# Novel Method for Generating Structure-Based Pharmacophores Using Energetic Analysis

Noeris K. Salam,*,[†] Roberto Nuti,[‡] and Woody Sherman[†]

Schrödinger, Inc., 120 West 45th Street, 29th Floor, New York, New York 10036, and Dipartimento di Chimica e Tecnologia del Farmaco, Università di Perugia, Via del Liceo 1, 06123 Perugia, Italy

We describe a novel method to develop energetically optimized, structure-based pharmacophores for use in rapid *in silico* screening. The method combines pharmacophore perception and database screening with protein−ligand energetic terms computed by the Glide XP scoring function to rank the importance of pharmacophore features. We derive energy-optimized pharmacophore hypotheses for 30 pharmaceutically relevant crystal structures and screen a database to assess the enrichment of active compounds. The method is compared to three other approaches: (1) pharmacophore hypotheses derived from a systematic assessment of receptor−ligand contacts, (2) Glide SP docking, and (3) 2D ligand fingerprint similarity. The method developed here shows better enrichments than the other three methods and yields a greater diversity of actives than the contact-based pharmacophores or the 2D ligand similarity. Docking produces the most cases (28/30) with enrichments greater than 10.0 in the top 1% of the database and on average produces the greatest diversity of active molecules. The combination of energy terms from a structure-based analysis with the speed of a ligand-based pharmacophore search results in a method that leverages the strengths of both approaches to produce high enrichments with a good diversity of active molecules.

## INTRODUCTION

Computer-aided drug design techniques such as ligand-based pharmacophore modeling and structure-based protein−ligand docking have become an integral part of drug discovery.[1] Research organizations in both industry and academia use these techniques to aid in the efficient discovery and design of active molecules.[2−4] Large compound databases can be screened computationally to reduce the number of compounds for bioassay screens, thereby saving time and resources.[5−7] Furthermore, computational techniques can be used in the lead optimization process after hits are identified. Ligand-based technologies, such as 2D fingerprint similarity searching, shape-based screening, and 3D-pharmacophore modeling are traditionally recognized as fast methods for screening large compound databases.[8−10] Structure-based approaches, on the other hand, are generally more computationally expensive but can lead to structural insights and have been shown to yield more diverse actives.[11]

Success stories in pharmacophore modeling are abundant in the literature,[12−14] where this technology has been effectively used as a means of both finding and optimizing active molecules. The general concept of pharmacophore modeling adheres to the premise that related chemical groups (i.e., hydrogen-bond donors/acceptors, aromatic rings, etc.) in spatially and geometrically similar arrangements can make comparable interactions with a receptor, thereby eliciting similar chemical and biological activity. When chemical knowledge of numerous active ligands is available, one can detect a common pharmacophore and subsequently generate a corresponding complementary model of the receptor's binding site based only on the ligands.[15,16] This model, or pharmacophore hypothesis, in principle represents the 3D arrangement of molecular features necessary for activity, including lipophilic, aromatic, hydrogen bonding, and ionic groups. A compound database can then be screened for molecules that are able to present 3D pharmacophore features that match the hypothesis. Pharmacophore modeling is most commonly performed when limited 3D structural information for the target is available. It is thus of particular relevance to drug discovery campaigns involving targets that are difficult to crystallize, such as ion channels, transporters, or G protein-coupled receptors (GPCRs).[12,17]

Screening a 3D database against a pharmacophore hypothesis is generally more computationally efficient than structure-based docking, which involves many energy evaluations as part of the conformational searching and scoring process. Recently, methods have emerged that attempt to capitalize on the speed of pharmacophore screening coupled with structure-based information by developing pharmacophore hypotheses derived from protein−ligand complexes.[18−20] These methods show promise and have been used to discover novel leads.[21] They typically use a simple contact score or a molecular mechanics scoring function to categorize protein−ligand interactions that can then be used to develop the pharmacophore model. We refer to such a model as a contact-based pharmacophore (c-pharmacophore). Using a scoring method with greater accuracy should, in principle, result in more relevant structure-based pharmacophores and thereby improved database-screening results. Thus, as part of this work we attempt to accurately characterize the protein−ligand interactions based on energetic contributions of each feature to binding such

* Corresponding author phone: (858)643-9000; e-mail: Noeris.Salam@ Schrodinger.com. Current address: Schrödinger, Inc., 8910 University Center Lane, Suite 270 San Diego, CA 92122.
† Schrödinger, Inc.
‡ Università di Perugia.

that energetically favorable features are incorporated into the pharmacophore in preference to energetically weaker features.

For cases where structural information for the target is available, it is common to use docking for virtual screening and lead optimization. A large number of docking programs exist that have been described in detail elsewhere.[22−26] Most programs treat the ligand flexibly and the receptor rigidly. However, a key complicating factor in docking is accounting for the possibility of receptor conformational changes, and, in many cases, the rigid receptor model is not sufficient to produce accurate poses. Increasingly sophisticated protocols have been developed in recent years that explicitly account for conformational sampling of the receptor either during the docking process or in an iterative fashion coupling the ligand docking and protein sampling.[27−31] Most of the work on treating receptor flexibility has gone into accurately predicting the ligand binding mode and protein conformation, whereas little work has gone into incorporating receptor flexibility into virtual screening. However, docking to ensembles of protein conformations has proven to be useful in recovering diverse active molecules that fit into different protein conformations, resulting in improved virtual screening enrichments.[28,32]

In the structure-based part of the protocol we use the docking program Glide[26] to refine ligand−receptor crystal structures and the Glide XP scoring function to obtain an energetic description of each complex. Glide has been well validated for producing accurate ligand poses[26] and good database enrichment.[33,34] The protocol we developed does not involve fully flexible docking but instead uses Glide to refine the cognate ligand, and therefore there is no need to explicitly treat the protein flexibly in the initial stage of the protocol. The pharmacophore screening stage, however, does allow for matching of molecules that may not ideally fit the binding site. The Glide XP scoring function used in this work accounts for more complex energy terms than traditional molecular mechanics or empirical scoring functions such as ChemScore.[35] For example, hydrophobic enclosure is detected when lipophilic ligand atoms are enclosed on opposite sides by lipophilic protein atoms, producing a large energetic reward as a consequence of the release of unfavorable water molecules that were previously in the hydrophobic environment. Furthermore, Glide XP recognizes common medicinal chemistry motifs such as $\pi...\pi$ and $\pi...$cation interactions.[34] Other structural recognition motifs include single or correlated hydrogen bonds in hydrophobically enclosed environments. These, and a number of other common medicinal chemistry recognition terms, are described in detail in the original Glide XP work.[34]

In this work, we describe a novel protocol for generating energy-optimized pharmacophores (e-pharmacophores) based on mapping of the energetic terms from the Glide XP scoring function onto atom centers. Beginning with a ligand−receptor complex we refine the ligand pose, compute the Glide XP scoring terms, and map the energies onto atoms. Then, pharmacophore sites are generated, and the Glide XP energies from the atoms that comprise each pharmacophore site are summed. The sites are then ranked based on these energies, and the most favorable sites are selected for the pharmacophore hypothesis. Finally, these e-pharmacophores are used
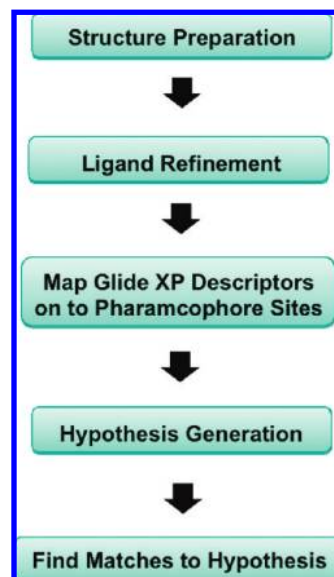


**Figure 1.** Overview of workflow.

as queries for virtual screening. The protocol is summarized in Figure 1 and is described thoroughly in the Methods section.

To test the e-pharmacophore methodology, we screened a set of 30 diverse crystal structures of pharmaceutically relevant protein targets (Figure 2) and present database enrichments. In addition to the traditional enrichment factor, we also developed a diversity-weighted enrichment factor that is useful to simultaneously assess both the number of actives recovered and the diversity spanned by the retrieved actives relative to the diversity of all of the actives. This is particularly important in pharmaceutical drug discovery, as there is substantial intellectual property value in finding compounds that are dissimilar to known actives. Furthermore, a method that is very good at identifying compounds that span a wide range of chemical families can be complementary to traditional ligand-based methods that generally retrieve many similar compounds. In-depth analyses of several systems are presented to highlight the key strengths and weaknesses of the method. The enrichments are compared to another hybrid structure-based/ligand-based method in which a systematic analysis of contacts is performed (c-pharmacophores) as well as two other widely used approaches for virtual screening: docking - a structure-based technique, and 2D fingerprint similarity searching - a ligand-based technique. Overall, the enrichments and the diversity-weighted enrichments are better for this method than for the other methods studied here.

