BHB: A Simple Knowledge-Based Scoring Function to Improve the Efficiency of Database Screening

Miklos Feher,*,# Eugen Deretey, and Samir Roy

SignalGene Inc., 2-335 Laird Road, Guelph, Ontario N1G 4P7, Canada

Received February 20, 2003

A new knowledge-based scoring function was developed in this work to facilitate the rapid ranking of ligands in databases. The acronym of the method is BHB based on the descriptors it utilizes: buriedness, hydrogen bonding, and binding energy. Receptor buriedness is a measure of how well molecules occupy the binding pocket in comparison to known high-affinity ligands or, alternatively, whether they have contact with identified residues in the pocket. The possibility of hydrogen bond formation is checked for selected residues that are recognized as being important in the binding of known ligands. The approximate binding energy is calculated from the thermodynamic cycle using the optimized bound and free solvent conformations of the ligand−receptor system. The information necessary for the scoring function can ideally be gleaned from the 3D structure of the receptor−ligand complex. Alternatively, the descriptors can be derived from the 3D structure of the unbound receptor, provided this receptor has a known ligand that binds to the given site with nanomolar activity. We show that the new scoring functions provide up to 12 times improvement in enrichment compared to the popular commercial docking program GOLD.

INTRODUCTION

Assessing ligand orientation and the strength of ligand−receptor interactions has become an essential component of modern drug discovery. There are many methods, approaches, and programs available to achieve this goal. In the process of investigating ligand binding, two tasks are normally distinguished:[1] finding the best conformation and orientation of the ligand in the binding pocket (usually referred to as the "docking problem"), and quantifying the quality of fit between the ligand and receptor (the so-called "scoring problem"). Commercial docking programs are generally built to perform both tasks. Different approaches for ligand docking have been recently reviewed.[2] A critical issue for current docking software is the quality and accuracy of the scoring functions available.[2−5] There are several popular scoring functions, such as ChemScore,[6] DrugScore,[7] PMF,[8] Score,[9] and Ludi.[10] These are applied either individually or in combination in a consensus scheme.[4,5,11,12]

Both the docking and the scoring steps utilize scoring functions, although these have very different requirements.[3] Scoring in the docking step needs to be fast and error tolerant, optimized to prioritize different poses of the same ligand correctly. Hence, these are not necessarily expected to give accurate affinity predictions or to rank different ligands. In contrast, the aim at the scoring stage is to obtain some measure of the binding energy or a separation of actives and inactives. Despite this, most commercial docking programs apply the same scoring function to both steps (e.g., refs 13−16). The applied scoring functions in these programs are optimized for proper ligand placement, good affinity prediction, separation of actives and inactives, or some kind of compromise between these requirements. Our experience has shown that programs optimized to be efficient in separating binders and nonbinders often fail to dock many of the active ligands correctly.

We propose that databases can be better enriched with actives if different scoring functions are applied for the two steps. In the first (docking) step, molecules have to be docked in the correct position. If a ligand is not docked correctly, the score based on this incorrect position cannot be expected to be correct.[17] Our experiments showed that the GOLD docking software was capable of reproducing the crystal pose of cocrystallized ligands in all cases tested. This is not surprising because its scoring function was optimized to predict binding positions[13,18−20] and was shown to outperform other packages in this regard.[1,5] Hence, GOLD[13] was selected as our docking engine with its built-in fitness function used for scoring different ligand positions. For the second (scoring) step, we used the simple scoring function developed in this work with the aim of optimum database enrichment.

Most of the cited scoring functions attempt to assess ligand−receptor interactions by estimating the free energy of their interaction. Our scoring function is knowledge-based in the sense that some information about receptor-binding in the specific receptor system is incorporated into it. The major factors in our scoring function are two important properties of the ligand−receptor system: the full spatial complementarity of the ligand and the receptor (described as "buriedness") and the occurrence of hydrogen bonding with a list of specific residues. Both of these properties ensure that ligands binding on the outside surface of the protein are excluded. Hence, it was unnecessary to include any specific measure for surface accessibility, such as the

* Corresponding author phone: (519) 823-9088; fax: (519) 823-9401; e-mail: miklos.feher@signalgene.com.

Current address: Neurocrine Biosciences 10555 Science Center Drive, San Diego, CA 92121-1102. Phone: (858)658-7600; fax: (858)658-7601; e-mail: mfeher@neurocrine.com.

10.1021/ci030006i CCC: $25.00
Published on Web 07/01/2003

descriptor described by Stahl.[21] In addition, our scoring function also includes general terms describing van der Waals and hydrogen bonding interactions between the ligand and the receptor that are components of the GOLD scoring function. Our simple scoring function can then be refined by incorporating an approximate binding energy to further improve enrichment. The aim of this work was to test the viability of this approach, as well as the improvement it brings in comparison to using a commercial docking program such as GOLD.

METHODS

The small molecule modeling tasks in this work were performed with the Molecular Operating Environment (MOE) modeling suite.[22] Molecules were first protonated/deprotonated according to their pK_a in an aqueous environment (using MOE's wash functionality, manually adjusted for certain functional groups to achieve agreement with predicted pK_a values[23]). The structures were then optimized using the MMFF94 force field[24] including GB/SA solvation terms.[25,26] The protein structures were downloaded from the Brookhaven database.[27] Simple structural errors (e.g., residues with missing or extra atoms and missing peptide bonds) in the protein were repaired and then the whole protein was protonated/deprotonated.[22] In the case of histidine residues on the surface of the binding pocket, both neutral tautomeric forms were considered. Molecules were initially docked to both of the corresponding protein structures to establish whether either of these alone would explain the binding of known actives better. For the estrogen receptors 1ERR and 3ERT it was found that this does occur, as the donated hydrogen in one case was much less accessible to the ligands.

Molecules were docked with the GOLD program[13,20] using 20 independent runs with 10 000 genetic algorithm (GA) operations on a single population of 100 individuals. Weights for crossover, mutation, and migration were set to 100, 100, and 0, respectively. (These parameters corresponded to the "7−8 times speed-up" setting in GOLD.) The active site radius from the center of mass of the ligand was chosen to be 20 Å in all cases. These parameters represented a compromise between quality of ligand placement and speed. In the case of protein kinases, the new version of the docking program, GOLD 2.0, was also tested. This allows the specification of selected hydrogen bonding contacts via the new "Protein H-Bond Constraint" option. In all cases, the constraint weight was set at 10 for each specified hydrogen bond. This approach considers only those molecules, conformations, and poses in which such contacts are satisfied. For the 1FVV and 1BL7 receptor, using constraints was the only way of achieving correct ligand placement for all the reference molecules.

Novel Descriptors for the Binding Site. To characterize ligand−receptor binding, novel descriptors were developed in this work. These fall into three categories: properties characterizing the ligand or receptor surfaces, predictors of hydrogen bonding, and methods for approximating the binding energy. It was realized early in the project that none of these sets of descriptors are meaningful if molecules are docked in incorrect binding modes. Our docking engine, GOLD,[13] placed all known actives into the pocket correctly for all receptor systems investigated. On the basis of this validation test it was inferred that other potential (but unknown) actives would probably also be positioned correctly.

Having studied the fitness scores generated by GOLD during the docking process, it was found that molecules with fitness scores less than 30 either failed to dock into the pocket correctly or had very high strains. Hence, only molecules with fitness scores above 30 were considered further. The docked molecules were then evaluated using different descriptors that were developed using MOE's SVL language.[22]

Our initial idea for a "ligand buriedness" function was based on the method of Stahl and Böhm[28] to describe the absolute or percentage surface area of the ligand buried into the receptor pocket. However, instead of devising a grid on the protein surface and counting grid points near the ligand, we utilized environment-dependent surface functions in SVL.[22] Those areas of the ligand that had residue atoms present at a distance $\leq$4.5 Å from ligand atoms were considered as buried. The molecular surface was calculated in the presence and absence of the receptor using closely packed spheres of 0.3 Å radius to define the accessible surface. The ligand buriedness ratio was obtained as the ratio of the buried ligand surface to the full surface of the molecule.

