Assessment of a Novel Scoring Method Based on Solvent Accessible Surface Area Descriptors

Sara Núñez,*,§ Jennifer Venhorst,§ and Chris G. Kruse

Research Laboratories, Solvay Pharmaceuticals, C. J. van Houtenlaan 36, 1381 CP Weesp, The Netherlands

Received December 4, 2009

A novel scoring algorithm based on unique solvent accessible surface area (SASA) descriptors was comparatively evaluated for its database enrichment potential against the virtual screening (VS) methods GOLD and Glide. Several protein test cases, including adenosine deaminase and estrogen receptor alpha, were used for the evaluation. The structure-based VS method GOLD was used to generate the protein−ligand docking poses. These docking poses were then postprocessed with a protein−ligand interaction fingerprint metric. Next, the SASA descriptors were computed for each ligand and its respective protein in their bound/unbound states; a Bayesian model was learned with SASA descriptors and subsequently used to score the remaining ligands in the screening databases. Early database enrichments using SASA descriptors were found comparable or superior to those of GOLD and Glide. Moreover, SASA descriptors display an outstanding robustness to produce satisfactory early enrichments for a large variety of target classes. Based on these encouraging results, these novel topological descriptors constitute a valuable in silico tool in hit finding practices.

1. INTRODUCTION

Hit finding (HF) and lead optimization (LO) efforts rely largely on the in silico prediction of the strength of interactions between a ligand and its receptor. Much of the drive for using computational methods arises from increased pressure to reduce the costs involved in high-throughput screening (HTS) and in ensuing synthetic efforts involved in the HF and LO discovery stages, ultimately resulting in more candidates in the developmental pipeline in shorter times.[1–6] To this end, extensive research in the computational community has focused on the development and assessment of algorithms that accurately predict binding free energies.[7–9] Here we will summarize the status quo of the structure-based computational methods used currently to predict binding free energies. (i) Free energy methods: These are most accurate for predicting binding free energies.[10] The most utilized method is free energy perturbation (FEP), which makes use of force fields coupled to molecular dynamics (MD) or Monte Carlo (MC) simulations. These are computationally intensive methods but can accurately predict relative binding free energies of closely related molecules. (ii) Approximate free energy methods: These approaches were developed in an attempt to reduce computation times without a major loss of accuracy in the prediction.[11,12] The two main approaches are the linear interaction energy (LIE) method and the molecular mechanics-Poisson−Boltzmann surface area (MM-PBSA) method. The LIE method computes binding free energies using ensemble average energies obtained from MD/MC simulations. The underlying approximation is that the energy expression is simplified into parametrized equations that contain a van der Waals (vdW) term, an electrostatic term that accounts for the loss of ligand−solvent interactions, and a solvent accessible surface area (SASA) term for nonpolar solvation. The main drawback of the LIE method is that it requires calibration using a set of experimentally determined ligand−protein binding free energies and prior knowledge of the binding mode of the ligands. Moreover, this method can only be applied to compute relative free energies of binding of structurally related ligands. In contrast, the MM-PBSA method computes binding free energies by conducting separate simulations (ligand, protein, and complex). The binding free energy is then computed by subtracting the energies of the protein and ligand from that of the complex. The MM-PBSA method approximates the binding free energy through an energetic expression that contains a MM term, an electrostatic term, a nonpolar solvation term, and an entropy term. MM-PBSA has been shown to improve early enrichment rates; however, these computations have to be carefully set up to prevent failure.[13] An alternative to decreasing the computational demand of MM-PBSA calculations is to apply the method as a scoring function for pregenerated ligand−protein complexes through docking studies. (iii) Docking methods: These are widely used and can predict the binding mode and binding free energy of ligands in a high-throughput manner.[14] Docking methods combine a search algorithm to identify ligand poses in the binding site of a protein and a parametrized scoring function to quantify the energetic contributions of the binding event. Alas, the errors associated with these docking methods can be large, giving only a comparative ranking of the database molecules. In addition, there is a high degree of variability in virtual screening results between both the protein targets and the docking protocols; some targets have excellent early recovery results for all docking protocols, while others have excellent early enrichment for only specific docking tools. Thus far, only binding free energy predictions arising from the use of single scoring function methods have been

* Corresponding author. E-mail: sara.nunez@solvay.com.
§ These authors contributed equally to this paper.

10.1021/ci9004628
Published on Web 03/31/2010

presented. However, the combination of several scoring functions may offer benefits over the use of a single algorithm. Consensus scoring may thus be a robust approach for improving binding free energy predictions.[15]

All the methods presented above recognize that changes in the solvent accessible surface area during the ligand binding event are of great importance and give direct insights into the role of the solvent and the hydrophobicity in the binding event. In the current investigation, SASA descriptors were used as the sole means of scoring database ligands; the underlying postulate in using these topological descriptors to score poses is that pose generators, like GOLD,[16] are capable of correctly sampling the binding site cavity and generating reliable poses, but they are not accurate enough at scoring them for a wide range of protein targets. Our approach makes use of these correct docking poses and ranks them using novel SASA descriptors of both the target and the ligand in their bound/unbound states. In addition, because the SASA-based scoring function cannot intrinsically identify crucial protein−ligand interactions, we have incorporated knowledge of experimentally derived key protein−ligand interactions through the application of molecular interaction fingerprints[6] (IFPs) to improve database enrichment.

