Rapid Context-Dependent Ligand Desolvation in Molecular Docking

Michael M. Mysinger and Brian K. Shoichet*

Department of Pharmaceutical Chemistry, University of California, San Francisco, California

Received May 27, 2010

In structure-based screens for new ligands, a molecular docking algorithm must rapidly score many molecules in multiple configurations, accounting for both the ligand's interactions with receptor and its competing interactions with solvent. Here we explore a context-dependent ligand desolvation scoring term for molecular docking. We relate the Generalized-Born effective Born radii for every ligand atom to a fractional desolvation and then use this fraction to scale an atom-by-atom decomposition of the full transfer free energy. The fractional desolvation is precomputed on a scoring grid by numerically integrating over the volume of receptor proximal to a ligand atom, weighted by distance. To test this method's performance, we dock ligands versus property-matched decoys over 40 DUD targets. Context-dependent desolvation better enriches ligands compared to both the raw full transfer free energy penalty and compared to ignoring desolvation altogether, though the improvement is modest. More compellingly, the new method improves docking performance across receptor types. Thus, whereas entirely ignoring desolvation works best for charged sites and overpenalizing with full desolvation works well for neutral sites, the physically more correct context-dependent ligand desolvation is competitive across both types of targets. The method also reliably discriminates ligands from highly charged molecules, where ignoring desolvation performs poorly. Since this context-dependent ligand desolvation may be precalculated, it improves docking reliability with minimal cost to calculation time and may be readily incorporated into any physics-based docking program.

INTRODUCTION

Molecular docking is widely used to computationally screen large chemical libraries for small molecules that complement receptors of known structure.[1−4] Complementarity and ranking are calculated by a scoring function that is either based on empirically fit descriptors,[5−9] knowledge-based potential functions,[10−12] or physics-based terms.[13−17] Physics-based scoring functions borrow forcefield derived terms, such as van der Waals (vdW) and electrostatics, to calculate the protein−ligand interaction energy.[18,19] More problematic has been the representation of bulk water, which solvates the unbound ligand and protein and differentially solvates the protein−ligand complex. Proper treatment of this net desolvation is essential, especially in polar or charged systems. This is because a large desolvation penalty opposes a large electrostatic interaction energy, while the much smaller net difference has a strong influence on binding affinity.[20−22]

Desolvation is often subdivided into a polar electrostatic component and a nonpolar component due to dispersion, hydrophobicity, and cavitation.[23] At a higher level of theory, when more computational power is available to analyze each ligand pose, molecular mechanics combined with Generalized-Born (GB) or Poisson−Boltzmann (PB) theory correlates to experimental solvation and occasionally even to binding affinities.[24−29] However, these methods remain at least 5 orders of magnitude slower than docking, which routinely screens more than 10^6 molecules, each in 10^3-10^5 configurations, highlighting the need for desolvation methods fast enough to be relevant to docking.[30]

Early docking methods neglected desolvation entirely and thus favored large and highly charged molecules, since extra size and polarity increase the interaction energy without a compensating penalty for desolvation.[31] To remedy these problems without sacrificing speed, different approximations of desolvation have been investigated. A common approach estimates the desolvation penalty as proportional to the solvent-accessible surface area scaled by atomic solvation parameters.[11,32−34] Whereas surface area is often approximated to be proportional to the nonpolar component of desolvation, it fails to capture the electrostatic component properly. An alternative to surface area is provided by volume-based Gaussian envelopes,[35,36] though these methods are only loosely related to electrostatics. In a recent effort to capture desolvation empirically, 1179 adjustable solvation parameters were fit to three different levels of potentials of mean force, where two levels were unique to every pair of 17 atom types.[37] Although well explored, both the variability in these atom-type approaches and the increasing number of parameters reflect the difficulty in representing the underlying physical terms. Another empirical approach approximates desolvation penalties based on displacement of ordered waters in the binding site,[8,38−40] though calculating the energetic cost or advantage of such displacement remains an area of active research.

The challenge of using physics-based desolvation in docking stems from the difficulties in calculating the energies accurately enough to balance what are often large magnitude terms and doing so fast enough to be relevant. Thus other investigators have tried decomposing the electrostatics of

* Corresponding author e-mail: shoichet@cgl.ucsf.edu. Corresponding author address: Department of Pharmaceutical Chemistry, University of California San Francisco, 1700 Fourth Street, Box 2550, San Francisco, CA 94158-2330.

10.1021/ci100214a
Published on Web 08/24/2010

ligand binding into components arising from partial ligand desolvation, partial receptor desolvation, and screened protein−ligand interactions. A faster implicit solvent model was recently generated by combining the Lorentz-Debye-Sack theory with a first-shell approximation, using the fraction of solvent-accessible surface area to scale empirical covalent radii.[41] Though promising, this method has a nonlinear relationship with what are thought to be more accurate Poisson−Boltzmann energies, and its performance in docking screens remains to be fully explored. Majeux et al. estimate both ligand and receptor desolvation for fragments by combining the Coulomb field approximation with a first-shell approximation, using empirical scaling to reduce the impact of those approximations.[42] Interestingly they also use grid-based numerical integration over volume elements, but for flexible ligands this has been slow as it requires a new ligand desolvation grid for each conformation.

In prior desolvation methods from our group[43] we modeled bound ligand as fully desolvated, approximating the ligand desolvation penalty as the full transfer free energy from high-to-low dielectric computed by the AMSOL program.[44] Whereas this removes the bias toward highly charged molecules, the magnitude of the desolvation penalty is often overestimated, being ideal only when the ligand is completely surrounded by the protein so that it resembles being fully immersed in organic solvent. A preferable idea would be to scale the full desolvation contribution of each atom, proportional to how much it is embedded in the low-dielectric medium (i.e., protein). This scaling factor is offered by Generalized-Born theory, where each atom has an effective Born radius that encapsulates geometric, context-dependent information about that atoms relative desolvation.[45] We begin by relating the GB effective Born radii for every ligand atom to a fractional desolvation and then use this fraction to scale an atom-by-atom decomposition of the full transfer free energy. To make this fast enough for docking, we precompute the fractional desolvation on a regular lattice grid for positions in the protein binding site. Correspondingly, we precalculate the atom-by-atom transfer free energy for each molecule in the docking library. This results in a rapid context-dependent ligand desolvation model for molecular docking.

The exact representation of the dielectric boundary can have more influence on the results than the difference between GB and PB,[46] but its explicit incorporation into molecular docking would be prohibitively expensive.[27] We therefore investigated a treatment of the local solvent-excluded volume (SEV) that captures the first-shell effects of solvent exclusion yet can be precalculated. This context-dependent desolvation term approaches the full transfer free energy as the ligand becomes fully engulfed in protein and approaches the exact GB energy in the limit of a single ligand atom. While these are attractive theoretical properties, we needed to investigate if we can balance desolvation and interaction energy accurately enough to allow for enrichment of ligands in docking screens. Thus we measured the virtual screening performance of several ligand desolvation approximations on the 40 DUD benchmarking targets against their property-matched decoys, in an effort to reduce background database bias.[47] To protect against database extrema, we also screened the ligands against highly charged molecules pulled from the free ZINC database.[48] To better account for early enrichment when measuring virtual screening performance, we introduce a new LogAUC metric. We finally investigate screening performance in two model binding sites against experimentally confirmed, as opposed to presumed, decoys.[43,49]

METHODS

Fractional Ligand Desolvation Scoring Term. In GB theory,[45] the effective Born radius α_i of a given atom can be expressed in terms of the normal atomic radii a_i and a volume integral over all low dielectric space outside the atom itself, where the contribution of each occupied volume element dV diminishes with the distance r between it and atom i as follows

$$\frac{1}{\alpha_i} = \frac{1}{a_i} - \frac{1}{4\pi}\int_{in,r>a_i} \frac{1}{r^4} dV \tag{1}$$

A simple rearrangement allows us to express the fractional desolvation D_i in terms of the volume integral

$$D_i = 1 - \frac{a_i}{\alpha_i} = \frac{a_i}{4\pi}\int_{in,r>a_i} \frac{1}{r^4} dV \tag{2}$$

Checking boundary cases, when α_i approaches a_i then D_i goes to 0 indicating the atom is completely surrounded by solvent, and conversely as α_i approaches ∞ then D_i goes to 1 indicating complete desolvation. We can numerically integrate to find the fractional desolvation by constructing a three-dimensional grid and using the equation

$$D_i = \frac{a_i dV}{4\pi} \sum_k^{in,r>a_i} \frac{1}{r_{ik}^4} \tag{3}$$

where dV is the volume of one grid element, and the summation runs over grid points inside the low dielectric region (Figure 1A).