Similarly to Stahl and Böhm,[28] we found this ligand buriedness to be a powerful filter for molecules that are not fully immersed into the pocket. However, it was found that this function is less useful as a component of a scoring function because fully buried molecules received high scores even if they were docked outside the known binding site. On the other hand, properly binding ligands that reached into the solvent were unjustifiably discriminated against. To resolve these issues, the idea of "receptor buriedness" (in contrast to "ligand buriedness") was developed in this work. We considered as "buried" those areas of the receptor that had ligand atoms within a specified distance. The accessible surface area was determined by generating closely packed spheres on the surface and calculating their total exposed area by triangulation.[22] Based on the estrogen and androgen receptor examples, the best results could be obtained if the distance and sphere radius were 4.5 Å and 0.3 Å, respectively. It is important to note that the sphere radius of 0.3 Å was obtained from the optimization of enrichment, and hence the calculated surface area is different from the water accessible area that would have been obtained by using 1.4 Å as the sphere diameter. In the cases of both the ligand buriedness and receptor buriedness, increasing the sphere size to measure surface area tended to smear out differences between compounds of different activity, whereas smaller sphere sizes tended to overemphasize insignificant receptor features.

Another surface-area measure tested as a descriptor of ligand−receptor binding was the hydrophobic contact surface. It was defined as the receptor surface area of hydrophobic side-chains in proximity (i.e., within 4.5 Å) to hydrophobic ligand atoms. The hydrophobic contact surface ratio was calculated as the ratio of this quantity to the full receptor contact surface (i.e., the buried receptor surface), using the same distance and sphere radius terms as for the other functions. Note that this hydrophobic contact surface is not equivalent to hydrophobic terms often applied in

scoring functions in the literature. In general, hydrophobic interactions arise between nonpolar ligand and receptor atoms. On the receptor side, these nonpolar atoms are often in hydrophobic side-chains but nonpolar backbone atoms also contribute to such interactions. The descriptor tested in this work incorporated only interactions between hydrophobic side-chain residues and nonpolar ligand atoms. This was done as general aspects of hydrophobicity were already described by other terms (such as the buriedness and binding energy terms).

The best description for the occurrence of hydrogen bonds was also investigated. To preclude hydrogen bonding to residues outside the pocket, a list of desired hydrogen bond contacts was determined for each receptor. This list comprised the hydrogen bonding residues that were thought to be important for all known active ligands based on the crystal structure, structure−activity relationship data, and mutagenesis studies from the literature. Hydrogen bonds were detected automatically using published geometric criteria,[29] considering the distance and angular requirements between different atom and residue types. The detection routine was based on MOE's pro_Contacts function. The total hydrogen bond count was divided by the number of considered hydrogen bonding residues, so that in case of a higher number of residues, their contribution would not overwhelm the total score.

The last set of novel filter functions approximated the binding energy of molecules. The binding energy was calculated from the energies of the complex, the unbound ligand in solution, and the unbound solvated rigid receptor: $E_{\text{bind}} = E^{\text{cpx}} - (E^{\text{lig}} + E^{\text{prot}})$. The full energy terms and their derivation has been given elsewhere.[30] Because the GOLD program only varies ligand torsional degrees of freedom during docking, molecules were first optimized in the field of the receptor, treating the receptor as a rigid entity. The energy of the complex was obtained from this optimized geometry. The energy of the free ligand was calculated by finding the local minimum nearest to the conformation in the complex. For purposes of speed, an accelerated version of this energy calculation was also tested. In this case, optimization in the field of the receptor was carried out assuming vacuum conditions (i.e., continuum solvation was switched off). Only the final energy calculation was performed with solvation. This substantially accelerated the optimization process, and in the overwhelming majority of cases the calculated binding energy differed by less than 10% from the one calculated fully in the solvent. As it is questionable whether the use of continuum solvation is justified within the pocket, all energy calculations in this work were subsequently performed using the accelerated version of the method.

Some more accurate energy estimates were also considered. For example, the energy of the unbound ligand could also be obtained during the screening process from a stochastic conformational search in the solvent. Inclusion of partial receptor flexibility in a predefined radius around the ligand was also evaluated.[30] This was achieved by treating entire residues in a given radius around the ligand as flexible. Best results were achieved by using a radius of 4.0 Å. A somewhat more accurate method of treating receptor flexibility used a sequence of minimization steps of the ligand, the receptor in close proximity to the ligand, and finally both.[30] Although these more accurate energy calculations somewhat improved ligand selection (as measured by a factor of about 2 in database enrichment), the time-consuming nature of these calculations made these methods impractical for database screening.

As an approximate binding energy can be calculated, the question may be asked why additional descriptors are necessary. Unfortunately, there is a simple conceptual problem with this property. In the case of a small ligand that occupies the center of the pocket and establishes no contact with the receptor, the binding energy calculated in the above manner will be 0 kcal/mol. Given the large errors in the binding energy due to several different factors (e.g., full or partial rigidity of the receptor, the neglect of entropic terms, errors in the molecular mechanics energy expressions, unknown position of water molecules, etc.), molecules only partially filling the pocket can have negative binding energies and will not be separated from the actives. Even inactive molecules can have negative binding energies, provided that after docking the polar and nonpolar surfaces of the ligand and the receptor are matched and the molecule is sufficiently small. Despite this, our calculated binding energy correctly predicts reduced activity for those molecules that have internal strain, mismatches of pharmacophoric points or hydrophobic surfaces, and those that make bad van der Waals contacts with the receptor. For this reason, the binding energy was found most useful in combination with the other descriptors.

In addition to the descriptors developed in this work, other terms were also evaluated. It was found that the strengths of external van der Waals and hydrogen bond interactions from the GOLD scoring function carry information that is complementary to our descriptors. Whereas our buriedness and hydrogen bonding descriptors express interactions for a set of pre-defined important residues, the external terms from GOLD describe all van der Waals and hydrogen bond interactions occurring between the ligand and the receptor. Clearly, interactions somewhat away from where the reference ligands bind will also contribute to the strength of binding, even if such interactions are insufficient for binding on their own. We can view the simultaneous use of these two sets of descriptors for van der Waals and hydrogen bonding interactions as a weighting scheme for those binding events that occur inside the known binding pocket.

Database Enrichment. One aim of scoring molecules in a database is to reduce the number of compounds that must be tested experimentally. The efficiency of virtual screening can be described by considering the improvement between selecting molecules at random from the original database and from the one enriched by the scoring function. The role of enrichment factors is to quantify the difference between the original set and the enriched set that should be richer in hits.

There are different descriptions in the literature for enrichment. One often applied method is to plot the percentage of active molecules recovered (i.e., identified as active) as a function of the percentage of the ranked original dataset (e.g., ref 31). Although such plots fully quantify the behavior of molecules for a given scoring function, and are hence ideal for comparing the performance of different functions for a

single receptor, it is cumbersome to make comparisons among different receptor systems. In this work we chose to use primarily numerical descriptors to describe database enrichment. One such measure was defined by Pearlman and Charifson:[32]

$$EF_{PC} = \frac{HITS_{sampled}/N_{sampled}}{HITS_{total}/N_{total}} \quad (1)$$

where HITS is the number of experimentally best binding hits and N is the number of compounds. This factor describes how much richer the selected fraction is in actives than the initial database. Using this definition, a random set will yield $EF = 1$. Such a factor can also be plotted as a function of the fraction selected from the database[32] or can be expressed numerically for given fractions of the database.