Here we present a novel scoring algorithm based solely on SASA descriptors and a comparative evaluation against the well-accepted energy-based scoring functions of GOLD and Glide.[17] Our results show that the topological scoring algorithm based on SASA descriptors is equal to or better than GOLD and Glide in early recovery success. In conclusion, the outstanding performance of SASA descriptors, evident for a wide variety of targets and independent of parameter-tweaking, makes this robust scoring method an attractive tool for hit finding efforts.

2. METHODOLOGY

A. Solvent Accessible Surface Area Descriptors. Solvation effects play an important role in the binding of a ligand to a protein. These effects include the desolvation of both the ligand and the protein cavity and the rearrangement of solvent molecules. These effects can be calculated by both implicit and explicit solvent models. Here, the nonpolar contributions to the free energy of solvation of every atom in a molecule are assumed to be proportional to the calculated SASA. The SASA of a molecule is defined as the area of the surface over which a water molecule can be placed while making vdW contact with each atom, and not penetrating any other atom. By comparing ligand and protein SASA terms in their unbound/bound states, estimates of both the desolvation energy required for binding and the resulting buriedness of polar and nonpolar surface areas are obtained. An analysis of a database of crystallized protein−ligand complexes demonstrated that the latter is also a key determinant for efficient ligand binding.[18]

In this study, the SASA of the ligand and protein binding site in their bound/unbound states was computed for each complex using a python script developed by Schrödinger, Inc. based on the statistical algorithm described by Wodak et al.[19] In short, a hard sphere with a radius of 1.4 Å was used to represent the water molecule, and the binding cavity of each protein was defined by a 5 Å radius from the bound ligand. Using the internal chemical feature typing program Chemfeatures in Maestro,[17] the individual contributions to the total SASA (Å^2) were generated for the individual atom types, leading to the following receptor/ligand SASA descriptors for the ligand and protein in their unbound/bound states: hydrophobic (H), aromatic (R), hydrogen-bond acceptor (HBA), hydrogen-bond donor (HBD), positive (P), and negative (N). Chemfeatures definitions are not mutually exclusive; in order to maintain correct summation of the individual SASA terms to the total SASA, the following scheme was used: (i) atoms that were both hydrophobic and aromatic were typed as aromatic; atoms being positive and donor were typed as positive; (ii) atoms that were both negatively charged acceptors were typed as negative; (iii) atoms that were either aromatic donors or acceptors were classified as donors and acceptors, respectively; and (iv) atoms with donor and acceptor propensities were classified as donor in the presence of a hydrogen atom, and otherwise classified as an acceptor. The required computing time for each complex (including unbound SASA contributions) was 1 min per complex on 1 AMD Opteron Core of a 2.2 GHz dual-core processor.

B. Bayesian Model Learning and Validation. Bayesian modeling is a probabilistic method that identifies the most likely category for a given observation by modeling the underlying distributions of the features across categories.[20] Two categories were used in this study: 'active' and 'inactive' (baseline) compounds. For each test case, naïve Bayes classifiers were learned with a training set of active/decoy/inactive compounds. Once learned, these models were used to rank a test set of molecules generating a probability score for the likelihood that a given ligand belonged to the 'active' category.

C. Model Building with SASA Descriptors: sc-PDB Test Case. In order to assess whether SASA descriptors are capable of discriminating 'active' ligands from 'inactive' ones in large databases, a number of protein test cases were used in combination with Bayesian modeling as described next. First, a Bayesian model was learned with SASA descriptors of high and low affinity protein−ligand complexes extracted from the sc-PDB database.[21] The sc-PDB is an annotated collection of druggable binding sites from the Protein Data Bank[22] and contains abundant structural, physicochemical, and pharmacodynamic information for complexes deposited in the Protein Data Bank. High-/low-affinity complexes (i.e., actives/inactives, respectively) were collected from the sc-PDB in the following manner: Active compounds were those having $EC_{50}/IC_{50}/pK_i \geq 8.0$, and inactive compounds were those with $EC_{50}/IC_{50}/pK_i \leq 6.0$. Additionally, only noncovalent complexes were accepted, and a crystal resolution of $\leq$2.8 Å was imposed for all complexes to ensure optimal accuracy. After the preliminary filtering, the total number of high- and low-affinity compounds in the data set was 260 and 271 compounds, respectively. The inherent diversity of the ligands in the data set was evaluated using physicochemical descriptors and compared to the Prous Science Integrity database (Supporting Information).[23] Subsequently, SASA descriptors were generated for the ligands, proteins, and corresponding complexes using a python script developed by Schrödinger Inc. The required computing time was 2 h on four 2.2 GHz dual-core Opteron AMD 64bit processors. The Bayesian model was learned with Pipeline Pilot[24] using a diverse selection