To approximate ligand desolvation we previously assumed that the ligand becomes fully desolvated in the protein binding site,[31,43] thus treating the desolvation penalty as the full transfer free energy from high-dielectric solvent to low-dielectric protein, as computed by AMSOL.[44] However, if we partition this full transfer free energy to the individual ligand atoms, obtaining $\Delta G_i^{\text{trans}}$, then we can multiply by the fractional desolvation D_i of each ligand atom to estimate the total ligand desolvation penalty as

$$\Delta G_{desolv}^{L} = \sum_i^{ligand} \Delta G_i^{trans} D_i \tag{4}$$

This is an attractive way to account for ligand desolvation during molecular docking since the atomic desolvation penalties may be precomputed, as can the fractional desolvation at any position in the binding site. During docking all that is required is multiplying the atomic transfer free energies by the fractional desolvation interpolated from the precalculated scoring grid.

Fractional Ligand Desolvation Implementation. When the ligand and protein low-dielectric volumes do not overlap substantially, we can decompose the fractional desolvation into ligand, receptor, and solvent-excluded volume (SEV) components. We can readily precalculate the receptor com-

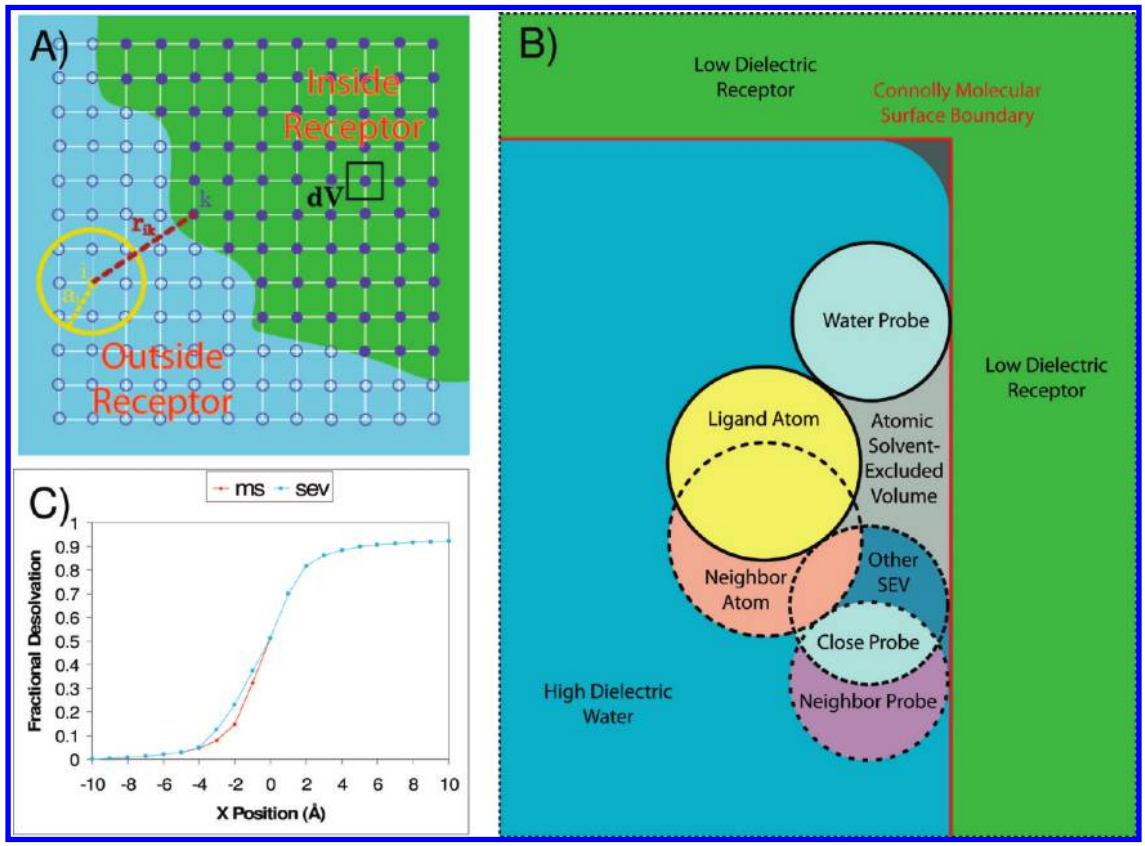

Figure 1. Volume-based desolvation implementation: A) Illustration of fractional desolvation numerical integration scheme: a_i = radius of ligand atom i, r_{ik} = distance from ligand atom i to volume element k. B) Incorporating solvent-excluded volume: At close distance a solvent-excluded volume (SEV) forms between protein (green) and ligand (yellow) where water probes (light blue) are unable to fit. We propose precalculating the atomic SEV region (gray) while neglecting the neighboring SEV region (darker blue). C) Fractional desolvation plotted versus the distance inside (+X) or outside (-X) an idealized solid low-dielectric slab: Atomic SEV method (cyan) plotted versus using only the receptor's molecular surface (red).

ponent and use a first-shell approximation to precalculate the SEV component. Admittedly, the portion of any ligand atom's desolvation that depends on the position of *other* ligand atoms remains expensive to compute, because it varies between ligand conformations. When we multiply fractional desolvation by atomic transfer free energy (eq 4), we therefore neglect the ligand component. This can lead to errors in some circumstances, but since this term is included in the AMSOL-derived[44] transfer free energies, expressing desolvation as a percentage of the initial transfer free energy partly compensates for this deficit (Table SI6 gives an atom-by-atom estimate of this error for one challenging molecule). With these approximations, the fractional desolvation at every point depends only on the geometry of the receptor and can be precalculated.

We initially used the dielectric boundary of the receptor alone to compute the fractional desolvation grid. However, solvation energies can strongly depend on the dielectric boundary, even more than the exact type of implicit solvation model used.[46] The ideal dielectric boundary for ligand desolvation is the molecular surface of the protein−ligand complex, but calculating this on-the-fly at the time of docking would be prohibitive, so we investigated a faster method. Upon ligand binding, a solvent-excluded volume, or gap, forms between the ligand and receptor where even a single water molecule cannot fit (Figure 1B). This SEV affects the fractional desolvation of the ligand but generally depends on the pose of the ligand in the protein which is not known until the ligand is docked. To enable fractional desolvation precalculation, we assume for each ligand atom that we can independently account for the SEV due to it alone (gray region) while neglecting the SEV due to other atoms in the ligand (darker blue region). With this atomic SEV correction, our method reproduces the exact Generalized-Born energy in the limiting case of single ligand atom. But even for molecules, neglecting distal excluded volume elements has only a modest effect since this term decays at $1/r^4$.

The numerical integration of the fractional desolvation is encoded in a DOCK accessory program sevsolv. Precomputing the atomic SEV correction involves recalculating the molecular surface for each ligand atom position in the grid and then numerically integrating using that new surface. The program computes the receptor molecular surface using the inkblot algorithm,[50] which generates evenly spaced points on a sphere at the solvent-accessible surface (SAS) distance (vdW + probe radius) using the golden section spiral algorithm, before removing inaccessible surface points inside other receptor atoms. The grid inside the SAS is tentatively assigned to low-dielectric, and then a water probe is placed at each accessible SAS point and the region inside each probe is reassigned back to high-dielectric. This effectively blots back to the molecular surface starting from the accessible SAS points. For the atomic SEV we keep track of how many receptor accessible SAS points are marking each grid point as solvent accessible, then for each grid position the ligand atom is placed there, and a cubing algorithm finds all newly inaccessible receptor surface points. If none are found, then this ligand atom position is either completely surrounded by

solvent or completely buried in protein, and there is no atomic SEV correction. Otherwise, we find the ligand's accessible SAS points and increment the marks of the grid points within a probe radius. For the newly inaccessible receptor SAS points we decrement the marks of grid points within probe radius and make the grid point low-dielectric if no marks remain to cover it. Finally, we blot the region inside the ligand vdW as low-dielectric and compute the numerical integral over this new SEV surface. The current implementation takes under half an hour to generate the atomic SEV desolvation grid on a single 2008 CPU core for any of the 40 targets in DUD.

Here we use a 1.4 Å water probe radius and 1000 SAS points per atom. The integration cutoff is 10 Å with a grid resolution of 0.5 Å. These choices were made as a compromise between speed and accuracy; the cost of the calculation rises as the third power of each of these terms. Since the cost of precalculating the desolvation grid is only paid once, before docking, should one want more accuracy it is available. For instance, for a probe atom buried in a 48 Å cube—a case where one might see maximal effects of truncation—the desolvation penalty changed by 3% in going from 0.5 Å to 0.33 Å resolution, and by an additional 1% to going to 0.167 Å resolution. Correspondingly, the penalty changed by 6% in going from a 10 Å to a 20 Å cutoff (Figure SI4). The input radii were set to 1.6, 1.65, 1.9, 1.9, 1.9, 0, and 1.6 Å for O, N, C, S, P, H, and other protein atoms, respectively. Unless noted otherwise, the assumed Born radius for all ligand atoms is 1.4 Å, a value intermediate between hydrogens and heavy atoms, and was chosen to enable one fractional desolvation grid to be used for all ligand atoms during docking; other implementations may be imagined where different desolvation grids are precalculated for each ligand atom type.