The ratio of recovered actives in a subset of the database ($HITS_{sampled}/HITS_{total}$) will be referred to as the level of recovery. Equation 1 can be used to quantify the process of database enrichment at any given level of recovery. Ideally, it would be interesting to know the percentage of the database that needs to be tested in order to find all actives (i.e., $HITS_{sampled} = HITS_{total}$). This case will be referred to as full recovery. In practice it is conceivable that a small percentage of actives will get bad scores due to factors such as incorrect placement. For practical purposes during virtual screening, it is sufficient to find the majority of actives. The level of recovery that is suitable depends on the application. In this study, we calculated enrichments at three points: full, 80%, and 50% recoveries. In the case of a small number of active ligands, we chose the number of recovered actives so that the percentage rounded to 50% or 80%.

A problem with the above definition of database enrichment is that it only considers the performance of the worst ligand in the sampled set. Thus, in case of full recovery there is no difference in the enrichment factors between having the active ligands scoring as 1st, 2nd, 3rd, and 50th, compared to 47th, 48th, 49th, and 50th, although the first outcome is obviously more desirable. One way to resolve this issue is by considering the average percentile rank of the actives using the following formula:[33]

$$EF = \frac{50\%/APR_{sampled}}{HITS_{total}/HITS_{sampled}} \quad (2)$$

where $HITS_{sampled}$ is the number of known actives sampled as above, whereas $APR_{sampled}$ is the average percentile rank of these actives. If the actives are uniformly distributed over the entire database, the average percentile rank will be 50% and the enrichment will be 1 as above. However, in contrast to eq 1, the enrichment calculated from eq 2 will also increase if the actives are concentrated toward the top of the database. This is desirable, because the goal is to find actives as early as possible in the ranked database.[33]

The enrichment in different databases can be best tested by constructing decoy sets that contain inactive drug-size molecules that have general properties similar to those of the actives. The decoy set used in this work[34] originally consisted of 888 molecules from the CMC database and 229 ligands extracted from PDB cocrystallized complexes. These were selected so that they have properties similar to drug molecules.[34] Of these, 3 ligands had to be removed because the GOLD program repeatedly failed to dock them. Repeated occurrences of molecules were also removed from the database, leaving 1056 unique molecules. These were protonated/deprotonated according to their natural state at physiological pH using the wash function in MOE[22] and then minimized using the MMFF94 force field[24] with continuum solvation.[25,26] For each receptor system studied, known actives of that receptor were added to this database and the enrichment was calculated for different recovery rates of these actives. The list of known actives for all receptors is given in Table 1. The studied receptor systems included nuclear receptors (1E3G, 1I37, 1ERR, and 3ERT), protein kinases (1BL7 and 1FVV), various enzymes (1KIM, 2QWK, and 1JRL) and a transport protein (1ABE).

This approach has one inherent problem. The calculated enrichments are only correct if it can be assumed that all the decoy ligands are inactive. Because many of these decoys were obtained from known receptor−ligand complexes, this assumption is unlikely to be correct given the fact that many molecules may interact with a number of receptor systems. However, it has to be assumed that the number of active decoy ligands is small and hence the calculated enrichments are good approximations to the true enrichment factors.

Calculating numerical values for enrichment in this work is primarily for comparative purposes between receptor systems. There are, however, difficulties in comparing the results directly with those of other works. A major difficulty is in the choice of the known actives, especially when the receptor conformation varies substantially during binding with different ligands. One such example is the estrogen receptor (vide infra). It has two major conformational states: one for binding agonists and one for antagonists. Clearly, major differences are expected between different programs depending on whether the active sets include only antagonists or a mixture of agonists and antagonists. Another issue that makes comparisons difficult is that some works include identical ligands under different names.[33] Although having actives present multiple times might be very useful to establish the reproducibility of the experiment from different starting points, it provides an incorrect assessment of the amount of enrichment. In this work we ensured that the selection of actives was as diverse as possible so that the strengths and weaknesses of the scoring function, and indeed the docking−scoring combination, could best be assessed.

Development of a Scoring Function. To test the performance of the developed descriptors for database enrichment, two different approaches were followed. In the first process, these properties were used as filters, applied in a sequential manner. At each step, only those molecules that passed the predefined limits were kept. It was found early on in the work that different receptor systems required different cutoff-values. Although an optimized set of cutoff values for any given receptor provided similar kinds of enrichments as the scoring functions described below, cutoff values that worked for all receptors had to be very generous. This resulted in relatively poor enrichments. Hence, this approach was abandoned and only the development of scoring functions was further pursued.

The first step in the development of a scoring function was to establish which properties contributed most to the

Table 1. Active Ligands Used to Judge the Enrichment for Different Receptor Systems[a]

receptor	reference ligand	reference H bonds	active ligands
		Androgen Receptors (refs 36, 38)	
1I37	testosterone	ARG93	metribolone
			trenbolone
1E3G	metribolone	ARG752	RU-59063
			RU-56187
			dihydrotestosterone hydroxyflutamide
			testosterone
		Estrogen Receptors (refs 1, 39, 40)	
1ERR	β-estradiol	ARG394	β-estradiol
		GLU353	EM-343
3ERT	β-estradiol	ARG394	ICI-164384
		GLU353	LY-326315
			RU-58668
			ZK-11901
			LY-357489
			raloxifene
			sumitomo-biphenol
			4-hydroxytamoxifen
			nafoxidene
			DES
			genistein
		Influenza Neuraminidase (refs 41, 42)	
2QWK	oseltamivir	ARG118	oseltamivir
		ARG152	GR210729
		ARG292	4-amino-neu5-ac2en
			neu5-ac2en
			neu5-ac2
		Arabinose Binding Protein (ref 43)	
1ABE	*l*-arabinose	ASP90	*d*-galactose
		GLU14	*d*-fucose
		ASN232	α-*l*-arabinose
		ASN205	β-*l*-arabinose
		HSV Thymidine Kinase (refs 39, 44)	
1KIM	deoxythymidine	GLN125	penciclovir
		TYR101	ganciclovir
			deoxythymidine
			idoxuridine
			acyclovir
		Inosine Monophosphate Dehydrogenase (ref 37)	
1JR1	inosine-monophosphate	SER68	IMP
		SER329	6-chloro-IMP
		SER388	ribavarin monophosphate
		GLY366	mizoribine monophosphate
		Cyclin Dependent Kinase 2 (refs 45−47)	
1FVV	1KE7 ligand	GLU 81	1FVV_LIG
		LEU 83	1KE6_LIG
		ASP 86	1KE7_LIG
			LIG_16
			LIG_ 23
		P38 MAP Kinase (refs 48−51)	
1BL7	SB220025	LYS53	SB220025
		ASP168	E−9
			E−32
			E−33
			E−39
			C-51
			A-13
			A-17
			A-25
			A-29
			RWJ67657

[a] Structures of these ligands are provided separately as Supporting Information.

final enrichment. This was determined by calculating enrichments corresponding to eq 2 for full recovery and checking how the change in enrichment correlated with changes in the weighting of the given descriptor for three test cases: 1ERR, 1I37, and 1BL7. It was found that receptor buriedness, hydrogen bonding, and the binding energy strongly affected the derived enrichments, whereas ligand buriedness only had a minor effect and the hydrophobic surface area had no significant effect at all. Hence, the first three quantities were taken into the scoring function. In addition, the sum of the

strengths of external van der Waals interaction and hydrogen bonds from the GOLD scoring function were also incorporated into our scoring function. Scoring functions were developed both with and without the binding energy. It was found that although the binding energy is an important descriptor and its use significantly improves enrichments, scoring functions without it could be run much faster and still achieved better enrichments than the docking program alone.

