

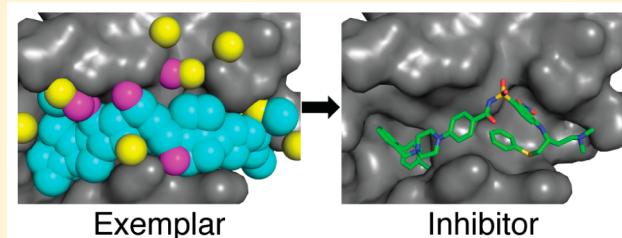
Ultra-High-Throughput Structure-Based Virtual Screening for Small-Molecule Inhibitors of Protein–Protein Interactions

David K. Johnson[†] and John Karanicolas^{*,†,‡}

[†]Center for Computational Biology, and [‡]Department of Molecular Biosciences, University of Kansas, 2030 Becker Drive, Lawrence, Kansas 66045-7534, United States

 Supporting Information

ABSTRACT: Protein–protein interactions play important roles in virtually all cellular processes, making them enticing targets for modulation by small-molecule therapeutics: specific examples have been well validated in diseases ranging from cancer and autoimmune disorders, to bacterial and viral infections. Despite several notable successes, however, overall these remain a very challenging target class. Protein interaction sites are especially challenging for computational approaches, because the target protein surface often undergoes a conformational change to enable ligand binding: this confounds traditional approaches for virtual screening. Through previous studies, we demonstrated that biased “pocket optimization” simulations could be used to build collections of low-energy pocket-containing conformations, starting from an unbound protein structure. Here, we demonstrate that these pockets can further be used to identify ligands that complement the protein surface. To do so, we first build from a given pocket its “exemplar”: a perfect, but nonphysical, pseudoligand that would optimally match the shape and chemical features of the pocket. In our previous studies, we used these exemplars to quantitatively compare protein surface pockets to one another. Here, we now introduce this exemplar as a template for pharmacophore-based screening of chemical libraries. Through a series of benchmark experiments, we demonstrate that this approach exhibits comparable performance as traditional docking methods for identifying known inhibitors acting at protein interaction sites. However, because this approach is predicated on ligand/exemplar overlays, and thus does not require explicit calculation of protein–ligand interactions, exemplar screening provides a tremendous speed advantage over docking: 6 million compounds can be screened in about 15 min on a single 16-core, dual-GPU computer. The extreme speed at which large compound libraries can be traversed easily enables screening against a “pocket-optimized” ensemble of protein conformations, which in turn facilitates identification of more diverse classes of active compounds for a given protein target.



INTRODUCTION

The concept of a pharmacophore dates back at least a century: it is traditionally attributed to Paul Ehrlich, who recognized that certain parts of molecules were responsible for their biological activity.¹ This concept was modernized 50 years later, shifting away from chemical groups and toward a more abstract notion of chemical forces in three-dimensional space.² The IUPAC now defines a pharmacophore as “the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response”.³

Pharmacophores enable design of small molecules capable of presenting specific functional moieties to elicit a desired biological response, and for decades they have been used to inspire medicinal chemists’ development of new analogues.^{4–6} Because they describe the spatial arrangement of critical interactions with a receptor, pharmacophores can also be used as templates for computational screens seeking to identify ligands containing functional groups positioned to recapitulate these interactions.

The first computed example of a modern pharmacophore is attributed to Lemont Kier, who recognized the spatial similarity of (modeled) three-dimensional geometries of various muscarinic receptor agonists.⁷ Presently, a broad assortment of computational tools can be used to define pharmacophores in distinct ways.^{8–16} The first pharmacophore-building algorithms drew information from the ligand alone: such approaches begin by finding a consensus structural alignment of multiple active compounds and, then, seek to identify shared functional groups in this set.¹¹ More recently, development of tools such as LigandScout¹⁶ allow key interactions to instead be defined from one or more crystal structures of a receptor with assorted ligands bound—here again, identifying features shared by multiple ligands to build a consensus pharmacophore.

More recent efforts have focused on building pharmacophore models from protein structures alone, solved without any bound ligand in the active site. These approaches typically begin by docking an assortment of small (chemically diverse)

Received: September 15, 2015

probe molecules into the active site, then evaluating the interactions with the protein that these probes make.^{9,12,15} Individual interactions presented by different probe molecules are then combined into a consensus pharmacophore and used as a template to identify larger compounds that simultaneously recapitulate the interactions from multiple probes. As an alternative, other approaches instead define desirable three-dimensional properties of candidate ligands using the “negative image” of the binding pocket.^{10,13}

Pharmacophores have been applied extensively to many diverse targets, including enzymes,^{17–20} G protein-coupled receptors,^{21–23} and transporters.^{24–26} In each of these cases, the protein target has evolved to bind some natural small-molecule partner: already this suggests that the chemical space of prospective hits may be similar to that of the natural binding partner(s).²⁷ In most such cases one or more of the natural ligand(s) are known, so the task that remains entails identifying alternate molecules that recapitulate the key interactions of these natural ligand(s).

In some cases, however, an important biological target is not evolved to bind *any* natural small-molecule ligand: these include protein–protein interactions, protein–RNA interactions, and others. Moreover, representatives from this target class include well-validated targets for cancer, viral and bacterial interactions, and autoimmune disorders.^{28–34} Given the lack of a natural small-molecule binding partner, it is not evident *a priori* in which regions of chemical space prospective inhibitors may reside—or even whether the protein surface is indeed druggable (or rather, “ligandable”³⁵) with *any* small molecule at all.³⁶ Until one or more small-molecule inhibitors have been identified, there are (of course) no active compounds to use as templates for building a pharmacophore model. As a further additional challenge, crystal structures of inhibitors bound to protein–protein inhibitors have shown that in some cases the binding pockets on the protein surface are transient, appearing neither in the unbound structure nor in that of the protein-bound complex.³⁷

