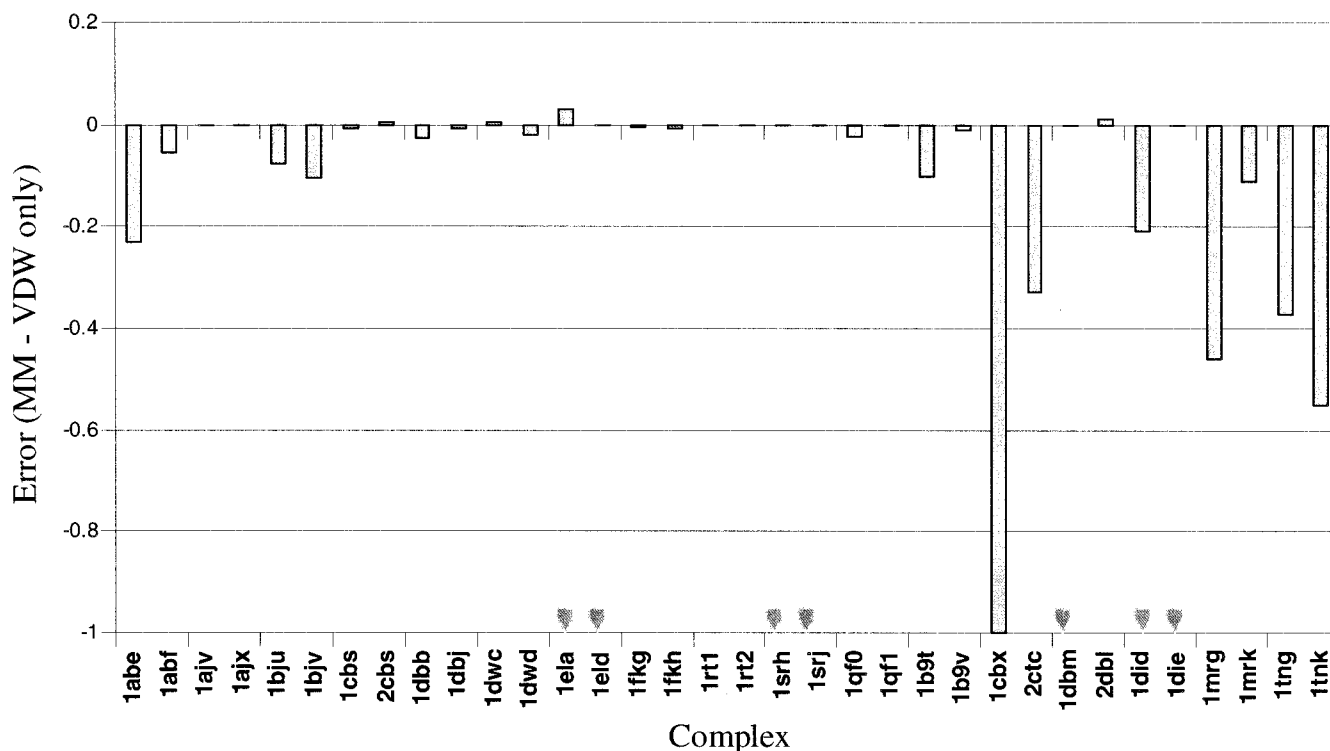


Figure 5. Error differences in going from the MM to the VDW only function with exact coordinates. Complexes where the MM function fails are marked at the bottom of the table.

having a contribution between 20 and 60% of the total interaction energy in small hydrogen-bonded clusters.⁴⁸ For these reasons, hydrogen bonds are difficult to model accurately with a fixed set of monopole-based atomic charges. In fact, it is not the first time it has been found that hydrogen bonds need special treatment when dealing with intermolecular interactions in biomolecules. For example, van der Vaart et al.⁴⁹ in their energy decomposition analysis of folding simulations of a small hydrophobic protein found that inclusion of charge-transfer and polarization effects were crucial in the stabilization of the system, and that charge transfer occurs almost exclusively through hydrogen bonds. Likewise, Hassan et al.,⁴⁵ during their development of implicit solvent models, found that successful folding simulations of small peptides were only possible after special parametrization of interactions involving hydrogen bonds, which needed consideration of directional effects.⁵⁰