## METHODS

**Structure Preparation.** Coordinates for each structure were taken from the RCSB Protein Data Bank (PDB)[36] and prepared using the Protein Preparation Wizard, which is part of the Maestro software package (Maestro, v8.5, Schrödinger, LLC, New York, NY). Bond orders and formal charges were added for heterogroups, and hydrogens were added to all atoms in the system. To optimize the hydrogen bond network, His tautomers and ionization states were predicted, 180° rotations of the terminal $\chi$ angle of Asn, Gln, and His residues were assigned, and hydroxyl and thiol hydrogens were sampled. Water molecules in all structures were removed.
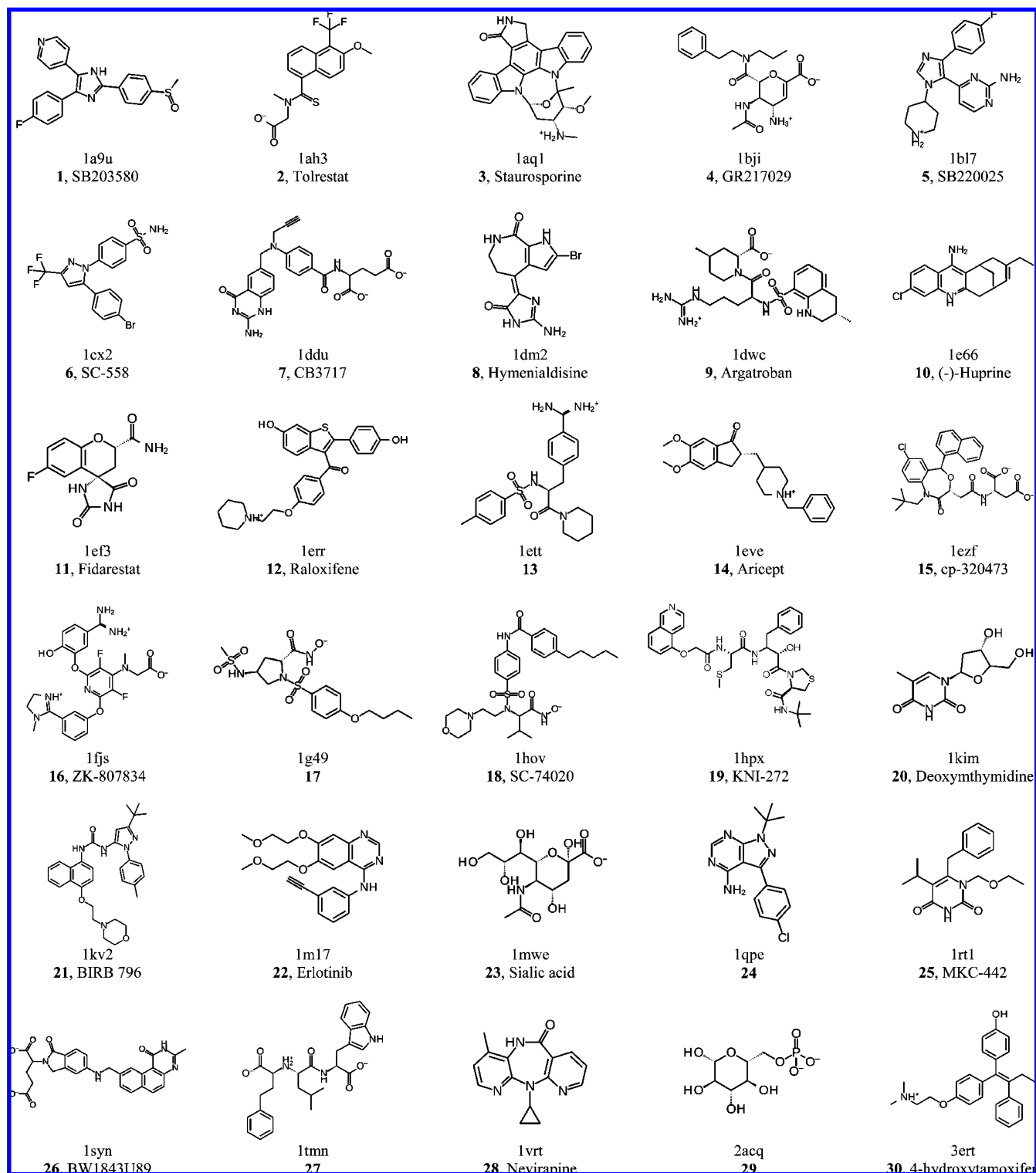
**Figure 2.** Structures of compounds considered in this study. PDB code and compound name.

For structures with missing side-chain atoms, the refinement module in Prime (Prime, v2.0, Schrödinger, LLC, New York, NY) was used to predict their conformations. For each structure, a brief relaxation was performed using an all-atom constrained minimization carried out with the Impact Refinement module (Impref) (Impact v5.0, Schrödinger, LLC, New York, NY) using the OPLS-2005 force field to alleviate steric clashes that may exist in the original PDB structures. The minimization was terminated when the energy converged or the rmsd reached a maximum cutoff of 0.30 Å.

**Ligand Docking/Refinement.** Glide energy grids were generated for each of the prepared complexes. The binding site was defined by a rectangular box surrounding the X-ray ligand. Ligands were refined using the "Refine" option in Glide, and the option to output Glide XP descriptor information was chosen (Glide v5.0, Schrödinger, LLC, New York, NY). Default settings were used for the refinement and scoring.

**Hypothesis Generation for Energy-Optimized Structure-Based Pharmacophores (e-Pharmacophores).** *Stage 1.* Starting with the refined X-ray ligand, pharmacophore sites were

ENERGETICALLY OPTIMIZED PHARMACOPHORE MODELING

*J. Chem. Inf. Model., Vol. 49, No. 10, 2009* **2359**
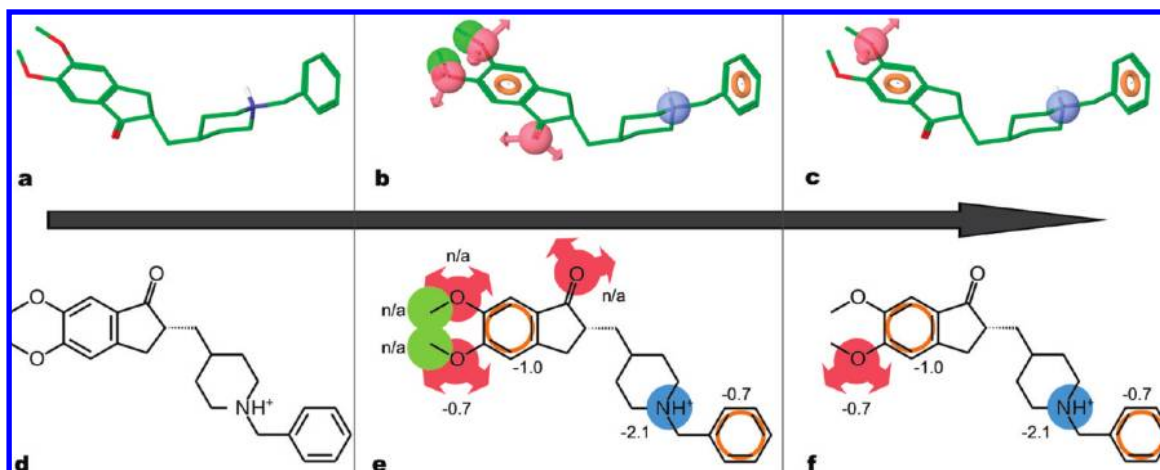


**Figure 3.** Example of energy-optimized hypothesis for Aricept, **14**, in PDB 1eve. Arrow represents the evolution from the initial ligand (a,d) to the ligand with all sites (b,e) to the ligand with the most energetically favorable sites (c,f). Pink sphere/circle = hydrogen bond acceptor, green sphere/circle = hydrophobic group, orange ring = aromatic ring, and blue sphere/circle = positive ionizable.

automatically generated with Phase (Phase, v3.0, Schrödinger, LLC, New York, NY) using the default set of six chemical features: hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobe (H), negative ionizable (N), positive ionizable (P), and aromatic ring (R). While Phase treats most cationic groups exclusively as positive ionizable, we chose to represent primary and secondary amines from guanidinium and amidine groups (such as the ligands in 1dwc, 1ett, and 1fjs) with a complementary set of hydrogen bond donors. Hydrogen bond acceptor sites were represented as vectors along the hydrogen bond axis in accordance with the hybridization of the acceptor atom. Hydrogen bond donors were represented as projected points, located at the corresponding hydrogen bond acceptor positions in the binding site. Projected-points allow the possibility for structurally dissimilar active compounds to form hydrogen bonds to the same location, regardless of their point of origin and directionality. Figure 3 shows an example of this process applied to Aricept, an acetylcholinesterase inhibitor aimed at slowing the progression of Alzheimer's symptoms.