Despite these challenges, pharmacophore models have still successfully been applied to protein–protein interactions, in certain cases. Most notably, several groups have recently developed related computational approaches that select some of the side chains from the natural protein partner (the “hotspot” or “anchor” residues) and, then, use this collection of functional groups to build a pharmacophore.^{38–45} However, this approach comes with several important limitations. First, the inhibitors designed to match such pharmacophores may be limited in the chemical space they sample, given that they are built from mimicry of canonical amino acid side chains. Second, their binding affinity may also be limited, given that they make use of only a subset of the natural binding partner’s interactions. Finally, this approach cannot take advantage of conformational changes at the protein surface that might allow binding of small molecules that do not resemble the natural binding partner; as noted earlier, such conformational changes are frequently observed in crystal structures of inhibitors bound at protein interaction sites.³⁷

In our previous work, we developed a biasing potential that can be used to create ensembles of low-energy pocket-containing conformations, starting from an unbound protein structure that lacks a surface pocket suitable for small-molecule binding.⁴⁶ Others have shown that docking-based virtual screens can be improved by docking to an ensemble of receptor conformations, rather than to a single static protein

structure.^{47–55} Thus, we surmised that the inclusion of many surface pockets—derived from “pocket optimization” biased simulations⁴⁶—would facilitate identification of inhibitors beyond those that simply mimic the natural binding partner.

In order to quantitatively compare the shape and chemical features between pairs of surface pockets, we then developed the concept of pocket “exemplars” and used them to compare pockets to other pockets in order to predict the selectivity of active compounds between related protein family members.⁵⁶ Briefly, an exemplar is built from a protein structure by determining the locations of hydrogen bond donors/acceptors within the pocket that would form idealized interactions with the receptor, then filling the remainder of the pocket with a series of spheres that simply occupy volume. In essence, an exemplar is the negative image of the protein pocket: it represents the “ideal” ligand that would complement a given pocket, albeit in the absence of physical chemical constraints (i.e., the bond connectivity is neither specified nor implied).

By representing pockets through their “idealized” complementary ligands, we found that we could make accurate comparisons of pocket similarity.⁵⁶ Here, we extend this approach further: instead of simply comparing exemplars to other exemplars, as done in previous work, we demonstrate that these exemplars can also be used directly as templates for very rapid pharmacophore-based virtual screening. The speed also enables us to screen against ensembles of conformations from the “pocket optimized” simulations, which had previously only been used to assess druggability and selectivity. Because exemplars are built directly from the protein structure, this approach is particularly well suited for regimes in which no natural binding partner for the protein target is available, or, as in the case of protein–protein interactions, the natural binding partner does not provide a suitable template from which to build a conventional pharmacophore.

■ COMPUTATIONAL APPROACH

Conformational sampling of the protein and exemplar-generation are implemented in the Rosetta software suite;⁵⁷ Rosetta is freely available for academic use (www.rosettacommons.org). Virtual screening using the exemplar was carried out using the FastROCS^{58,59} and ROCS^{59,60} software. The relevant command lines used to run both Rosetta and FastROCS/ROCS are included as *Supplementary Information*.

In addition to the command-line program included in Rosetta, we have provided a tool to generate an exemplar from an input PDB file through the ROSIE web interface.⁶¹ This tool is available at: http://rosie.rosettacommons.org/make_exemplar.

Our approach for screening a chemical library using exemplars is summarized schematically in Figure 1; each step is described in detail below.

Using Pocket Optimization to Build Ensembles of Pocket-Containing Conformations. To generate ensembles of pocket containing conformations, we use a Monte Carlo sampling technique implemented in the Rosetta software suite;⁵⁷ this local move set was originally designed for refining templates in comparative modeling applications⁶² and includes both side chain and backbone moves among its degrees of freedom. In order to enrich our sampling with low-energy conformations that include surface pockets, we include a biasing term in the energy function that favors conformations in which a surface pocket is present.⁴⁶ We have previously shown

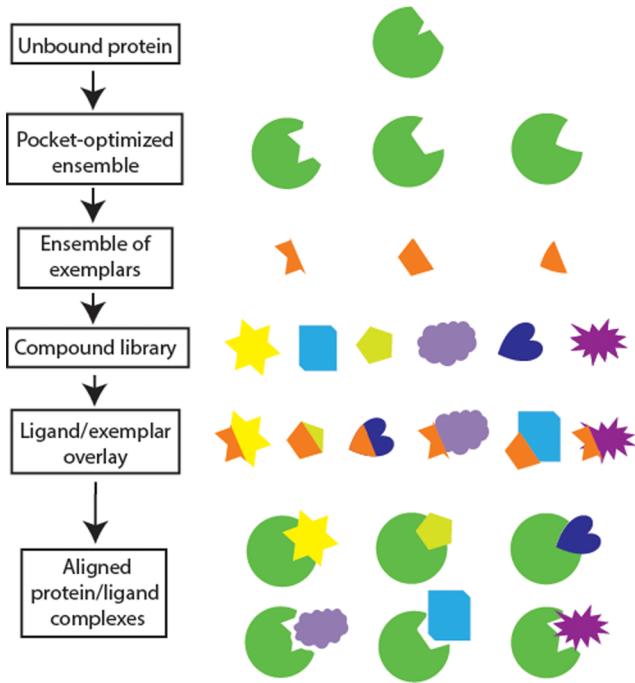


Figure 1. Overview of the “exemplar screening” pipeline. The protein-interaction site of an unbound protein often lacks a suitable surface pocket for small-molecule binding. Through a series of biased pocket optimization simulations, a collection of low-energy pocket-containing conformations of the target protein is assembled. Each of these conformations is then used to build a unique exemplar: a map of the perfect ligand that would complement this surface pocket. The exemplars described in this study incorporate both the pocket shape and the locations of hydrogen bonding groups; the latter are not included in this schematic. Each member of a compound library is then rapidly overlaid with each exemplar, and compounds are ranked on the basis of how well they recapitulate the properties of their preferred (most-similar) exemplar. Because the exemplars are derived from the shape of the protein surface pocket, alignment of ligands to exemplars then allows trivial reconstruction of the protein–ligand complex, using the protein conformation from which the exemplar was derived.