In the case of the PMF function, MR appeared as the most important descriptor either with or without volume filter. Although we have not made a specific calculation, we would predict that simpler descriptors such as atom counting would provide good discrimination results. The best models had MR and L/BS as the best descriptors to classify docking results, obtaining a discriminant equation (Figure 8b) with a classification power of 77%, a Wilks' Λ of 0.43, a F -test of 19.8, and a P -value of 0.000 01 (Tables 4–6). Significance was also tested with the scrambled models, which gave an average classification power of only 22%, a Wilks' Λ of 0.95, F -test of 0.67, and a P -value of 0.66 (Table 6). In a further effort to determine the importance of dispersion effects, we made an additional computational experiment in which the MR was divided by the molecular volume, to isolate the term related to refractive index. A new discriminant

analysis run was performed, but this time removing first from the analysis those variables correlated with volume. The model obtained is shown in Table 7, where it can be seen that although poorer in quality than the previously discussed models is, nevertheless, significant. As stated previously, MR has been used traditionally in QSAR as an estimator of bulk and dispersion interactions. Since failures in docking tend to occur for the molecules with the smallest value of MR in our set, the discriminant model suggests that dispersion effects are not entirely correct in the PMF function when these interactions are weak. These interactions are modeled in MM force fields through the attractive part of the Lennard-Jones potential, precisely the part of the potential being replaced by the PMF function. Our analysis suggests that a correct modeling of both excluded volume and attractive interactions is difficult if statistical potentials are directly mixed with physically based force fields; simple addition is not warranted. A possible explanation for this effect can be found in the lack of direct correspondence between elementary interaction energies and the "effective" energies derived from potentials of mean force.⁵¹ Consider a pure hard sphere liquid in which there are no interactions between spheres except for the excluded volume. The PMF between particles in this liquid is nonzero, in fact it has an attractive minima at the contact distance, even though the potential energy is zero beyond the contact distance. Using the derived PMF and putting back directly the excluded volume would not reproduce the real energy surface for the system, it would actually create a distorted surface, with a tendency to maximize the number of contacts between spheres, i.e., it would create an artificial surface tension in the system. Couplings of this sort have been observed in computer simulations of liquids⁵¹ as well as in lattice models of