*Stage 2.* Each pharmacophore feature site is first assigned an energetic value equal to the sum of the Glide XP contributions of the atoms comprising the site (Figure 3b,e). This allows sites to be quantified and ranked on the basis of these energetic terms. Glide XP descriptors include terms for hydrophobic enclosure, hydrophobically packed hydrogen bonds, hydrophobically packed correlated hydrogen bonds, electrostatic rewards, $\pi...\pi$ stacking, $\pi...$cation, and other interactions.[34] ChemScore[35] hydrogen bonding and lipophilic atom pair interaction terms are included when the Glide XP terms for hydrogen binding and hydrophobic enclosure are zero. Sites with less than half of the heavy atoms contributing to the pharmacophore feature are excluded from the final hypothesis. Thus, if only two heavy atoms in a six-member ring exhibit energetic interactions, the ring is not considered as a pharmacophore feature. Variations on this rule, such as including features with fewer than half of the heavy atoms exhibiting energetic interactions, were tried and tested during the course of this work and ultimately did not show any improvements in enrichments (results not shown). The final procedure reported here leads to a final energy optimized hypothesis (e-pharmacophore), for example, as shown in Figure 3c,f.

**Hypothesis Generation for Contact-Based Structure-Based Pharmacophores (c-Pharmacophores).** *Stage 1.* The steps for generating the initial set of all pharmacophore sites are the same as in the in the *Stage 1* section above for e-pharmacophores.

*Stage 2.* Pharmacophore features were chosen for the final hypothesis based on a systematic analysis of receptor−ligand contacts. The contacts considered included hydrophobic or aromatic groups contacting the receptor, hydrogen bonds as measured in Maestro v8.5 (Maestro, v8.5, Schrödinger, LLC, New York, NY) or ionic interactions. For hydrogen-bonds, we used a O/N...H maximum distance of 2.5 Å, a minimum donor angle of 120.0°, and a minimum acceptor angle of 90.0°.

**Receptor-Based Excluded Volumes.** Excluded volumes include all atoms within 5 Å of the refined ligand for each target. Each excluded volume sphere has a radius corresponding to 50% of the van der Waals radius of the receptor atom. The radius scaling is used to help model receptor flexibility and lessen the effects of minor steric clashes. To further account for protein flexibility, receptor atoms less than 1.5 Å from the ligand are ignored. The same set of excluded volumes was used in both the e-pharmacophore and c-pharmacophore approaches.

**Database Creation.** An initial data set of 1000 druglike decoys, with an average molecular weight of 400 Da (the "dl-400" data set), was employed for all targets except thymidine kinase (1kim), which has a small binding site, and therefore we used a set of 1000 decoys with an average molecule weight of 360 Da (the "dl-360" data set) to more accurately represent the actives for this target. Both the dl-360 and dl-400 ligand decoy sets are available for download (http://www.schrodinger.com/glide_decoy_set) and are also included in the Supporting Information. The property distributions of these data sets have been well characterized,[26] and they have been employed in several retrospective virtual screens.[33,34]

We supplemented the dl-360 and dl-400 data sets with the known binders for each target. The compounds are similar to those used in previous screens.[26,33,34] Each active compound exhibits affinities better than 10 μM, with the exception of active neuraminidase ligands, which are relatively weak binders that do not have activities better than

10 $\mu$M. For cases with more than 30 actives, we performed hierarchical clustering based on linear fingerprints using the program Canvas (Canvas v1.0, Schrödinger LLC, New York, NY) to reduce the total number of actives to 20. This was done to avoid saturation of actives relative to decoys, as has been discussed by Truchon et al.,[37] and to present a consistent ratio of actives to decoys for each screen.

Database molecules were prepared using LigPrep (LigPrep v2.2, Schrödinger LLC, New York, NY) with Epik (Epik v1.6, Schrödinger, LLC, New York, NY) to expand protonation and tautomeric states at 7.0 ± 2.0 pH units. Conformational sampling was performed on all database molecules using the ConfGen search algorithm.[26] ConfGen samples conformations based on a heuristic search algorithm and energetic evaluations to efficiently explore diversity around rotatable bonds, flexible ring systems, and nitrogen inversions. We employed ConfGen with the OPLS_2005 force field and a duplicate pose elimination criterion of 1.0 Å rmsd to remove redundant conformers. A distance-dependent dielectric solvation treatment was used to screen electrostatic interactions. The maximum relative energy difference of 10.0 kcal/mol was chosen to exclude high-energy structures. Using Phase, the database was indexed with the automatic creation of pharmacophore sites for each conformer to allow rapid database alignments and screening.

**e-Pharmacophore Database Screening.** For the e-pharmacophore approach, explicit matching was required for the most energetically favorable site, conditional on it scoring better than −1.0 kcal/mol. Multiple sites were included in cases where more than one site scored equally as the top site. This occurred only for the 1aq1 and 1kim ligands, where the hydrogen bond donor and acceptor sites that define a correlated hydrogen bond pair scored equally and were both required. Screening molecules were required to match a minimum of 3 sites for hypotheses with 3 or 4 sites and a minimum of 4 sites for hypotheses with 5 or more sites. Distance matching tolerance was set to 2.0 Å as a balance between stringent and loose-fitting matching alignment. Database hits were ranked in order of their Fitness score,[15] a measure of how well the aligned ligand conformer matches the hypothesis based on rmsd site matching, vector alignments, and volume terms. The Fitness scoring function is an equally weighted composite of these three terms and ranges from 0 to 3, as implemented in the default database screening in Phase.

**c-Pharmacophore Database Screening.** The c-pharmacophore screen was similarly performed to that of e-pharmacophore. Hypotheses were limited to seven sites. An identical scoring function and set of excluded volumes were used in the c-pharmacophore as in the e-pharmacophore. The only difference in these two approaches is the actual features chosen for the pharmacophore hypothesis.

**Glide Docking.** Database ligands were docked into the binding sites of the 30 receptors with Glide 5.0 (Glide v5.0 Schrödinger, LLC, New York, NY) utilizing the standard precision (SP) scoring function to estimate protein−ligand binding affinities. Glide SP is faster and more tolerant to suboptimal fits than Glide XP, making it better for comparison in this work. The center of the Glide grid was defined by the position of the cocrystallized ligand. Default settings were used for both the grid generation and docking. This included a scaling of the ligand vdW radii by 0.8 to partially account for suboptimal fits that could be accommodated by minor receptor movements.

**2D Screening.** Linear fingerprints similar to Daylight[38] were generated using Canvas (Canvas v1.1, Schrödinger LLC, New York, NY) with Daylight invariant atom types. Fingerprints were generated for each of the 30 X-ray ligands as well as all members of the database (actives and decoys). Database screening was then performed for each target using the Tanimoto metric to assess similarity. Database compounds were sorted by their similarity score in order to compute enrichments.

**Measures of Enrichment.** We employed the enrichment factor (EF) for recovering a fraction of the known actives after a fraction of the database has been screened. EF(X%) is defined as the fraction of the actives recovered after X% of the decoy database has been screened (eq 1). This value gives the enrichment over random that a compound recovered is an active rather than a decoy

$$EF = \frac{F_a}{F_d} \tag{1}$$

In eq 1, $F_a$ is the fraction of active compounds recovered after the fraction $F_d$ of decoys has been screened. For this study, we focused primarily on EF(1%), the enrichment in the top 1% of the decoys. A maximum enrichment is therefore 100 if all of the actives (i.e., $F_a = 1$) are found within the top 1% of the decoys ($F_d = 0.01$).

A second enrichment metric — the Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC)[37] — was also used as a way to ensure that the results and conclusions were significant. BEDROC is a generalization of the receiver operating characteristic (ROC)[39] that addresses the "early scoring problem" by Boltzmann-weighting the hits based on how early they are retrieved. We used both $\alpha = 20.0$ and $\alpha = 160.9$ for the comparison. The value of $\alpha = 20.0$ is suggested in the original paper by Truchon et al. as a reasonable choice for virtual screening evaluations and corresponds to 80% of the total score being accounted for in the top 8% of the database. We have also found a value of $\alpha = 160.9$ to be useful in virtual screening, which corresponds to 80% of the score being accounted for in the top 1% of the database. We found a very high correlation between EF and BEDROC values in this study ($R^2 = 0.96$ for EF(1%) and BEDROC($\alpha = 20.0$); see Supporting Information Figure S4), and therefore we opted to use to the less complicated and more commonly used EF metric throughout this work.