that this approach creates ensembles of pocket-containing conformations with energies that overlap those from the corresponding unbiased simulations, implying low energetic cost to open the particular pockets we observe.^{46,56} For each protein of interest we carried out biased simulations, collected the final output structure from each of 1000 independent trajectories, and retained only those with energies comparable to those observed in analogous unbiased simulations.^{46,56}

In order to focus the resulting pockets on the intended surface (the protein interaction site, in this case), the energetic “bonus” associated with pocket formation is awarded only if the pocket is in direct contact with one or two predetermined “target” residues. For this study we used an automated tool to determine the target residues,⁵⁰ provided that a crystal structure of the protein–protein complex was available. In one case (Brd4) the crystal structure of the protein–protein complex included noncanonical amino acids at the protein interaction site, so for this protein we chose the target residues manually.

Generating Exemplars. Starting from the conformation generated by the pocket optimization simulation, we defined the surface pocket as described previously.⁴⁶ Briefly, the protein and its inflated van der Waals surface were mapped to a local grid centered on the target residues, and the pocket was

identified by searching for linear segments of solvent bounded by the protein. The “deep” pocket volume was defined by removing the most solvent-exposed portion of the pocket.

We then generated an exemplar from each pocket, as described previously.⁵⁶ Briefly, we used the protein conformation to identify the location of idealized hydrogen bond donors and acceptors that overlap with any of the pocket, but do not clash with the protein. The remainder of the deep pocket volume was then filled in using carbon atoms, placing these such that they were completely contained in the pocket with minimal overlap, and requiring no less than 1.5 Å between the centers of atom pairs.

For virtual screening, we sought to use several representative exemplars from the ensemble of pocket-optimized output structures. To select these representatives, we used Tanimoto scores computed with the ROCS software^{49,59} as a measure of exemplar similarity⁵⁶ and computed the pairwise similarity of all the exemplars. We then grouped together similar exemplars using hierarchical clustering, and excluded from consideration any clusters with less than 10 members (in cases for which there were fewer than 200 exemplars being clustered, we relaxed this criterion to require only 5 members). From each cluster, we advanced two representative exemplars: the cluster centroid, and the exemplar from the lowest-energy conformation in the cluster.

Building Small-Molecule Libraries for Screening. In the benchmark experiments described here, we compiled collections of active and decoy compounds (as described in the *Results* section). We then used the OMEGA software^{63,64} (with default parameters) to generate up to 100 low-energy three-dimensional conformations for each compound. The active compounds were identified using the TIMBAL database^{65,66} when available and the CHEMBL database⁶⁷ when the target was not present in TIMBAL. In the cases where the targets were not present in TIMBAL nor CHEMBL, the identity of the crystallographic ligand was used (but not the crystallographic conformation).

Active compounds were retained only if they had K_i or K_d less than 10 μM, and molecular weight between 200 and 800 Da. In the cases where these criteria were met by multiple members of a chemical series (2D Fingerprint Tanimoto similarity > 0.7), only the most potent compound was retained. In order to make use of compounds for which only IC₅₀ values were reported (instead of K_i or K_d), these values were scaled by a factor of 0.5 so that they could be compared on equal footing with K_i or K_d values.⁶⁸

Exemplar Screening. Starting from an exemplar, FastROCS^{58,59} was used to rapidly overlay each compound in the library onto the exemplar. To generate full receiver-operating characteristic curves, the top-scoring conformer overlay of each compound was then rescored (without optimization) using ROCS.^{59,60} For the large-scale virtual screening application presented later, only the top 0.75% scoring compounds for a particular exemplar were rescored using ROCS: in a virtual screen, often one reranks only the top-scoring compounds from the initial ranking.

Using ROCS, we computed the Tversky score for each overlay: this asymmetric comparison evaluates the extent to which the ligand subsumes/matches the exemplar features (shape and hydrogen bond donors/acceptors) *without* penalizing the ligand for including “extra” features (for example, if the ligand volume extends outside that of the exemplar, or if the ligand includes additional hydrogen bond donors/acceptors not

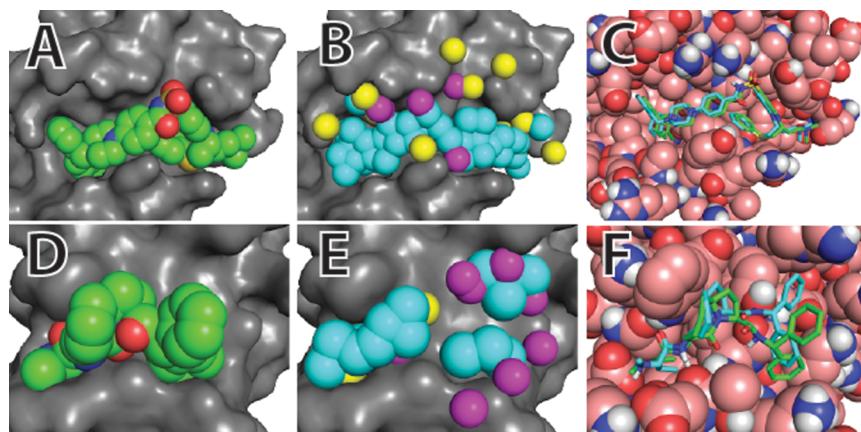


Figure 2. Defining exemplars. (A) Bcl-xL (gray surface) is shown in complex with small-molecule inhibitor ABT-737 (spheres). (B) This conformation of Bcl-xL was used to construct an exemplar comprised of hydrogen bond donors (yellow), hydrogen bond acceptors (magenta), and nonpolar atoms that fill the remainder of the pocket volume (cyan). The exemplar is in essence a map of the ligand that would perfectly complement the pocket. (C) Upon alignment to the exemplar, the predicted pose of ABT-737 (cyan) very closely matches the pose observed in the crystal structure (green) (ligand RMSD 0.93 Å). (D) XIAP (gray surface) is shown in complex with a small-molecule inhibitor (spheres). (E) This conformation of XIAP was used to construct an exemplar. (F) Upon alignment to the exemplar, the predicted pose of this inhibitor (cyan) recapitulates many interactions from the pose observed in the crystal structure (green), including specific hydrogen bonds (ligand RMSD 2.61 Å).

matched to any in the exemplar).⁶⁹ We note that Tversky scores depend on properties of the exemplar; for this reason, one cannot simply compare raw Tversky scores for ligands overlaid with different exemplars. When screening with multiple exemplars, therefore, we expressed each ligand overlay as a Z-score for the exemplar used (i.e., normalized to the mean and standard deviation of the scores from all compounds overlaid to this exemplar). Each compound was assigned a final score corresponding to the best Z-score drawn from any of the exemplars in the set.