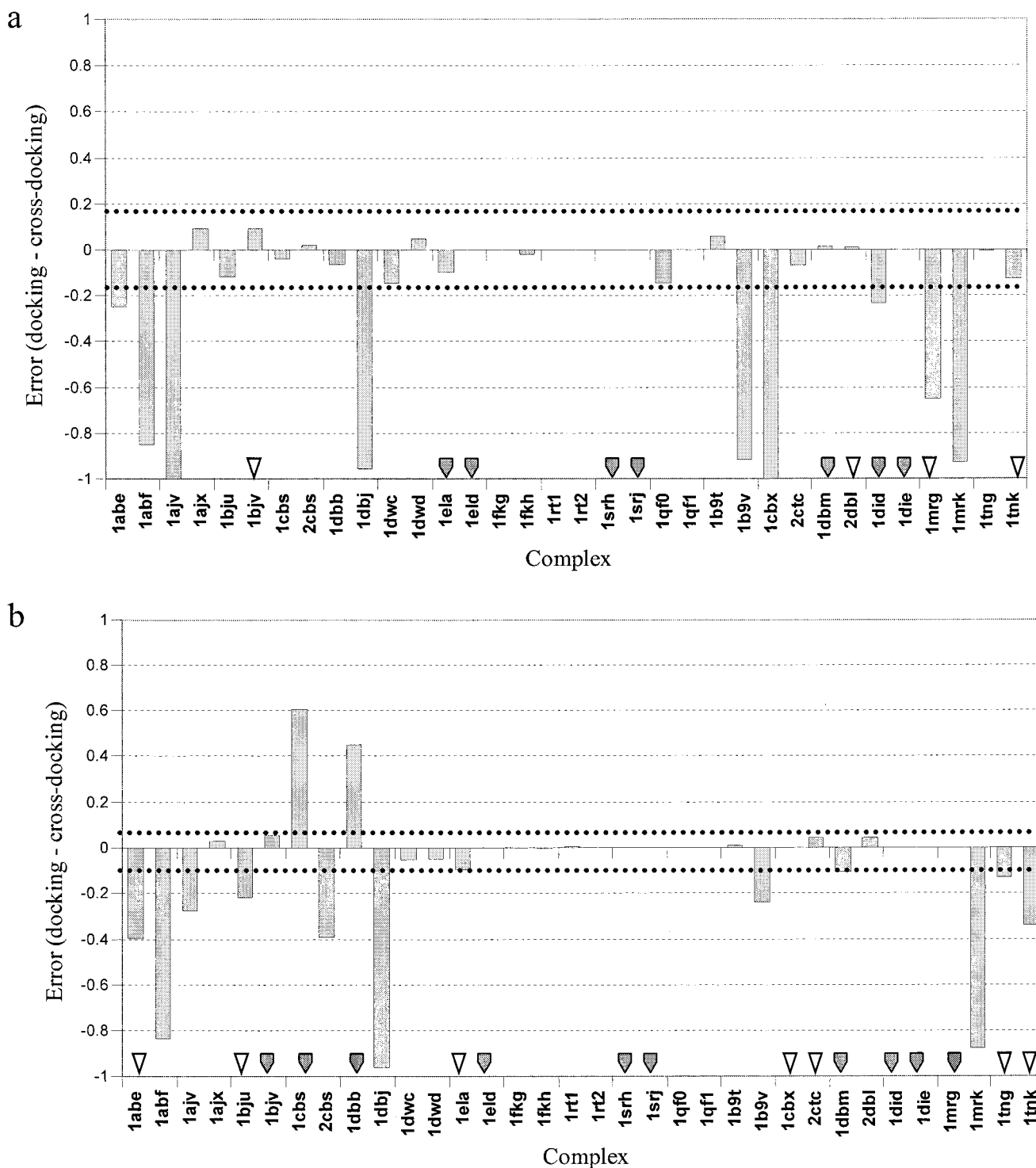


Figure 6. Error differences in going from “exact” docking to cross-docking. Dotted lines show the cutoffs selected to define the success of the function in each case (see Linear Discriminant Analysis). Complexes for which the docking function fails in the “exact” coordinates case are highlighted in the figure with unfilled (20–50% error) or filled (50–100% error) marks. (a) MM; (b) PMF.

proteins,^{51,52} and it can be expected that similarly complex effects also operate in protein–ligand systems.

It is interesting to note that L/BS is a descriptor playing opposite effects within each scoring function (Table 3). MM has a tendency to fail if the ligand is not filling the binding site, which seems to be related, as discussed previously, to inaccuracies in treatment of hydrogen bonds and desolvation effects, particularly when the molecule binds at the protein/solvent interface. On the other hand, PMF fails when the ligand has

extensive steric interactions with the protein, suggesting a poor description of dispersion effects (hard core repulsion is treated in the same way with both potentials), as is indeed pointed out in the analysis by the presence of MR. However, and possibly due to a better consideration of hydrogen bond interactions, PMF allows small molecules (with respect to the binding site) to be correctly docked. This is in agreement with observations by Mitchell et al.,⁵³ who showed a good correspondence between hydrogen bond energies as

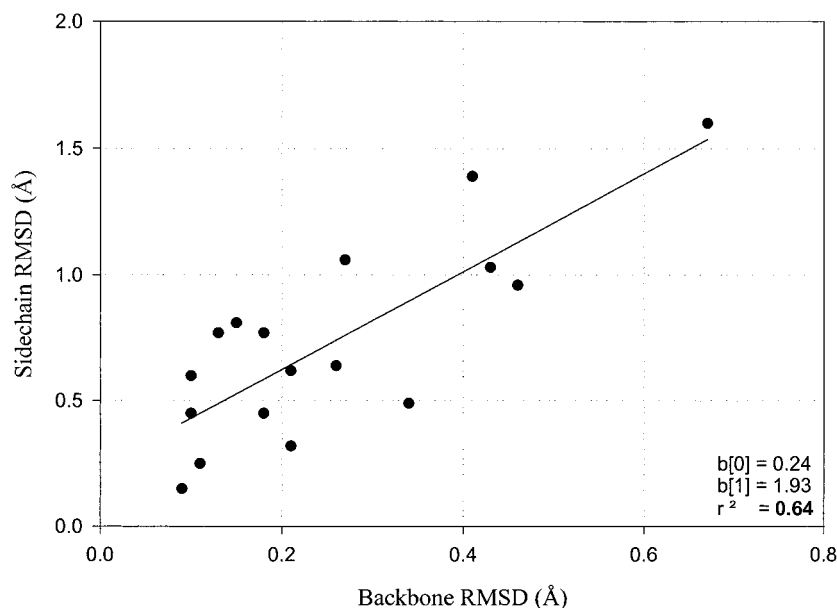


Figure 7. Correlation between the RMSD observed in backbone and in side chains for every pair of protein structures in our set. Residues in a 8 Å sphere around the binding site of the ligand were selected for the calculation. A reasonable correlation is found ($r^2 = 0.64$), with a slope of 1.93.

determined by quantum chemical methods and those determined by statistical potentials.

