PROTEINS: Structure, Function, and Genetics 34:4–16 (1999)

RESEARCH ARTICLES

Ligand Solvation in Molecular Docking

Brian K. Shoichet,[1]* Andrew R. Leach,[2] and Irwin D. Kuntz[2]*
[1]*Department of Molecular Pharmacology and Biological Chemistry, Northwestern University Medical School, Chicago, Illinois*
[2]*Department of Pharmaceutical Chemistry, University of California, San Francisco, California*

ABSTRACT **Solvation plays an important role in ligand-protein association and has a strong impact on comparisons of binding energies for dissimilar molecules. When databases of such molecules are screened for complementarity to receptors of known structure, as often occurs in structure-based inhibitor discovery, failure to consider ligand solvation often leads to putative ligands that are too highly charged or too large. To correct for the different charge states and sizes of the ligands, we calculated electrostatic and non-polar solvation free energies for molecules in a widely used molecular database, the Available Chemicals Directory (ACD). A modified Born equation treatment was used to calculate the electrostatic component of ligand solvation. The non-polar component of ligand solvation was calculated based on the surface area of the ligand and parameters derived from the hydration energies of apolar ligands. These solvation energies were subtracted from the ligand-receptor interaction energies. We tested the usefulness of these corrections by screening the ACD for molecules that complemented three proteins of known structure, using a molecular docking program. Correcting for ligand solvation improved the rankings of known ligands and discriminated against molecules with inappropriate charge states and sizes. Proteins 1999;34:4–16.** © 1999 Wiley-Liss, Inc.

Key words: solvation; molecular docking; database search; structure-based drug design; computer-aided drug design

INTRODUCTION

Given the structure of a biological receptor, it should be possible to design or discover molecules that will bind to it. Using atomic resolution structures and computational techniques, investigators have attempted to design[1–3] or discover[4–8] novel inhibitors for biological receptors. Such putative ligands have been selected for their complementarity to the structure of the receptor. When the energy of the solvated state is not considered in these calculations, the ligands that are selected often bear high formal charge or are larger than expected. This is particularly true when comparing many potential ligands that differ in polarity and size.

The binding affinity of a ligand for a receptor depends on the interaction free energy of the two molecules relative to their free energies in solution:

$$\Delta G_{bind} = \Delta G_{interact} - \Delta G_{solv,L} - \Delta G_{solv,R} \quad (1)$$

where $\Delta G_{interact}$ is the interaction free energy of the complex, $\Delta G_{solv,L}$ is the free energy of desolvating the ligand and $\Delta G_{solv,R}$ is the free energy of occluding the receptor site from solvent. Various methods have been proposed to evaluate or to estimate these terms; the problem is difficult because the energy of each component on the right hand side of Equation 1 is large while the difference between them is small.

The most accurate way to calculate relative binding energies is with free-energy perturbation techniques.[9–11] These techniques are usually restricted to calculating the differential binding of similar compounds, and require extensive computation.[12–14] In novel inhibitor design and discovery, many different candidate molecules are evaluated, making free energy perturbation impractical as an initial screen.

Methods that consider many different possible ligand-receptor complexes are necessarily less accurate than perturbation techniques. The free energy of interaction ($\Delta G_{interact}$) is usually approximated as an enthalpy and is calculated with a receptor potential function, often derived from molecular mechanics.[15–17] Ligand and receptor solvation contributions ($\Delta G_{solv,L}$ and $\Delta G_{solv,R}$) are usually calculated as free energies; efforts range from empirical scales[18,19] to parameterization for specific functional groups [20–24] to detailed theoretical treatments.[25–27] Several

Grant sponsor: PhRMA Foundation; Grant sponsor: Animal Alternatives Program of Procter and Gamble; Grant sponsor: National Institutes of Health; Grant number: GM31497; Grant sponsor: Science and Engineering Research Council (UK) under the NATO postdoctoral fellowship program.

Andrew R. Leach's present address is Department of Chemistry, Glaxo-Wellcome, Southampton, Hampshire SO9 5NH, United Kingdom.

*Correspondence to: Brian K. Shoichet, Department of Molecular Pharmacology and Biological Chemistry, Northwestern University Medical School, 303 East Chicago Avenue, Chicago, IL 60611–3008. E-mail: b-shoichet@nwu.edu, or to Irwin D. Kuntz, Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143–0446. E-mail: kuntz@cgl.ucsf.edu.

Received 3 February 1998; Accepted 27 August 1998

investigators have implicitly included solvation in their affinity calculations by parameterizing the affinity of individual functional groups based on the binding of sets of known ligands to different receptors.[28–32] This overcomes the problem of subtracting the explicitly calculated solvation and interaction energies, both of which are typically large. The energies used in the parameterized functional group methods have no necessary relationship to discrete atomic terms, such as electrostatic and van der Waals forces. We are only interested in such atomic terms here, and will consider only force-field-based calculations of affinities.

Several authors have described force fields that consider the bound and solvated states. Amino acids have been parameterized using experimental vapor-partitioning coefficients[20] for the ECEPP potential function[23] to allow for solvation correction. Similar corrections have been used in ligand discovery methods[2,33,34] These methods successfully predicted new ligands and also the structures of ligand-receptor complexes[33] It seems clear that the solvation correction terms improved the evaluation of the putative complexes. On the other hand, the relative magnitudes of the solvation and interaction terms in these studies remained problematic, as the experimental partitioning numbers are often small compared with the interaction energies one might expect with a force field such as ECEPP or AMBER. Also, the experimental solvation numbers used by these authors are best characterized, in the context of a macromolecular potential function, for peptides and nucleic acids, and less well characterized for other small organic molecules. Many inhibitor design algorithms, including several of our own, have ignored solvation and have used only the interaction energy for evaluating complexes, treating it as a proxy for the free energy of binding of different putative ligands or functional groups[1,3,5,6,35–37]

Here we attempt to balance protein-ligand interaction energies with ligand solvation energies in docking studies of a molecular database. The molecular docking program DOCK was used to screen the Available Chemical directories (ACD) database for compounds that complemented the enzymes thymidylate synthase (TS), the L99A mutant of T4 lysozyme, and dihydrofolate reductase (DHFR). TS binds pyrimidine nucleotides bearing one or two formal negative charges; the "cavity" mutant of T4 lysozyme binds neutral aromatic hydrocarbons; DHFR binds pteridines bearing either one or no positive formal charge. By choosing these three receptors, we hoped to span a range of ligand charge and size, testing the solvation corrections in very different receptor environments. The ACD includes known ligands for each receptor, but most of molecules in the ACD are not thought to bind to these enzymes.

We corrected the electrostatic interaction energy ($\Delta G_{elec,interact}$) with a ligand electrostatic solvation energy ($\Delta G_{elec,L,solv}$) and the van der Waals component of the interaction energy ($\Delta E_{vdw,interact}$) with a non-polar component of ligand solvation ($\Delta G_{np,L,solv}$). Dividing the solvation and interaction terms into electrostatic and non-electrostatic components is not rigorously correct,[13] but the approximation is often reasonable.[24] The binding energy score equation becomes:

$$\Delta G_{bind} = \Delta G_{elec,interact} + \Delta E_{vdw,interact} - \Delta G_{elec,L,solv} - \Delta G_{np,L,solv} \quad (2)$$

Equation 2 lacks many of the terms that are considered important for determining binding affinities. Since we will be concerned only with relative binding affinities of the ligands, some of these other terms will cancel. We assume that: every ligand pays the same configurational entropy cost for binding; the receptor adopts the same conformation in each complex; every ligand desolvates the receptor equally; the ligand is completely desolvated on binding; each ligand has only one bound conformation. These assumptions limit the accuracy of our results. The energies returned by the docking calculations are prone to deviations, sometimes enormous deviations, from true binding energies, and we will refer to energy "scores" rather than simply energies in recognition of this. Still, for inhibitor discovery and database screening applications, the first task to perform is to screen out unlikely ligands, and highlight likely ones. Equation 2 significantly improves our ability to do so, over the case where solvation is ignored, as we will show.

METHODS

Approach

We first outline the general procedure for the DOCK database screens and then provide the values of the variables that were used in the various calculations.