Similarly, the atom-by-atom partitioning of the transfer free energy can be precalculated and stored in the ligand database prior to docking. Here we further assume that the atomic transfer free energies are similar between all ligand conformers. This assumption could have a substantial effect on overall solvation energies for highly flexible or charged molecules; this is an area of ongoing research. To compute atom-by-atom transfer free energies we perform two AMSOL calculations using the SM5.42R solvation model with the semiempirical AM1 Hamiltonian,[44] subtracting the results in water from those in hexadecane (dielectric = 2.06). From this procedure we obtain separate polar and nonpolar atomic transfer free energies, and the DOCK accessory mol2db stores them into the input flexibase[51] as found in ZINC.[48]

We add this ligand desolvation scoring term to DOCK 3.5.54, resulting in the following scoring function

$$E_{score} = E_{vdW} + E_{elec} + \Delta G^{L}_{desolv} \quad (5)$$

where the vdW term is based on the AMBER united-atom force field,[52] and the electrostatics term is based on PB calculations by DELPHI,[50] as described previously.[13] The fractional desolvation is trilinearly interpolated from the precomputed desolvation scoring grid for each ligand atom position and multiplied by the polar and nonpolar desolvation terms read from the ligand flexibase.

DUD Database Docking. To reduce enrichment bias and focus on the screening performance of molecular docking

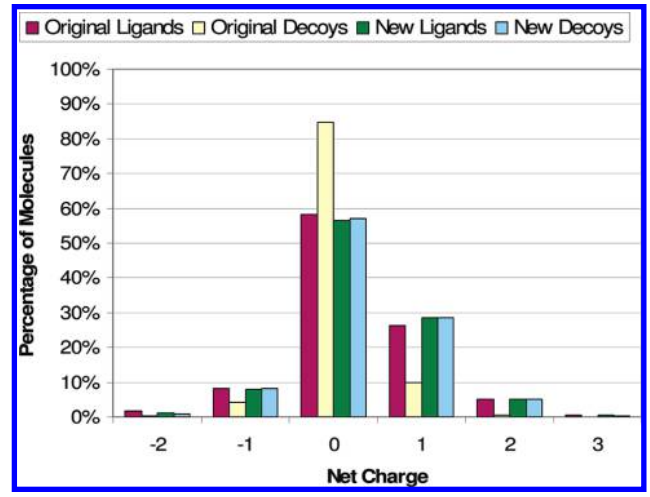

Figure 2. DUD charge distribution: Percentage of new and original DUD molecules that have each given net charge. Over 40% of new and original ligands are charged, but only 15% of the original decoys are, while the new decoys closely match the charge distribution of the ligands.

itself, the DUD benchmarking set property-matches decoys to known ligands.[47] In using the 40 DUD targets to examine enrichment performance of ligand desolvation methods, we found that we had to property-match ligands and decoys on net formal charge, a term which was originally neglected. Without net charge, the original DUD decoys were much more neutral than the ligands (Figure 2). To overcome this deficit we recalculated DUD, matching the original ligands to new computational decoys. We property-matched on the original five physical terms plus net molecular charge, creating an automated procedure that may be applied to any user specified set of ligands (M.M.M. and John J. Irwin, unpublished results).

The ligands and decoy sets of this charge-matched DUD database were built into input dockable flexibases by an updated ZINC[48] procedure. To remove input structure bias, the molecules were converted to smiles by Openeye's OEChem library,[53] stereochemically enumerated and built into 3-D structures by Molecular Networks' Corina,[54] protonated and tautomerized by Schrödinger's Ligprep,[55] and then run through two AMSOL[44] calculations to generate partial charges and atomic desolvation contributions as mentioned previously. Corina[54] was run again to generate aliphatic ring puckers, and each ring is selected in turn to serve as the starting rigid fragment for conformational enumeration by Openeye's Omega.[56] The conformational ensemble and atomic desolvation were combined by mol2db[51] to generate the input docking hierarchies.

Input parameters for docking and grid generation were as described previously,[47] except as noted. We preferentially used the semiautomated preparation of the receptor for the 13 targets where it was available. The bump limit was set to allow up to one steric clash, with a DISTMAP grid resolution of 0.333 Å. We performed 50 steps of simplex minimization for the best initial pose. In the electrostatic potential map calculations using DELPHI,[50] the exterior and interior dielectrics were set to 78.5 and 2.06, respectively, while the salt concentration was 0.145 M and the ion exclusion radius was 2 Å.

LogAUC Virtual Screening Metric. Virtual screening performance is typically evaluated using enrichment or

receiver operator characteristic (ROC) plots,[57] which measure the prioritization of annotated ligands versus known or presumed decoys among the highest ranking molecules. Often one wishes to overweight molecules that rank among the very top of the "hit-list", as these are the ones most likely to be tested for activity in an actual prospective experiment. Several approaches have been introduced to do so;[58,59] here we adopt a variation of the standard ROC plot that, rather than plotting percentage of decoys found vs percentage of annotated ligands, plots the $\log_{10}$ of the decoys found on the x-axis, which preferentially weights early enrichment.

The area under the curve (AUC) is a well-regarded metric to summarize the overall performance of a virtual screening campaign as a single number.[57] While ROC AUC can be formulated in alternate ways, it can mechanically be constructed by integrating under the ROC curve; and interpreted as the fraction of the area under the ROC curve over the area under the best possible ROC curve, which happens to be 1. By analogy, in the case of our semilog ROC plot, we can construct the same fraction of the area under the current log curve over the area under the perfect log curve and define that fraction as the LogAUC. The lone wrinkle is that the area under the perfect log curve is, in general, infinite. However, if we are practical and limit our focus to a region of log space we can actually measure, say above a certain threshold λ, then the perfect log area is finite in that region.

Formally, we define LogAUC_λ, where the log area computations run from λ to 1.0, and in this paper we refer to $\text{LogAUC}_{0.001}$ as simply LogAUC. For integrating the area under the log curve we use an analytical formula derived by log transforming an individual trapezoidal segment. The y-intercept (b) of the line for a given trapezoidal segment is

$$b_i = y_{i+1} - x_{i+1}\left(\frac{y_{i+1} - y_i}{x_{i+1} - x_i}\right) \tag{6}$$

LogAUC_λ is then calculated as follows

$$\text{logAUC}_\lambda = \frac{\sum_i^{\text{where } x_i \geq \lambda} \left(\frac{y_{i+1} - y_i}{\log_e 10}\right) + b_i(\log_{10} x_{i+1} - \log_{10} x_i)}{\log_{10}(1.0/\lambda)} \tag{7}$$

From parallel reasoning based on semilog ROC plots, Clark and Webster-Clark construct the pROC AUC metric,[58] which is similar to the numerator of LogAUC except that the area integration is over horizontal bars instead of vertical trapezoids. The advantage of constructing LogAUC as a fraction over the ideal area is that the choice of base for the logarithm is irrelevant, because changing base simply results in a constant that cancels between numerator and denominator. Also, by explicitly defining the area of interest using λ and integrating vertically, we are able to avoid the singularity at $x_i = 0$ encountered in pROC. More importantly the fixed integration area means we can compare LogAUC values across databases of different sizes and across targets with different ratios of actives to inactives. While the recent NSQ_AUC metric[60] also recognizes the need to normalize based on a perfect ROC curve, the multiplicative random curve based rescaling makes the metric harder to rationalize. In contrast LogAUC_λ is simply the percentage of the total area below a semilog ROC curve plotted from λ to 1, which intuitively feels like linear AUC, except that it is derived from the semilog plot. The random line on a semilog ROC plot (dashed line, Figure 4) occupies only a sliver of the total area, and indeed its LogAUC is just 14.5%. In order to more easily see the performance above random enrichment, we report the "adjusted LogAUC" as the LogAUC minus this random LogAUC, so that positive values mean enrichments better than random and the maximum enrichment occurs at 85.5% adjusted LogAUC.

Charge Outliers. As noted previously,[31] a substantial problem of neglecting desolvation is that highly charged molecules pollute the top hit lists of polar targets. Since the revised DUD is property-matched on net charge, the enrichment of a given target's ligands versus its own decoys fails to interrogate the ability of the desolvation penalty to penalize these highly charged hit list polluters. To address this issue, we pulled 500 random molecules from ZINC with each net charge of −4, −3, −2, +2, +3, and +4 to create a charge outlier database of 3000 molecules. We then compared the enrichment of the ligands for each target against this background database of random highly charged molecules, to see how competitive these random charge outliers were under the various scoring functions.

Pose Fidelity. To check how ligand desolvation affected docked ligand poses, we investigated pose fidelity across the 114 protein−ligand complexes in the DOCK5 benchmark.[61] The automated DOCKBlaster[62] procedure was used to prepare all calculations beginning with just the pdb code for 84 out of 114 targets, while the ligand three-letter ligand identifier was needed to build 16 additional targets. For the following targets manual intervention was needed, typically to fix ligand or cofactor parameters: 1FLR, 1IMB, 1AOE, 4COX, 1ETT, 3CPA. The updated ligand flexibase building procedure described above in the DUD Database Docking section was used to generate input ligand databases, except that Schrödinger's Epik[63] replaced Ligprep[55] for protomer and tautomer generation.