A diversity-base enrichment factor (DEF) was formulated in this work and computed for each screen (eq 2). The DEF was designed to penalize enrichment scores where the recovered actives have low structural diversity among each other. The DEF is the EF times the ratio of the diversity of the actives recovered to the maximum diversity (i.e., minimum similarity) of the full active set. Similarity is calculated as described in 2D Screening

$$DEF = EF \times \frac{1 - \min(Sim_{Hits})}{1 - \min(Sim_{Actives})} \tag{2}$$

**Computation Times.** All calculations were run on 2.33 GHz Xeon processors. The average time for Glide grid

ENERGETICALLY OPTIMIZED PHARMACOPHORE MODELING

*J. Chem. Inf. Model., Vol. 49, No. 10, 2009* **2361**

**Table 1.** Targets Studied in This Work

| PDB code | description | no. actives |
|---|---|---|
| 1e66 | acetylcholinesterase | 25 |
| 1eve | acetylcholinesterase | 25 |
| 1ah3 | aldose reductase | 36 |
| 1ef3 | aldose reductase | 36 |
| 2acq | aldose reductase | 36 |
| 1aq1 | cyclin dependent kinase-2 | 20 |
| 1dm2 | cyclin dependent kinase-2 | 20 |
| 1cx2 | cyclooxygenase-2 | 33 |
| 1ddu | *E. coli* thymidylate synthase | 20 |
| 1syn | *E. coli* thymidylate synthase | 20 |
| 1m17 | EGFR tyrosine kinase | 20 |
| 1fjs | factor Xa | 20 |
| 1hpx | HIV-1 protease | 15 |
| 1rt1 | HIV-1 reverse transcriptase[b] | 20 |
| 1vrt | HIV-1 reverse transcriptase | 20 |
| 1err | human estrogen receptor | 20 |
| 3ert | human estrogen receptor | 20 |
| 1qpe | lymphocyte-specific tyrosine kinase | 20 |
| 1hov | matrix metalloproteinase-2 | 20 |
| 1g49 | matrix metalloproteinase-3 | 20 |
| 1bji | neuraminidase | 20 |
| 1mwe | neuraminidase | 20 |
| 1a9u | p38 MAP kinase | 20 |
| 1bl7 | p38 MAP kinase | 20 |
| 1kv2 | p38 MAP kinase[a] | 20 |
| 1ezf | squalene synthase | 10 |
| 1tmn | thermolysin | 10 |
| 1dwc | thrombin | 19 |
| 1ett | thrombin | 19 |
| 1kim | thymidine kinase | 7 |

[a] DFG-out binding site. [b] Allosteric binding site - NNRT.

**Table 2.** Number of Pharmacophoric Sites for Each Ligand Prior and Post Energy-Optimization Site Selection for the Set of 30 Targets

| reference ligand | total no. of available sites | no. of optimized sites | hypothesis[a] | no. of sites to match | no. of must match sites[b] |
|---|---|---|---|---|---|
| **1**, 1a9u | 9 | 6 | AAHRRR | 4 | 1 |
| **2**, 1ah3 | 7 | 4 | HNRR | 3 | 1 |
| **3**, 1aq1 | 15 | 7 | ADDDRRR | 4 | 2 |
| **4**, 1bji | 12 | 6 | ADDNPR | 4 | 1 |
| **5**, 1bl7 | 13 | 7 | AADDHRR | 4 | 1 |
| **6**, 1cx2 | 10 | 7 | ADDHRRR | 4 | 0 |
| **7**, 1ddu | 14 | 5 | NNRRR | 4 | 0 |
| **8**, 1dm2 | 12 | 7 | ADDDDHR | 4 | 1 |
| **9**, 1dwc | 16 | 7 | ADDDDHN | 4 | 0 |
| **10**, 1e66 | 9 | 6 | DHHHRR | 4 | 1 |
| **11**, 1ef3 | 11 | 6 | AAADHR | 4 | 1 |
| **12**, 1err | 12 | 7 | ADDPRRR | 4 | 0 |
| **13**, 1ett | 10 | 7 | ADHHPRR | 4 | 0 |
| **14**, 1eve | 10 | 4 | APRR | 3 | 1 |
| **15**, 1ezf | 12 | 7 | AHNNRRR | 4 | 1 |
| **16**, 1fjs | 22 | 7 | DDDNPRR | 4 | 1 |
| **17**, 1g49 | 11 | 4 | ADNR | 3 | 1 |
| **18**, 1hov | 15 | 4 | DNRR | 3 | 1 |
| **19**, 1hpx | 21 | 6 | AAADDR | 4 | 0 |
| **20**, 1kim | 9 | 6 | AADDDR | 4 | 2 |
| **21**, 1kv2 | 14 | 7 | AADDHRR | 4 | 1 |
| **22**, 1m17 | 12 | 5 | AARRR | 4 | 1 |
| **23**, 1mwe | 14 | 6 | AADDDN | 4 | 1 |
| **24**, 1qpe | 8 | 4 | ADRR | 3 | 1 |
| **25**, 1rt1 | 8 | 4 | DHRR | 3 | 1 |
| **26**, 1syn | 13 | 7 | ADDNRRR | 4 | 1 |
| **27**, 1tmn | 12 | 7 | ADDDHNN | 4 | 1 |
| **28**, 1vrt | 9 | 4 | HHRR | 3 | 1 |
| **29**, 2acq | 13 | 3 | AAN | 3 | 0 |
| **30**, 3ert | 8 | 7 | ADHPRRR | 4 | 0 |

[a] A = h-bond acceptor; D = h-bond donor; H = hydrophobic group; N = negative ionizable group; P = positive ionizable group; R = aromatic ring. [b] For must match sites see Figure 4.

generation and XP refinement is approximately 5 min. This is needed only once for the starting ligand−receptor complex in order to develop the e-pharmacophore. The Phase pharmacophore screens were performed on 20 processors and took approximately 1.5 min for the set of 1000 decoy compounds, translating to 2 s per compound. Glide SP docking times were approximately 15 s per ligand, and Canvas fingerprint similarity screens were approximately 0.005 s per ligand.

## RESULTS AND DISCUSSION

The e-pharmacophore method developed here, which combines aspects of structure-based and ligand-based techniques, was assessed on a diverse set of 30 well-characterized complexes representing 19 pharmaceutically relevant targets (Table 1). Of this set, 20 complexes were taken from previous docking studies,[26,33,34] and an additional 10 complexes, spanning 5 new targets, were added (aldose reductase, *E. coli* thymidylate synthase, matrix metalloproteinase-2 and -3, and squalene synthase). Several targets are represented by two or more PDB crystallized structures with varying degrees of binding site differences. The most significant changes between crystal structures of the same target are seen in the allosteric pockets of p38 MAP kinase[40] and in non-nucleoside reverse transcriptase (NNRTI) bound complexes of HIV reverse transcriptase (HIV-RT).[41] In the following sections, we first show examples related to the steps involved in developing pharmacophore hypotheses from mapping Glide XP energetic terms onto pharmacophore sites. We then present the performance of this method on 30 crystal structures for enriching active compounds from a database.

Specific examples are given to highlight the strengths and weaknesses of the method.

**Pharmacophore Development.** Incorporating protein−ligand contacts into ligand-based pharmacophore approaches has been shown to produce enhanced enrichments over using ligand information alone.[42] The method described here attempts to take a step beyond simple contact scoring by incorporating structural and energetic information using the scoring function in Glide XP.[34] Table 2 lists the total number of pharmacophore sites for each ligand prior to energy-based site selection for the set of 30 targets. On average, there are 12 sites per hypothesis, many of which do not appear to be directly involved in protein−ligand interactions. Figure 3b,e shows an example of the acetylcholinesterase inhibitor Aricept (14, 1eve) with all eight pharmacophore sites, whereas Figure 3c,f shows only the four sites selected by the e-pharmacophore method. The energetically favorable sites encompass the specific interactions of Aricept's (14) dimethoxyindanone, piperidine, and benzyl moieties, which previous experimental studies have established as being responsible for high binding affinity and selectivity.[43] Across all 30 systems, the protocol was able to reduce the average number of sites per ligand by more than half, to 5.8, which makes the hypotheses more practical for screening (see Table 2 for the number of sites in each hypothesis). Figure 4 shows the final e-pharmacophores for each of the 30 systems studied here.
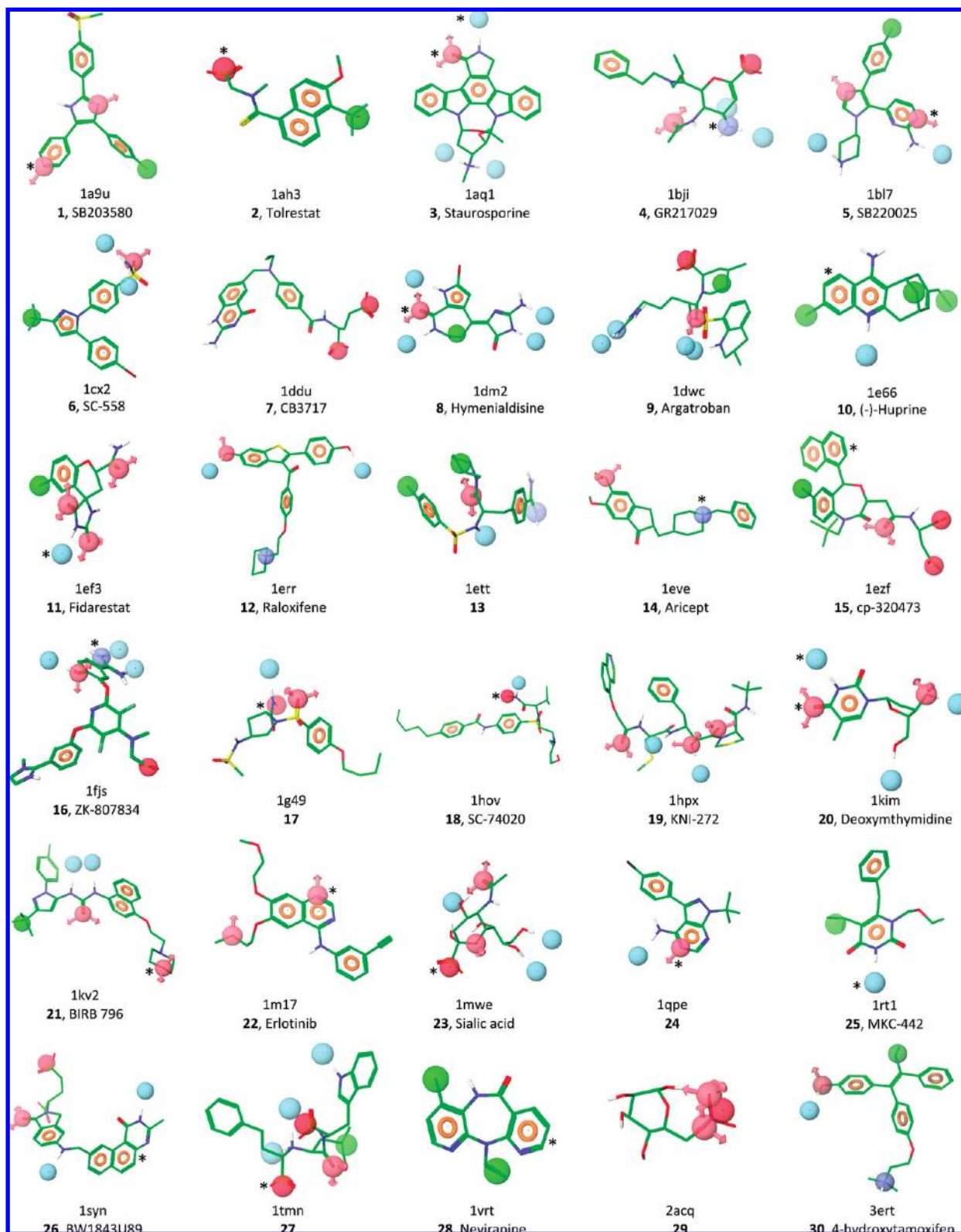
**Figure 4.** Energy-optimized hypotheses for all 30 structures considered in this study. Pink sphere/circle = hydrogen bond acceptor, green sphere/circle = hydrophobic group, orange ring = aromatic ring, light-blue = hydrogen bond donors blue sphere/circle = positive ionizable. Note: Excluded volumes are not shown. *Denotes the "Must Match" site.