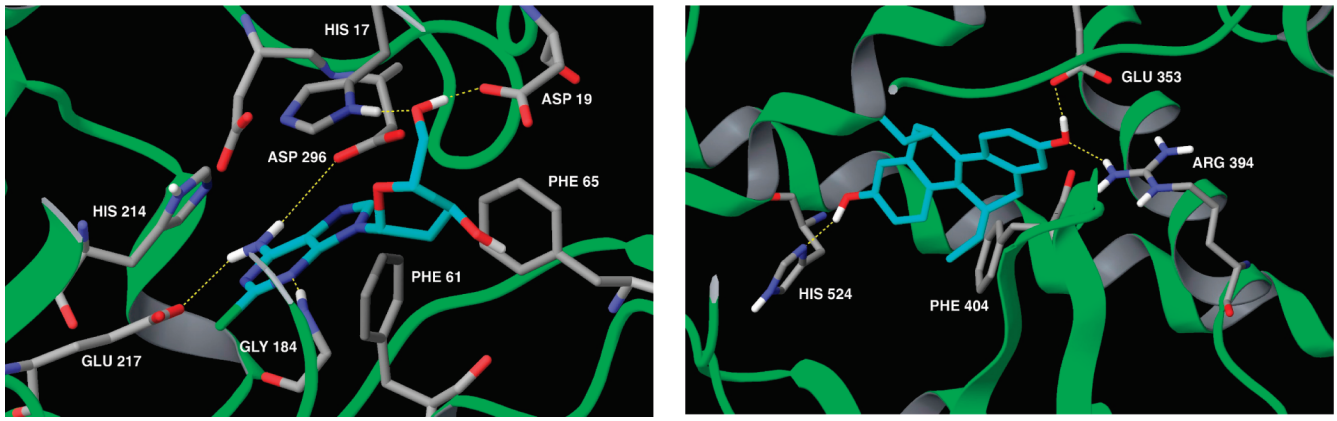

Figure 1. Binding sites of adenosine deaminase, ADA (left) and estrogen receptor alpha, ERα (right). The reference ligands are colored in blue, and protein residues that are actively involved in ligand binding are annotated accordingly.

of 230 actives and an equal number of inactive compounds. All individual SASA contributions for ligand/protein in their bound/unbound states were used for model building. Thereafter, this Bayesian model was used to rank the remaining compounds in the sc-PDB data set using the normalized probability.

D. Model Building with SASA Descriptors: ADA and ERα Test Cases. Verdonk et al recently highlighted in their works that significant differences in physicochemical properties between ligands and decoys can lead docking enrichments to appear artificially good.[25] Accurate validation of any novel molecular descriptor and/or VS method thus demands an appropriate choice of data sets to correctly evaluate the predictive power of a given algorithm.[26] With this objective in mind, two test cases were chosen from the Directory of Useful Decoys (DUD),[27] a publicly available database of physically matched decoys and ligands that provides a more stringent decoy criterion with which to evaluate the performance of SASA descriptors in their database enrichment capability. We also investigated two additional data sets in which the aforementioned decoys were replaced with an identical number of inactive compounds collected from our proprietary compound stock. This so-called inactive data set had similar Lipinski descriptors[28] to the true ligands, but they differed substantially in two-dimensional (2D) similarity, whereas the decoys were intimately related to the true ligands in their similarity metric.[29] The reason for this approach was to assess the performance of SASA descriptors for two typical drug discovery scenarios, that is, the hit finding and the lead optimization drug discovery stages, respectively.

The first test case was adenosine deaminase (ADA), an enzyme involved in purine metabolism which irreversibly deaminates adenosine. Many crystal structures of ADA in complex with assorted inhibitors exemplify its rather polar active site (Figure 1). The second test case was estrogen receptor alpha (ERα), a nuclear hormone receptor which activates transcription upon binding of the hormone 17β-estradiol or other ERα agonists. Several crystal structures of ERα in complex with different ligands reveal this nuclear receptor binding site as being considerably hydrophobic (Figure 1). The choice of very disparate targets for the underlying analysis was intended to identify possible biases which may exist in the virtual screening algorithms with regard to the nature of the ligand database and/or protein−ligand interactions.

Table 1. Eight-Bit Molecular Interaction Fingerprint used in This Study for the Postprocessing of the Docking Poses

bit vector position	protein atom type	ligand atom type	interaction
1	hydrophobe	hydrophobe	hydrophobic
2	aromatic	aromatic	face-to-face
3	aromatic	aromatic	edge-to-face
4	donor	acceptor	hydrogen bond
5	acceptor	donor	hydrogen bond
6	cation	anion	ionic
7	anion	cation	ionic
8	any	any	protein−ligand clash