RESULTS

Exemplars Map the Binding Pocket of Known Ligands. In this study, we use an exemplar to describe the shape and chemical composition of a protein surface pocket. The exemplar is a collection of atoms that captures the ideal shape and chemical properties that a ligand would exhibit if it were to perfectly complement the pocket: thus, we expect that exemplars may naturally serve as templates for pharmacophore screening.

The exemplars we will use here employ the same three atom types defined in our original conception:⁵⁶ an atom type to define locations at which hydrogen bond donors should be placed, an atom type to define locations at which hydrogen bond acceptors should be placed, and a nonpolar atom type used to fill the remainder of the pocket volume. By construction, then, the exemplar maps out the shape of the pocket and also indicates the key locations for moieties that will participate in hydrogen bonding interactions.

As a starting point, we used the crystal structure of Bcl-xL bound to ABT-737, a compound that inhibits binding of BH3 peptides at this site. In this protein–inhibitor complex, ABT-737 occupies a long, deep channel along the protein surface (Figure 2a). We removed the ligand from this structure, and used the protein structure to generate its corresponding exemplar; indeed, we find that the corresponding exemplar has similar size and shape as the inhibitor itself (Figure 2b). We reiterate that the exemplar is built from the protein structure alone, without direct knowledge of the ligand: here, the size and shape of the exemplar matches that of the ligand, because this

ligand is a natural match for the size and shape of the protein surface pocket. We note that the exemplar includes both a hydrogen bond acceptor and a hydrogen bond donor in close proximity, in the region occupied by the inhibitor's sulfonamide group; this arises from the fact that Bcl-xL includes a tyrosine hydroxyl group at this location on the protein surface, allowing complementary ligands to present either a donor or an acceptor.

To examine whether this exemplar indeed unambiguously encodes the location and orientation of this inhibitor, we built a series of 200 low-energy conformations of ABT-737 and “hid” the crystallographic conformation among these. We then overlaid each conformation with our exemplar and evaluated the extent to which the ligand filled the exemplar volume and matched the locations of hydrogen bonding groups (see Computational Approach). The top-scoring model from this “pose prediction” experiment makes use of the crystallographic ligand conformation and places the ligand in essentially the same position and orientation as in the crystal structure (Figure 2c).

We turned next to the crystal structure of XIAP bound to a bicyclic peptidomimetic compound that disrupts binding of the Smac peptide (Figure 2d). We find that the exemplar built from this protein structure is much smaller (Figure 2e), consistent with the fact that the XIAP inhibitor is much smaller than ABT-737, and binds at a much shallower surface pocket. We further note that the exemplar is expected only to capture features that interact with the protein: thus, the exemplar excludes the solvent-facing aromatic ring and carbonyl group found in this inhibitor.

Once again we built a series of 200 low-energy conformations for this inhibitor, and included among these the crystallographic conformation. Upon overlaying each of these with the exemplar derived from the XIAP, we find that the top-scoring conformation does not precisely match the crystal structure (Figure 2f). The left side of the molecule includes two hydrogen bonds to the protein; the exemplar successfully includes these features, and indeed this part of the molecule is faithfully recapitulated in the predicted pose. The right side of the molecule includes a pair of aromatic rings, and

one of these faces solvent. Here, the exemplar includes a secondary pocket (at the top left) that is not filled by the inhibitor. Because of this, the top-scoring pose instead uses an alternate ligand conformation that allows this secondary pocket to be accessed. The lack of agreement between the crystallographic pose and the predicted pose, then, derives not from how the overlaid structures are scored, but rather from the fact that the exemplar included features of the protein surface that are not used in binding this particular ligand.

Encouraged by these results, we next proceeded to explore the utility of this exemplar-based method for virtual screening.

Virtual Screening Using a Ligand-Bound Protein Structure. The goal of a virtual screen is to distinguish the active compounds in a chemical library (true positives) from among the inactive compounds in the library. Previously we compiled a virtual screening benchmark for small-molecule inhibitors of Bcl-xL and XIAP;⁷⁰ the benchmark evaluates the ability of various methods to distinguish known inhibitors of these protein interactions from “decoy” compounds that are presumed to be inactive. Compounds active against either of these proteins were drawn from the TIMBAL database.^{65,66} To prevent bias from inclusion of many members of a chemical series, we removed any compound with 2D Fingerprint Tanimoto similarity⁷¹ closer than 0.70 to any other compound in the data set. This led to a total of 27 compounds active against Bcl-xL, and 14 compounds active against XIAP. To generate physicochemically matched decoy compounds, we drew all compounds included in the TIMBAL database described as inhibitors of *other* protein–protein interactions. We removed all decoy compounds with 2D Fingerprint Tanimoto similarity closer than 0.70 to any active compound, or to any other decoy compound. This led to a total of 328 decoy compounds for Bcl-xL, and 425 decoy compounds for XIAP. While the decoy compounds are not explicitly known to be inactive against these protein targets, we rely on their lack of similarity to known targets to minimize the likelihood of inadvertently including in the decoy set any compounds that are active.