The binding site of the protein is defined using spheres[38] that complement the molecular surface[39] of the protein or points in the site where ligand atoms are experimentally known to bind.[7] These spheres and points can be thought of as pseudo-atom positions onto which DOCK superimposes atoms of the database molecules to generate a ligand orientation in the binding site. For any given ligand, multiple orientations are generated, depending on the correspondence of the internal distances of its atoms with those of the receptor spheres.[35,38,40]

Once oriented in the site, the molecule is screened for steric complementarity. For orientations that pass this screen, an interaction energy is calculated based on electrostatic and van der Waals complementarity to the protein. The electrostatic potential of the protein is calculated using the finite-difference Poisson-Boltzmann method implemented in DelPhi.[41] The electrostatic component of the interaction energy comes from multiplying ligand partial atomic charges by the receptor potential at the atom positions of a given ligand orientation.[36] Partial atomic charges are calculated with the Gasteiger algorithm[42] implemented in the program SYBYL (Tripos Associates, 1991).[36] The van der Waals potential of the protein site is calculated with CHEMGRID.[36] The van der Waals component of the interaction energy comes from multiplying ligand van der Waals parameters, assigned by DOCK, by the receptor potential at the various atom positions of a given ligand orientation.[36] For any orientation of a mol-

ecule in a binding site, the interaction energy is:

$$\Delta G_{interact} = \sum_i q_i P_i + v_i P_v \quad (3)$$

Where q_i is the charge of atom i of the ligand, P_i is the electrostatic potential of the receptor at the position occupied by atom i, v_i is the van der Waals atomic parameter of atom i and P_v is the van der Waals potential at position i.

To calculate a free energy of binding, we subtract an electrostatic and a non-polar solvation energy from the interaction energy. The electrostatic component of ligand solvation is calculated with continuum electrostatic method of Rashin,[27,43] implemented in the program HYDREN. The non-polar solvation energy is also calculated by HYDREN and is derived from the surface area of the ligand.[25,27] Both electrostatic and non-polar terms are calculated once and stored in a look-up table. The energy score calculated by DOCK becomes:

$$E_{bind} = \sum_i q_i P_i + v_i P_v - E_{solv,elec} - E_{solv,np} \quad (4)$$

This energy is used to rank the database molecules in their complexes with the protein.

Ligand Solvation

Continuum electrostatic methods for calculating molecular electrostatic hydration energies derive from the Born equation:

$$\Delta G_{solv} = (q^2/2r)(1/D_0 - 1/D_w) \quad (5)$$

Where q is charge, r is the radius of the charged group, D_0 is the dielectric of the phase to which the ligand is being transferred and D_w is the dielectric of water. This equation, with several correction factors, has been used to accurately calculate the solvation energies of spherically distributed charges.[44] Recent work has extended the approach to non-spherical charged and polar groups.[25,27,45,46] The method used here[27,43] employs a reaction field approach. The solvent polarization charge induced by a point charge embedded within a sphere of defined dielectric is calculated. The enthalpy is the interaction energy between the induced polarization charge and the original charge of the atom. The equations used have the same form as the Born equation (Equation 5).[43] A boundary element method is used to define the border between the dielectric of the atom and that of the solvent. Different parts of the solute are represented as continuum dielectrics, or as sets of atoms that are characterized by isotropic atomic polarisabilities. The dielectric boundary is identified with the molecular surface and is determined using the program MS.[39] This surface is divided into discrete smooth boundary elements. The charge distribution of the solute is represented by point charges. Polarization effects are represented by polarization charge densities at dielectric boundaries and by induced dipoles at polarisable atoms. The polarization charge density is assumed to be constant within each boundary element. This enables a system of linear equations to be established that are solved to give the polarization charge densities from which various thermodynamic values can be derived. The method can treat charged or partially charged atoms. The basic approach was attractive to us because it is a continuum dielectric method that seemed physically consistent with DelPhi, which we use to calculate the receptor potential. The particular implementation in HYDREN was attractive because the code was available to us, the program required few external parameters, and could be applied to the large numbers of disparate compounds found in molecular databases. Several other implementations of continuum solvation approaches have been published.[25,26,45,46] Although they differ in the treatment of the solvent boundary and in several correction terms, each appears to give similar results for similar molecules; presumably any of these methods could be used to calculate the solvation energies similar to those used here.

The non-polar component is calculated based on the surface area of the ligand.[43] For non-polar solutes, the hydration enthalpy may be modeled using the size of the cavity created in the solvent to accommodate the ligand, and the surface area of the ligand in the solvent. For most ligands the cavitation term scales as the square of the size of the ligand, i.e., with the surface area.[43] Thus the overall energy of hydration of a non-polar ligand can be related to the surface area of that ligand. If one can split the hydration of a ligand into a polar and a non-polar term, then for all ligands the non-polar component of hydration should scale linearly with the surface area of the ligand. Rashin and co-workers[43] have used this relation to calculate non-polar components of solvation for ligands of arbitrary shape by fitting experimental hydration enthalpies for non-polar ligands to their calculated surface areas. The non-polar component of solvation enthalpy is calculated by HYDREN[43] as:

$$\Delta H_{np} = 8638.59 - 96.518 \times \text{area} \quad \text{area} < 131.11 \text{ Å}^2$$

$$\Delta H_{np} = -621.48 - 25.890 \times \text{area} \quad \text{area} > 131.11 \text{ Å}^2$$

Where ΔH_{np} is the non-polar hydration enthalpy in cal/mol of a molecule with the cavity surface area in Å^2. Here too, similar corrections and methods for calculating them have been proposed by other authors[25]—we presume that these methods would do as well as the one used here.

Parameters Used in All Docking Calculations

All calculations were performed with DOCK3.5, modified to read-in and correct for the pre-calculated electrostatic and van der Waals components of solvation. The polar and non-polar close contact limits used in the steric grids were 2.4 and 2.8 Å.[35] The AMBER united atom charge set, distributed with DelPhi, was used for all receptor electrostatic calculations. All heteroatom hydro-

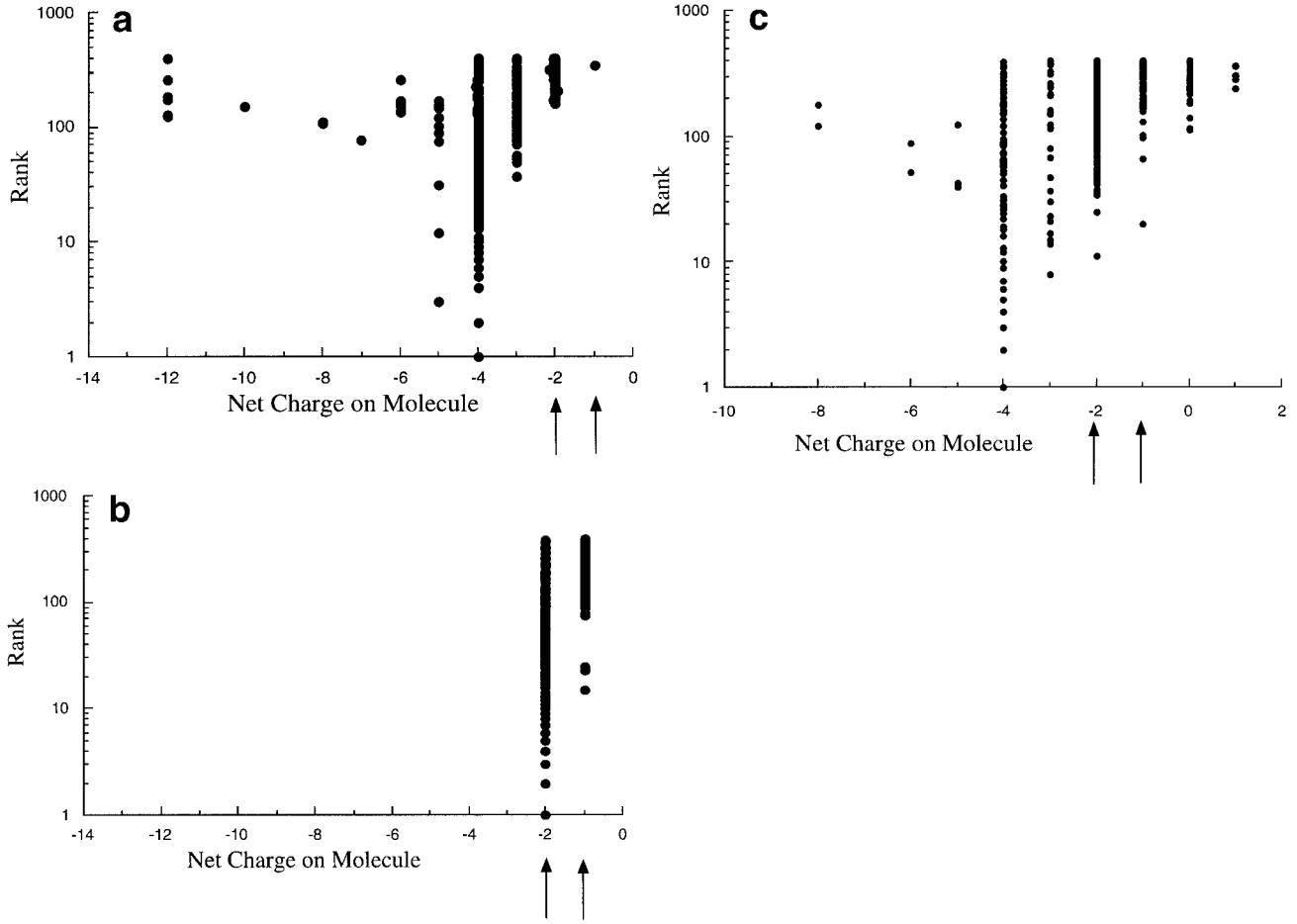

Fig. 1. The variation of DOCK rank with ligand charge for TS. The top ranking 400 molecules out of 153,536 screened by DOCK are shown. The arrows mark the charge state of known nucleotide ligands of TS. **a.** Solvation *uncorrected* search. **b.** Solvation *corrected* search. **c.** Solvation *uncorrected* search using a protein dielectric of 20 instead of 2.

gens used in the DelPhi calculations were placed using the EDIT program in AMBER, unless otherwise noted. CHEMGRID was used to calculate a van der Waals potential for the enzymes using standard parameters.[36] Chemical labeling was used[47] in the matching calculation. This involves labeling site positions or atoms by chemical properties to speed the docking calculation. Here, five labels were employed: positive, negative, hydrogen-bond donor, hydrogen-bond acceptor, and neutral. Except in the DHFR calculation, where one bound-water was included, we did not include structural water molecules or counterions in either the electrostatic or steric calculations of enzyme potential grids.