It was found early in this work that linear terms in the scoring function are insufficient if high enrichments are sought. We decided to introduce a cubic term containing the buriedness in conjunction with the original linear term, as this was the simplest nonlinear function that retained the parity of the linear function. Cubic terms for other properties were deemed nonessential, as they provided only minor improvements in enrichment. Hence the final scoring function contained four terms without the binding energy and five with the binding energy.

To develop a general scoring function that can be applied directly on any new dataset, the weights of receptor buriedness, hydrogen bonding, and the $vdW_{ext}+H_{ext}$ terms from GOLD were first optimized for four test cases (1I37, 1ERR 1KIM, and 1BL7). The weights of different properties in the scoring function were varied and the enrichment factors at each point were calculated. The optimal solution for the choice of weights was determined from contour plots, so that it provides the best compromise results for the four cases. These plots were generated using a Visual Basic program within MS Excel. The optimized scoring function was the following:

$$F_{\text{score}} = 20(\text{RB} + \text{RB}^3) + \left(\frac{10}{n}\right)\sum_i^n \text{HB}_i + 0.5(\text{H}_{\text{ext}} + \text{vdW}_{\text{ext}}) \quad (3)$$

where RB is the receptor buriedness, HB_i is an indicator variable for the *i*th specified hydrogen bond (1 if it exists and 0 otherwise), *n* is the total number of hydrogen bonding residues considered for the receptor, and H_{ext} and vdW_{ext} are terms describing van der Waals interactions and hydrogen bonding between the ligand and the receptor and are taken from the GOLD scoring function.[20] It is important to note that if the weights of H_{ext} and vdW_{ext} are set to zero (e.g., the docking was not done in GOLD and these terms are unavailable), the optimum weights of the remaining terms in eq 3 remain practically identical. The inclusion of the terms from GOLD improves enrichment by 1.2−2 times, depending on the receptor.

The coefficient for the binding energy term in the scoring function was optimized for four test cases (1I37, 1ERR, 1KIM, and 1BL7). Inclusion of the binding energy improves enrichment but it appears to have only a minor influence on the weights of individual terms in the scoring function. Thus, for the sake of simplicity, the weights of the other terms in eq 3 are not changed when the binding energy is included. This allows some simple pre-filtering to be performed before the time-consuming binding energy step, so that the energy has to be calculated for a smaller number of molecules. The generalized function is the following:

$$F_{\text{score}} = -0.05\text{BE} + 20(\text{RB} + \text{RB}^3) + \left(\frac{10}{n}\right)\sum_i^n \text{HB}_i + 0.5(\text{H}_{\text{ext}} + \text{vdW}_{\text{ext}}) \quad (4)$$

where BE is the calculated binding energy. The two general scoring functions given by eqs 3 and 4 were then validated using the rest of the receptors described below. It is expected that when scoring functions 3 and 4 are applied for a receptor not considered in this work, the obtained enrichments will be reasonable.

Somewhat better enrichments can be obtained by optimizing the coefficient weights of the descriptors for each receptor being studied. As above, this optimization was carried out by varying the coefficients in eqs 3 and 4 around the optima and generating contour plots using a Visual Basic program. The range of values considered for the coefficients was chosen in such a manner that the maximum and plateau regions of the surface were spanned. In practice, such an optimization can be achieved within 3−4 h for a new receptor, provided the calculated descriptors are available for a set of actives and the decoy ligands. In general, there are many optimized scoring functions that are equivalent with respect to the enrichments obtained and one of these was always arbitrarily selected. Because of the arbitrary nature of this selection and the fact that for a different set of active and decoy molecules these solutions will no longer be equivalent, the coefficients are not given here. The enrichments obtained using these optimized functions were always the same or superior to the ones with the generalized function. Thus, the use of the optimized scoring function is preferable to the generalized one provided the necessary information is available to generate the optimized coefficients. This optimization is possible if the descriptors for the reference system can be generated and a set of actives is available to measure enrichments. Ideally, the descriptors can be generated from the 3D structure of the complex as shown in this study. Nonetheless, the 3D structure of the unbound receptor and a nanomolar active known to bind to the given site should usually be sufficient to define the receptor buriedness and the required hydrogen bond contacts, in which case the knowledge of the 3D structure of the ligand−receptor complex is unnecessary for the use of the scoring function.

The most time-consuming step in the process is the actual docking with GOLD. The scoring of the database is rapid in comparison: it took about 22 s to score a randomly selected set of 100 ligands of the 1E3G receptor without the binding energy. When the binding energy is included, the calculation on the same molecules took almost 14 min on the same computer. These speeds correspond to 400 000 and 10 000 molecules screened a day on a PC with a 1.67 GHz Athlon processor running RedHat Linux. This latter set of calculations can be speeded up by pre-filtering the molecules on the basis of van der Waals volume, receptor buriedness, and whether the already calculated terms in the scoring function make it likely that the molecule has a reasonable likelihood of activity. The speed enhancement depends on the source of molecules, but in case of random ACD structures and the androgen receptor it amounts to about 15−

30% without any penalties in the results for the top scoring compounds.

Issues Related to the Scoring Function for Individual Receptor Systems. *(a) Estrogen Receptors 1ERR and 3ERT.* The 1ERR receptor is a homodimer and only the first chain of the PDB structure was used in the present study. The choice of the active set is somewhat controversial in this case. It is known[35] that the estrogen receptor has two major conformational states: one for binding agonists and one for antagonists. In the agonist conformation, the receptor is enclosed and hence is unable to bind the much larger antagonist compounds. On the other hand, the antagonist conformation is capable of binding both agonist and antagonist ligands as it has the requisite pharmacophoric arrangement and hydrophobic contacts. (Clearly, this is a hypothetical case because on binding agonists the receptor conformation is known to be different from the antagonist case.) When testing database enrichments, often only antagonists are selected in the test set.[1,33] On the other hand, if the interest is in compounds that bind to the estrogen receptor, the active set might contain both antagonists and agonists. It is quite natural that this latter choice of ligands should lead to worse enrichments. In this work, both scenarios were tested. Using the buriedness function our method is capable of considering only those parts of the ligand that are, in terms of binding, common to both agonists and antagonists. This was achieved by referencing receptor buriedness to the small agonist β-estradiol. Hydrogen bonds to two residues, ARG 394 and GLU353, were used for scoring. One problematic compound was nafoxidene because it has only a hydrogen bond acceptor in the head region and hence cannot possibly hydrogen bond to both of the above residues. The docked nafoxidene molecules indeed formed this hydrogen bond in the 1ERR cavity. In the 3ERT cavity, however, because of the way GOLD positioned the methoxy group during the test, our hydrogen bond detection routine could not identify it as such. This, however, does not justify the removal of nafoxidene from the set of actives, as it is unknown whether a similar problem occurs with any of the decoy molecules. Note that when hydrogen bond constraints are enforced this problem is resolved: in repeated experiments the top docked solution for nafoxidene always established this hydrogen bond. However, to maintain consistency of the methodology, only unconstrained runs were utilized for this receptor.

(*b*) *Androgen Receptors 1I37 and 1E3G.* The receptors tested were 1I37, cocrystallized with testosterone, and 1E3G, cocrystallized with metribolone. (1I37 is incorrectly labeled in the PDB as if it contained the cocrystallized ligand dihydrotestosterone.) ARG93 in 1I37 and the equivalent ARG752 in 1E3G were chosen as the characteristic hydrogen-bonding residues. It has been shown that steroids lacking the ability to form this hydrogen bond have a 100-fold loss in affinity to the androgen receptor in binding assays.[36] Other hydrogen bonding residues are expected to be less crucial for activity based on the structure of potent nonsteroidal compounds such as RU-56187.[36] In the selection of active ligands it was ensured that both steroidal and nonsteroidal ones are represented.