The crystal structures of ADA and ERα were obtained from the ZINC database.[27] Hydrogen atoms were added, and the positions of hydrogen atoms involved in polar interactions were subsequently minimized to optimize their hydrogen-bond interactions using the protein preparation module in Maestro.[17] Protein residues were inspected, and the tautomeric states of histidines, hydroxyl group orientations, and protonation states of titratable residues were adjusted accordingly. Relevant decoy libraries for the two targets under investigation in this study were used and further enriched with active ligands obtained from both the ZINC and Prous Science Integrity databases. The screening database of ADA consisted of 821 decoys and 46 active ligands, and that of ERα consisted of 2355 decoys and 67 active ligands. The ADA and ERα data sets were then docked into their respective proteins using GOLD and Glide, as described next. With regards to GOLD, the binding pocket of each target was defined from the crystallographic coordinates of the ligand, taking into account any residue within 15 Å of any atom of the ligand. Docking runs were performed keeping the protein rigid and using the default settings, which provided baseline performance by assuming no prior experience on the part of the user with the target or software, thereby eliminating the possibility that user subjectivity could bias the results. No imposed geometrical constraints were used, and the GoldScore function was used to obtain a maximum of 20 poses of each ligand. With regard to the virtual screening settings of Glide SP, rigid protein docking was carried out using the 'single precision' (SP) mode.[17] The remaining options were kept as default, and no constraints were used. The Glide SP score was used to obtain the best pose for each ligand in the database. In parallel, in order to assess the competence of SASA descriptors as a

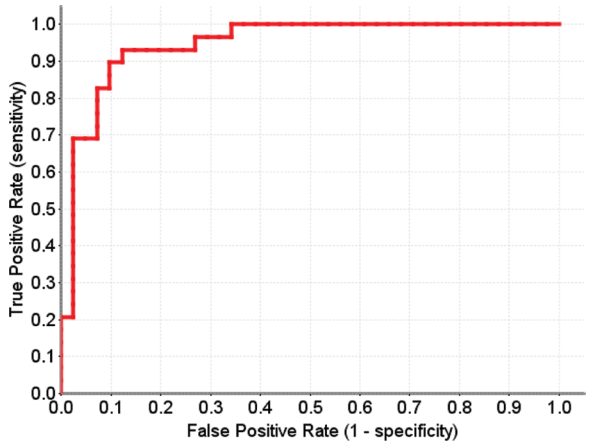

Figure 2. ROC curve for the sc-PDB test case using a naïve Bayes model learned with SASA descriptors.

scoring method, GOLD was used to generate 20 docking poses, as previously described, and these poses were subsequently assessed with a proprietary postprocessing method based on molecular protein−ligand interaction fingerprints (IFPs) to discard the number of nonrelevant GOLD docking poses. IFPs are simple bit strings that convert 3D information on protein−ligand interactions into 1D bit vector representations that are useful as a postprocessing tool in docking studies.[6]

Table 1 illustrates the topology of the IFP used in this study. Each 'bit vector position' typifies a classic protein−ligand interaction, and the interaction of each binding site residue with the ligand is described through a characteristic eight-bit fingerprint. Concatenation of the various eight-bit strings gives rise to the total IFP, giving a detailed picture of the intermolecular interactions between the protein binding site residues and the docked ligand. Docking poses that displayed more than two clashes with the protein were discarded; our experience with docking practices for different targets has proven this number to be an adequate threshold.

In addition, several protein−ligand interaction constraints were used based on acquired knowledge on the binding mode of reference ligands. For ADA, a hydrogen-bond interaction between GLU217 and/or ASP19 in each docking pose was imposed.[30] With regards to ERα, a hydrogen-bond interaction between GLU353 and ARG394 in each docking pose was imposed.[31] These filters discard all docked compounds that lack a hydrogen-bond interaction with these two essential binding residues. Subsequently, SASA descriptors were computed with Maestro for the remaining database docking poses of both ADA and ERα, that is, for each ligand and protein in their bound and unbound states. The required computing time was 48 h on four 2.2 GHz dual Opteron core AMD 64bit processors. The respective Bayesian models were learned using an average of 5% actives and 5% inactives/decoys. The Bayesian models were then used to rank the remaining compounds in the ADA/ERα data sets using the normalized probability.

3. RESULTS AND DISCUSSION

A. Bayesian Model Learned with SASA Descriptors: sc-PDB Test Case. In order to assess whether SASA descriptors are capable of discriminating 'active' ligands from 'inactives' in large databases, a number of protein test cases were used in this study. In the first, a Bayesian model was learned with 260 and 271 high- and low-affinity ligands, respectively, extracted from the sc-PDB database. The inherent diversity of the ligands in the sc-PDB data set was evaluated using physicochemical descriptors and compared to Prous Science Integrity. The comparison shows a reasonable overlap between the sc-PDB data set and Prous Science Integrity, so it is reasonable to state that the chemical space sampled by the sc-PDB data set is rather high (Supporting Information). Subsequently, SASA descriptors were generated for the ligands and the relevant proteins in their bound/unbound states using a python script developed by Schrödinger, Inc., and a Bayesian model was learned with Pipeline Pilot using a diverse selection of actives and the same number of inactives. This Bayesian model was used to rank the remaining compounds in the sc-PDB data set using the normalized probability. Figure 2 shows the receiver operating characteristic (ROC) curve for the sc-PDB test set. The area under the curve (AUC) in ROC plots is a graphical representation of the true positive rate versus the false positive rate, and it is suitable to evaluate the virtual screening performance where one needs to discriminate between active/inactive compounds.[32]