We used a dielectric of 2 for the protein interior[26] and 78[27] for the water phase in the DelPhi calculations. The internal and external dielectrics in the hydration calculations were also set to 2 and 78.[27] In the DelPhi calculation the probe size was set to 1.4 Å, in the HYDREN calculations the probe radius was 0.8 Å.[27] Atomic van der Waals radii for the protein and the ligand were taken from Rashin.[27] In the DelPhi calculation, the ionic exclusion radius was set to 2 Å and the ionic molarity was set to 0.1 M. The proper values of ligand and protein dielectrics, probe, van der Waals, and ionic radii are active areas of research; we have not tried to optimize these terms.

In the receptor potential calculation three-step focusing[41] was used with protein containment iteratively set to 20, 60, and 90 percent within a 65^3 Å^3 lattice. In the HYDREN calculations the maximum number of iterations was 10, convergence was set to 0.001 with low and high density surface numbers set to 2 and 10. Density numbers were automatically set lower for large ligands that exceeded array bounds when running the programs.

To consider the possibility that using a higher dielectric constant might obviate the need for a formal solvation correction term, we performed a calculation with TS that used a dielectric constant of 20 instead of 2. In this calculation, no solvation correction term was applied. We also conducted calculations on the ligand-bound and unbound conformations of TS and DHFR, to investigate the effect that conformational change would have on the magnitude of the interaction energies.

Except where noted, all database searches used the same 153,536 compound subset of the 1995/2 release of the ACD.[48] These molecules were selected based on our ability

TABLE I. Number of Known Ligands in the Top 400 Ranked Molecules, Out of 153,536 Molecules Searched, With and Without Solvation

Enzyme	Known ligands in top 400 with solvation correction	Known ligands in top 400 without solvation correction	Rank of characteristic ligands with solvation correction	Rank of characteristic ligands without solvation correction
TS	21	11	7[b]	285[b]
			144[c]	575[c]
L99A	28 (74)[a]	0 (39)[a]	29[d]	175[d]
			141[e]	459[e]
DHFR	13 (37)	3 (7)[a]	2[f]	48[f]
			172[g]	2047[g]

[a]Number of molecules that correspond to known ligands (close analogs for which binding data could not be found).
[b]Pyridoxal phosphate.
[c]dUMP.
[d]Indene.
[e]Toluene.
[f]2,4-Diaminopteridine.
[g]2,4-Diamino-6,7-dimethylpteridine.

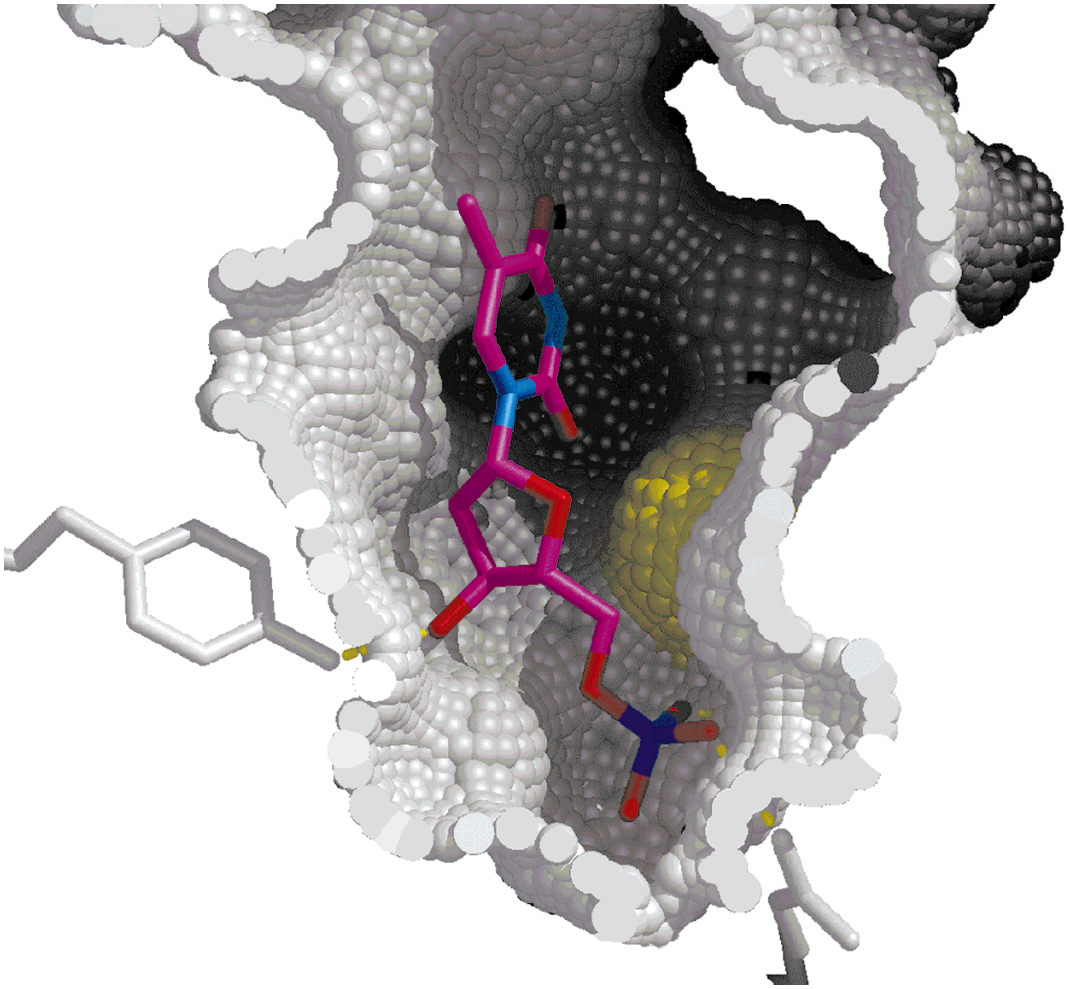

Fig. 2. The docked orientation of thymidine monophosphate in the molecular surface of the TS. As seen in crystallographic complexes with nucleotides and TS, Tyr261 makes a hydrogen bond with the O3' hydroxyl of the ligand (2.8Å), and Arg178' makes a hydrogen bond with a phosphate oxygen (2.6Å). The conformation of the pyrimidine moiety differs from crystallographic complexes of nucleotides with TS owing to the lack of ligand flexibility in the current docking algorithm.

TABLE II. Energy Scores of Known Inhibitors from Docking Searches Against the Bound and Free Conformations of TS and DHFR

Ligand	Enzyme	Energy in bound conformation of enzyme (kcal/mol)	Energy in unbound conformation of enzyme (kcal/mol)
dUMP	TS	−97	−8.4
Pyridoxal phosphate	TS	−118	−16.8
2,4-Diaminopteridine	DHFR	−26	−7.6
6-Methylpterin	DHFR	−18.7	−11.9

to calculate partial atomic charges,[36] and included most of the molecules in the ACD-3D.

TS Docking Calculations

We used the structure of TS from *Lactobacillus casei* determined in the presence of phosphate.[49] The phosphate was deleted from the site to allow database molecules to fit. To allow for the close approach of nucleotides to the catalytic Cys198, the Sg of this residue was deleted from the structure used to calculate potential energy maps. Docking spheres were generated as described.[35] We used a node limit of five and bin sizes and overlaps of 0.2 Å for the receptor, a bin size of 0.2 Å, and an overlap of 0.1 Å for the ligand. The distance tolerance for ligand atom-receptor sphere matching (dislim) was 1.5 Å. Thirty steps of rigid body minimization were conducted for the docked molecules.[50]

To compare the effects of enzyme conformational change on the calculated interaction energies, we also performed docking calculations on the TS from *Escherichia coli* in its unbound conformation (in the absence of nucleotide, PDB accession number 3tms) and in its bound, ternary form (in the presence of dUMP and a folate inhibitor, PDB accession number 1syn). The coordinates of dUMP from 1syn were used as matching spheres in both searches; for the calculation against 3tms this was achieved by RMS-fitting the Cα coordinates of 3tms onto the 1syn structure. All other parameters were as described above.