An additional issue arose with the 1E3G receptor due to the rigid-receptor approximation at the docking stage, as the binding pocket of this receptor cannot properly accommodate the potent androgen ligands testosterone and dihydrotestosterone. From the comparison of the 1E3G and 1I37 receptors it was established that this problem is caused by the position of the side chain of the MET745 residue. On changing the C−S−C−C dihedral angle to the same value as in 1I37, testosterone and all other ligands dock well into the receptor.

(*c*) *Wild-Type Tern N9 Influenza Virus Neuraminidase (2QWK).* This enzyme was cocrystallized with GS-4071 (oseltamivir), which was used as a reference ligand for the calculation of buriedness. Three hydrogen bonds were common to all actives (involving residues ARG118, ARG152, and ARG292). These residues are distributed around the cavity and thus were expected to provide good selectivity for high-affinity ligands.

(*d*) *L-Arabinose-Binding Protein (1ABE).* The protein was cocrystallized with α- and β-L-arabinose and the latter was used as the reference ligand for the calculation of receptor buriedness. Four amino acid residues were selected for hydrogen-bonding (ASP90, GLU14, ASN232, and ASN205). Only a small number of ligands are known for this receptor but these have high specificity.

(*e*) *Herpes Simplex Virus Type I Thymidine Kinase (1KIM).* The cocrystallized ligand in the complex was deoxythymidine and it was used for the calculation of buriedness. This ligand and all other known ones establish hydrogen bonds with GLN125 and TYR101 and consequently these residues were used in the scoring function.

(*f*) *Inosine Monophosphate Dehydrogenase (1JR1).* The published receptor is a homodimer and its first chain was used for docking. The receptor was cocrystallized with mycophenolic acid (MPA) and inosine-monophosphate (IMP), the latter was used to define the occupied volume for the calculation of buriedness. The molecule has two known binding pockets: the competitive (IMP site) and the noncompetitive (MPA site). In this study, the competitive site was used. Therefore ligands were selected that were known to bind to this site and not to the noncompetitive one. Furthermore, only phosphorylated forms of the active molecules were considered because that is the form in which ligands are known to bind to the receptor.[37] All known ligands hydrogen-bond to four residues (SER329, SER388, SER68, and GLY366) and hence all these were used in the scoring function.

(*g*) *Human P38 MAP Kinase (1BL7).* The receptor was cocrystallized with SB220025 and this was used as a reference ligand to define receptor buriedness. For considering hydrogen bonding, the two residues common to all known actives, LYS53 and ASP168, were selected. Unlike in other receptors, at the usual docking speed (the "7−8 times speed-up" setting in GOLD) ligands were not docked reproducibly in the correct pose. This problem was solved in two different ways. One solution was to run GOLD at its base speed ("standard default settings"). In this case only two ligands were sometimes docked in an improper orientation. The alternative was to run at the usual speed (the "7−8 times speed-up" setting in GOLD) but to constrain pharmacophore interactions to include LYS53 and MET109, in which case all known ligands docked correctly in a reproducible manner.

(*h*) *Cyclin-Dependent Kinase CDK2 (1FVV).* The receptor used in this study is a dimer and its first chain was used in

Table 2. Comparison of the Performance of Different Scoring Functions for 50% Recovery[a]

	theoretical maximum	individually optimized		general formula		GOLD	
		with BE	without BE	with BE	without BE	fitness	$vdW_{ext} + H_{ext}$
3ERT							
antagonists	89	74	67	74	67	33	43
mixed	72	61	53	61	52	24	31
1ERR							
antagonists	89	89	89	89	89	89	89
mixed	72	72	72	72	72	67	69
1E3G[b]	121	93 (39)	51 (21)	64 (19)	61 (28)	14 (5)	6 (3)
1I37	121	87	32	55	26	15	6
2QWK	160	159	119	159	87	31	28
1ABE	176	176	88	176	106	15	23
1KIM	159	136	38	106	35	13	10
1JR1	177	177	177	132	106	106	76
1FVV	159	159	159	136	119	33	7
1BL7							
without constr.	83	83	79	83	76	23	12
with constr.	83	83	83	83	83	79	53

[a] The enrichment factor EF was defined by eq 2. The general formula for the scoring functions was given in eqs 3 and 4. The enrichments calculated using the fitness function in GOLD and the ones calculated using the quantity $vdW_{ext} + H_{ext}$ are also given. The theoretical enrichments correspond to the actives being ranked as the top solutions. See text for further details. [b] Values in brackets were obtained without modifying the torsion angle in MET745. See text for details.

Table 3. Comparison of the Performance of Different Scoring Functions for 80% Recovery[a]

	theoretical maximum	individually optimized		general formula		GOLD	
		with BE	without BE	with BE	without BE	fitness	$vdW_{ext} + H_{ext}$
3ERT							
antagonists	95	81	70	81	67	30	38
mixed	75	65	54	65	53	13	23
1ERR							
antagonists	95	95	95	95	92	40	88
mixed	75	75	73	73	71	21	46
1E3G[b]	130	40 (25)	15 (13)	22 (14)	12 (10)	9 (4)	4 (3)
1I37	130	57	24	25	17	4	3
2QWK	170	154	121	106	89	34	27
1ABE	198	198	108	170	119	20	32
1KIM	170	130	45	113	40	12	12
1JR1	199	199	199	149	132	92	92
1FVV	170	170	170	113	106	40	8
1BL7							
without constr.	87	52	35	35	24	21	10
with constr.	87	87	74	82	63	68	42

[a] The enrichment factor EF was defined by eq 2. The general formula for the scoring functions was given in eqs 3 and 4. The enrichments calculated using the fitness function in GOLD and the ones calculated using the quantity $vdW_{ext} + H_{ext}$ are also given. The theoretical enrichments correspond to the actives being ranked as the top solutions. See text for further details. [b] Values in brackets were obtained without modifying the torsion angle in MET745. See text for details.

the study. The docking was performed using hydrogen bond constraints for GLU81 and LEU 83. This was necessary as many test ligands could not be docked reproducibly into their correct position without using constraints or slowing down the screening process. The constrained residues and ASP86 were used as the hydrogen bonds applied in the scoring function. The docked ligand extracted from the 1KE7 receptor complex was used to define receptor buriedness because it occupied the common space of all high-affinity ligands inspected.

RESULTS AND DISCUSSION

To characterize our generalized and individually optimized scoring functions both with and without the binding energy terms, the databases containing the decoy molecules and the selected actives were docked and scored for 10 different receptor systems and the enrichment factors were calculated at recovery rates of 50, 80, and 100%. The results are summarized in Tables 2−4, respectively. These were compared with the enrichments obtained by using only the GOLD fitness function and also with those obtained using the quantity $vdW_{ext} + H_{ext}$ from GOLD that has recently been shown to perform better than the original GOLD fitness function.[19] Tables 2−4 present these results, together with the maximum enrichment possible for each system (i.e., when the active ligands added to the decoy set are ranked as the best predicted ligands).