We started from the crystal structures of Bcl-xL and XIAP each bound to a known inhibitor; to avoid biasing our benchmark, we excluded this particular inhibitor from among the active compounds in our library. We then used the exemplars generated from these two protein structures (Figure 2) to overlay and rank each compound in the corresponding compound library: the compounds that most closely match the exemplar are inferred to be the active compounds in the set. We also used three other docking tools to rank these compounds: DARC,⁷⁰ DOCK 6,^{72,73} and AutoDock4.⁷⁴ Parameters used to run each of these three programs are described fully in our previous work.⁷⁰ Finally, in addition to these docking methods we used a ligand-based method, ROCS,^{30,49,59} to evaluate the three-dimensional similarity of each compound to the conformation of the known inhibitor in the crystal structure.

From the rank order of each compound, we generated receiver operator characteristic (ROC) curves for each of these methods, for both Bcl-xL and for XIAP (Figure 3). For a given method, each point on the plot corresponds to a different ranking at which compounds are be considered active/inactive (i.e., what fraction of the library will be classified as active). For a given cutoff, the *y*-axis defines the percentage of active compounds that have been (correctly) classified as active (true positives) and the *x*-axis defines the percentage of inactive compounds that have been (incorrectly) classified as active

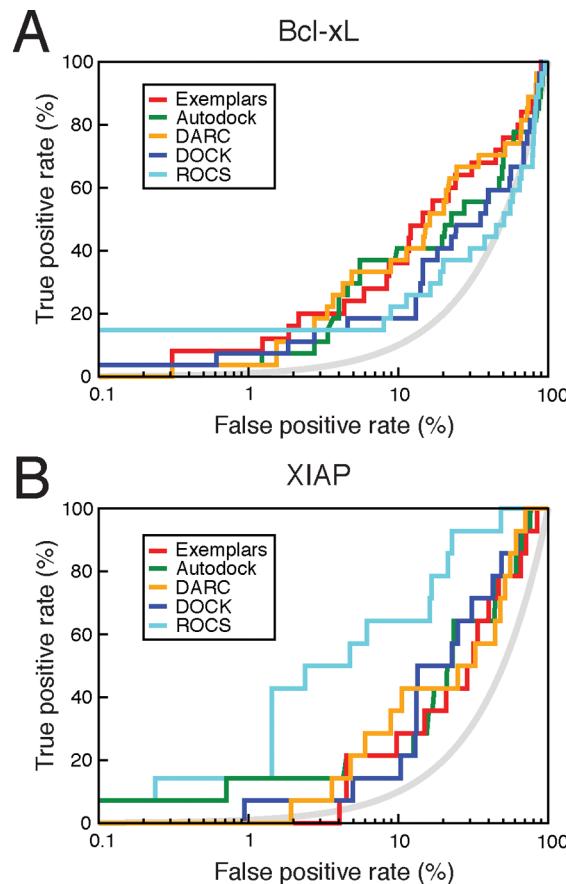


Figure 3. Exemplar screening using a ligand-bound protein structure. This receiver operating characteristic (ROC) plot compares the performance of exemplar screening to three structure-based docking methods (DARC, DOCK 6, and AutoDock4), along with a ligand-based method (ROCS). Data are presented on a semilog plot, to emphasize “early” performance of each method; the gray curve indicates performance of a random predictor. This benchmark experiment involves distinguishing (A) 27 compounds active against Bcl-xL from among 328 decoy compounds and (B) 14 compounds active against XIAP from among 425 decoy compounds. In both cases, decoy compounds were drawn from the TIMBAL database.^{65,66}

(false positives). For a purely random classifier, the percentage of both active and inactive compounds will rise at an equal rate (gray curve).

We found that all five of these methods outperformed the random classifier, in both the Bcl-xL benchmark and the XIAP benchmark. The performance of ROCS in the XIAP benchmark was particularly impressive: we expect that this arises because many of the active compounds in the XIAP set are peptidomimetics, and thus can be recognized by virtue of the pattern of hydrogen bonds that is shared with the template ligand. Meanwhile, the performance of ROCS when applied to Bcl-xL is not quite as impressive, since this set of inhibitors uses a more diverse range of interactions.

Overall, we find that the exemplar screening method performed at a similar level in these benchmarks as the three other receptor-based methods to which we compared it.

Virtual Screening Using a Pocket-Optimized Unbound Structure. Though one was provided for each of the previous benchmarks, at the outset of many real virtual screening scenarios there is no available inhibitor-bound crystal structure from which to draw the protein conformation. While

screening against an unbound conformation leads to diminished performance for most target classes,⁷⁵ this is a particularly acute problem when seeking inhibitors of protein–protein interactions since these protein surfaces are particularly likely to require conformational changes (relative to the unbound protein) in order to bind the inhibitor.³⁷ As noted earlier, we have developed a pocket optimization method that—starting from the unbound protein structure—uses biased sampling to efficiently explore low-energy conformations that contain surface pockets suitable for small molecule binding.^{46,56} We therefore sought to extend this virtual screening benchmark to a regime in which no inhibitor-bound protein conformation is available, and one must instead screen using a pocket-optimized conformation of the target protein.

From pocket optimization simulations of Bcl-xL and XIAP, we therefore extracted the lowest-energy pocket-containing conformations (832 conformations for Bcl-xL and 354 conformations for XIAP) and used these to carry out the same screening experiment described earlier. We omitted ROCS from this comparison: given that we approach this experiment with no presumed knowledge of any other inhibitors, one would not have a template for use with a ligand-based screening method. Among the four receptor-based methods, we find that both exemplar screening and DARC exhibit similar performance as in the previous experiment (Figure 4). In contrast, both DOCK 6 and AutoDock4 show dramatically diminished performance in this experiment.