DHFR Docking Calculations

The structure of DHFR from *E. coli* bound to methotrexate was used (PDB[51] structure 3dfr[52]). The methotrexate was deleted from the structure to allow database molecules to fit into the site. The locations of methotrexate atoms and those of folate fit into the site from a folate complex with DHFR were used as proxies for receptor sphere centers.[47] The docking calculation used bin sizes of 0.5 and 0.5 Å for the ligand and the receptor, with no overlaps and a distance tolerance (dislim) of 1.0 Å. We included the structurally conserved water 253 from the 3dfr structure as part of the protein for the DISTMAP and DelPhi calculations.[47] Also, the Oγ hydrogen dihedral angle of Thr116, which is set to a default value of 180 degrees by AMBER, was rotated to 120 degrees, allowing this residue to better complement pteridine ring nitrogens.[47] No heavy atoms were moved in making this change. Five hundred steps of rigid body minimization were conducted for the docked molecules.[50]

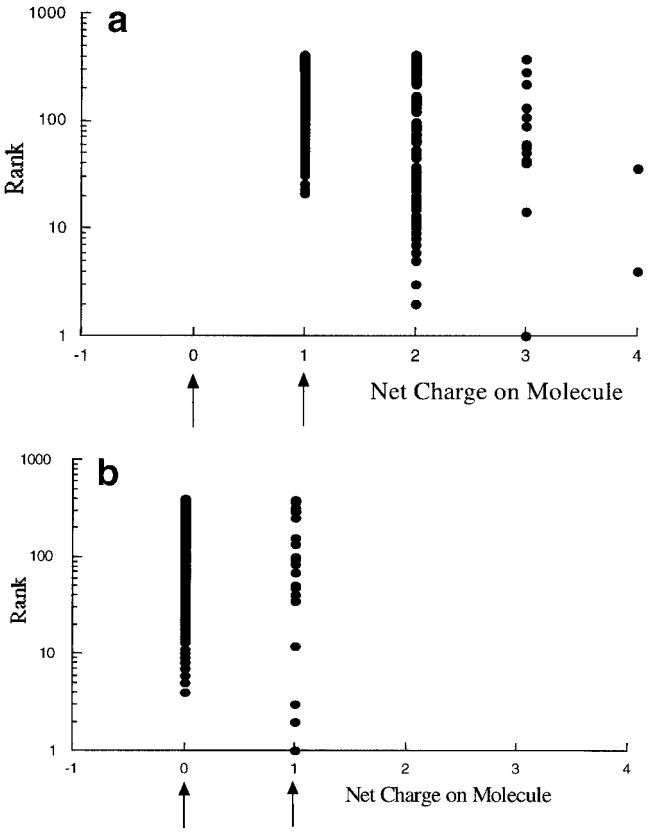

Fig. 3. The variation of DOCK search rank with ligand charge for DHFR. The top ranking 400 molecules out of 153,536 screened by DOCK are shown. The arrow marks the charge state of known ligands for this site. **a.** Solvation *uncorrected* search. **b.** Solvation *corrected* search.

To compare the effects of enzyme conformational change on the calculated interaction energies, we also performed docking calculations on DHFR in its unbound conformation, in the absence of methotrexate (PDB accession number 6dfr). This calculation proceeded as described above.

T4 Lysozyme Docking Calculations

We used the benzene-bound complex structure of L99A (PDB code 181L). The ligand was deleted from the structure to allow database molecules to fit into the site. The atomic positions of other known ligands for the L99A binding site, determined in complex to lysozyme by X-ray crystallography, were fit into this structure by overlapping common receptor atoms. Ligand atoms that did not make steric contacts with L99A in its conformation bound to benzene were used as proxies for receptor sphere centers. Several spheres from the SPHGEN program were also used; in total, 39 potential atom sites were used for these calculations. A distance tolerance (dislim) for matching of 0.75 Å was used. Bin sizes and overlaps were set at 0.2 for both ligand and receptor. One hundred steps of rigid body minimization were conducted for the docked molecules.[50]

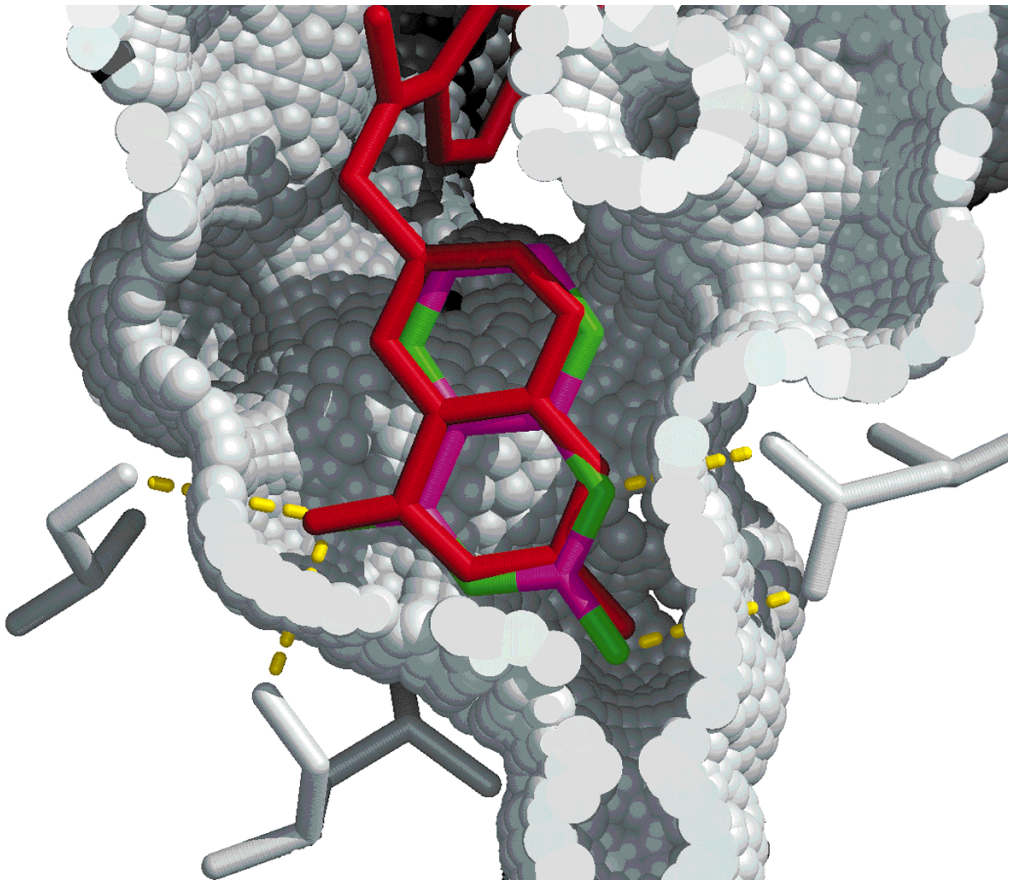

Fig. 4. The docked geometry of 2,4-diaminopteridine (carbons magenta, nitrogen green) overlaid upon the crystallographic configuration of methotrexate (red) in the molecular surface of the DHFR binding site. The surface of the site has been z-clipped to show the ligands. Asp26, Leu4, and Ala97 are shown making hydrogen bond interactions with 2,4-diaminopterindine.

RESULTS

Database searches were run against TS, DHFR, and T4 lysozyme with and without ligand solvation correction. Each database molecule was fit by DOCK into the binding site defined by the spheres or pseudo-atoms. Orientations were evaluated for electrostatic and van der Waals fit.[36] In solvation corrected calculations, the electrostatic and nonpolar components of solvation were subtracted from the electrostatic and van der Waals interaction energies to get a binding energy score. These solvation energy terms were pre-calculated for each molecule before the docking calculation, and were stored in a look-up table.

Docking Screens Against TS

The molecules of the ACD were screened for complementarity to the nucleotide-binding site of TS. This site binds pyrimidine monophosphates and monophosphate esters with K_d values in the 0.1–100 μM range.[53] Several other monophosphates also inhibit the enzyme, including pyridoxal phosphate[54] and phenolphthalein monophosphate (BKS, unpublished results). The enzyme does not bind nucleotide di- or triphosphates. When solvation was not considered, the compounds with the best interaction energies had high net negative formal charges, ranging from −2 to −12, with the majority having a formal charge of −4 (Figure 1a). This calculation was repeated but the cost of desolvating each ligand was now considered. This resulted in top ranking compounds with net formal charges of −2 or −1 (Figure 1b). When solvation correction was used in the docking calculation, the ranks of known ligands were improved compared to when solvation correction was not applied (Table I, Figure 2). For example, in the solvation corrected search, pyridoxal phosphate ranked 7th and dUMP ranked 144th out of 153,536 molecules searched. When solvation was not corrected for, these molecules ranked 255th and 562nd, respectively. When solvation was not considered, nucleotide di- and triphosphates scored well and ranked highly in the dock calculation. This reduced the relative ranks of the nucleotide monophosphates. For example, thymidine 5'-triphosphate was ranked second in the solvation uncorrected calculation but 82,570th in the solvation corrected calculation. All of the di- and triphosphates scored poorly in the solvation corrected calculation, as is appropriate.