Model Binding Site Docking. To test enrichment performance versus true experimental decoys, rather than the presumed DUD decoys, we docked against two model cavity sites where nonbinders have been experimentally confirmed. The hydrophobic pocket introduced into T4 lysozyme by the Leu99→Ala substitution (L99A) accepts small apolar ligands. Previously, 73 ligands and 64 experimental decoys have been determined for this cavity.[30,43,64−67] In contrast, the cavity introduced into cytochrome *c* peroxidase by the Trp191→Gly substitution (W191G) accepts charged ligands. Previously, 61 ligands and 26 experimental decoys have been determined for this charged cavity.[30,49,68,69] We analyzed ROC curves for the ligands against either the experimental decoys alone or the experimental decoys seeded into a fragment-like subset of ZINC filtered only by molecular weight (30−250).

RESULTS

Here we compare the performance of three ligand desolvation approximations: "none", "full", and "sev". "None" lacks any scoring penalty for ligand desolvation, serving as a crude lower bound; while "full" uses the entire transfer free energy from water to hexadecane as the ligand desolvation penalty, serving as an upper bound. Our ligand

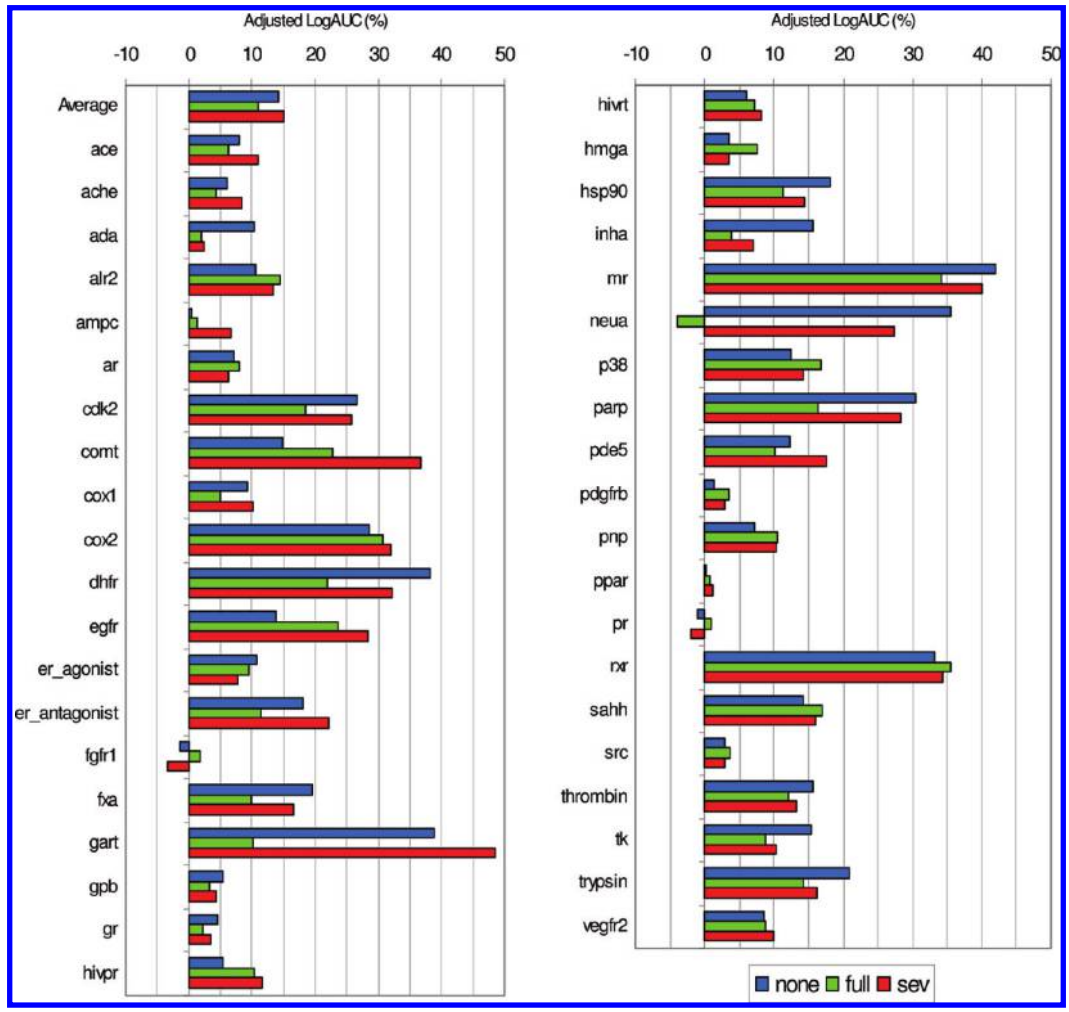

Figure 3. DUD enrichment comparison: Comparison of adjusted LogAUC against DUD decoys over the 40 DUD targets. The docking calculations are identical except that "none" contains no ligand desolvation term, "full" uses the full ligand transfer free energy from water to organic solvent as the desolvation term, and "sev" uses context-dependent ligand desolvation while accounting for the local solvent-excluded volume as a ligand atom approaches the receptor.

desolvation term takes the per-atom contributions from the full desolvation penalty and scales them by the fractional desolvation each ligand atom experiences in Generalized-Born theory. The "sev" method computes fractional desolvation while further accounting for the solvent-excluded volume (gap) that arises between a single ligand atom and the receptor at close distances.

To check that the fractional desolvation implementation is reasonable, we first looked at a low-dielectric slab, infinite in the *yz*-plane, covering half of all space with the dielectric boundary occurring at $x = 0$ Å. Plotting the fractional desolvation in this idealized system using just the receptor's molecular surface (ms) or the atomic SEV method (sev) shows several key features (Figure 1C). The fractional desolvation approaches zero far away from the slab but approaches a value slightly less than 1 (0.92) deep inside the slab. As complete desolvation can only be achieved when buried in an infinite slab,[70] the lower value is due to cutting off the numerical integral at 10 Å (Figure SI4B). When straddling the dielectric wall at $x = 0$ Å, the fractional desolvation is close to the envisioned one-half (0.51). Also, the "ms" and "sev" results are identical except for the region between 0 and 4 Å away from the slab, where the solvent-excluded volume between ligand and slab increases the fractional desolvation, as expected.

Next we used the 40 diverse protein targets in the DUD benchmark[47] to examine the overall virtual screening performance of our context-dependent ligand desolvation terms. DUD combines known ligands for each of the targets with decoys designed to be a challenging test of the predictive value added by docking. To remove enrichment biases due to simple physical properties, the decoys are property-matched to the ligands; and to ensure they are not actual binders, the decoys are selected to be topologically dissimilar from the ligands. Despite this physical property matching, we found it necessary to eliminate an additional property bias that is critical when looking at ligand desolvation, that of net molecular charge. When net charge was not property-matched, as in the original DUD, the resulting enrichments were artificially inflated for targets with highly charged ligands using no ligand desolvation and conversely inflated for targets with neutral ligands using full desolvation. We therefore recalculated the DUD decoys, retaining the same ligands, but adding net-charge as an additional matched property. While only 15% of the original DUD decoys were charged,