Previously, we found that pocket-optimized conformations often include small structural differences relative to the corresponding ligand-bound conformations, but nonetheless often capture features of the binding pocket with high fidelity;⁵⁶ this, in turn, may explain the performance observed in this benchmark. At high resolution, the pocket-optimized conformations do not fit the active compounds quite as well, because the detailed positioning of certain protein groups is not optimal since the simulations were performed without knowledge of any particular inhibitor: this presents a problem for methods such as DOCK 6 and AutoDock4, which evaluate the complementarity between individual intermolecular contacts. In contrast, the faithful recapitulation of the crude features of the binding pocket allows methods that focus instead on identifying ligands that match overall shape and chemical complementary of the pocket itself (DARC and exemplar screening) to perform just as well in this regime.

Virtual Screening Using an Ensemble of Pocket-Optimized Unbound Structures. An important advantage of the exemplar screening method presented here is that the underlying ligand/exemplar overlays (using ROCS) prove much faster than any of the explicit docking approaches included in our study. In fact, OpenEye recently developed a newer implementation (“FastROCS”) that takes advantage of multiple CPU and GPU cores, and is even faster than ROCS.^{47,59} Using this new implementation, we find that screening 6 million compounds against a single exemplar, depending on the number of hit compounds written to disk, can be completed in as few as 15 min on a single 16-core, dual-GPU computer.

Given the speed of this approach, it now becomes feasible to screen a large compound library against *every* exemplar derived from an ensemble of pocket-optimized protein conformations, not just the exemplar from the lowest-energy conformation. Nonetheless, for computational efficiency we instead elected to

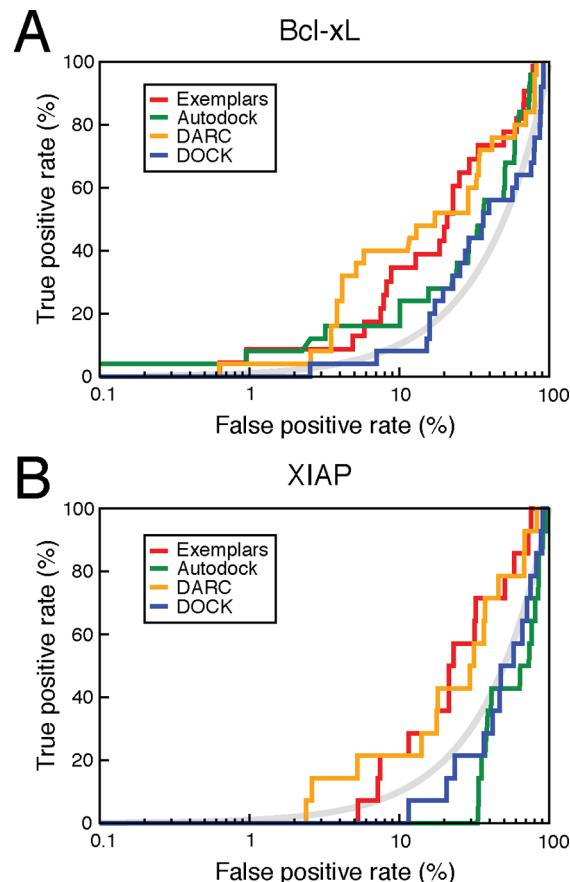


Figure 4. Exemplar screening using a pocket-optimized unbound structure. This benchmark experiment entails discrimination of the same active versus decoy compounds as in Figure 3. This time, however, the protein conformation used for screening was taken from pocket optimization simulations. Data are again presented on a semilog plot, to emphasize the early performance of each method; the gray curve indicates performance of a random predictor. This benchmark experiment is again applied to (A) Bcl-xL and (B) XIAP.

identify a representative subset of exemplars for screening. To this end we used the same overlay approach (in ROCS) to evaluate the pairwise similarity of the exemplars, and then used this measure as the basis for hierarchical clustering. For Bcl-xL, we used 10 distinct exemplar clusters, and for XIAP we used 6 distinct exemplar clusters. From each cluster we then extracted the exemplar from the lowest-energy conformation, and the exemplar corresponding to the cluster centroid. This led to a collection of 15 representative exemplars for Bcl-xL and 9 representative exemplars for XIAP.

We then used each of these exemplars to once again screen our benchmark libraries for Bcl-xL and XIAP. For a given exemplar, we computed the mean and standard deviation of the scores from all overlaid ligands and used this to express the ligand/exemplar overlap as a Z-score. We then searched over all the exemplars to find the best Z-score for each compound, and used these as the basis for ranking the compounds. Overall, we find that this “ensemble” exemplar approach yields similar performance in this experiment as using an exemplar from the unbound structure, or using an exemplar from the single lowest-energy pocket-optimized protein conformation (Figure 5).

Ultra-High-Throughput Virtual Screening Using an Ensemble of Pocket-Optimized Unbound Structures.

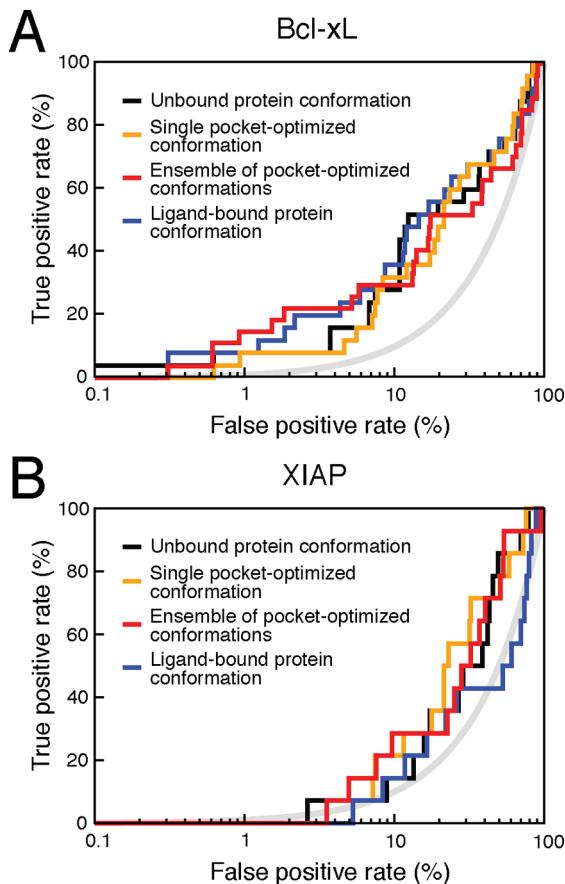


Figure 5. Exemplar screening using an *ensemble* of pocket-optimized unbound structures. Screens were carried out using the same benchmark sets as in Figures 3 and 4, for both (A) Bcl-xL and (B) XIAP. Here we compare performance when screening using exemplars generated from the unbound protein structure, from a single pocket-optimized conformation, from an ensemble of pocket-optimized conformations, or from an inhibitor-bound protein structure. The gray curve indicates performance of a random predictor.