To consider the possibility that using a higher dielectric constant might obviate the need for a formal solvation correction term, we performed a calculation with TS that used a protein dielectric constant of 20 instead of 2 (Figure

1c); the water dielectric was kept at 78. In this calculation, no solvation correction term was applied. The higher dielectric constant resulted in high-ranking ligands with reduced energy scores. The rankings of compounds resembled those from the calculation performed with a dielectric of 2 and no solvation. Compounds with high formal charges were ranked better than known ligands that had lower formal charges.

To investigate the effect of conformational change on the magnitude of the interaction energies, we also conducted calculations against a ligand-bound form of TS and a ligand unbound form of the enzyme. Both calculations were solvation corrected. For easy comparison we chose the TS from *E. coli* bound to dUMP and a folate inhibitor (PDB code 1syn) and the same enzyme unbound to any ligand except for phosphate ion; the enzyme has been determined to good resolution in both forms. Known monophosphate inhibitors ranked well in both searches, but their energies of interaction were much reduced in the screen against the unbound receptor (Table II). Fewer known inhibitors were found in the unbound conformation of TS than in the bound conformation of TS.

Docking Screens Against DHFR

The molecules of the ACD database were screened for complementarity to the folate-binding site of DHFR. This site is known to bind neutral and charged diaminopteridines and diaminoquinazolines with disassociation constants that vary from 10 mM to less than 1 nM.[55] When the docking calculations did not correct for ligand solvation, the top ranking ligands bore formal charges ranging between +1 to +4 (Figure 3a). This calculation was repeated but the electrostatic and van der Waals solvation terms for each ligand were now considered (Figure 3b). The best ranking compounds in this second search were either neutral or bore a charge of +1, consistent with the known ligand binding data.[55] When ligand solvation was considered in the docking calculations, thirteen of the top 400 molecules corresponded to known pteridine ligands (Table I, Figure 4). Among these thirteen, both neutral and positive species were found, as appropriate for this site. When ligand solvation was not corrected for, only 3 known ligands were found among the top 400 molecules, and each of these was charged. Analogs of known ligands were also found for this site. These included molecules such as 2,4-diamino-6-hydroxymethylpteridine, which is a close analog of the known ligand 2,4-diamino-6-formylpteridine, but for which binding data could not be found. More of these "reasonable" analogs were found for the docking calculation that corrected for ligand solvation than for the docking calculation that ignored ligand solvation.

To investigate the effect of conformational change on the magnitude of the interaction energies, we conducted calculations against a ligand-unbound conformation of DHFR. This calculation was solvation corrected. As in the TS search, known inhibitors still ranked well against the unbound conformation of the receptor, but their energies of interaction were much reduced (Table II). Fewer known inhibitors were found in the unbound conformation DHFR.

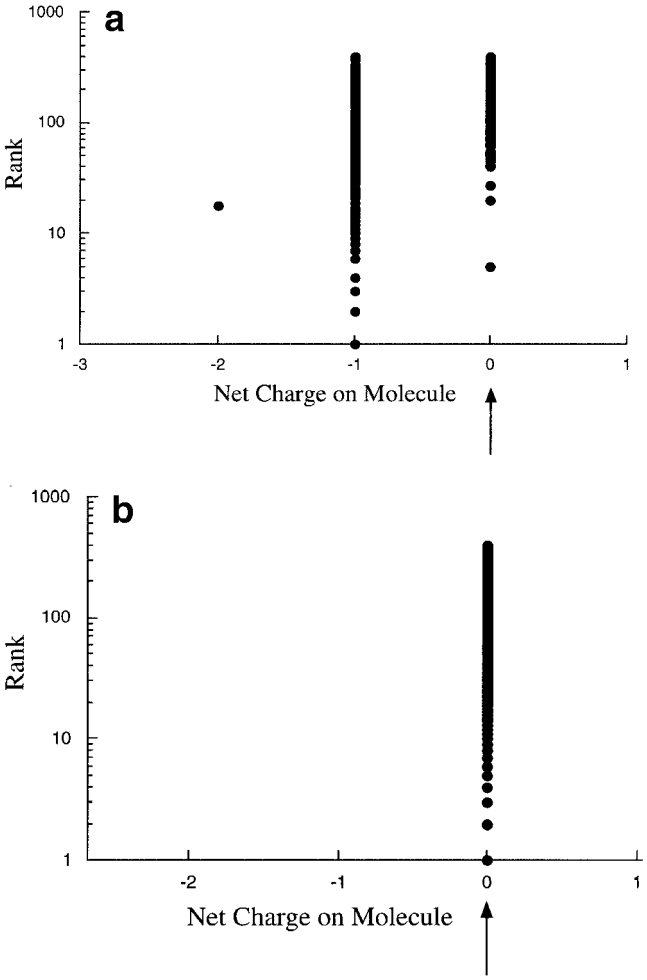

Fig. 5. The variation of DOCK search rank with ligand charge for the core binding site of the mutant lysozyme L99A. The top ranking 400 molecules out of 153,536 screened by DOCK are shown. The arrow marks the charge state of known ligands for this site. **a.** Solvation *uncorrected* search. **b.** Solvation *corrected* search.

Docking Screens Against L99A

A docking search was conducted for the cavity binding site of the L99A mutant of T4 lysozyme. The site was created by substituting a leucine in the core of the enzyme for an alanine.[56] This resulted in a cavity in the enzyme that was buried from water. The cavity binds small, neutral ligands with affinities that range from 10 to 1,000 μM;[57] charged molecules, or overly polar molecules, have not been observed to bind to this site. The molecules of the ACD database were docked into the cavity without correcting for solvation. The best ranking compounds bore formal net charges of −1 (Figure 5a). This calculation was repeated but ligand solvation was now considered (Figure 5b). The best ranking compounds in this second search were neutral, consistent with the known ligand binding data.[57] When solvation correction was used, 28 molecules of the top 400 corresponded to known ligands (Table I, Figure 6). If we include close analogs of known ligands that have not been tested for binding to date, this number rises to 102 molecules in the top 400. For instance, DOCK finds

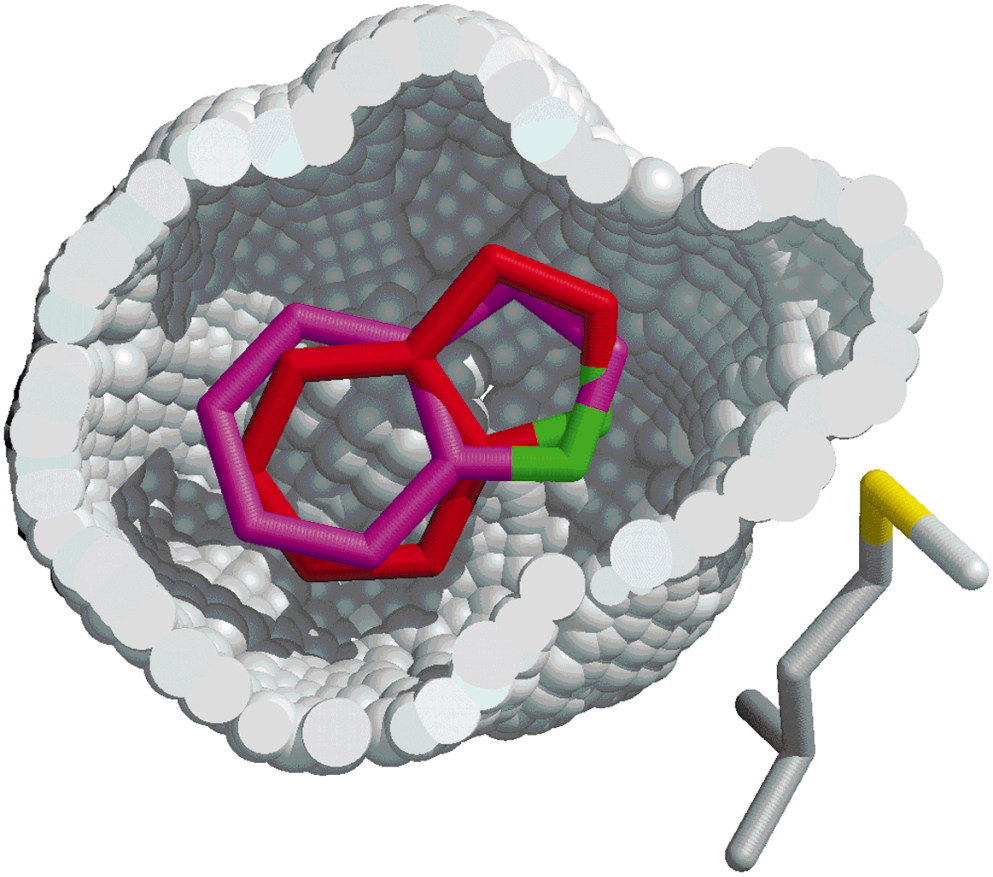

Fig. 6. The docked geometry of indole (magenta), from the database screen, overlaid upon the crystallographic configuration of indole (red) in the molecular surface[39] of the L99A binding site. In both structures the nitrogen is colored green. The surface of the site has been z-clipped to show the ligands. Met102, which makes a hydrogen-bond interaction with the indole nitrogen, is shown at right.