Cross-docking results were not evaluated by linear discriminant analysis because the number of members on each family was not sufficient to allow the derivation of a reliable discriminant model. This imbalance precludes obtaining reliable information from the LDA analysis, because it breaks down one of the basic requirements to evaluate the significance of the LDA results: the assumption of equal variances and covariances for both groups of data. We stress that the imbalance does not seem to be due to artifacts created by the rigid body approach, but rather it seems to be the result of the considerably larger overlap of nativelike and nonnativelike energy spectra in the cross-docking case. We note that this phenomenon has been observed also by Murray et al.⁴² in their cross-docking experiments using PRO_LEADS, and thus it does not seem to be related to peculiarities of our docking implementation.

Conclusions

A procedure to quantitatively evaluate docking functions has been presented, which has the important feature of capturing information about binding energy landscapes. To be able to exhaustively search the relevant conformational space for a statistically significant number of cases, docking has been restricted to the rigid body case, and the six relevant degrees of freedom have been conveniently discretized. The method has been used to study MM and statistical potentials. Linear discriminant analysis has then been applied to these numerical results in order to find physicochemical descriptors able to separate successes from failures, with the purpose of obtaining deeper physical insight about the main energetic factors at play in the docking process. We caution that the docking results, and consequently our linear discriminant analysis, can be somewhat dependent upon details of the implementation of the potentials, particularly the VDW terms, as

Table 3. Effect of Changing the Group Classification Cutoff in the Linear Discriminant Analysis Results

cutoff (%)	count		linear discriminant anal				
	fail	OK	descrip- tor	descriptor power (%)	Wilks' Λ	F -test	P -value
10	10	15	(a) MM (with Volume Filter, See Text)				
			SL	76	0.41	6.9	1.0×10^{-3}
			LP				
			LIBS				
			L/HB				
			SL	84	0.4	7.4	8.0×10^{-4}
			LP				
			L/BS				
			L/HB				
			SL	100	0.35	9.1	2.0×10^{-4}
LP							
L/BS							
L/HB							
SL	100	0.35	9.0	1.0×10^{-4}			
LP							
L/BS							
L/HB							
SL	100	0.34	9.5	2.0×10^{-4}			
LP							
DHB							
MR							
SL	100	0.48	5.4	4.0×10^{-3}			
LP							
PL							
DHB							
10	20	14	(b) PMF				
			MR	77	0.43	19.8	1.0×10^{-5}
			L/BS				
			MR	71	0.49	16.0	1.0×10^{-5}
			L/BS				
			MR	53	0.56	12.1	1.0×10^{-4}
			L/BS				
			MR	41	0.67	31.0	2.5×10^{-3}
L/BS							
MR	41	0.84	5.9	2.1×10^{-2}			
L/BS							
MR	53	0.89	3.7	6.3×10^{-2}			
L/BS							

^a Descriptors: SL, solvation; LP, log P ; L/BS, percentage of the free binding site volume occupied by the ligand volume; L/HB, surface area of the ligand in Å²/hydrogen bond; MP, polarizability; DHB, all possible ligand hydrogen bonds – ligand hydrogen bonds accomplished on enzyme; MR, molar refractivity.

Table 4. Description of the Best Linear Discriminant Models Found

descriptor ^a	mean			std dev			significance	
	fail	OK	overall	fail	OK	overall	<i>F</i> -ratio ^b	<i>P</i> -value
(a) MM (with Volume Filter, See Text)								
SL	80.79	41.85	55.87	19.12	37.45	36.91	8.55	1.0 102
LP	5.84	4.06	4.7	3.12	2.78	2.97	11.28	3.0 103
L/BS	29.01	35.66	33.27	8.84	7.7	8.58	3.69	7.0 102
L/HB	58.3	82.37	73.71	12.5	48.97	41.11	4.55	5.0 102
(b) PMF								
MR	64.8	116.63	86.14	28.51	26.16	37.52	39.01	1.0 106
L/BS	38.45	36.34	37.58	12.69	8.08	10.94	6.02	2.0 102