noted above, a key advantage of the exemplar screening method is its speed. This is important, because it allows one to carry out screens using multiple receptor conformations, and/or screen against very large compound libraries. As a demonstration of this capability, we compiled from the ZINC database⁷⁶ a collection of ~8.33 million drug-like compounds to use as decoys for a large screening experiment. Once again, we hid among these decoy compounds either the 24 compounds active against Bcl-xL or the 14 compounds active against XIAP.

We screened the compound library using each of the 15 representative exemplars from pocket-optimized conformations of Bcl-xL or the 9 representative exemplars from pocket-optimized conformations of XIAP. The scale of this screening experiment would make it extremely time-consuming using any of the other three receptor-based (docking) methods included in our previous benchmark experiments. Using exemplar screening, however, screening more than 8 million compounds against each of these two protein targets—with multiple receptor conformations for each—required a wall time (total execution time) of only 7 h on a single (dual GPU) computer.

Results from this large-scale screen are compiled into Figure 6. In both cases, exemplar docking against an ensemble of protein conformations identifies active compounds far more effectively than a random predictor (Figure 6a), with the area

under the curve of each ROC plot (AUC) at 0.793 and 0.848 for Bcl-xL and XIAP, respectively. Screening against the Bcl-xL ensemble identified 50% of the (chemically distinct) active compounds within the top 1% of the library, while screening against the XIAP ensemble identified 29% of the active compounds within the top 1% of the library. The early enrichment—a key measure for virtual screening because only a small fraction of the total library is expected to actually be tested experimentally—is also impressive: in both cases the top-ranking 0.01% of the library has more than 750-fold enrichment of active compounds relative to the complete library (Figure 6b). This exciting result suggests that—with very modest computational expense—much of this large screening database can be eliminated from further consideration, leaving a much smaller library that is considerably enriched with active compounds. Notably, performance in this screen is worse when the exemplar from the single lowest-energy pocket-containing conformation is utilized in this screen, instead of exemplars from a small ensemble of low-energy conformations (Figure 6a): this observation highlights the benefit of screening against multiple receptor conformations, which in turn is enabled by the very fast exemplar screening approach.

Examination of the top-ranking complexes involving compounds known to be active reveal compelling binding modes for these compounds (Figure 6c), the structures of which have not (to date) been solved in complex with their protein partners. Meanwhile, the top-ranking complexes involving decoy compounds also exhibit compelling binding modes (Figure 6d); as expected from their match to the underlying exemplars, each of these combine exquisite shape complementarity with multiple intermolecular hydrogen bonds. It is useful to recall that the decoy compounds in this experiment are merely presumed inactive and have not been explicitly tested against Bcl-xL or XIAP: it is quite possible that the top-ranking compounds from this experiment may indeed exhibit activity toward their respective protein targets.

Ultra-High-Throughput Virtual Screening for a Variety of Protein–Protein Interaction Targets. Previously we compiled a set of 18 complexes, each corresponding to a crystal structure of a small-molecule inhibitor bound at a protein interaction site;⁷⁷ in 11 of these cases the binding site is only composed of the biological unit (i.e., the stoichiometry of binding is clearly 1:1), and a structure of the unbound protein is also available. None of the proteins in the set are homologues of one another; for proteins in which more than one inhibitor-bound structure was available, we retained only the structure involving the most potent inhibitor. These 11 complexes comprise the majority of available nonredundant structures of small-molecule inhibitors of protein–protein inhibitors and are presented in Table S1. For each of the inhibitor-bound structures, we removed the ligand and generated an exemplar from the inhibitor-bound conformation. For each of the unbound protein structures, we built an ensemble of pocket-optimized conformations; upon clustering, we retained between 9 and 27 distinct conformations for each (depending on the number of high-population clusters that were identified) and used these exemplars for our subsequent screening benchmark. As described above, we collected a nonredundant and diverse set of inhibitors for each target protein, drawing from the TIMBAL^{65,66} and CHEMBL⁶⁷ databases in addition to the PDB. These known inhibitors were added to the collection of