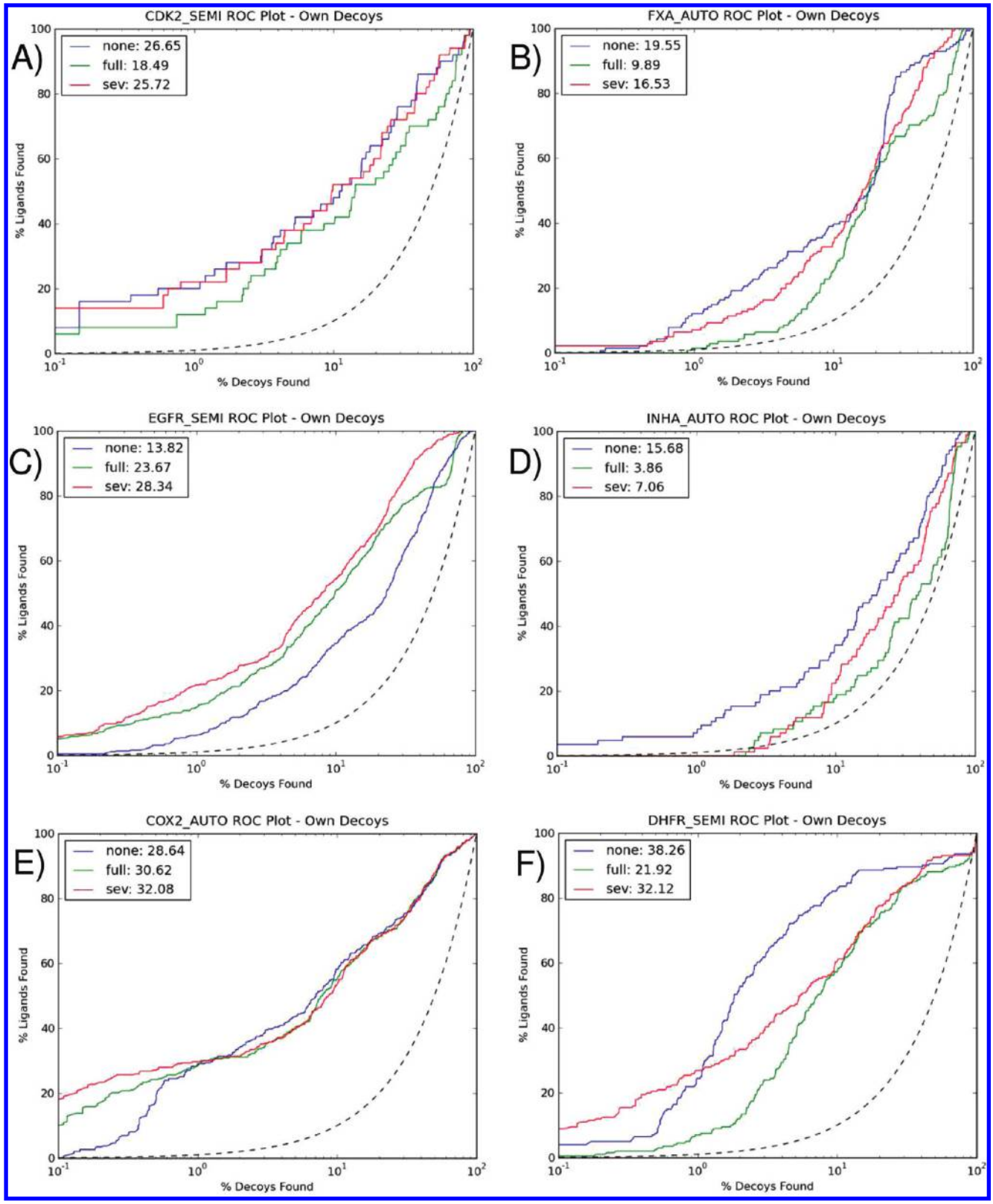

Figure 4. Selected DUD enrichment plots: Individual semilog ROC plots highlight the adaptability of "sev" context-dependent desolvation, which properly tracks no desolvation in both panels A) cdk2 and B) fxa and properly tracks full desolvation in panel C) egfr. In the cases where it does not track the best curve, it typically takes intermediate values as shown in panel D) inha. Early enrichment is often improved the most, as shown in panels E) cox2 and F) dhfr.

around 40% of the new decoys are now charged, and the entire charge distribution of the new decoys now matches the ligands closely (Figure 2). This updated DUD database is available free of charge at http://dud.docking.org.

With new DUD decoys in hand, we can compare the virtual screening performance of the three different approximations to ligand desolvation. After docking, we compute ROC curves for the DUD ligands versus their

Table 1. Enrichments against Matched Decoys over the 40 DUD Targets

		Ligand Desolvation Type		
Adjusted LogAUCs	**Abbreviation**	none	full	sev
Target	*Average*	*14.3*	*10.9*	*15.0*
angiotensin-converting enzyme	ace[c]	7.9	6.4	11.1
acetylcholinesterase	ache[c]	6.1	4.3	8.5
adenosine deaminase	ada[e]	10.4	2.0	2.3
aldose reductase	alr2[b]	10.7	14.4	13.4
AmpC beta-lactamase	ampc[e]	0.5	1.3	6.7
androgen receptor	ar[d]	7.1	7.9	6.3
cyclin-dependent kinase 2	cdk2[a]	26.7	18.5	25.7
catechol O-methyltransferase	comt[b]	14.9	22.8	36.8
cycloxygenase-1	cox1[c]	9.2	5.0	10.1
cycloxygenase-2	cox2[d]	28.6	30.6	32.1
dihydrofolate reductase	dhfr[a]	38.3	21.9	32.1
epidermal growth factor receptor	egfr[b]	13.8	23.7	28.3
estrogen receptor agonists	er_agonist[d]	10.8	9.4	7.7
estrogen receptor antagonists	er_antagonist[a]	18.0	11.5	22.1
fibroblast growth factor receptor kinase	fgfr1[e]	-1.5	1.8	-3.4
factor Xa	fxa[a]	19.6	9.9	16.5
glycinamide ribonucleotide transformylase	gart[a]	38.8	10.1	48.6
glycogen phosphorylase β	gpb[e]	5.4	3.3	4.3
glucocorticoid receptor	gr[e]	4.6	2.2	3.5
HIV protease	hivpr[b]	5.4	10.3	11.6
HIV reverse transcriptase	hivrt[d]	6.0	7.1	8.1
hydroxymethylglutaryl-CoA reductase	hmga[e]	3.5	7.5	3.5
heat shock protein 90	hsp90[a]	18.1	11.4	14.5
enoyl ACP reductase	inha[a]	15.7	3.9	7.1
mineralocorticoid receptor	mr[a]	42.0	34.1	40.0
neuraminidase	neua[a]	35.5	-4.0	27.3
p38 mitogen activated protein	p38[b]	12.6	16.7	14.2
poly-ADP-ribose polymerase	parp[a]	30.5	16.4	28.2
phosphodiesterase 5	pde5[c]	12.4	10.2	17.6
platelet derived growth factor receptor kinase	pdgfrb[e]	1.4	3.6	3.0
purine nucleoside phosphorylase	pnp[b]	7.1	10.5	10.3
peroxisome proliferator activated receptor γ	ppar[e]	0.2	0.8	1.2
progesterone receptor	pr[e]	-0.9	0.9	-2.0
retinoic X receptor α	rxr[d]	33.2	35.6	34.3
S-adenosyl-homocysteine hydrolase	sahh[d]	14.2	17.0	16.1
tyrosine kinase SRC	src[e]	2.9	3.6	2.9
thrombin	thrombin[d]	15.7	12.1	13.2
thymidine kinase	tk[a]	15.4	8.7	10.2
trypsin	trypsin[a]	20.9	14.2	16.2
vascular endothelial growth factor receptor	vegfr2[d]	8.6	8.9	9.8

[a] No desolvation performs much better than full. [b] Full desolvation performs better than none. [c] No and full desolvation are similar, but "sev" outperforms both. [d] All desolvation models perform similarly. [e] All desolvation models perform poorly.

own decoys, using the adjusted LogAUC to summarize the performance of each curve. LogAUC equally weights the area under the curve in the decade between 0.1% and 1% with the other two decades with upper bounds of 10% and 100%. This results in a metric with the attractive features of AUC, but biased toward early enrichment, and compares favorably to other early enrichment metrics (see Methods).

The results for the no desolvation ("none") and full desolvation ("full") are highly target dependent, whereas the context dependent desolvation shows competitive enrichment across most targets (Figure 3, Table 1). In one set of targets (highlighted blue) where no desolvation outperforms full desolvation, which typically have charged or open binding sites, context-dependent ligand desolvation generally tracks no desolvation. For instance in poly-ADP-ribose polymerase with its very open binding pocket, no desolvation gives 30.5% adjusted LogAUC, outperforming full desolvation at 16.4%, but context-dependent "sev" solvation falls just behind no desolvation at 28.2% adjusted LogAUC. In another set of targets (highlighted green) where full desolvation outperforms no desolvation, which typically have neutral or enclosed binding sites, context-dependent ligand desolvation adapts to track the better performing full desolvation. For instance in HIV protease with its mostly enclosed binding pocket, no desolvation performs worst at 5.4% adjusted LogAUC, full desolvation performs better at 10.3%, while "sev" performs best at 11.6%. While no and full desolvation show considerable variation in enrichment from target-to-target, often switching from best to worst, context-dependent desolvation is more consistent across targets.