^a Descriptors (see Methods): SL, solvation; LP, log *P*; L/BS, percentage of binding site volume occupied by ligand; L/HB, surface area of the ligand in Å²/hydrogen bond; MR, molar refractivity. ^b *F*-ratio is used to assess the significance of removing the corresponding descriptor.³⁴

Table 5. Classification Tables of the Linear Discriminant Functions Obtained with Each Docking Function: MM after Volume Filter (See Text for Details)

		predicted					
		MM			PMF		
obsd	fail	fail	OK	total	fail	OK	total
	fail	8	1	9	18	2	20
	OK	1	15	16	2	12	14
	total	9	16	25	20	14	34

Table 6. Comparison of the Best Linear Discriminant Model Found for Each Docking Function (Observed) with the Mean Result from Scrambled Models (Randomized) (Standard Deviations Also Shown in Parentheses)

potentials		descriptor power (%)	Wilks' Λ	<i>F</i> -test	<i>P</i> -value
MM	observed	84	0.40	7.4	8.0 104
	randomized	39(18)	0.78(0.16)	1.2(0.8)	0.48(0.30)
PMF	observed	77	0.43	19.8	1.0 105
	randomized	22(8)	0.95(0.05)	0.7(1.0)	0.66(0.27)

it has been shown that the VDW implementation can affect docking performance.⁵⁴ There is the risk that rigid body docking, together with the use of an underlying grid and a hard VDW potential could in some cases penalize nativelike solutions, particularly in complexes displaying a strong complementarity. In fact, this is likely the case of the 1srh complex. However, comparison of our results, particularly with the MM potential, with other state-of-the-art docking codes, suggests that

these effects have played a minor role. In any case, further studies with slightly modified conditions are required to fully confirm our conclusions, which we describe in what follows.

We find that the global minimum (if properly found) is a necessary, but not sufficient, condition to judge a docking function. Deeper insight, particularly about the stability and kinetic accessibility of the solution, is obtained if the complete docking energy surface is taken into account. In this regard, the CEP plot seems to be an adequate tool. Using CEPs, we find that MM force fields seem to be superior to statistical potentials when exact receptor coordinates are used. However, none of the functions appear good enough to resolve effectively the docking problem when there are coordinate shifts in the receptor, as occurs in cross-docking. In this case both functions show about the same performance.

Linear discriminant analysis studies of CEP results allowed us to conclude that a sophisticated treatment of desolvation effects and hydrogen bond interactions appears to be required in the implementation of MM potentials for docking. On the other hand, a detailed consideration of steric interactions, with a careful treatment of dispersive forces, seems to be needed when using statistical potentials derived from a structural database. Part of the difficulty in supplementing PMF potentials with physically based terms could reside in the fact that even "perfect" statistical potentials cannot recover exactly the underlying physical potential energy

Table 7. Best Linear Discriminant Model Obtained by Excluding Volume Related Descriptors (See Text for Details)

(a) Descriptors Selected and Influence in the Model								
descriptors ^a	mean			std dev			significance	
	fail	OK	overall	fail	OK	overall	<i>F</i> -ratio ^b	<i>P</i> -value
MR/VO	0.27	0.29	0.29	0.05	0.01	0.04	1.6	2.1 101
SL	71.83	42.94	57.38	60.45	35.89	51.10	3.21	8.3 102
HB	4.29	6.41	5.35	1.83	2.48	2.40	7.57	9.9 103
(b) Classification Table								
		predicted PMF						
		fail	OK	total	fail	OK	total	
obsd	fail	12	5	17	OK	3	14	17
	total	15	19	34				
(c) Discriminant Performance								
potentials		descriptor power (%)		Wilks' Λ	<i>F</i> -test	<i>P</i> -value		
PMF		53		0.68	4.5	9.7 × 10 ⁻²		

^a Descriptors: MR/VO, molar refractivity divided by volume; SL, solvation; HB, all possible ligand hydrogen bonds. ^b *F*-ratio is used to assess the significance of removing the corresponding descriptor.