The dominant feature statistics of the Bayesian model of the sc-PDB test case were the *ligand hydrophobic unbound* and *receptor hydrophobic unbound* SASA terms (Figure 3). Specifically, active compounds in the sc-PDB data set differentiate from inactives in that they show a higher hydrophobic surface in their unbound states. Similarly, their

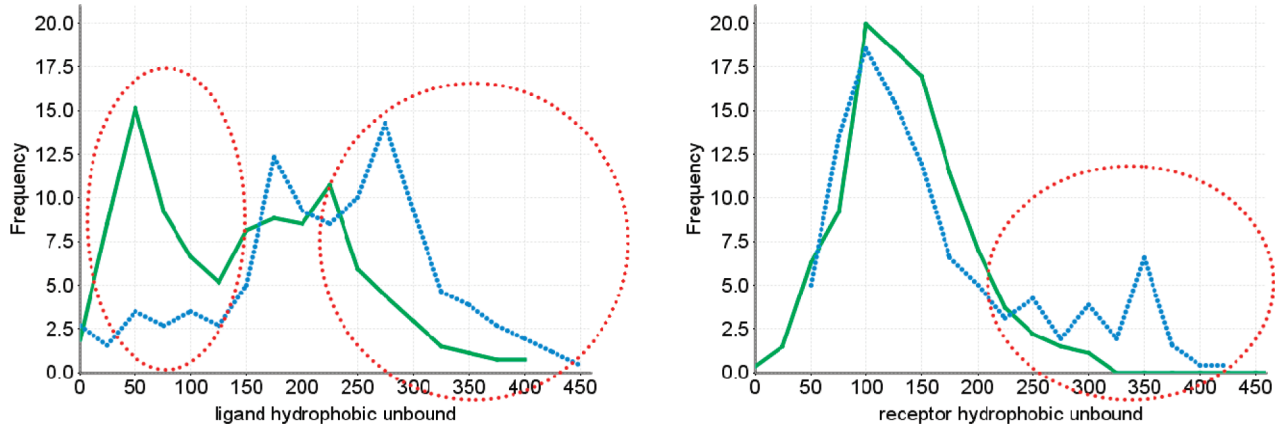

Figure 3. Dominant features of the Bayes model learned with SASA descriptors using the sc-PDB data set. Actives are shown in blue, and inactives are in green. The ligand and receptor hydrophobic SASA term is shown to correlate with activity.

Table 2. AUC and Database Enrichments at 1, 5 and 10% of Database Screened using GOLD, Glide, and Naïve Bayes Classifiers Learned with SASA Descriptors for the ADA-Decoys, ADA-Inactives, ERα-Decoys, and ERα-Inactives Data Sets

test case/method	SASA ROC area	SASA TP% at 1% FP	SASA TP% at 5% FP	SASA TP% at 10% FP	GOLD ROC area	GOLD TP% at 1% FP	GOLD TP% at 5% FP	GOLD TP% at 10% FP	Glide ROC area	Glide TP% at 1% FP	Glide TP% at 5% FP	Glide TP% at 10% FP
ADA-decoys	0.83	17	33	59	0.85	39	59	65	0.76	11	41	59
ADA-inactives	0.92	59	78	82	0.93	22	67	76	0.86	37	61	70
ERα-decoys	0.73	16	46	61	0.61	6	12	18	0.86	35	64	73
ERα-inactives	0.89	64	67	70	0.61	0	6	12	0.89	43	67	79