that chlorobenzene fits well into the site. It is not known whether this molecule actually binds to L99A, but both iodobenzene and fluorobenzene bind, and it seemed reasonable to include chlorobenzene among the 74 molecules that are listed as close analogs of known ligands. When solvation correction is not used, the number of ligands in the top 400 drops to 0, or 39 if analogs of known ligands are included. Without solvation correction, the ranks of the known ligands are reduced (Table I). For example, indene, which binds with a K_d of 190 μM, was ranked of 29th in the solvation corrected calculation, but ranked only 1482nd in the solvation uncorrected calculation. Similarly, toluene, which binds with a K_d of 100 μM, was ranked 141st in the solvation corrected screen but ranked only 3,806 in the screen where solvation was not considered. When ligand solvation was neglected, many of the known ligands were replaced in the top ranking 400 molecules with charged or polar isosteres.

To address the effect of non-polar solvation on the docking calculations, the molecules of the ACD were docked into DHFR, with and without the non-polar solvation term (Eq. 4). When non-polar solvation was not considered (Figure 7a), the 400 most complementary molecules in the database were typically larger than when non-polar solvation was considered (Figure 7b). When non-polar solvation was neglected, the top ranking molecules included portions that had few or no contacts with the enzyme, leaving them exposed to solvent (Figure 8). When non-polar solvation was included, the best ranking molecules were more completely surrounded by the enzymes, with few moieties exposed to solvent.

DISCUSSION

Including ligand solvation in molecular docking dramatically changes the relative ranking of compounds in database screens. Since it is this effect that will most interest the general reader, we will consider it first. We will then discuss the absolute energies that are returned by the docking calculations, and several of the approximations that we have made in this implementation of ligand solvation.

Ligand Rankings

When ligand solvation is not considered in molecular docking, there is no penalty for placing a charged ligand atom in a region where the receptor potential only weakly complements it. In this situation, a highly charged molecule will receive a better interaction energy than a true ligand. The true ligand, bearing less formal charge, will have a less favorable interaction energy with the receptor

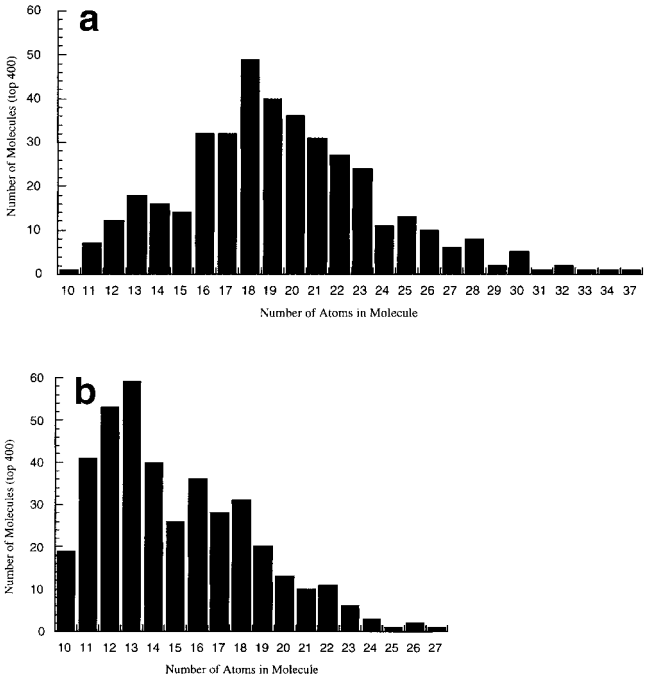

Fig. 7. The effect of correcting for non-polar solvation in a screen of the ACD against DHFR. The distributions of number of atoms within the top 400 top scoring molecules out of 153,536 screened are shown. **a.** Without considering non-polar solvation. **b.** Including non-polar solvation. The number of atoms reported on the x-axis exclude hydrogens.

potential. This result does not depend on the dielectric properties of the enzyme and binding site (e.g., Figure 1c.); dampening the electrostatic interaction energy does not obviate the need to consider ligand desolvation. Neglecting the desolvation of highly charged molecules in database searches results in reduced relative ranks for true ligands.

Considering the cost for electrostatic desolvation lowers the scores of highly charged molecules relative to the known inhibitors. When a charged molecule moves from water to a binding site it exchanges a high dielectric for a low dielectric environment. This increases its self-energy. In contrast to the database screens that did not correct for ligand solvation (Figures 1a, 3a, and 5a), most of the top scoring ligands in the solvation corrected screens have the correct charge state (Figures 1b, 3b and 5b). By considering the cost of moving a charged species from a high to a low dielectric environment, the bias toward molecules bearing high charge is eliminated.

When non-polar solvation is not considered, the top scoring molecules are typically larger than they should be. In the docked complexes, these molecules often have fragments that are poorly complemented by the binding site (Figure 8). Here again, no cost is assessed for removing the compound from the solvent, and so even loose contacts make for a better interaction energy. This biases the calculation toward larger molecules. When non-polar solvation is considered, molecules that make few favorable interactions with the enzyme are disfavored relative to molecules that are well complemented by the binding site. The non-polar solvation term acts as a balance to the van der Waals term in the interaction energy, leading to complexes with a higher proportion of interacting surfaces. The molecules found in such complexes more often correspond to known ligands than when non-polar solvation is ignored.

In summary, ignoring ligand solvation in screens of diverse molecular databases leads to pathologies in docking calculations. Neglecting the electrostatic component of ligand solvation energy results in compounds with high formal charges that rank higher than the known inhibitors for these enzymes (Figures 1, 3, and 5). Neglecting the non-polar component of ligand solvation biases the results towards larger compounds that, overall, complement the binding site worse than the known, smaller ligands.

Energies

In the scoring function used here (Eq. 4), a van der Waals energy derived from the AMBER potential[36] is added to an electrostatic energy calculated using the DelPhi potential.[36] Ligand solvation is subtracted using a Born-derived energy and a surface area-derived non-polar term. The AMBER and DelPhi potentials were derived independently, as were the HYDREN solvation terms (although the HYDREN and DelPhi energies are based on the same physical considerations). Though each method attempts to model physical interactions from first-principles, it is easy to imagine how the individual energy terms might not have the same magnitudes. Since we have made no effort to weight any term in Eq. 4, it is appropriate to ask how well the terms balance in the final binding energy, and how well the predicted energies, as opposed to the rankings, correspond with experiment.

In docking screens against the bound conformation of TS, typical interaction energies, after correcting for solvation, for high-scoring ligands were on the order of −100 kcal/mol (−97 kcal/mol for dUMP). In docking screens against the bound conformation of DHFR, typical energies for high-scoring ligands were on the order of −25 kcal/mol (−26 kcal/mol for 2,4-diaminopteridine). The magnitudes of these energies are unreasonably high. Partly this owes to inaccuracies in the scoring scheme, to which we shall return, and the inherent problem of subtracting two large numbers to get a small one. Partly, the high magnitudes of the interaction energies arises because we are not considering the cost of desolvating the enzymes prior to ligand binding. In effect, we are docking to naked binding sites that have been stripped of their solvating waters. Also, we are docking to enzymes that are in their ligand-bound conformations, without considering the energy cost of changing conformations from the unbound to the bound forms.