As can be seen from these results and also in Figure 1, the enrichments obtained with the individually optimized scoring functions are close to the theoretical ones in many examples. It is unreasonable to expect that this maximum is reached for all receptor systems, as some ligands in the decoy set may also be active for some receptors. The opposite is also possible: a molecule may fulfill the requirements for binding in the docked conformation and may still be inactive. This may arise from the lack of sufficient solubility, the fact

Table 4. Comparison of the Performance of Different Scoring Functions for Full Recovery[a]

	theoretical maximum	individually optimized		general formula		GOLD	
		with BE	without BE	with BE	without BE	fitness	$vdW_{ext} + H_{ext}$
3ERT							
antagonists	97	43	32	35	24	11	22
mixed	76	36	24	30	22	9	13
1ERR							
antagonists	97	94	89	85	74	17	46
mixed	76	64	51	60	47	8	12
1E3G[b]	133	36 (10)	13 (9)	18 (8)	10 (7)	6 (4)	4 (3)
1I37	133	39	23	20	14	4	3
2QWK	177	106	95	55	65	38	28
1ABE	211	211	176	163	124	24	38
1KIM	177	115	46	115	41	11	14
1JR1	212	212	212	163	132	73	92
1FVV	177	166	166	102	95	39	9
1BL7							
without constr.	89	24	21	14	13	15	8
with constr.	89	79	55	64	49	45	29

[a] The enrichment factor EF was defined by eq 2. The general formula for the scoring functions was given in eqs 3 and 4. The enrichments calculated using the fitness function in GOLD and the ones calculated using the quantity $vdW_{ext} + H_{ext}$ are also given. The theoretical enrichments correspond to the actives being ranked as the top solutions. See text for further details. [b] Values in brackets were obtained without modifying the torsion angle in MET745. See text for details.

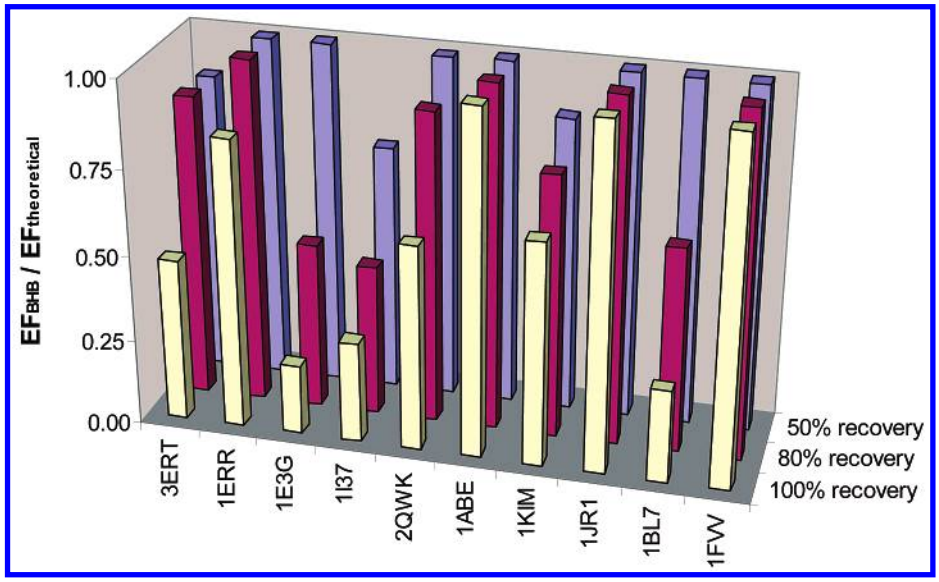

Figure 1. Ratio of the observed enrichment factor to the theoretical maximum value for different receptors and recovery rates. The BHB scoring functions were optimized for each receptor and include the binding energy. The enrichment factors were calculated using eq 2.

that it does not reach the receptor (such as in a cell-based assay), or if it simply cannot elicit the required cellular response for detection.

The justification for using the BHB scoring function as an add-on after docking is the improvement it provides in enrichment compared to the GOLD fitness function alone, or to the score $vdW_{ext} + H_{ext}$ from GOLD.[19] This improvement for different receptors is depicted in Figure 2. Tables 2−4 indicate that there is a substantial improvement in comparison to either score in GOLD. The improvement depends on the recovery rate and the receptor and is generally between 2 and 12 times. The example of the 1BL7 and 1FVV receptor shows that constraining pharmacophore contacts, as has been introduced recently as an option in GOLD, holds great promise for improving enrichments. It contributed to a major improvement in enrichment with the GOLD scores alone and also with the BHB scoring function. Using Tables 2−4 it can also be seen that the $vdW_{ext} + H_{ext}$ score from GOLD performs better than the fitness score from GOLD only in about half of the cases studied. However, the individually optimized BHB score including binding energy performed much better than either of the GOLD scores.

Including the energy term in the scoring function led to an improvement in enrichment for every database used in this study. There is only one exception: at full recovery and the general scoring function, there is a slight decrease in enrichment for the 2QWK receptor. For the majority of recovery rates and receptors, the improvement resulting from including the binding energy is less than a factor of 2. This is an important consideration, as database screening is about 40 times faster when the binding energy is excluded. However, there are exceptions, such as the 1KIM receptor, where the inclusion of the binding energy brings more improvement.

It is obvious that the individually optimized scoring function should provide better or equal enrichments than the general scoring function. As Tables 2−4 indicate, this improvement is generally moderate. Hence, although best

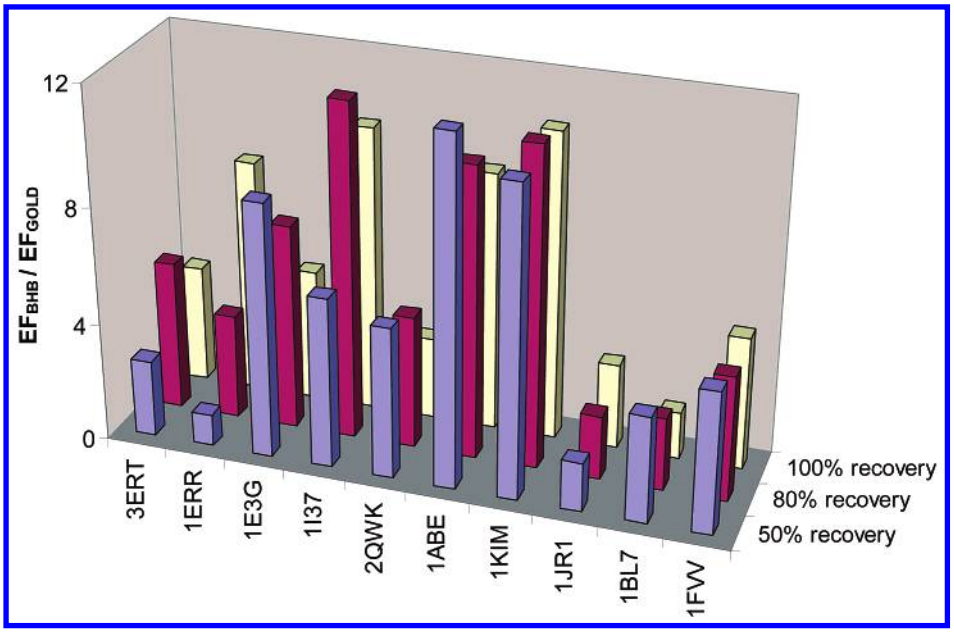

Figure 2. Ratio of the enrichment factors with the BHB scoring function to that with GOLD for 50%, 80%, and 100% recovery rates. The BHB scoring functions were optimized for each receptor and include the binding energy. The enrichment factors were calculated using eq 2.

results can be achieved by optimizing the scoring function for the receptor being studied, sufficiently good results can also be obtained simply by using the general scoring function. This is important both because the latter procedure requires much less knowledge of the receptor (e.g., there is no need to devise a dataset to measure enrichments) and because screening can start without actually having to go through an optimization process.