**Enrichments.** The enrichment results for all 30 targets using the e-pharmacophore method are compared to c-pharmacophores, docking, and 2D similarity in Figure 5 and discussed in detail below. Figure 5a shows the average enrichment factor in the top 1% of the database for each method. In general, the e-pharmacophores developed in this

work outperform the other methods, with an average EF(1%) of 40.9 compared to 34.5, 27.0, and 36.0 for 2D fingerprints, c-pharmacophores, and Glide SP, respectively.

While the e-pharmacophore approach outperforms the other methods on average, the standard deviation of the averages (Figure 5b) indicates a large variability in the computed
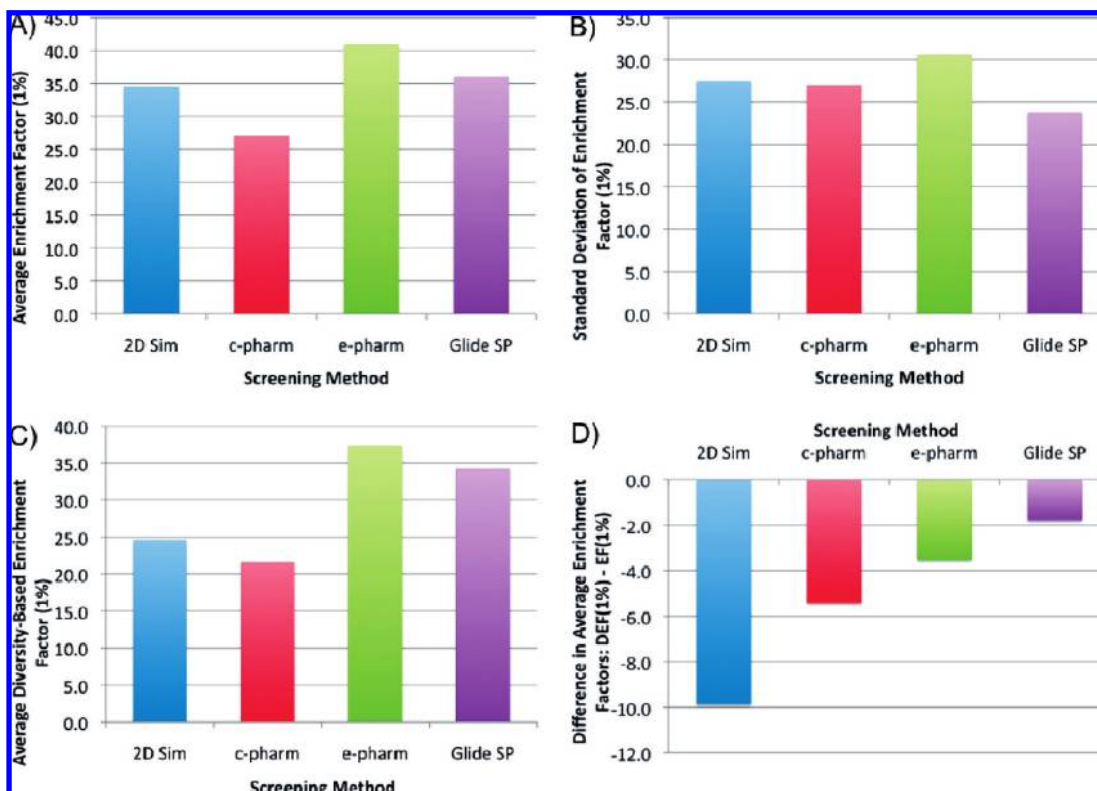
**Figure 5.** Comparison of 2D linear fingerprint searching (blue), c-pharmacophores (red), e-pharmacophores (green), and Glide SP docking (purple). **A)** Average enrichment factor for top 1% (EF(1%)) of database. **B)** Standard deviation of average EF(1%). **C)** Diversity-base enrichment factor (DEF) of the top 1% of database. **D)** Difference between the EF(1%) and the DEF(1%). (For further details see S2 of the Supporting Information.)

enrichments. For example, e-pharmacophores produce the best average enrichment and the most cases with EF(1%) greater than 50 (11/30), but Glide SP has more cases with EF(1%) greater than 10 (28/30 vs. 26/30); see Table 4. The fact that e-pharmacophores produce the most number of cases with very high enrichments (i.e., EF(1%) $\geq$ 50) where as Glide SP produces the least number of cases with poor enrichment (i.e., EF(1%) $\leq$ 10) suggests that combining the various methods could yield more consistent results than either of the two methods alone, although not necessarily better on average.

Relatively poor enrichments are observed for the three aldose reductase structures (1ah3, 1ef3, and 2acq), as seen in Figure 6 (results for each target are included in Supporting Information S1). In particular, the hypothesis generated from 2acq yields the worst enrichment out of the 30 systems, with an EF(1%) of 2.8. The 2acq ligand is comprised of a glucose molecule phosphorylated at carbon 6 (29, glucose-6-phosphate), and the energy-optimized hypothesis only defines pharmacophore sites on the phosphoric acid. While the acidic bioisostere has been established as an important characteristic of aldose reductase inhibitors,[44] the pharmacophore does not have any other features to define specificity and therefore cannot distinguish aldose reductase binders from any other acid-containing compound. This problem arises because the method described here relies upon a ligand that contains enough pharmacophoric information to generate a hypothesis that can retrieve other active molecules. Glucose-6-phosphate (29), being a substrate of aldose reductase,[45] is not a tight binder and would not be considered druglike by standard criteria. Therefore, it is of little surprise that a suboptimal hypothesis is derived, and poor enrichment results are