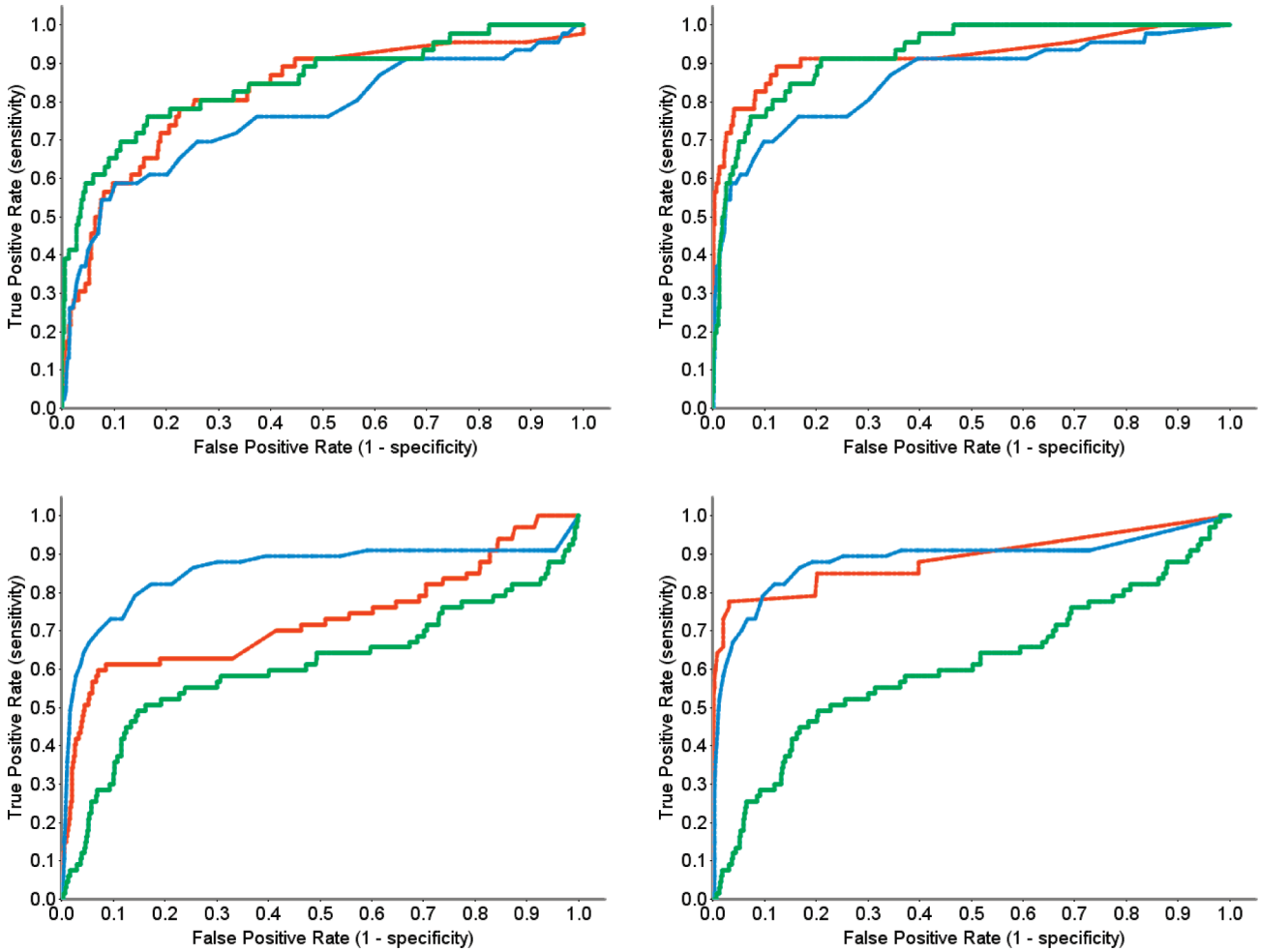

Figure 4. ROC curves showing database enrichments using GoldScore (green), Glide SP (blue), and naïve Bayes classifiers learned with SASA descriptors (red). ADA-decoys (left) and ADA-inactives (right) are shown on the top panel; and ERα-decoys (left) and ERα-inactives (right) test cases in the bottom panel.

protein binding sites also tend to be more hydrophobic than binding sites of inactive compounds. Though further individual contributions to the SASA total term were also found to be descriptive of activity in this data set, they were found to be smaller.

B. Bayesian Model Learned with SASA Descriptors: ADA and ERα Test Cases. Two test cases, ADA and ERα were additionally chosen to further evaluate the performance of SASA descriptors in database enrichment. We investigated two different data sets for baseline compounds: the 'decoy' and 'inactive' data sets. The latter baseline data set has similar Lipinski descriptors to 'active' ligands, but they differ substantially in 2D similarity, whereas the decoys are intimately related to the true ligands in similarity metric. SASA descriptors were calculated with Maestro for the ligands and proteins in their bound and unbound states. Bayesian models were then learned using an average of 5% actives and 5% inactives/decoys for ADA/ERα. Thereafter, the Bayesian models were used to rank the database compounds in the data sets using the normalized probability.

Table 2 shows the database enrichments of ADA and ERα using GOLD, Glide, and a Bayes classifier learned with SASA descriptors. It can be observed that the AUC and database enrichments using the topological SASA descriptors are comparable or better than the energy-based scoring function of GOLD and Glide (Figure 4 and Table 2). Interestingly, database enrichments were superior for the inactive data sets than those of the decoy data sets. This can be rationalized by the fact that active compounds tend to have similar SASA terms to structurally related decoys, which gives rise to a higher false positive rate. For this reason, the authors recognize that SASA descriptors may be