It is interesting to compare the trends observed for different receptors and recovery rates when the individually optimized scoring function is applied. For some receptors, the enrichment values decrease at increasing recovery rates (e.g., 1E3G and 1I37). This happens when most of the actives are recovered as the best ligands, whereas the remaining actives are located further down the prioritized list. In this case, a smaller recovery rate means that an incomparably better set of ligands is considered and this is expected to improve the enrichment value. For other databases (e.g., 1JR1 and 1ABE) the enrichment factor decreases for increasing recovery rates. This arises because of the small number of active compounds in these databases which causes the ratio $HITS_{total}/HITS_{sampled}$ in eq 2 to dominate the enrichment factor over the rest of terms. However, the actual enrichment for both receptors was equal to the theoretical one at all three recovery rates. Close to theoretical enrichments were achieved also for the 1FVV and 1KIM receptors.

For the estrogen receptors the enrichments were near the theoretical values in almost all cases, although the results are somewhat worse when a mixed list of agonists and antagonists are docked. The small agonists get worse scores than the larger antagonists, primarily due to the vdW_{ext} contribution in the scoring function. This arises as a larger interacting surface area translates into a greater vdW_{ext} score. In the case of the 3ERT receptor, missing the hydrogen bond in nafoxidene after docking (vide supra) contributed to reaching only half the theoretical enrichment at full recovery.

For the two androgen receptors high enrichments were achieved at 50% recovery but the enrichments decreased for higher recovery rates. A closer look at the ligands that score near the top reveals that they indeed look similar to known androgens. It is worth noting that the corresponding enrichments for the GOLD program alone are substantially lower, between 10 and 20% of those obtained with the BHB scoring function. In case of the 1E3G receptor, significant improvements were achieved when the torsion angle for MET745 was changed to a value similar to that in the 1I37 receptor, as this facilitated a better fit of ligands such as testosterone and dihydrotestosterone.

Ligands were docked to the 1BL7 receptor in two different ways: using an unconstrained docking similar to all other ligands and one in which two residues, LYS53 and MET109, were constrained to have pharmacophore contact with the ligands. (Those ligands where such contact was impossible were automatically discarded.) The results indicate that with pharmacophore constraints the enrichments are near the theoretical values. Without applying such constraints, the enrichments change from the theoretical value (at 50% recovery) to about a quarter of that (at full recovery).

CONCLUSIONS

A novel knowledge-based scoring function, BHB, based on the properties buriedness, hydrogen bonding and binding energy, was developed and tested in this work. It uses information on the receptor that can be obtained easily provided a cocrystallized structure or a set of active ligands is available. (The structure of the unbound receptor and some active ligands may be sufficient in many cases.) It was developed for prioritizing ligands that had already been docked. In this work, it was applied in conjunction with the GOLD docking program, providing great improvements in comparison to the enrichments with GOLD alone. As the BHB scoring function uses very simple descriptors to predict binding, it is best used for ranking large virtual libraries with a small number of binders, rather than for prioritizing in a small focused library of related actives. The scoring function

was optimized for the best enrichment of actives, so that only a small portion of a database should have to be tested.

Our scoring function can be used in conjunction with any docking program, provided the molecules were docked in the correct position. As the enrichment is measured on the docked and scored ligands, the final enrichments depend also on the quality of docking, as the two steps are inextricably linked. As shown above, scoring is very rapid and thus the ultimate speed of database screening is still determined by the docking program.

Our scoring program appears to provide consistently excellent enrichments independently of the studied receptor. As knowledge about the receptor under study is utilized in scoring, a potential danger is that it might limit novelty or diversity by discriminating against ligands binding slightly differently. However, from the tests undertaken in this work, this does not appear to be the case. If the area of the receptor where ligands might conceivably bind is known, buriedness can be defined using all the affected residues, hence allowing for different binding interactions within the pocket. The converse is also true: if only a small portion of the receptor pocket is known to be important that portion can be used to test for receptor buriedness, thus improving selectivity. The calculated binding energy is independent of the position where ligands and receptor interact and thus has no bearing on diversity. The only factor that can potentially discriminate against novel ligands is the hydrogen-bonding descriptor. However, if this is a concern, its negative impact can be limited by specifying only those residues that are known to be crucial for activity. The stepwise optimization of the scoring function ensures that even if no hydrogen bonding is defined, reasonable enrichments can be achieved, albeit worse than with it. In addition, normalization in the scoring function ensures that even if a long list of hydrogen bonds is specified, the relative weight of hydrogen bonding will not change. On the other hand, those solutions in which the molecule binds to a different pocket are clearly discriminated against.

ACKNOWLEDGMENT

Dr. Jonathan Schmidt is acknowledged for his support, useful discussions, and critical reading of the manuscript. Dr. Tom Halgren from Schrodinger Inc. is gratefully acknowledged for supplying his set of decoy ligands used in quantifying enrichment for different receptors and also for initial discussions that contributed to initiating this study.

Supporting Information Available: Table of ligands (5 pp; PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) 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.

(2) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein-small molecule docking methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151−166.

(3) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.

(4) 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.

(5) Nicodeme. P.; Didier, R. ConsDock: A New Program for the Consensus Analysis of Protein−Ligand Interactions. *Proteins* **2002**, *47*, 521−533.

(6) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−45.

(7) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein−ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337−356.

(8) 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.

(9) Rognan, D.; Lauemoller, S. L.; Holm, A.; Buus, S.; Tschinke, V. Predicting binding affinities of protein ligands from three-dimensional models: application to peptide binding to class I major histocompatibility proteins. *J. Med. Chem.* **1999**, *42*, 4650−4658.

(10) Böhm, H. J. The development of a single empirical scoring function to estimate the binding constant for a protein−ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243−256.

(11) Clark, R. D.; Strizev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graph. Model.* **2002**, *20*, 281−295.

(12) Wang, R.; Wang, S. How Does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1422−1426.

(13) Jones, G.; Wilett, 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.

(14) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(15) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.

(16) Glide 2.0. Schrodinger Inc.: Portland, OR, 2002.

(17) Abagyan, R.; Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **2001**, *5*, 375−82.

(18) Jones, G.; Willett, P.; Glen, R. C. Moelcular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *J. Mol. Biol.* **1995**, *245*, 43−53.

(19) Nissink, J. W. M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A New Test Set for Validating Predictions of Protein−Ligand Interactions. *Proteins* **2002**, *49*, 457−71.

(20) GOLD, versions 1.2 and 2.0. Cambridge Crystallographic Data Centre: Cambridge, U.K., 2002.

(21) Stahl, M. Modifications of the scoring function in FlexX for virtual screening applications. *Perspect. Drug Discovery Des.* **2000**, *20*, 83−98.

(22) Molecular Operating Environment, version 2002.03. Chemical Computing Group Inc.: Montreal, PQ.

(23) ACD logD suite, version 6.00. Advanced Chemistry Development Inc.: Toronto, ON.

(24) Halgren, T. A. The Merck Molecular Force Field. I. Basis, Form, Scope, Parametrization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.

(25) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(26) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A* **1997**, *101*, 3005−3114.

(27) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucl. Acid Res.* **2000**, *28*, 235−242.

(28) Stahl, M.; Böhm, H.-J. Development of filter functions for protein−ligand docking. *J. Mol. Graph. Model.* **1998**, *16*, 121−132.

(29) Stickle, D. F.; Presta, L. G.; Dill, K. A.; Rose, G. D. Hydrogen bonding in globular proteins. *J. Mol. Biol.* **1992**, *226*, 1143−1159.

(30) Schmidt, J. M.; Mercure, J. A.; Feher, M.; Dunn-Dufault, R.; Peter, M. G.; Redden, P. R. De Novo Design, Synthesis and Evaluation of Novel Non-Steroidal High Affinity Ligands for the Estrogen Receptor. *J. Med. Chem.* **2003**, *46*, 1408−1418.