**Table 3.** Native RMSD, Enrichment Factor at 1% of the Decoy Database Screened, and BEDROC[37] at $\alpha = 20.0$ and 160.9

| reference ligand and PDB | rmsd | EF(1%) | BEDROC $\alpha = 20.0$ | BEDROC $\alpha = 160.9$ |
|---|---|---|---|---|
| **1**, 1a9u | 0.8 | 35.0 | 0.48 | 0.65 |
| **2**, 1ah3 | 0.3 | 8.3 | 0.33 | 0.31 |
| **3**, 1aq1 | 0.3 | 25.0 | 0.39 | 0.45 |
| **4**, 1bji | 0.5 | 35.0 | 0.43 | 0.61 |
| **5**, 1bl7 | 1.6 | 62.1 | 0.73 | 0.93 |
| **6**, 1cx2 | 1.7 | 30.3 | 0.44 | 0.68 |
| **7**, 1ddu | 3.1 | 65.0 | 0.68 | 0.86 |
| **8**, 1dm2 | 0.9 | 10.0 | 0.25 | 0.25 |
| **9**, 1dwc | 2.1 | 12.5 | 0.18 | 0.30 |
| **10**, 1e66 | 0.6 | 8.0 | 0.17 | 0.28 |
| **11**, 1ef3 | 1.0 | 11.1 | 0.15 | 0.42 |
| **12**, 1err | 0.7 | 80.0 | 0.86 | 0.69 |
| **13**, 1ett | 1.1 | 6.3 | 0.16 | 0.16 |
| **14**, 1eve | 2.6 | 56.0 | 0.60 | 0.75 |
| **15**, 1ezf | 0.9 | 80.0 | 0.87 | 0.95 |
| **16**, 1fjs | 6.9 | 25.0 | 0.28 | 0.44 |
| **17**, 1g49 | 0.8 | 25.0 | 0.32 | 0.45 |
| **18**, 1hov | 2.4 | 20.0 | 0.35 | 0.42 |
| **19**, 1hpx | 3.4 | 10.0 | 0.20 | 0.10 |
| **20**, 1kim | 0.7 | 100.0 | 1.00 | 1.00 |
| **21**, 1kv2 | 1.3 | 20.0 | 0.23 | 0.31 |
| **22**, 1m17 | 1.4 | 100.0 | 1.00 | 1.00 |
| **23**, 1mwe | 0.8 | 75.0 | 0.78 | 0.87 |
| **24**, 1qpe | 0.7 | 65.0 | 0.65 | 0.52 |
| **25**, 1rt1 | 1.1 | 20.0 | 0.30 | 0.50 |
| **26**, 1syn | 2.6 | 45.0 | 0.56 | 0.73 |
| **27**, 1tmn | 2.7 | 80.0 | 0.89 | 0.55 |
| **28**, 1vrt | 0.6 | 30.0 | 0.50 | 0.55 |
| **29**, 2acq | 3.2 | 2.8 | 0.04 | 0.14 |
| **30**, 3ert | 1.9 | 85.0 | 0.95 | 0.93 |
| **Mean** | 1.6 | 40.9 | 0.49 | 0.56 |
| **Median** | 1.1 | 30.2 | 0.43 | 0.53 |

**Table 4.** Fraction of Cases with EF(1%) Better than 10 and 50

| screening method | % ≥ 10 | % ≥ 50 |
|---|---|---|
| 2D Similarity | 80.0 | 23.3 |
| c-Pharmacophore | 73.3 | 20.0 |
| e-Pharmacophore | 86.7 | 36.7 |
| Glide SP | 93.3 | 30.0 |

obtained using this compound to derive the e-pharmacophore. In comparison, aldose reductase e-pharmacophores derived from tolrestat (1ah3, (2); $IC_{50} = 35-50$ nM)[46] and fidarestat (1ef3, (11); $IC_{50} = 9$ nM)[47] give better enrichment results (EF(1%) of 8.3 and 11.3, respectively). Both of these hypotheses include hydrophobic and aromatic features in addition to an acid, giving them more specificity than the hypothesis from 2acq.

In addition to the three aldose reductase crystal structures described above from our original set, a new hypothesis derived from IDD594 (2i16, (31); IC50 = 0.03 $\mu$M)[44] produces a significantly improved EF(1%) of 25.0 (Figure 6). The hypothesis has many of the features from tolrestat (2) and fidarestat (11), which have slightly different pharmacophore features. For example, fidarestat (11) lacks the negative ionizable feature that defines the carboxylic acid from glucose-6-phosphate, tolrestat (2), and many other aldose reductase inhibitors. Instead, fidarestat (11) contains two hydrogen bond acceptors defined by two carbonyls of the spirohydantoin ring, which acts in lieu of a carboxylic acid. On the other hand, tolrestat (2) lacks the aromatic and hydrophobic groups defining the fluorobenzene of fidarestat (11) and other aldose reductase inhibitors.[47] The hypothesis from IDD549 (31) contains these features in common with fidarestat (11), tolrestat (2), and glucose-6-phosphate (29) and, therefore, may justify the better enrichment. Thus, selecting a reference ligand that contains more of the relevant features and incorporating structure−activity relationship considerations can aid in choosing the crystal structure that will lead to higher enrichments.

**RMSDs.** A major advantage of using a 3D structure-based method over a 2D method is that it can present insight into the binding mode for an active molecule. As seen in Table
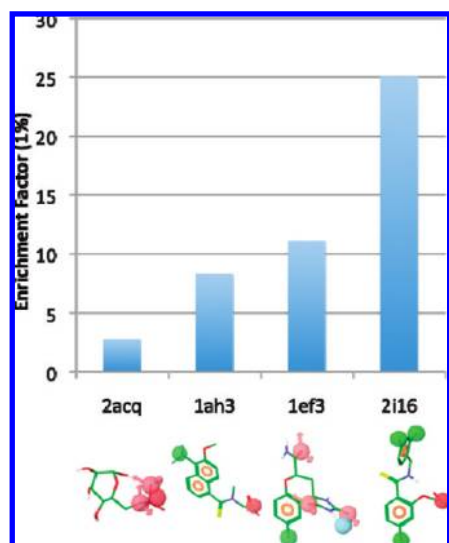


**Figure 6.** Database enrichments for aldose reductase structures: 2acq, 1ah3, 1ef3, and 2i16. The latter was not part of the original structures studies in this work.

3, the rmsd to the X-ray structure for the native ligand aligned to the pharmacophore is generally good, with an average of 1.6 Å. Furthermore, 70% of the cases have an rmsd less than 2.0 Å and 87% of the cases are less than 3.0 Å. In some cases where the alignments were poor (greater than 3.0 Å) there were either solvent exposed groups not involved in direct interactions with the protein (i.e., HIV-1 protease, 1hpx) or a ligand with some degree of pharmacophoric symmetry that could adopt a flipped binding mode (i.e., factor Xa, 1fjs). The relatively high number of cases with low RMSDs show that while the energy-optimized hypotheses have a greatly reduced number of features, they have the features necessary to generate good alignments. Figure 7 shows examples of top active hits aligned with the X-ray ligand.

**Effect of Excluded Volumes.** Excluded volumes were added to the hypotheses for a number of reasons. First, queries based on relatively simple pharmacophore hypothesis (that contain three or four features) and no excluded volume spheres yield very large hit lists.[48,49] Addition of excluded volume spheres are a proven means of reducing these numbers.[50] Thus, we expect the inclusion of receptor-based excluded volumes to help reduce false positives by eliminating inactive compounds that cannot simultaneously match the hypothesis and avoid clashing with the receptor. Indeed, across the entire set of 30 targets, we find a drop in the average enrichment EF(1%) from 40.9 to 34.3 when excluded volumes are not used (Figure 8). However, the average results are not sufficient to conclude that the uses of receptor-based excluded volumes will always improve enrichments. Only 16 of the 30 cases show improvements with excluded volumes, while the remaining 14 show degradation in enrichments (Figure 8). This suggests special consideration needs to be taken when implementing receptor-based excluded volumes, as discussed below.

An important factor related to the use of receptor-based excluded volumes is the plasticity of the receptor in question and the potential induced-fit effects that the active ligands have upon binding. For example, the use of excluded volumes may degrade results in binding sites where protein flexibility is needed to accommodate the majority of active ligands. Of the targets that show significant degradations in enrichments when using receptor-based excluded volumes, HIV-1 RT, Factor Xa, CDK2, and aldose reductase have been shown to produce better cross-docking results when flexibility is included.[27] While our method attempts to model moderate receptor flexibility through the use of radii scaling and ignoring receptor atoms close to the ligand, the data presented in Figure 8 suggest that the method for determining excluded volumes may require improvements when applied to flexible binding sites.