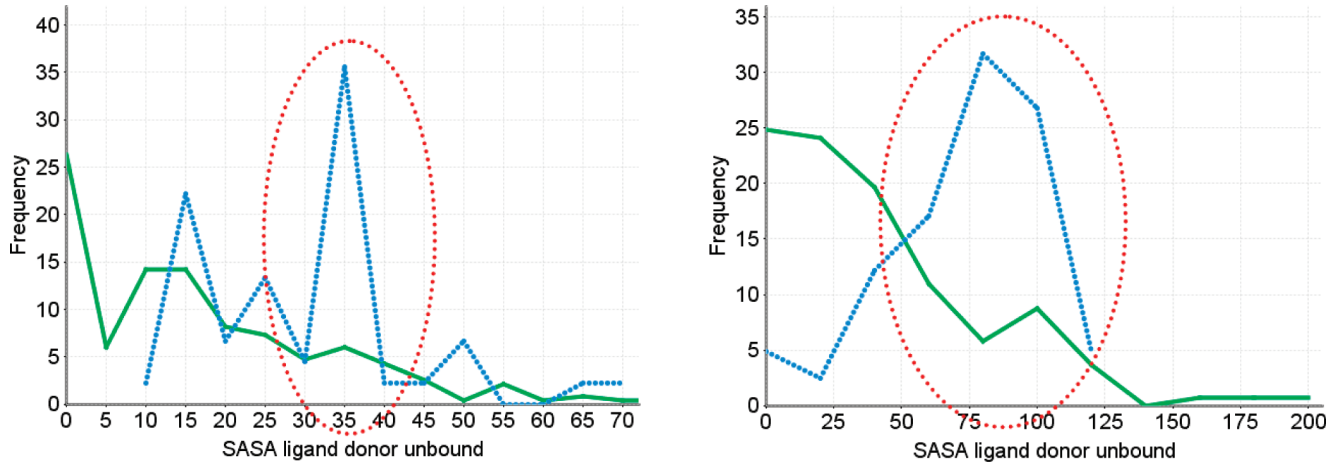

Figure 5. Dominant features of the Bayes model learned with SASA descriptors using the ERα (left) and ADA (right) data sets. Actives are shown in blue and inactives are in green. The *ligand donor unbound* SASA term has been shown to correlate well with activity.

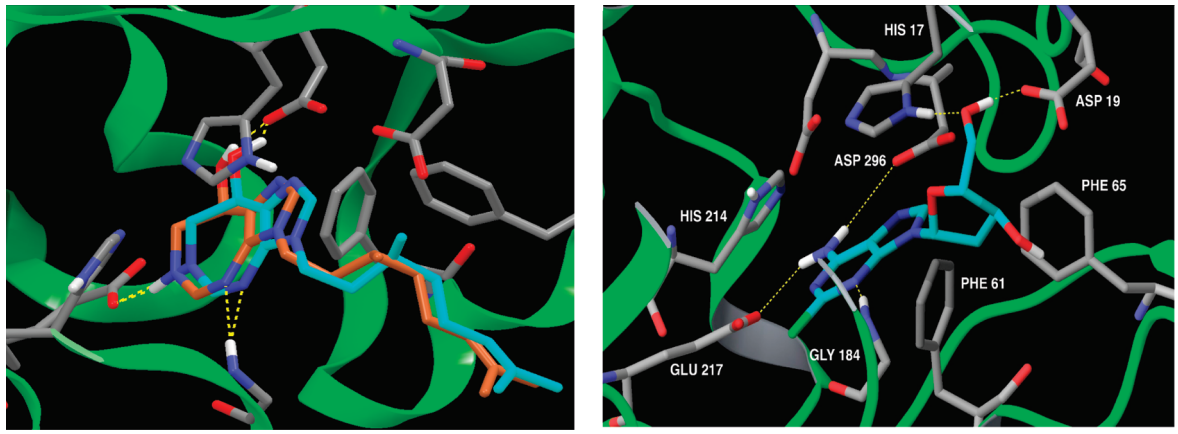

Figure 6. Active ADA ligand (left) and reference ADA ligand (right).

most useful for hit finding exercises. Further improvement of SASA descriptors for their use in lead optimization studies is warranted.

It is worth pointing out that the positive enrichments observed with the SASA method are not an artifact of the IFP postprocessing. GOLD and Glide enrichments in combination with identical postprocessing did not lead to improvements in database enrichments of the energy-based VS methods. Thus, the combination of ligand−protein complementarity provided by the IFP postprocessing, coupled with the 'filling of the binding site' measure provided by the SASA descriptors, provides a unique method for improving early database enrichment.

Feature statistics of the SASA Bayesian models for ADA and ERα were extracted to better understand the putative dominant forces leading the binding event. For ADA, the most dominant features which distinguish active from inactive compounds were the *ligand positive unbound* and *ligand donor unbound* SASA terms. For ERα, both the *ligand donor unbound* and the *receptor positive bound* SASA terms were instrumental in discriminating active compounds from baseline, which correlates well with the pharmacophoric points of ADA and ERα (Figure 5).

To illustrate the added benefit of SASA descriptors in database ranking, a direct comparison was made for a given ADA active ligand that demonstrated a docking pose comparable to that of the active ligands in the training set (Figure 6). For this ligand, the docking score (GoldScore) was low, and the SASA score was high. In other words, the naïve Bayesian model learned with SASA descriptors was able to identify the key features in the SASA descriptors of this ligand that match the training set ligands and that satisfactorily classified it as active. On the other hand, the scoring function of GOLD was not able to score well the ligand. This inefficient scoring most likely arises from the overestimation of the hydrogen-bond contribution to the binding free energy of active ADA ligands, thus resulting in a false negative in the GoldScoring scheme for active ligands having a less extensive hydrogen-bond network.