By itself, docking to the bound, pre-organized conformation of the enzymes can have a large effect on calculated interaction energies. When the docking calculations were repeated against the unbound conformations of TS and DHFR, the energy scores of the high-ranking ligands were globally reduced and those of known inhibitors moved much closer to their true binding energies (Table II). In the

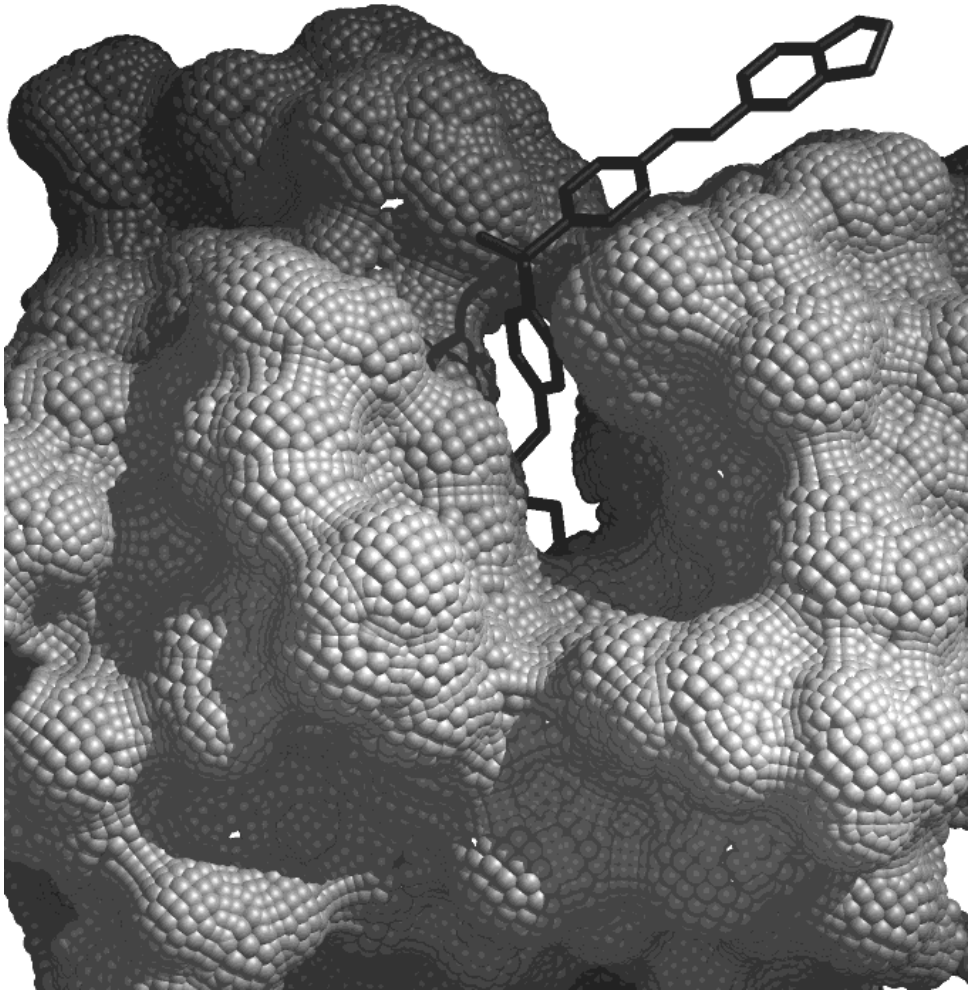

Fig. 8. The effect of non-polar ligand solvation on the size of high-scoring ligands. 4,4'-sulfonyl-bis(N-(5-indanylmethylene)aniline) (dark gray) is shown in bound to the molecular surface of DHFR. This molecule ranked 239th when non-polar solvation was not used, and 10,910th when non-polar solvation was used. All molecular graphics were rendered with neon in MidasPlus;[61] all molecular surfaces were calculated with MS.[39]

bound conformations of the enzyme active sites, residues are in ideal positions to complement the ligand. In the unbound conformations of the enzymes several residues relax away from their ligand-bound conformations. For instance, in TS two of the four arginines that interact with the nucleotide phosphate in the ternary form of the enzyme have moved out of the phosphate-binding site in the unbound form. This leads to lower interaction energies with phosphate groups in the docking calculations. It also leads to some known ligands not being identified. In the calculation against the unbound conformation of DHFR, the known ligand 2,4-diamino-6,7-dimethylpteridine has a positive (unfavorable) binding energy score whereas in the bound conformation calculation this ligand was highly ranked with a favorable energy. We note that efforts to parameterize empirical scoring schemes for docking calculations sometimes use experimental structures of ligand-enzyme complexes, which are of course in the bound conformations of the enzymes. This may lead to biases in the parameter sets.

Compared to the DHFR and TS sites, the L99A binding site allows us to consider the magnitude of errors of our energy terms with less ambiguity. This site is completely buried from solvent, and no ordered waters have been observed in this site by X-ray crystallography. There is little conformational adaptation by the site to the binding of small ligands such as benzene (for larger alkyl benzenes some accommodation is observed.[57] The L99A binding site is thus in many ways a naked binding site. The energies that we calculate for this site are more likely to reflect affinities and not only rankings.

The DOCK energies for known ligands are four to five kcal/mol greater in magnitude than those determined experimentally (Table III). There are several possible explanations for this difference. The calculations do not correct for lost degrees of rotational and translational freedom on binding, nor do they consider gains in vibrational entropy of the system on ligand binding; the calculations ignore hydrophobicity. We have not investigated how these terms add up. It is clear that we are not penalizing for the desolvation of hydrogen bonding groups adequately. For instance, the calculated electrostatic component of desolvating phenol is -0.84 kcal/mol, whereas that for desolvating the isosteric toluene is -0.13 kcal/mol. Apart from absolute errors, the differences in these energies inadequately accounts for the cost of desolvating the polar

TABLE III. Comparing Experimental and Computed Energies for Ligands Binding in the L99A Site

Compound	DOCK energy solvation uncorrected (kcal/mol)	DOCK energy solvation corrected (kcal/mol)	Experimental energy (kcal/mol)
Indene	−18.0	−10.5	−5.13
o-Xylene	−17.2	−9.7	−4.6
Indole	−17.1	−9.5	−4.89
Toluene	−17.1	−9.3	−5.52
Benzene	−15.0	−8.7	−5.19

hydroxyl of phenol. This allows phenol to achieve a DOCK energy score of −9.4 kcal/mol, when in fact this ligand was not observed to bind to the cavity.[57]

Several algorithm changes might improve performance. Our failure to adequately penalize neutral polarity may stem from the use of an inductive method for calculating partial atomic charges.[42] Using quantum mechanically-derived partial atomic charges may improve matters.[27] Until recently this has been unfeasible for the large database of ligands that we wished to consider, but improvements in hardware and software will make this more practical in the future. Another gross approximation in the method is the subtraction of the full solvation energy of the ligand from each orientation's interaction energy (Eq. 2). This over-penalizes ligands, since even fully buried polar groups retain some interaction with solvent.[26] Calculating desolvation penalties that reflected the degree of burial for each orientation of each ligand would improve matters. Of course, this might be computationally expensive. Our reliance on fixed ligand conformations is another source of error in this work—result improve when we allow for the ligand conformational flexibility.[58]

Even with such changes our energy calculations will remain very approximate. Calculating the absolute magnitude of interaction energies remains an unsolved problem for screens against large databases of diverse molecules, such as the ACD, due to the many degrees of geometrical and chemical freedom and to inaccuracies in the force-fields.[13] It may be interesting to explore more computationally intensive scoring schemes for a small number of docked complexes. Progress has been reported in calculating the absolute magnitudes of binding affinities for ligands whose complex to a receptor has been determined to high resolution.[59,60] Although these methods are too computationally intensive to apply to the entire database, they may be useful for detailed calculations against a smaller set of putative ligands after the initial screen. We would be pleased to provide such pre-screened, high-scoring docked ligands, in their complexes with various enzymes, to investigators interested in testing more detailed energy evaluation schemes against the fairly diverse set of ligands returned by structure-based database screens.

Even in the absence of more intensive, detailed energy evaluation schemes, it is clear that fairly simple considerations can dramatically improve our ability to distinguish likely from unlikely ligands. One such is balancing the calculated interaction energy between a ligand and a receptor with a ligand solvation term. If this term is ignored, calculations comparing the binding affinities of dissimilar ligands will be biased towards overly charged and overly large molecules. Correcting for solvation helps us to recognize inhibitors, familiar and yet to be discovered, in database screens of receptors of known structure.

ACKNOWLEDGMENTS

We thank C. Corwin, E. Meng, D. Bodian, and R. Lewis for enlightening conversations, and D. Lorber, B. Beadle, and A. Patera for reading this manuscript. This research was partly supported by the PhRMA Foundation and the Animal Alternatives Program of Procter and Gamble (to BKS), NIH grant GM31497 (to IDK) and the Science and Engineering Research Council (UK) under the NATO postdoctoral fellowship program (to ARL). We thank MDL Inc. (San Leandro, CA) for supplying the Available Chemicals Database.