**Diversity.** In pharmaceutical drug discovery it is not only important to find active compounds but also to find diverse compounds. This increases the probabilities of avoiding existing patents and provides more candidates with different properties to consider for lead optimization. To assess both the enrichment and diversity of active molecules retrieved, we developed a metric we call diversity-based enrichment factor (DEF; see Methods). The purpose of the DEF is to combine, in a single metric, both the fraction of actives recovered and the structural diversity of those actives molecules. Figure 5c compares the average DEF performance
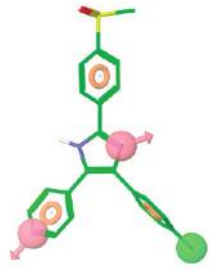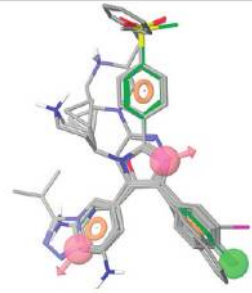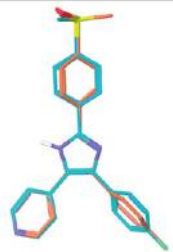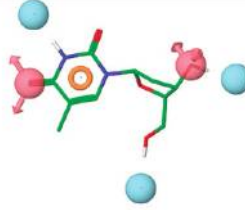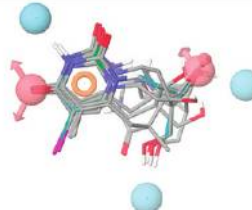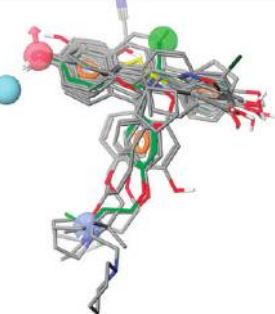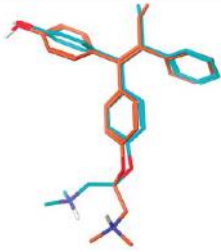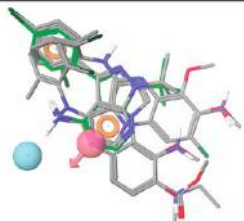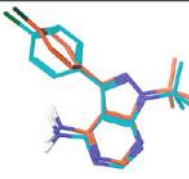
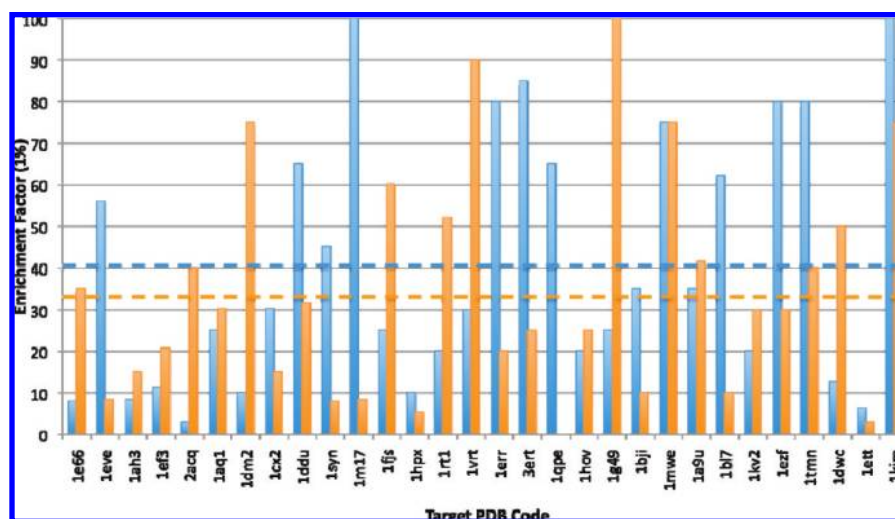| Reference Ligand and PDB | Pharmacophore Hypothesis | Active Hits in Top 1% | Alignment pose (cyan) with crystal pose (orange) |
|---|---|---|---|
| 1, 1a9u | | | RMSD = 0.85 Å |
| 20, 1kim | | | RMSD = 0.69 Å |
| 30, 3ert | | | RMSD = 1.39 Å |
| 24, 1qpe | | | RMSD = 0.67 Å |

**Figure 7.** Example of alignments to hypotheses.



**Figure 8.** Enrichment factors for top 1% of database using energy-optimized pharmacophores (e-pharmacophore), with (blue) and without (orange) receptor-based excluded volumes. Average EF(1%) are shown as dashed lines.

of the e-pharmacophore approach (green) to c-pharmacophores (red), Glide SP docking (purple), and 2D linear fingerprint (blue). The e-pharmacophore approach produces better DEF results than any of the other three methods. Comparing the DEF(1%) to the EF(1%), the pharmacophore method is penalized by 3.5 units for not retrieving the full diversity of actives (Figure 5d). This can be compared to penalties in the other methods of 9.9, 5.4, and 1.8 for the 2D fingerprint, c-pharmacophore, and docking approaches. The docking method is impacted the least when accounting for the diversity of actives retrieved, which is in agreement with other studies that have shown docking to be good at retrieving diverse compounds.[11] On the other hand, as might be expected, the 2D fingerprint similarity method retrieves the most similar compounds and is penalized the most by the DEF.

CONCLUSION

A novel method for generating energetically optimized, structure-based pharmacophores has been presented. The method, referred to here as generating e-pharmacophores, unites the benefits of two computational strategies: the computational efficiency of ligand-based pharmacophore screening and the accuracy of scoring from structure-based docking. By decomposing the energetics from Glide XP on an atomistic level, it was possible to assign a score to each pharmacophore site in order to prioritize the importance of sites predicted to be responsible for binding. The top scoring sites based on energetic ranking were used to generate a pharmacophore hypothesis that was then used for screening a database.

The e-pharmacophore methodology was applied to 30 crystal structures. *In silico* screening showed this method produced higher enrichments of active molecules than three other virtual screening approaches: contact-based pharmacophores (c-pharmacophores), Glide SP docking, and 2D ligand similarity. Furthermore, use of this method resulted in retrieving more diverse compounds than either the c-pharmacophore or 2D similarity approaches, which is a highly desirable characteristic in virtual screening. While the studies performed here were run with ligands from cocrystal structures, the general methodology is applicable to ligand poses from any source. For example, e-pharmacophores can be generated in cases where experimental information about ligand structure is unavailable, not suitable for pharmacophore development (i.e., non druglike), or where it is desirable to retrieve diverse compounds that are not biased by the cocrystallized ligand. It is possible that the best results will be obtained with an e-pharmacophore generated from docking a potent (or perhaps efficient) ligand for a given target, although that was not shown here and will be the focus of future work. A related study was recently published on the application of the e-pharmacophore methodology to develop receptor-based pharmacophores derived from docked fragments.[51] The strategy was shown to produce viable hypotheses and good database enrichments and retrieve a diverse set of hits.

While the database enrichment results presented in this work are good, there are still improvements that can be made. For example, the use of receptor-based excluded volumes improved results in some cases but not in others. This highlights one of the most important challenges in computational drug design: adequately accounting for receptor flexibility. A potential avenue for future work is the development of a procedure to determine when excluded volumes should or should not be used. Furthermore, it may be possible to develop rules for the generation of excluded volumes to allow for more overlap in cases of flexible receptors. Another important consideration is the choice of crystal structure when multiple structures are available. In this work and others[32,52] it was shown that different crystal structures of the same target can produce significantly different results. It will be important to develop automated methods for characterizing multiple receptors to identify the best structures for screening. In many cases, the best results may be obtained by considering a structural ensemble.[32]

In the work presented here we used the default pharmacophore Fitness scoring function in Phase,[15] which includes rmsd site matching, vector alignment, and volume overlap. While the good average database enrichments produced in this work are a sign that the scoring function is accurately characterizing important parts of the molecules, there may be additional terms that could further improve the scoring. For example, the ligand strain energy incurred upon binding was not explicitly accounted for in the scoring function, although the generation of conformations for the database was limited to only energetically reasonable states. Still, if one could accurately estimate the conformational cost associated with binding (both enthalpic strain end entropic restriction) it might be possible to improve the results. Another contribution missing from the scoring function is the ligand desolvation penalty. While it is normally impractical to estimate this term in a ligand-based method due to the lack of information about the bound state, in the work presented here the receptor structure is known, and therefore it should be possible to estimate a desolvation penalty for each aligned ligand using an implicit solvent model or an empirical function. To our knowledge such an approach has not previously been published; however, the ability to penalize buried polar or charged groups should improve the results by eliminating false positives. One difficulty with including the ligand strain and desolvation penalty is that these are both energy-based terms, whereas the Phase Fitness score is a unitless quantity based on a sum of different terms related to the alignment. It would require some degree of fitting to determine an adequate way to combine the energy-based terms with the Phase Fitness score. This is being explored and will be the focus of future work.

Overall, the e-pharmacophore method presented here produces good database enrichments, high diversity of retrieved actives, and fast database screening. Furthermore, the automated rules for generating the e-pharmacophores remove the need for human decisions in determining the pharmacophore hypothesis, although user information can be inserted into the process by manual inclusion/removal of features or changing the features that are required to match. Such information may be derived through experimental means or from other computational methods. The e-pharmacophore method represents an advance in the state of the art for generating pharmacophores by including energetic terms derived from a structure-based scoring function and represents an attractive alternative or complementary approach to other virtual screening techniques.

Energetically Optimized Pharmacophore Modeling

*J. Chem. Inf. Model.*, Vol. 49, No. 10, 2009 **2367**

REFERENCES AND NOTES

(1) Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813–8.

(2) Bembenek, S. D.; Keith, J. M.; Letavic, M. A.; Apodaca, R.; Barbier, A. J.; Dvorak, L.; Aluisio, L.; Miller, K. L.; Lovenberg, T. W.; Carruthers, N. I. Lead identification of acetylcholinesterase inhibitors-histamine H3 receptor antagonists from molecular modeling. *Bioorg. Med. Chem.* **2008**, *16*, 2968–73.

(3) Cheng, J. F.; Zapf, J.; Takedomi, K.; Fukushima, C.; Ogiku, T.; Zhang, S. H.; Yang, G.; Sakurai, N.; Barbosa, M.; Jack, R.; Xu, K. Combination of virtual screening and high throughput gene profiling for identification of novel liver X receptor modulators. *J. Med. Chem.* **2008**, *51*, 2057–61.

(4) Siddiquee, K.; Zhang, S.; Guida, W. C.; Blaskovich, M. A.; Greedy, B.; Lawrence, H. R.; Yip, M. L.; Jove, R.; McLaughlin, M. M.; Lawrence, N. J.; Sebti, S. M.; Turkson, J. Selective chemical probe inhibitor of Stat3, identified through structure-based virtual screening, induces antitumor activity. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 7391–6.