For some targets both no and full desolvation perform similarly, while "sev" desolvation outperforms both (highlighted red). Other targets are uninformative, either due to poor performance across the board (highlighted purple) or due to similar performance regardless of the ligand desol-

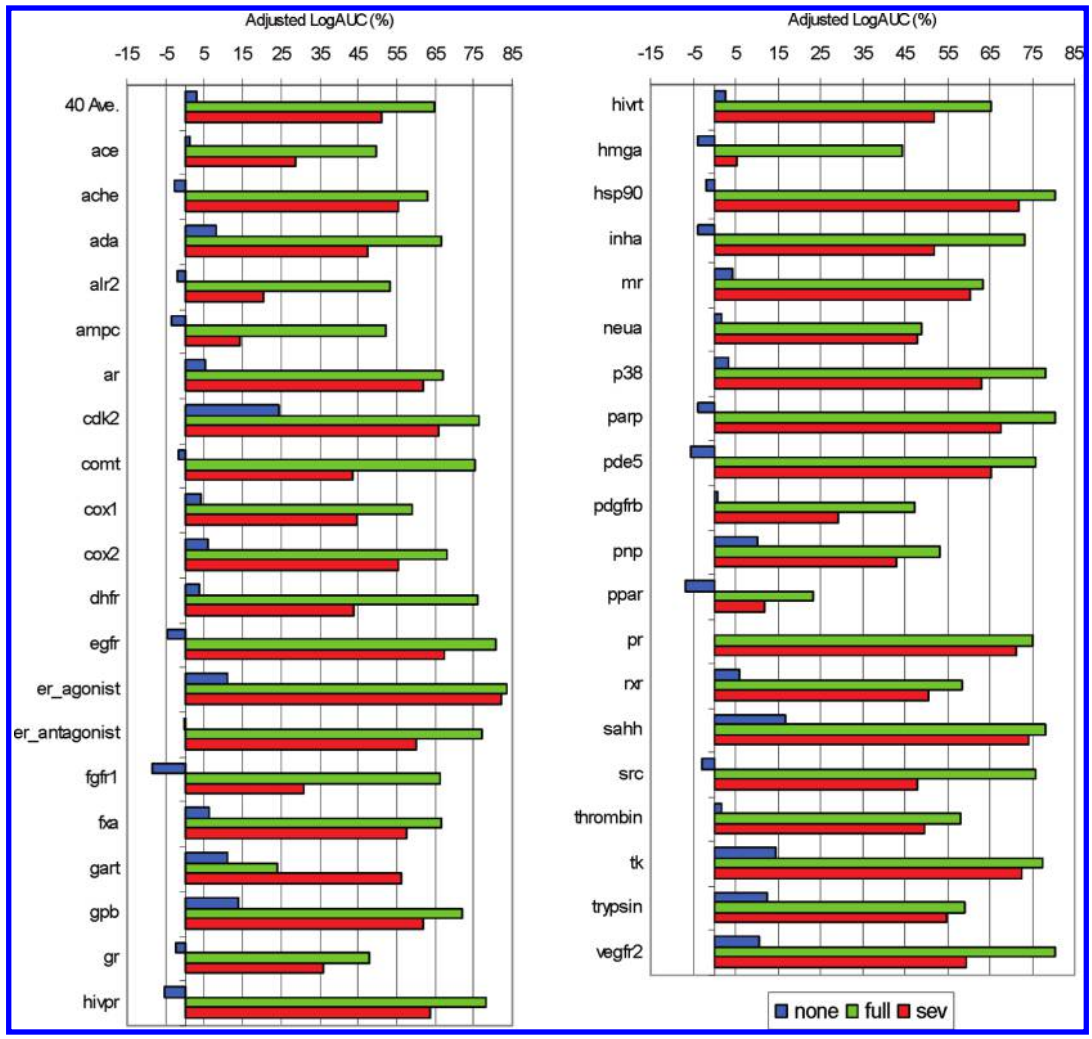

Figure 5. Enrichment versus random charged outliers: Enrichment of the DUD ligands compared to a background of highly charged molecules selected randomly from ZINC. The smaller the ligand desolvation penalty, "full" > "sev" > "none", the less we discriminate against the random charge outliers and thus get worse enrichments.

vation method (highlighted black). Averaged over all 40 targets, full desolvation performs worst with an average of 10.9% adjusted LogAUC, while no desolvation averages a much better 14.3% and the atomic solvent-excluded volume method performs best with 15.0% adjusted LogAUC. We also include a table with the comparison of adjusted LogAUC to traditional AUC, plus ROC-based enrichment factors at 1 and 10% (Table SI1).

Context-dependent ligand desolvation tracks no desolvation across the entire curve in those targets where no desolvation performs well (Figure 4), such as cyclin-dependent kinase 2 (panel A) and factor Xa (B), and tracks full desolvation across the entire curve in systems where it performs well, such as epidermal growth factor receptor (C), consistent with its ability to perform well in different receptor contexts. While context-dependent desolvation does not always match the best curve, it is usually able to at least interpolate between no and full ligand desolvation, as demonstrated in enoyl ACP reductase (D). In fact, context-dependent desolvation is particularly adept at improving early enrichment, as shown in cycloxygenase-2 (E) and dihydrofolate reductase (F). Semilog ROC plots of all 40 targets are also included (Figure SI3). We further analyzed how the ranks of the ligands and decoys change upon switching from no desolvation to "sev" desolvation (Figure SI5). For some targets, the difference between ranking with and without solvation correlates highly for ligands and decoys, while for other targets there is so much scatter as to defy generalization. However, where there is scatter in the effect on ligands this is mirrored in the effect on decoys.

Looking at the screening performance against property-matched decoys removes background database biases, allowing us to truly compare ligand desolvation methods. However, it fails to capture how they respond to extreme molecules, of the sort one might find when running large database screens. Ligand desolvation is particularly sensitive to differences in net charge, so we constructed a set of charge outliers having net charges from -4 to -2 and from $+2$ to $+4$, choosing 500 molecules of each type randomly from the ZINC database.[48] We docked the charge outliers to the 40 DUD targets, computed ROC curves against the charge outliers, and summarized them by their adjusted LogAUC (Figure 5). In sharp contrast to the performance against DUD decoys, where no desolvation performed competitively for many targets, its performance against charge outliers is very poor. Unfettered by a desolvation term, no desolvation often enriches the highly charged molecules more than the annotated ligands because the electrostatic interaction energy grows stronger as the ligands become more highly charged. By the same token, full desolvation performs well against

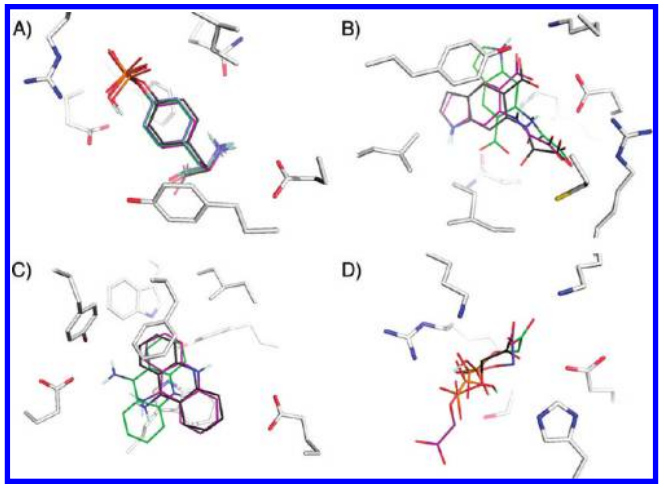

Figure 6. Pose fidelity: Over most targets all three desolvation methods give similar poses, as in A) 1PTV. Occasionally context-dependent "sev" desolvation improved poses, as in B) 1C83 and C) 1ACJ. For a few cases "sev" desolvation reduces pose fidelity, as in D) 1PDZ. The carbon atoms are colored as follows: receptor (white), crystallographic ligand (dark gray), no desolvation (blue), full desolvation (green), "sev" desolvation (purple).

charged outliers, much better than it does against DUD decoys, because it overweights the ligand desolvation penalty. Context-dependent desolvation significantly enriches true ligands over these highly charged outliers, often tracking the superior discriminatory power of full desolvation, without overweighting the ligand desolvation penalty.

To examine how well the ligand desolvation approximations affect docked ligand pose, we used the 114 crystal-structures in the DOCK5 benchmark. While we were able to begin docking with all 114 structures, one target (1HSL) consistently failed during sphere matching and was discarded. The best scoring pose was below the usual threshold of 2 Å rmsd in 61 (none), 61 (full), and 54 (sev) of the remaining targets. In contrast, the gap from full desolvation to "sev" in number of successful predictions at thresholds of 1.5 Å and 2.5 Å is 1 and 4 targets instead of 7. In most targets the poses are visually similar regardless of the ligand desolvation approximation used (Figure 6). In the protein tyrosine phosphatase 1B (PTP1B) structure 1PTV (panel A), all three desolvation approximations lead to docked poses that closely match the crystallographic pose. However, the phosphate moiety remains singly protonated with "sev" and "full" desolvation. In another PTB1B structure 1C83 (B) and the acetyl cholinesterase structure 1ACJ (C), "sev" desolvation torques polar groups to more exposed positions, causing the crystallographic agreement to improve. In a few cases such as 1PDZ (D), "sev" desolvation spins the molecule out of alignment with the crystal pose. Overall, pose fidelity with context-dependent ligand desolvation is similar to full or no desolvation.