(31) Hindle, S. A.; Rarey, M.; Buning, C.; Lengauer, T. Flexible docking under pharmacophore type constraints. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 129−149.

(32) Pearlman, D. A.; Charifson, P. S. Improved Scoring of Ligand-Protein Interactions Using OWFEG Energy Grids. *J. Med. Chem.* **2001**, *44*, 502−511.
(33) http://www.schrodinger.com/docs/impact2.0/pdf/tech_notes/tech_notes.pdf, (last accessed 16.10.2002).
(34) The decoy set was kindly provided by Dr. Tom Halgren, Schrodinger Inc., for the evaluation of the Glide program and its comparison with other approaches.
(35) Anstead, G. M.; Carlson, K. E.; Katzenellenbogen, J. A. The estradiol pharmacophore: Ligand structure − estrogen receptor binding affinity relationships and a model for the estrogen receptor binding site. *Steroids* **1997**, *62*, 268−298.
(36) Singh, S. M.; Gauthier, S.; Labrie, F. Androgen Receptor Anatagonists (Antiandrogens): Structure Activity Relationships. *Curr. Med. Chem.* **2000**, *7*, 211−247.
(37) Sintchak, M. D.; Nimmesgern, E. The structure of inosine 5′-monophosphate dehydrogenase and the design of novel inhibitors. *Immunopharmacology* **2000**, *47*, 163−184.
(38) Bauer, E. R. S.; Daxenberger, A.; Petri, T.; Sauerwein, H.; Meyer, H. H. D. Characterisation of the affinity of different anabolics and synthetic hormones to the human androgen receptor, human sex hormone binding globulin and to the bovine progestin receptor. *Acta Pathol. Microbiol. Immunol. Scand.* **2000**, *108*, 838−846.
(39) Hardman, J. G., Limbird, L. E., Eds. *Goodman & Gilman's: The Pharmacological Basis of Therapeutics*, 10th ed.; McGraw-Hill Medical: New York, 2001; Chapter 58, pp 1597−1629.
(40) Harris, H. A.; Bapat, A. R.; Gonder, D. S.; Frail, D. E. The ligand binding profiles of estrogen receptors alpha and beta are species dependent. *Steroids* **2002**, *67*, 379−384.
(41) McKimm-Breschkin, J. L.; Sahasrabudhe, A.; Blick, T. J.; McDonald, M.; Colman, P. M.; Hart, G. J.; Bethell, R. C.; Varghese, J. N. Mutations in a Conserved Residue in the Influenza virus Neuraminidase Active Site Decrease Sensitivity to Neu5Ac2en-derived Inhibitors. *J. Virol.* **1998**, *72*, 2456−2462.
(42) Taylor, N. R.; Cleasby, A.; Singh, O.; Skarzynski, T.; Wonacott, A. J.; Smith, P. W.; Sollis, S. L.; Howes, P. D.; Cherry, P. C.; Bethell, R.; Colman, P.; Varghese, J. Dihydropyrancarboxamides related to zanamivir: a new series of inhibitors of influenza virus sialidases. 2. Crystallographic and molecular modeling study of complexes of 4-amino-4H-pyran-6-carboxamides and sialidase from influenza virus types A and B. *J. Med. Chem.* **1998**, *41*, 798−807.
(43) Vermersch, P. S.; Tesmer, J. G.; Lemon, D. D.; Quiocho, F. A. A Pro to Gly Mutation in the Arabinose-binding Protein Enhances and Alters Specificity. *J. Biol. Chem.* **1990**, *265*, 16592−16603.
(44) Champness, J. N.; Bennett, M. S.; Wien, F.; Visse, R.; Summers, W. C.; Herdewijn, P.; Clerq, E.; Ostrowski, T.; Jarvest, R. L.; Sanderson, M. R. Exploring the active site of herpes simplex virus type-1 thymidine kinase by X-ray crystallography of complexes with aciclovir and other ligands. *Proteins* **1998**, *32*, 350−361.
(45) Hardcastle, I. R.; Golding, B. T.; Griffin, R. J. Designing Inhibitors of Cyclin-Dependent Kinases. *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 325−348.
(46) Davis, S. T.; Benson, B. G.; Bramson, H. N.; Chapman, D. E.; Dickerson, S. H.; Dold, K. M.; Eberwein, D. J.; Edelstein, M.; Frye, S. V.; Gamp, R. T., Jr.; Griffin, R. J.; Harris, P. A.; Hassell, A. M.; Holmes, W. D.; Hunter, R. N.; Knick, V. B.; Lackey, K.; Lovejoy, B.; Luzzio, M. J.; Murray, D.; Parker, P.; Rocque, W. J.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. Prevention of chemotherapy-induced alopecia in rats by CDK inhibitors. *Science* **2001**, *291*, 134−137.
(47) Bramson, H. N.; Corona, J.; Davis, S. T.; Dickerson, S. H.; Edelstein, M.; Frye, S. V.; Gampe, R. T., Jr.; Harris, P. A.; Hassell, A.; Holmes, W. D.; Hunter, R. N.; Lackey, K. E.; Lovejoy, B.; Luzzio, M. J.; Montana. V.; Rocque, W. J.; Rusnak, D.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): design, synthesis, enzymatic activities, and X-ray crystallographic analysis. *J. Med. Chem.* **2001**, *44*, 4339−58.
(48) Gallagher, T. F.; Seibel, G. L.; Kassis, S.; Laydon, J. T.; Blumenthal, M. J.; Lee, J. C.; Lee, D.; Boehm, J. C.; Fier-Thompson, S. M.; Abt, J. W.; Soreson, M. E.; Smietana, J. M.; Hall, R. F.; Garigipati, R. S.; Bender, P. E.; Erhard, K. F.; Krog, A. J.; Hofmann, G. A.; Sheldrake, P. L.; McDonnel, P. C.; Kumar, S.; Young, P. R.; Adams, J. L. Regulation of Stress-Induced Cytokine Production by Pyridinylimidazoles; Inhibition of CSBP Kinase. *Bioorg. Med. Chem.* **1997**, *5*, 49−64.
(49) Liverton, N. J.; Butcher, J. W.; Claiborne, C. F.; Claremon, D. A.; Libby, B. E.; Nguyen, K. T.; Pitzenberger, S. M.; Selnick, H. G.; Smith, G. R.; Tebben, A.; Vacca, J. P.; Varga, S. L.; Agarwal, L.; Dancheck, K.; Forsyth, A. J.; Fletcher, D. S.; Frantz, B.; Hanlon, W. A.; Harper, C. F.; Hofsess, S. J.; Kostura, M.; Lin, J.; Luell, S.; O'Neill, E. A.; Orevillo, C. J.; Pang, M.; Parsons, J.; Rolando, A.; Sahly, Y.; Visco, D. M.; O'Keefe, S. J. Design and synthesis of potent, selective, and orally bioavailable tetrasubstituted imidazole inhibitors of p38 mitogen-activated protein kinase. *J. Med. Chem.* **1999**, *42*, 2180−2190.
(50) Thurmond, R. L.; Wadsworth, S. A.; Schafer, P. H.; Zivin, R. A.; Siekierka, J. J. Kinetics of small molecule inhibitor binding to p38 kinase. *Eur. J. Biochem.* **2001**, *268*, 5747−5754.
(51) Adams, J. L.; Boehm, J. C.; Gallagher, T. F.; Kassis, S.; Webb, E. F.; Hall, R.; Sorenson, M.; Garigipati, R.; Griswold, D. E.; Lee, J. C. Pyrimidinylimidazole inhibitors of p38: cyclic N-1 imidazole substituents enhance p38 kinase inhibition and oral activity. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2867−2870.

CI030006I