(5) Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862–5.

(6) Dixon, S. L.; Smondyrev, A. M.; Rao, S. N. PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem. Biol. Drug Des.* **2006**, *67*, 370–2.

(7) Jacobsson, M.; Garedal, M.; Schultz, J.; Karlen, A. Identification of Plasmodium falciparum spermidine synthase active site binders through structure-based virtual screening. *J. Med. Chem.* **2008**, *51*, 2777–86.

(8) Putta, S.; Lemmen, C.; Beroza, P.; Greene, J. A novel shape-feature based approach to virtual library screening. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1230–40.

(9) Guner, O. F. History and evolution of the pharmacophore concept in computer-aided drug design. *Curr. Top. Med. Chem.* **2002**, *2*, 1321–32.

(10) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Similarity searching of chemical databases using atom environment descriptors (MOL-PRINT 2D): evaluation of performance. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1708–18.

(11) McGaughey, G. B.; Sheridan, R. P.; Bayly, C. I.; Culberson, J. C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J. F.; Cornell, W. D. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* **2007**, *47*, 1504–19.

(12) Wang, H.; Duffy, R. A.; Boykow, G. C.; Chackalamannil, S.; Madison, V. S. Identification of novel cannabinoid CB1 receptor antagonists by using virtual screening with a pharmacophore model. *J. Med. Chem.* **2008**, *51*, 2439–46.

(13) Schuster, D.; Nashev, L. G.; Kirchmair, J.; Laggner, C.; Wolber, G.; Langer, T.; Odermatt, A. Discovery of nonsteroidal 17beta-hydroxysteroid dehydrogenase 1 inhibitors by pharmacophore-based screening of virtual compound libraries. *J. Med. Chem.* **2008**, *51*, 4188–99.

(14) Neves, M. A.; Dinis, T. C.; Colombo, G.; Sa e Melo, M. L. Fast three dimensional pharmacophore virtual screening of new potent non-steroid aromatase inhibitors. *J. Med. Chem.* **2009**, *52*, 143–50.

(15) Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database searching: 1. Methodology and preliminary results. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 647–71.

(16) Guner, O.; Clement, O.; Kurogi, Y. Pharmacophore modeling and three dimensional database searching for drug design using catalyst: recent advances. *Curr. Med. Chem.* **2004**, *11*, 2991–3005.

(17) Brown, D. A.; Kharkar, P. S.; Parrington, I.; Reith, M. E.; Dutta, A. K. Structurally constrained hybrid derivatives containing octahydroben-zo[g or f]quinoline moieties for dopamine D2 and D3 receptors: binding characterization at D2/D3 receptors and elucidation of a pharmacophore model. *J. Med. Chem.* **2008**, *51*, 7806–19.

(18) Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **2005**, *45*, 160–9.

(19) Ortuso, F.; Langer, T.; Alcaro, S. GBPM: GRID-based pharmacophore model: concept and application studies to protein-protein recognition. *Bioinformatics* **2006**, *22*, 1449–55.

(20) Chen, J.; Lai, L. Pocket v.2: further developments on receptor-based pharmacophore modeling. *J. Chem. Inf. Model.* **2006**, *46*, 2684–91.

(21) Yang, H.; Shen, Y.; Chen, J.; Jiang, Q.; Leng, Y.; Shen, J. Structure-based virtual screening for identification of novel 11beta-HSD1 inhibitors. *Eur. J. Med. Chem.* **2008**.

(22) Oshiro, C. M.; Kuntz, I. D.; Dixon, J. S. Flexible ligand docking using a genetic algorithm. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 113–30.

(23) Rarey, M. Protein-ligand docking in drug design. In *Bioinformatics - From Genomes to Drugs*; Wiley-VCH: Heidelberg, 2001; Vol. I, pp 315−360.

(24) 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*, 727–48.

(25) Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46*, 499–511.

(26) 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*, 1739–49.

(27) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534–53.

(28) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337*, 209–25.

(29) Mizutani, M. Y.; Takamatsu, Y.; Ichinose, T.; Nakamura, K.; Itai, A. Effective handling of induced-fit motion in flexible docking. *Proteins* **2006**, *63*, 878–91.

(30) Moitessier, N.; Therrien, E.; Hanessian, S. A method for induced-fit docking, scoring, and ranking of flexible ligands. Application to peptidic and pseudopeptidic beta-secretase (BACE 1) inhibitors. *J. Med. Chem.* **2006**, *49*, 5885–94.

(31) Nabuurs, S. B.; Wagener, M.; de Vlieg, J. A flexible approach to induced fit docking. *J. Med. Chem.* **2007**, *50*, 6507–18.

(32) Rao, S.; Sanschagrin, P. C.; Greenwood, J. R.; Repasky, M. P.; Sherman, W.; Farid, R. Improving database enrichment through ensemble docking. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 621–7.

(33) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–9.

(34) 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*, 6177–96.

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

(36) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–42.

(37) 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.

(38) Shemetulskis, N. E.; Dunbar, J. B., Jr.; Dunbar, B. W.; Moreland, D. W.; Humblet, C. Enhancing the diversity of a corporate database using chemical database clustering and analysis. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 407–16.

(39) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P.; Bertrand, H. O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J. Med. Chem.* **2005**, *48*, 2534–47.

(40) Wang, Z.; Canagarajah, B. J.; Boehm, J. C.; Kassisa, S.; Cobb, M. H.; Young, P. R.; Abdel-Meguid, S.; Adams, J. L.; Goldsmith, E. J. Structural basis of inhibitor selectivity in MAP kinases. *Structure* **1998**, *6*, 1117–28.

(41) Hopkins, A. L.; Ren, J.; Esnouf, R. M.; Willcox, B. E.; Jones, E. Y.; Ross, C.; Miyasaka, T.; Walker, R. T.; Tanaka, H.; Stammers, D. K.; Stuart, D. I. Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT series reveal conformational changes relevant to the

design of potent non-nucleoside inhibitors. *J. Med. Chem.* **1996**, *39*, 1589–600.

(42) Tan, L.; Lounkine, E.; Bajorath, J. Similarity searching using fingerprints of molecular fragments involved in protein-ligand interactions. *J. Chem. Inf. Model.* **2008**, *48*, 2308–12.

(43) Kryger, G.; Silman, I.; Sussman, J. L. Structure of acetylcholinesterase complexed with E2020 (Aricept): implications for the design of new anti-Alzheimer drugs. *Structure* **1999**, *7*, 297–307.

(44) Howard, E. I.; Sanishvili, R.; Cachau, R. E.; Mitschler, A.; Chevrier, B.; Barth, P.; Lamour, V.; Van Zandt, M.; Sibley, E.; Bon, C.; Moras, D.; Schneider, T. R.; Joachimiak, A.; Podjarny, A. Ultrahigh resolution drug design I: details of interactions in human aldose reductase-inhibitor complex at 0.66 A. *Proteins* **2004**, *55*, 792–804.

(45) Harrison, D. H.; Bohren, K. M.; Ringe, D.; Petsko, G. A.; Gabbay, K. H. An anion binding site in human aldose reductase: mechanistic implications for the binding of citrate, cacodylate, and glucose 6-phosphate. *Biochemistry* **1994**, *33*, 2011–20.

(46) Sestanj, K.; Bellini, F.; Fung, S.; Abraham, N.; Treasurywala, A.; Humber, L.; Simard-Duquesne, N.; Dvornik, D. N-[5-(trifluoromethyl)-6-methoxy-1-naphthalenyl]thioxomethyl]- N-methylglycine (Tolrestat), a potent, orally active aldose reductase inhibitor. *J. Med. Chem.* **1984**, *27*, 255–6.

(47) Oka, M.; Matsumoto, Y.; Sugiyama, S.; Tsuruta, N.; Matsushima, M. A potent aldose reductase inhibitor, (2S,4S)-6-fluoro-2′, 5′-dioxospiro[chroman-4,4′-imidazolidine]-2-carboxamide (Fidarestat): its absolute configuration and interactions with the aldose reductase by X-ray crystallography. *J. Med. Chem.* **2000**, *43*, 2479–83.

(48) Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* **2000**, *43*, 2100–14.

(49) Steindl, T.; Langer, T. Influenza virus neuraminidase inhibitors: generation and comparison of structure-based and common feature pharmacophore hypotheses and their application in virtual screening. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1849–56.

(50) Greenidge, P. A.; Carlsson, B.; Bladh, L. G.; Gillner, M. Pharmacophores incorporating numerous excluded volumes defined by X-ray crystallographic structure in three-dimensional database searching: application to the thyroid hormone receptor. *J. Med. Chem.* **1998**, *41*, 2503–12.

(51) Loving, K.; Salam, N. K.; Sherman, W. Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. *J. Comput.-Aided Mol. Des.* **2009**.

(52) Sheridan, R. P.; McGaughey, G. B.; Cornell, W. D. Multiple protein structures and multiple ligands: effects on the apparent goodness of virtual screening results. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 257–65.