REFERENCES

1. Boobbyer DN, Goodford PJ, McWhinnie PM, Wade RC. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. J Med Chem 1989;32:1083–1094.
2. Moon JB, Howe WJ. Computer design of bioactive molecules: A method for receptor-based de novo ligand design. Proteins 1991;11: 314–328.
3. Miranker A, Karplus M. Functionality maps of binding sites: A multicopy simultaneous search method. Proteins 1991;11:29– 34.
4. Bartlett PA, Shea GT, Telfer ST, Waterman S. CAVEAT: A program to facilitate the structure-derived design of biologically active molecules. In: Roberts SM, editor. Molecular Recognition. London: Royal Society of Chemistry;1989. p 182–196.
5. DesJarlais RL, Seibel GL, Kuntz ID et al. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. Proc Natl Acad Sci USA 1990;87:6644–6648.
6. Lawrence MC, Davis PC. CLIX: A search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure. Proteins 1992;12:31–41.
7. Shoichet BK, Perry KM, Santi DV, Stroud RM, Kuntz ID. Structure-based discovery of inhibitors of thymidylate synthase. Science 1993;259:1445–1450.
8. Bodian DL, Yamasaki RB, Buswell RL, Stearns JF, White JM, Kuntz ID. Inhibition of the fusion-inducing conformational change of the influenza hemagglutinin by bensoquinones and hydroquinones. Biochemistry 1993;32:2967–2978.
9. Lybrand TP, Ghose I, McCammon JA. Hydration of chloride and bromide anions - determination of relative free-energy by computer-simulation. J Am Chem Soc 1985;107:7793–7794.
10. Bash PA, Singh UC, Langridge R, Kollman PA. Free energy calculations by computer simulation. Science 1987;236:564–568.
11. Warshel A, Sussman F, Hwang J-K. Evaluation of Catalytic Free Energies in Genetically Modified Proteins. J Mol Biol 1988;201: 139–159.
12. Straatsma TP, McCammon JA. Theoretical calculations of relative affinities of binding. Methods Enzymol 1991;202:497–511.
13. Shi YY, Mark AE, Wang CX, Huang F, Berendsen HJ, Gunsteren WFv. Can the stability of protein mutants be predicted by free energy calculations? Protein Eng 1993;6:289–295.
14. Pearlman DA Connelly PR. Determination of the differential effects of hydrogen bonding and water release on the binding of FK506 to native and Tyr82--Phe82 FKBP-12 proteins using free energy simulations. J Mol Biol 1995;248:696–717.
15. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comp Chem 1983;4: 187–217.
16. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P. A new force field for molecular mechanical

simulation of nucleic acids and proteins. J Am Chem Soc 1984;106: 765–784.
17. Gunsteren WFv, Berendsen HJC. GROMOS Library Manual. Groningen: Biomos; 1987.
18. Hansch C, Leo AJ. Substituent Constants for Correlation Analysis in Chemistry and Biology. New York: John Wiley & Sons; 1979. 339 p.
19. Nozaki Y, Tanford C. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. J Biol Chem 1971;246:2211–2217.
20. Wolfenden R. Affinities of amino acid side chains for solvent water. Biochemistry 1981;20:849–855.
21. Fauchere JL, Charton M, Kier LB, Verloop A, Pliska V. Amino acid side chain parameters for correlation studies in biology and pharmacology. Int J Pept Protein Res 1988;32:269–278.
22. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. Nature 1986;319:199.
23. Ooi T, Oobatake M, Nemethy G, Scheraga H. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc Natl Acad Sci USA 1987;84:3086–3090.
24. Tokarski JS, Hopfinger AJ. Prediction of ligand-receptor binding thermodynamics by free energy force field (FEFF) 3D-QSAR analysis: application to a set of peptidomimetic renin inhibitors. J Chem Inf Comput Sci 1997;37:792–811.
25. Still WC, Tempczyk A, Hawley RC, Hendrickson T. A Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127.
26. Gilson MK, Honig B. The inclusion of electrostatic hydration energies in molecular mechanics calculations. J Comput Aided Mol Des 1991;5:5–20.
27. Rashin AA. Hydration phenomena, classical electrostatics and the boundary element method. J Phys Chem 1990;94:1725–1733.
28. Andrews PR, Craik DJ, Martin JL. Functional group contributions to drug-receptor interactions. J Med Chem 1984;27:1648–1657.
29. Williams DH, Cox JPL, Doig AJ, Gardner M, Gerhard U, Kaye PT, Lal AR, Nicolls IA, Salte CJ, Mitchell RC. Towards the semiquantitative estimation of binding constants. Guides for peptide-peptide binding in aqueous solution. J Am Chem Soc 1991;113: 7020–7030.
30. Head RD, Smyte ML, Oprea TI, Waller CL, Green SM, Marshall GR. VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. J Am Chem Soc 1996;118: 3959–3969.
31. Bohm H-J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput Aided Mol Des 1994;8:243–256.
32. DeWitte RS, Shakhnovich EI. SMoG: de novo design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. J Am Chem Soc 1996;118: 11733–11744.
33. Wilson C, Mace JE, Agard DA. A computational method for designing enzymes with altered substrate specificity. J Mol Biol 1991;220:495–506.
34. Wilson C, Gregoret LM, Agard DA. Modeling side-chain conformation for homologous proteins using an energy-based rotamer search. J Mol Biol 1993;229:996–1006.
35. Shoichet B, Bodian DL, Kuntz ID. Molecular docking using sphere descriptors. J Comp Chem 1992;13:380–397.
36. Meng EC, Shoichet B, Kuntz ID. Automated docking with grid-based energy evaluation. J Comp Chem 1992;13:505–524.
37. Ho CMW, Marshall GR. FOUNDATION-A program to retrieve all possible structures containing a user defined minimum number of matching query elements from 3-dimensional databases. J Comput Aided Mol Des 1993;7:3–22.
38. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. J Mol Biol 1982;161:269–288.
39. Connolly ML. Solvent-accessible surfaces of proteins and nucleic acids. Science 1983;221:709–713.
40. Ewing TJA, Kuntz ID. Critical evaluation of search algorithms for automated molecular docking and database screening. J Comp Chem 1997;18:1175–1189.
41. Gilson MK, Honig BH. Calculation of electrostatic potentials in an enzyme active site. Nature 1987;330:84–86.
42. Gasteiger J, Marsili M. Electrostatic charges. Tetrahedron 1980; 36:3219.
43. Rashin AA, Namboodiri K. A simple method for the calculation of hydration enthalpies of polar molecules with arbitrary shapes. J Phys Chem 1987;91:6003–6012.
44. Rashin AA, Honig B. Reevaluation of the Born Model of Ion Hydration. J Phys Chem 1985;89:5588.
45. Jean-Charles A, Nichols A, Sharp K et al. Electrostatic contributions to solvation energies: Comparison of free energy perturbation and continuum calculations. J Am Chem Soc 1991;113:1454–1455.
46. Cramer CJ, Truhlar DG. AM1-SM2 and PM3-SM3 parameterized SCF solvation models for free energies in aqueous solution. J Comput Aided Mol Des 1992;6:629–666.
47. Shoichet BK, Kuntz ID. Matching chemistry and shape in molecular docking. Protein Eng 1993;6:723–732.
48. Guner OF, Hughes DW, Dumont LM. An integrated approach to three dimensional information management with MACCS-3D. J Chem Inf Comput Sci 1991;31:408–414.
49. Finer-Moore J, Fauman EB, Foster PG, Perry KM, Santi DV, Stroud RM. Refined structures of substrate-bound and phosphate-bound thymidylate synthase from *Lactobacillus casei*. J Mol Biol 1993;232:1101–1116.
50. Meng EC, Gschwend DC, Blaney JM, Kuntz ID. Orientational sampling and rigid-body minimization in molecular docking. Proteins 1993;17:266–278.
51. Bernstein FC, Koetzle TF, Williams GJB et al. The protein data bank: A computer-based archival file for macromolecular structure. J Mol Biol 1977;112:535–542.
52. Bolin JT, Filman DJ, Matthews DA, Hamlin RC, Kraut J. Crystal structures of E. coli and L. casei DHFR refined to 1.7 angstrom resolution. 1. General features and binding of methotrexate. J Biol Chem 1982;257:13650–62.
53. Santi DV, Danenberg PV. In: Blakely RL, Benkovic SJ, editors. Folates and Pterins. Vol. 1. New York: John Wiley & Sons, Inc.; 1984. p 345–398.
54. Santi DV, Ouyang TM, Tan AK, Gregory DH, Scanlan T, Carreras CW. Interaction of thymidylate synthase with pyridoxal-5'-phosphate as studied by UV/visible difference spectroscopy and molecular modeling. Biochemistry 1993;32:11819–11824.
55. Blaney JM, Hansch C, Silipo C, Vittoria A. Structure-activity relationships of dihydrofolate reductase inhibitors. Chem Rev 1984;84:333–407.
56. Eriksson AE, Baase WA, Wosniak JA, Matthews BW. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. Nature 1992;355:371–373.
57. Morton A, Matthews BW. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: Linkage of dynamics and structural plasticity. Biochemistry 1995;34:8576–8588.
58. Lorber DM, Shoichet BK. Flexible ligand docking using conformational ensembles. Protein Sci 1998;7:151–158.
59. Bardi JS, Luque I, Freire E. Structure-based thermodynamic analysis of HIV-1 protease inhibitors. Biochemistry 1997;36:6588–6596.
60. Luque I, Gomez J, Semo N, Freire E. Structure-based thermodynamic design of peptide ligands: Application to peptide inhibitors of the aspartic protease endothiapepsin. Proteins 1998;30:74–85.
61. Ferrin TE, Huang CC, Jarvis LE, Langridge R. The MIDAS Display Program. J Mol Graphics 1988;6:13–27.