Though efforts are made during DUD decoy generation to ensure they are topologically dissimilar from the ligands, there is no guarantee that the decoys do not, in fact, bind. To examine the screening performance against validated decoys, we employed two artificial binding cavities, a small hydrophobic pocket created in T4 lysozyme by the Leu99→Ala substitution (L99A) and a charged binding pocket created in cytochrome *c* peroxidase by the Trp191→Gly substitution (W191G). In the enclosed hydrophobic pocket of L99A, we would expect full desolvation to be a good approximation (Figure 7A), and indeed it gives the best LogAUC at 18.6%, with "sev" tracking it at 15.5%, and no desolvation at −1.5%. In the enclosed but charged pocket of W191G (Figure 7B), "sev" and full desolvation tie for the best LogAUC at 48.8%, with no desolvation lagging at 8.1% adjusted LogAUC. Compared to the high enrichments we typically see against larger databases,[30,49] overall enrichment against these experimental decoys is poor; but note that these are especially challenging decoys since most of them were experimentally tested due to strong scores in earlier docking screens.

DISCUSSION

At first blush it may seem surprising that docking programs ever discover new ligands for proteins, so many are the approximations made by their scoring functions. That they do so[1−4] reflects, at least partly, a cancellation of errors among approximations. Whenever a term is improved by making it physically more correct it is easy to imagine that the new model may perform *worse* than the old by upsetting this prior cancellation of errors. Thus it is comforting that in moving to a more realistic model of ligand desolvation the results against the DUD benchmark do not deteriorate. More convincingly, context-dependent desolvation confers improved stability to target-by-target and ligand-by-ligand variability. It is almost an aphorism in the field that different docking methods are better suited to different targets, and this behavior may be seen when comparing no and full desolvation (Figure 3), where no desolvation often better enriches targets with charged ligands and more open binding sites such as DHFR, GART, and PARP, while full desolvation often better enriches targets with neutral ligands and more enclosed binding sites such as EGFR, HIVPR, and P38. In contrast, context-dependent desolvation, though still varying in performance from target-to-target, always gives either the best or second-best enrichments and does not swing wildly from best to worst. Even though no desolvation outperforms full desolvation against matched decoys, against charged outliers no desolvation performs only marginally better than random. Conversely, full desolvation has extraordinarily high enrichments against charged outliers, while context-dependent desolvation approaches full desolvation in its ability to discriminate against them (Figure 5). This stability to variations in ligand preferences and binding site contexts reflects the better physical model in the context-dependent ligand desolvation penalties.

Even with this improved stability, it would be comforting if "sev" desolvation gave substantially better enrichment than ignoring desolvation entirely; currently its overall improvement is marginally significant but certainly not substantial. Inspection of particular cases suggests that desolvation leads to worse enrichments largely in targets that bind more positive ligands. This made us wonder about the charge distribution in these ligands, which had significant charge localized on the hydrogen atoms. We therefore explored using more reasonable radii for hydrogen and heavy atoms in our desolvation calculations, moving from treating all atoms as having a radius of 1.4 Å, to modeling hydrogens as size 1.0 and all heavy atoms as 1.8 Å. Encouragingly, enrichment improved substantially for the full DUD set, especially against targets recognizing positively charged

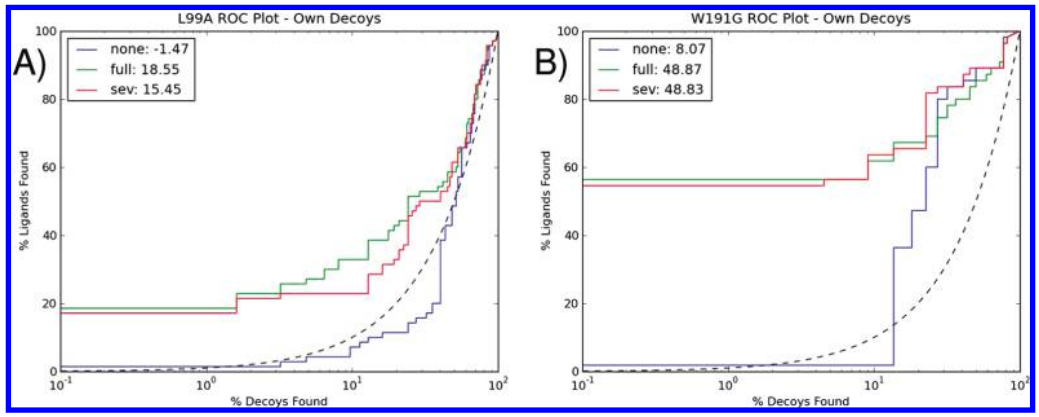

Figure 7. ROC for experimental decoys in model binding sites. ROC curves of known ligands versus experimental decoys for A) the small hydrophopic cavity of L99A T4 lysozyme and B) the negatively charged cavity in W191G cytochrome *c* peroxidase.

ligands (Figure SI2). The largest improvements occurred, for instance, in DHFR, INHA, and NEUA, whose ligands all have key positively charged moieties. In overall enrichment, the adjusted LogAUC gap between no desolvation, at 14.3%, and "sev" desolvation, at 15.0%, widens from 0.7 to 1.5% in the process of giving the hydrogen atoms their own desolvation grid, resulting in an overall average enrichment of 15.8 LogAUC. While this still leaves us with an overall performance improvement that remains small, it is comforting that as we move to better physical models the performance consistently improves without any fitting whatsoever. This need not be the case—we rely so heavily on cancellation of errors in docking, that one might easily imagine models that are physically more correct but that actually reduce performance by our standard enrichment metrics. Further improvement of our physical models, for instance by including receptor terms in desolvation and internal energy terms in the ligand, thus seems not only scholarly but also pragmatic.

Three results in this study may interest the specialist. First, we introduce a virtual screening metric, LogAUC, which tackles the "early enrichment" problem by computing the percentage of the ideal area that lies under the semilog enrichment curve. LogAUC shares many desirable characteristics with ROC AUC: as it is easily interpreted, robust, and independent of similar extensive variables.[57] Second, we introduce the atomic solvent-excluded volume (sev) method to deal with the often neglected low-dielectric region that forms in the gaps between ligand and protein (Figure 1B). This still allows precalculation of fractional desolvation while improving the desolvation penalty magnitude at no computational cost during docking, with only a moderate (~30 min) precalculation cost per protein. Third, we provide a revised version of the DUD database which corrects some problems with the initial release, such as the inclusion of "decoys" that were in fact ligands, and adds a sixth physical property term against which to balance ligands and decoys — net charge. We also provide the charged outlier sets, which can be challenging for any method that does not consider desolvation, explicitly or implicitly, and are meant to mimic molecules one will encounter in a large unbiased library screen, such as ZINC. The updated version of DUD is freely available at http://dud.docking.org.

Certain caveats merit airing. We multiply the fractional desolvation by the entire transfer free energy, and this is formally correct only for the self-energy terms of polar desolvation. Simple multiplication is just an approximation for both the Coulombic term and nonpolar desolvation. We additionally ignore the fractional desolvation due to the other ligand atoms. Exact treatment of these desolvation terms is possible, but as it would add substantially to the on-the-fly cost, it may be better left to subsequent rescoring. Along the same lines, we use atomic desolvation energies derived from a single ligand conformation to represent all conformations docked. Although this is reasonable for many molecules, there will be others where the solvation energy changes substantially from conformation-to-conformation. Whereas the calculation and storage costs of atomic solvation energies of every conformation of every ligand seemed daunting enough to ignore for this study, this term can be precalculated and so may merit further investigation. Finally, it should be clear that we only consider ligand desolvation here, and not its logical complement protein desolvation. Doing so will undoubtedly affect docking performance substantially, but this is a term that may be added without affecting the calculated ligand solvation term.

Notwithstanding these caveats, an important result of this work is the incorporation of a physically reasonable model of ligand desolvation with negligible cost to docking run time. Thus, the total time to screen 134,000 flexible molecules of the new DUD set against all 40 targets was 2446 CPU hours without correcting for desolvation, while the total time to prosecute this screen with "sev" desolvation was 3862 CPU hours. The method thus incorporates critical aspects of higher-level solvation theory, while maintaining the speed advantages of docking. This is achieved by precalculating most terms before docking, and as such this method should be applicable to any physics-based docking approach.

ACKNOWLEDGMENT

Supported by NIH grant GM59957. We thank Jens Carlsson, Ryan G. Coleman, and Magdalena Korczynska for reading this manuscript and especially thank Dr. John J. Irwin for many insightful conversations and help with DUD and ZINC.