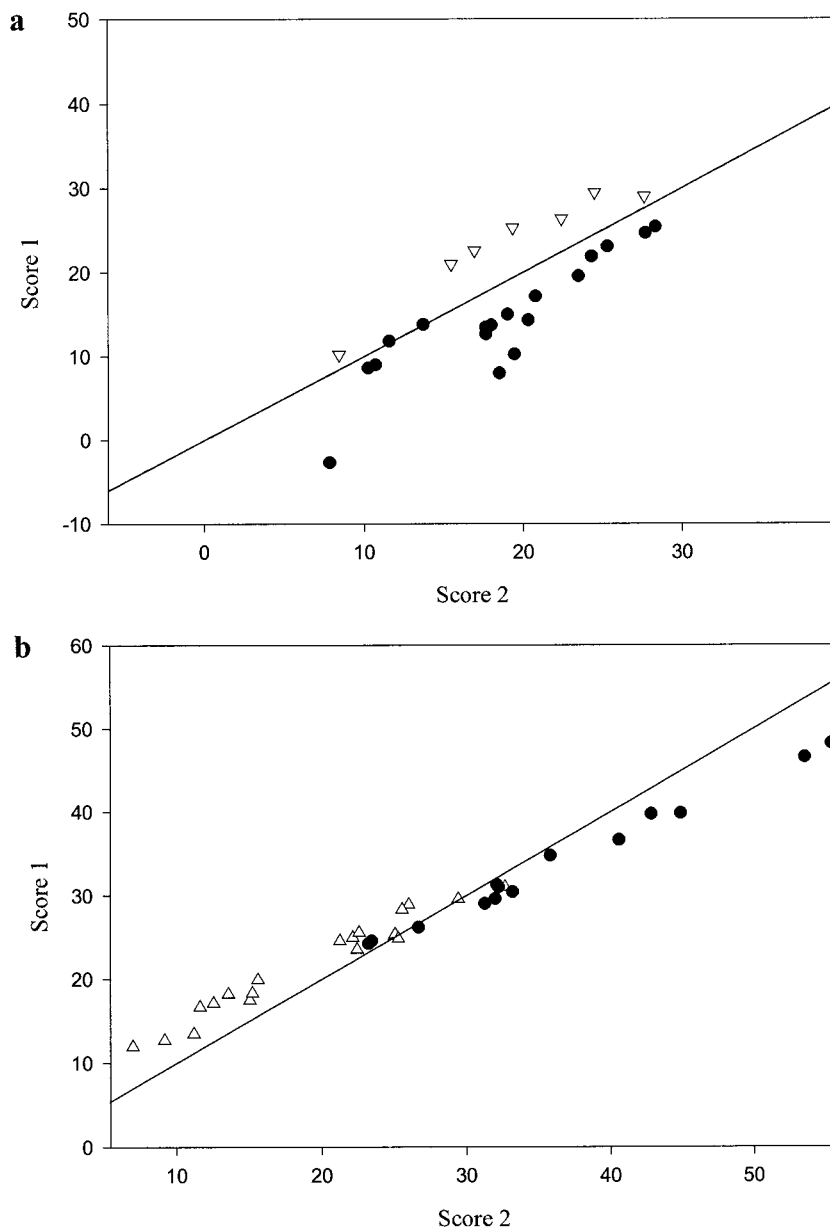


Figure 8. Linear discriminant function score plots of docking results. OK (circle), Fail (triangle). The line shows points with identical probability to be classified as any of the groups. (a) MM (volume filter); (b) PMF.

used to generate them. This study however suggests that both functions contain different but, in some ways, complementary information, which could in principle be mixed in the design of better docking potentials. For example, a correction of the hydrogen bond term by incorporating statistical potentials in MM seems obvious.

Finally, our results could be fruitfully exploited in the design and filtering of molecular databases in virtual screening or combinatorial docking. In this paper we show that it is possible to establish quantitative relationships between likelihood of docking success using a given docking function and simple physicochemical descriptors, readily computable from the molecular structure of the ligand alone or in combination with the active site being targeted, but in any case prior to any docking calculation. This opens the possibility of biasing virtual library design or filtering chemical databases by selecting those molecules with high probability of docking success according to the values of their physico-

chemical descriptors, and integrates nicely with recent proposals to design and tailor compound libraries on the basis of other parameters important in drug discovery, for instance predicted solubility, membrane permeability, or metabolic stability, using physicochemical descriptors.^{55,56} Also, and since a correct docking mode is a prerequisite for being able to predict correctly binding affinities, this is an interesting alternative to other proposals aimed at increasing the hit rate of active molecules, such as those based on the combination of scoring functions.^{57,58} It can be expected that molecular diversity of a library created through selection of highly probable dockings coming from different (separated) docking functions will be larger than that obtained by applying a single consensus scoring function. We will explore these issues in future research.