4. CONCLUSIONS

The current study presents a comparison of a novel topological scoring method based on SASA descriptors against the VS methods, Glide and GOLD, as to their early database enrichment efficiency. Early database enrichments using SASA descriptors were found comparable or superior to those of GOLD and Glide. Moreover, we have shown that the performance of GOLD and Glide is highly dependent on the protein test case; however, SASA descriptors display an outstanding robustness to produce satisfactory early enrichments for a large variety of target classes. Topological SASA descriptors constitute thus a fast and attractive in silico scoring method and show promise in virtual hit finding efforts in a wide variety of drug discovery projects. The future use

of SASA descriptors to aid in silico lead optimization efforts is warranted.

ACKNOWLEDGMENT

The authors acknowledge the help of Dr. Antoniou and Dr. McCormack for proof-reading the manuscript, Dr. Lange for useful discussions, and Schrödinger, Inc. for providing us with the SASA python script.

Supporting Information Available: A comparison of the ligand properties of the described sc-PDB data set and ligands available from Prous Science Integrity is included. Moreover, a list of the sc-PDB training and test set complexes is included. This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Jain, A. N. Virtual screening in lead discovery and optimization. *Curr. Opin. Drug Discovery Dev.* **2004**, *7*, 396–403.
(2) Venhorst, J.; Núñez, S.; Terpstra, J. W.; Kruse, C. G. Assessment of scaffold hopping efficiency by use of molecular interaction fingerprints. *J. Med. Chem.* **2008**, *51*, 3222–3229.
(3) Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11*, 580–594.
(4) Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862–865.
(5) Seifert, M. H.; Kraus, J.; Kramer, B. Virtual high-throughput screening of molecular databases. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 298–307.
(6) Marcou, G.; Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* **2007**, *47*, 195–207.
(7) Deng, Z.; Chuaqui, C.; Singh, J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. *J. Med. Chem.* **2004**, *47*, 337–44.
(8) Pearlman, D. A. Evaluating the molecular mechanics poisson-boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase. *J. Med. Chem.* **2005**, *48*, 7796–7807.
(9) Ferrari, A. M.; Degliesposti, G.; Sgobba, M.; Rastelli, G. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. *Bioorg. Med. Chem.* **2007**, *15*, 7865–7877.
(10) Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
(11) Aqvist, J.; Medina, C.; Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385–391.
(12) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–897.
(13) Pearlman, D. A. Evaluating the molecular mechanics poisson-boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase. *J. Med. Chem.* **2005**, *48*, 7796–7807.
(14) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-ligand docking: current status and future challenges. *Proteins* **2006**, *65*, 15–26.
(15) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109.
(16) 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, Version 4.0 was used in this study.
(17) *Maestro*, version 8.5; Schrödinger, L.L.C.: New York, 2009.
(18) Smith, R. D.; Hu, L.; Falkner, J. A.; Benson, M. L.; Nerothin, J. P.; Carlson, H. A. Exploring protein-ligand recognition with Binding MOAD. *J. Mol. Graphics Modell.* **2006**, *24*, 414–425.
(19) Wodak, S. J.; Janin, J. Analytical approximation to the accessible surface area of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1736–1740.
(20) Wilkinson, D. J. Bayesian methods in bioinformatics and computational systems biology. *Briefings Bioinf.* **2007**, *8*, 109–116.
(21) Kellenberger, E.; Muller, P.; Schalon, C.; Bret, G.; Foata, N.; Rognan, D. sc-PDB: an annotated database of druggable binding sites from the Protein Data Bank. *J. Chem. Inf. Model.* **2006**, *46*, 717–727.
(22) RCSB Protein Data Bank; http://www.pdb.org. Accessed October 28, 2009.
(23) Prous Science Integrity; http://www.prous.com/integrity. Accessed October 25, 2009.
(24) *PipelinePilot*, version 7.0.1; Accelrys, Inc.: San Diego, CA, 2008.
(25) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T.; Murray, C. W.; Taylor, R. D.; Watson, P. Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793–806.
(26) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.
(27) 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.
(28) Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235–249.
(29) Tanimoto, T. T. *IBM Internal Report*; IBM: Armonk, NY, November, 17, 1957.
(30) La Motta, C.; Sartini, S.; Mugnaini, L.; Salerno, S.; Simorini, F.; Taliani, S.; Marini, A. M.; Da Settimo, F.; Lavecchia, A.; Novellino, E.; Antonioli, L.; Fornai, M.; Blandizzi, C.; Del Tacca, M. Exploiting the pyrazolo[3,4-d]pyrimidin-4-one ring system as a useful template to obtain potent adenosine deaminase inhibitors. *J. Med. Chem.* **2009**, *52*, 1681–1692.
(31) Shiau, A. K.; Barstad, D.; Radek, J. T.; Meyers, M. J.; Nettles, K. W.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A.; Agard, D. A.; Greene, G. L. Structural characterization of a subtype-selective ligand reveals a novel mode of estrogen receptor antagonism. *Nat. Struct. Biol.* **2002**, *9*, 359–364.
(32) Hand, D. J.; Till, R. J. A simple generalization of the area under the ROC curve to multiple class classification problems. *Mac. Learn.* **2001**, *45*, 171–186.

CI9004628