Supporting Information Available: In Table SI1, we compare adjusted LogAUC to traditional AUC. In Figure SI2, we examine the effect of using separate desolvation grids for hydrogen and heavy atoms. In Figure SI3, we give all

40 ROC plots against property-matched DUD decoys. In Figure SI4, we show the effect of changing the grid spacing and integration cutoff. In Figure SI5, we compare the ligand and decoys ranks as "sev" desolvation is turned off or on. In Figure SI6, we estimate the error in the self-energy term due to neglecting the ligand component of fractional desolvation. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813–1818.
(2) Alvarez, J. C. High-throughput docking as a source of novel drug leads. *Curr. Opin. Chem. Biol.* **2004**, *8*, 365–370.
(3) Shoichet, B. K.; McGovern, S. L.; Wei, B.; Irwin, J. J. Lead discovery using molecular docking. *Curr. Opin. Chem. Biol.* **2002**, *6*, 439–446.
(4) Cavasotto, C. N.; Orry, A. J. Ligand docking and structure-based virtual screening in drug discovery. *Curr. Top. Med. Chem.* **2007**, *7*, 1006–1014.
(5) 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*, 425–445.
(6) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
(7) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
(8) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
(9) Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46*, 499–511.
(10) 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.
(11) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.
(12) Velec, H. F.; Gohlke, H.; Klebe, G. DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.* **2005**, *48*, 6296–6303.
(13) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated Docking with Grid-Based Energy Evaluation. *J. Comput. Chem.* **1992**, *13*, 505–524.
(14) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
(15) Abagyan, R.; Totrov, M.; Kuznetsov, D. Icm - a New Method for Protein Modeling and Design - Applications to Docking and Structure Prediction from the Distorted Native Conformation. *J. Comput. Chem.* **1994**, *15*, 488–506.
(16) McMartin, C.; Bohacek, R. S. QXP: powerful, rapid computer algorithms for structure-based drug design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 333–344.
(17) Dominguez, C.; Boelens, R.; Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737.
(18) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol.* **2008**, *153*, S7–S26.
(19) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.
(20) Shoichet, B. K. No free energy lunch. *Nat. Biotechnol.* **2007**, *25*, 1109–1110.
(21) Chang, C. E.; Gilson, M. K. Free energy, entropy, and induced fit in host-guest recognition: calculations with the second-generation mining minima algorithm. *J. Am. Chem. Soc.* **2004**, *126*, 13156–13164.
(22) Chen, W.; Chang, C. E.; Gilson, M. K. Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design. *Biophys. J.* **2004**, *87*, 3035–3049.
(23) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978–1988.
(24) Guimaraes, C. R.; Cardozo, M. MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J. Chem. Inf. Model.* **2008**, *48*, 958–970.
(25) Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P. Virtual screening against highly charged active sites: identifying substrates of alpha-beta barrel enzymes. *Biochemistry* **2005**, *44*, 2059–2071.
(26) Zou, X. Q.; Sun, Y. X.; Kuntz, I. D. Inclusion of solvation in ligand binding free energy calculations using the generalized-born model. *J. Am. Chem. Soc.* **1999**, *121*, 8033–8043.
(27) Liu, H. Y.; Kuntz, I. D.; Zou, X. Q. Pairwise GB/SA scoring function for structure-based drug design. *J. Phys. Chem. B* **2004**, *108*, 5453–5462.
(28) Wang, J.; Kang, X.; Kuntz, I. D.; Kollman, P. A. Hierarchical database screenings for HIV-1 reverse transcriptase using a pharmacophore model, rigid docking, solvation docking, and MM-PB/SA. *J. Med. Chem.* **2005**, *48*, 2432–2444.
(29) Grant, J. A.; Pickup, B. T.; Nicholls, A. A smooth permittivity function for Poisson-Boltzmann solvation methods. *J. Comput. Chem.* **2001**, *22*, 608–640.
(30) Graves, A. P.; Shivakumar, D. M.; Boyce, S. E.; Jacobson, M. P.; Case, D. A.; Shoichet, B. K. Rescoring docking hit lists for model cavity sites: predictions and experimental testing. *J. Mol. Biol.* **2008**, *377*, 914–934.
(31) Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. Ligand solvation in molecular docking. *Proteins* **1999**, *34*, 4–16.
(32) Wesson, L.; Eisenberg, D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.* **1992**, *1*, 227–235.
(33) van Dijk, A. D.; Bonvin, A. M. Solvated docking: introducing water into the modelling of biomolecular complexes. *Bioinformatics* **2006**, *22*, 2340–2347.
(34) Fernandez-Recio, J.; Totrov, M.; Abagyan, R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins* **2003**, *52*, 113–117.
(35) Stouten, P. F. W.; Frommel, C.; Nakamura, H.; Sander, C. An Effective Solvation Term Based on Atomic Occupancies for Use in Protein Simulations. *Mol. Simul.* **1993**, *10*, 97.
(36) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.
(37) Cerutti, D. S.; Jain, T.; McCammon, J. A. CIRSE: a solvation energy estimator compatible with flexible protein docking and design applications. *Protein Sci.* **2006**, *15*, 1579–1596.
(38) Schnecke, V.; Kuhn, L. A. Virtual screening with solvation and ligand-induced complementarity. *Perspect. Drug Discovery* **2000**, *20*, 171–190.
(39) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504–6515.
(40) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.
(41) Morreale, A.; Gil-Redondo, R.; Ortiz, A. R. A new implicit solvent model for protein-ligand docking. *Proteins* **2007**, *67*, 606–616.
(42) Majeux, N.; Scarsi, M.; Caflisch, A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins* **2001**, *42*, 256–268.
(43) Wei, B. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W.; Shoichet, B. K. A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol.* **2002**, *322*, 339–355.
(44) Li, J. B.; Zhu, T. H.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. Extension of the platform of applicability of the SM5.42R universal solvation model. *Theor. Chem. Acc.* **1999**, *103*, 9–63.
(45) Bashford, D.; Case, D. A. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.
(46) Scarsi, M.; Apostolakis, J.; Caflisch, A. Continuum electrostatic energies of macromolecules in aqueous solutions. *J. Phys. Chem. A* **1997**, *101*, 8098–8106.
(47) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.
(48) Irwin, J. J.; Shoichet, B. K. ZINC--a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
(49) Brenk, R.; Vetter, S. W.; Boyce, S. E.; Goodin, D. B.; Shoichet, B. K. Probing molecular docking in a charged model binding site. *J. Mol. Biol.* **2006**, *357*, 1449–1470.

(50) Gilson, M. K.; Sharp, K.; Honig, B. Calculating electrostatic interactions in biomolecules: method and error assessment. *J. Comput. Chem.* **1987**, *9*, 327–335.
(51) Lorber, D. M.; Shoichet, B. K. Hierarchical docking of databases of multiple ligand conformations. *Curr. Top. Med. Chem.* **2005**, *5*, 739–749.
(52) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; 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.
(53) *OEChem TK 1.7.2.4*; Openeye Scientific Software: Santa Fe, NM, 2009. www.eyesopen.com (accessed month day, year).
(54) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic 3-Dimensional Model Builders Using 639 X-Ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000–1008.
(55) *Ligprep 23110*; Schrödinger: New York, NY, 2009. www.schrodinger.com (accessed month day, year).
(56) *Omega 2.3.2*; Openeye Scientific Software: Santa Fe, NM, 2008. www.eyesopen.com (accessed month day, year).
(57) Nicholls, A. What do we know and when do we know it. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 239–255.
(58) Clark, R. D.; Webster-Clark, D. J. Managing bias in ROC curves. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 141–146.
(59) Truchon, J. F.; Bayly, C. I. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. *J. Chem. Inf. Model.* **2007**, *47*, 488–508.
(60) Katritch, V.; Rueda, M.; Lam, P. C.; Yeager, M.; Abagyan, R. GPCR 3D homology models for ligand screening: lessons learned from blind predictions of adenosine A2a receptor complex. *Proteins* **2010**, *78*, 197–211.
(61) Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. Development and validation of a modular, extensible docking program: DOCK 5. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 601–619.
(62) Irwin, J. J.; Shoichet, B. K.; Mysinger, M. M.; Huang, N.; Colizzi, F.; Wassam, P.; Cao, Y. Automated docking screens: a feasibility study. *J. Med. Chem.* **2009**, *52*, 5712–5720.
(63) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 681–691.
(64) Eriksson, A. E.; Baase, W. A.; Wozniak, J. A.; Matthews, B. W. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* **1992**, *355*, 371–373.
(65) Morton, A.; Baase, W. A.; Matthews, B. W. Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry* **1995**, *34*, 8564–8575.
(66) Su, A. I.; Lorber, D. M.; Weston, G. S.; Baase, W. A.; Matthews, B. W.; Shoichet, B. K. Docking molecules by families to increase the diversity of hits in database screens: computational strategy and experimental evaluation. *Proteins* **2001**, *42*, 279–293.
(67) Graves, A. P.; Brenk, R.; Shoichet, B. K. Decoys for docking. *J. Med. Chem.* **2005**, *48*, 3714–3728.
(68) Fitzgerald, M. M.; Musah, R. A.; McRee, D. E.; Goodin, D. B. A ligand-gated, hinged loop rearrangement opens a channel to a buried artificial protein cavity. *Nat. Struct. Biol.* **1996**, *3*, 626–631.
(69) Musah, R. A.; Jensen, G. M.; Bunte, S. W.; Rosenfeld, R. J.; Goodin, D. B. Artificial protein cavities as specific ligand-binding templates: characterization of an engineered heterocyclic cation-binding site that preserves the evolved specificity of the parent protein. *J. Mol. Biol.* **2002**, *315*, 845–857.
(70) Gilson, M. K.; Honig, B. The inclusion of electrostatic hydration energies in molecular mechanics calculations. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 5–20.

CI100214A