Acknowledgment. We thank Dr. Ingo Muegge for providing us with the statistical potentials needed in our studies, and Dr. Juanjo Lozano for helping us in

implementing them within the docking program. We also thank Dr. Marc Ceruso and Dr. Pere Constans for insightful discussions, and Dr. Federico Gago for carefully reading the manuscript. Support for this research has been provided by MSSM start-up funds (A.R.O.). C.P. was partially supported by a predoctoral fellowship from the Spanish Ministry of Education.

References

- Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening—an overview. *Drug Discov. Today* **1998**, *3*, 160–178.
- Knegtel, R. M. A.; Grootenhuys, P. D. J. Binding affinities and nonbonded interaction energies. *Perspect. Drug Discov. Des.* **1998**, *9*, 99–114.
- Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caffisch, A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins* **1999**, *37*, 88–105.
- Zou, X.; Sun, Y.; Kuntz, I. D. Inclusion of solvation in ligand binding free energy calculations using the Generalized-Born model. *J. Am. Chem. Soc.* **1999**, *121*, 8033–8043.
- Abagyan, R.; Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **1994**, *235*, 983–1002.
- Verkhivker, G.; K. K. A.; Freer, S. T.; Villafranca, J. E. Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Eng.* **1995**, *8*, 677–691.
- Wallqvist, A.; Jernigan, R. L.; Covell, D. G. A preference-based free-energy parametrization of enzyme-inhibitor binding. Applications to HIV-1-protease inhibitor design. *Protein Sci.* **1995**, *4*, 1881–1903.
- DeWitte, R. S.; Ishchenko, A. V.; Shakhnovich, E. I. SMOG: De novo design method based on simple, fast, and accurate free energy estimates. 2. Case studies in molecular design. *J. Am. Chem. Soc.* **1997**, *119*, 4608–4617.
- Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791–804.
- Mitchell, J. B. O.; Laskowski, R. A.; Alex, A.; Thornton, J. M. BLEEP—Potential of Mean Force Describing Protein-Ligand Interactions: I. Generating Potential. *J. Comput. Chem.* **1999**, *20*, 1165–1176.
- Ha, S.; Andreani, R.; Robbins, A.; Muegge, I. Evaluation of docking/scoring approaches: A comparative study based on MMP3 inhibitors. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 435–448.
- Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.
- Dixon, J. S. Evaluation of the CASP2 docking section. *Proteins* **1997**, *Suppl.* 198–204.
- Totrov, M.; Abagyan, R. Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins* **1997**, *Suppl.* 215–220.
- Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Arthurs, S.; et al. Deciphering common failures in molecular docking of ligand-protein complexes. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 731–751.
- Diller, D. J.; Verlinde, C. L. A critical evaluation of several global optimization algorithms for the purpose of molecular docking. *J. Comput. Chem.* **1999**, *20*, 1740–1751.
- Brenner, S. E.; Chothia, C.; Hubbard, T. J. Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6073–6078.
- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. et al. A Second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- Johnson, R. A.; Wichern, D. W. *Applied Multivariate Statistical Analysis*, 4th ed.; Prentice-Hall: Upper Saddle River, NJ, 1998.
- Case, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 5*; University of California: San Francisco.
- SYBYL *Molecular Modeling Software*, v. 6.6; Tripos Associates, I: St. Louis, MO.
- Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity. A rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228.
- Taylor, J. S.; Burnett, R. M. DARWIN: a program for docking flexible molecules. *Proteins* **2000**, *41*, 173–191.
- Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- David, L.; Luo, R.; Gilson, M. K. Ligand-Receptor Docking with the Mining Minima Optimizer. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 157–171.
- Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
- Makino, S.; Kuntz, I. D. Automated flexible ligand docking method and its application for database search. *J. Comput. Chem.* **1997**, *18*, 1812–1825.
- Muegge, I.; Martin, Y. C.; Hajduk, P. J.; Fesik, S. W. Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein. *J. Med. Chem.* **1999**, *42*, 2498–2503.
- Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An all atom force field for simulations of proteins and nucleic acids. *J. Am. Chem. Soc.* **1986**, *7*, 230.
- Bliznyuk, A. A.; E. Gready, J. Simple method for locating possible ligand binding sites on protein surfaces. *J. Comput. Chem.* **1999**, *20*, 983–988.
- Majeux, N.; Scarsi, M.; Caffisch, A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins* **2001**, *42*, 256–268.
- Lattman, E. E. Optimal sampling of the rotation function. *The molecular replacement method*; Gordon and Breach, Science Publishers Inc.: New York, 1972; pp 179–185.
- Nelder, J. A.; Mead, R. A simplex method for function minimization. *Computer J.* **1965**, *7*, 308–313.
- Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical recipes. The art of scientific computing (FORTRAN)*; Cambridge University Press: Cambridge, U.K., 1989.
- Bodor, N.; Gabanyi, Z.; Wong, C. K. A new method for the estimation of partition-coefficient. *J. Am. Chem. Soc.* **1989**, *111*, 3783–3786.
- Ghose, A. K.; Pritchett, A.; Crippen, G. M. Atomic physicochemical parameters for three-dimensional structure directed QSAR III: Modeling hydrophobic interactions. *J. Comput. Chem.* **1988**, *9*, 80–90.
- Miller, K. J. Additivity methods in molecular polarizability. *J. Am. Chem. Soc.* **1990**, *112*, 8533.
- HyperChem*; 6th ed.; HyperCube, Inc.: Waterloo, Canada.
- Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144–1149.
- Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **1995**, *8*, 127–134.
- Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins* **1998**, *33*, 367–382.
- Murray, C. W.; Baxter, C. A.; Frenkel, A. D. The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 547–562.
- Aroda, N.; Bashford, D. Solvation energy density occlusion approximation for evaluation of desolvation penalties in biomolecular interactions. *Proteins* **2001**, *43*, 12–27.
- Jeffrey, G. A. *An introduction to hydrogen bonding*; Oxford University Press: Oxford, U.K., 1997.
- Hassan, S. A.; Guarnieri, F.; Mehler, E. L. A general treatment of solvent effects based on screened Coulomb potentials. *J. Phys. Chem. B* **2000**, *104*, 6478–6489.
- Milet, A.; Moszynski, R.; Wormer, P. E. S.; Avoird, A. v. d. Hydrogen bonding in water clusters: Pair and many-body interactions from symmetry-adapted perturbation theory. *J. Phys. Chem. A* **1999**, *103*, 6811–6819.
- Vaart, A. v. d.; Merz, K. M., Jr. Divide and conquer interaction energy decomposition. *J. Phys. Chem. A* **1999**, *103*, 3321–3329.
- Stone, A. J. Computation of charge-transfer energies by perturbation-theory. *Chem. Phys. Lett.* **1993**, *211*, 101–109.
- Vaart, A. v. d.; Bursualaya, B. D.; Brooks, C. L., III; Merz, K. M., Jr. Are many-body effects important in protein folding? *J. Phys. Chem. B* **2000**, *104*, 9554–9563.
- Hassan, S. A.; Guarnieri, F.; Mehler, E. L. Characterization of hydrogen bonding in a continuum solvent model. *J. Phys. Chem. B* **2000**, *104*, 6490–6498.
- Shan, Y.; Zhou, H. X. Correspondence of potentials of mean force in proteins and in liquids. *J. Chem. Phys.* **2000**, *113*, 4794–4798.
- Zhang, L.; Skolnick, J. How do potentials derived from structural databases relate to 'true' potentials? *Protein Sci.* **2000**, *7*, 112–122.
- Mitchell, J. B. O.; Laskowski, R. A.; Alex, A.; Thornton, J. M. BLEEP—Potential of Mean Force Describing Protein-Ligand Interactions: II. Calculation of Binding Energies and Comparison with Experimental Data. *J. Comput. Chem.* **1999**, *20*, 1177–1185.

- (54) Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L., III. Assessing energy functions for flexible docking. *J. Comput. Chem.* **1998**, *19*, 1612–1622.
- (55) Crivori, P.; Cruciani, G.; Carrupt, P. A.; Testa, B. Predicting blood–brain barrier permeation from three-dimensional molecular structure. *J. Med. Chem.* **2000**, *43*, 2204–2216.
- (56) Labute, P. A widely applicable set of descriptors. *J. Mol. Graph. Mod.* **2000**, *18*, 464–477.
- (57) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109.
- (58) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759–4767.
- (59) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. E.; Brice, M. D.; et al. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535–542.

JM010141R