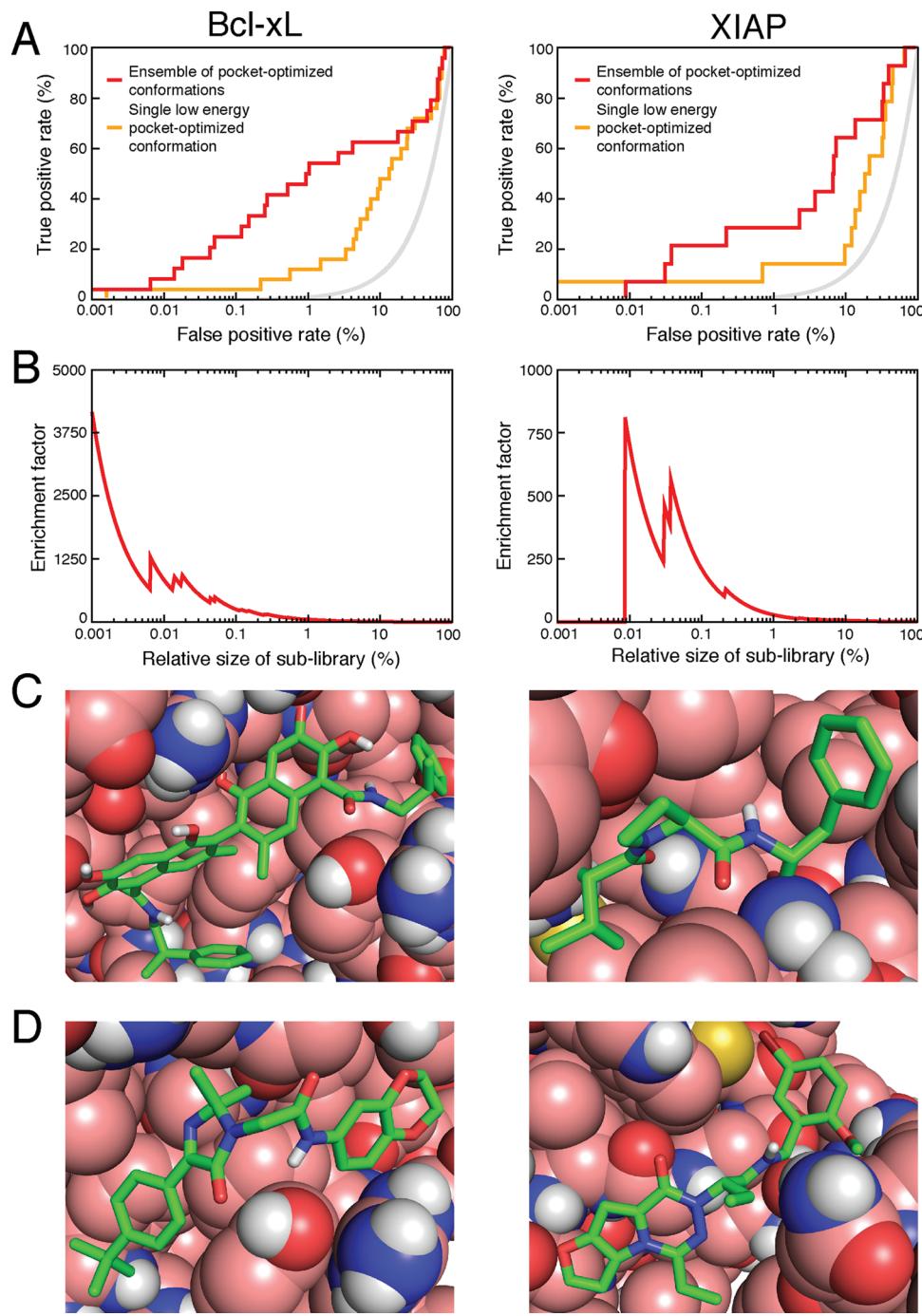


Figure 6. Ultra-high-throughput screening using an ensemble of pocket-optimized unbound structures. The same 27 compounds active against Bcl-xL and 14 compounds active against XIAP were now hidden among a library of 8.33 million drug-like compounds. This collection was screened using exemplars generated from an ensemble of pocket-optimized conformations. (A) This ROC plot compares the performance of exemplar screening when using an ensemble of pocket-optimized conformations (red) or the single lowest energy, pocket-optimized conformation (orange). The gray curve indicates performance of a random predictor. The BEDROC metric with $\alpha = 20$,⁸⁹ a key measure designed specifically to assess performance for early recognition, is 0.559 for Bcl-xL when screening against an ensemble of conformations but drops to 0.249 when screening against the single lowest-energy conformation. Similarly for XIAP, the BEDROC value drops from 0.419 to 0.058 when screening against a single conformation instead of an ensemble. (B) The enrichment factor indicates the extent to which active compounds are over-represented in “sublibraries” generated by extracting the top-scoring compounds from the original compound library, when screening against an ensemble of pocket-optimized conformations (red). Because this value corresponds to enrichment over the fraction of active compounds in the original compound library, a random predictor would have a value of 1 for all sublibraries. (C) Representative models of top-scoring complexes involving compounds known to be active against each of these protein targets, for which no structural information is currently available. (D) Representative models of top-scoring complexes involving decoy (presumed inactive) compounds. To date, these compounds have not been explicitly tested for activity against the corresponding protein targets.

~8.33 million drug-like compounds from the ZINC database,⁷⁶ which served as decoy compounds in this experiment.

Exemplar screens were carried out using either a crystallographic inhibitor-bound conformation, the single lowest-energy pocket-optimized conformation, or the ensemble of distinct pocket-optimized conformations after clustering. The rank of each active compound, relative to the set of decoy compounds, was then determined for each case: a summary of these data are presented in Tables 1, S2, and S3.

Table 1. Ultra-High-Throughput Virtual Screening for a Variety of Protein–Protein Interaction Targets, Using an Ensemble of Pocket-Optimized Conformations^a

protein name	best rank of known active	number of known actives in top 100	fraction of known actives in top 1%
HIV-1 integrase (87)	2	4	0.16
Mdm2 (34)	175	0	0.24
Bcl-xL (25)	10	1	0.08
BRD4 (20)	208209	0	0.00
XIAP-BIR3 (12)	45043	0	0.08
Grb2-SH2 (9)	3	3	0.78
Interleukin-2 (6)	1573	0	0.17
HPV E2 (4)	20111	0	0.50
Menin (3)	1	1	0.33
WDR5 (3)	613564	0	0.00
VHL (1)	25376	0	1.00

^aA collection of diverse known inhibitors for each protein target were hidden amongst a set of 8.33 million drug-like decoy compounds (the number of active compounds for each protein target is in parentheses). The compound library was then screened using multiple exemplars, corresponding to an ensemble of pocket-optimization simulations started from the *unbound* protein structure in each case. Corresponding data are also presented for screens carried out using an inhibitor-bound protein structure (Table S2) or using only the single top-scoring pocket-optimized conformation (Table S3).

Focusing first on the results from screening against an ensemble of pocket-optimized conformations (Table 1), the enrichment of active compounds is evident. Remarkably, for 4 of the 11 protein target, a known active compound is ranked in the top 10 (of more than 8 million compounds). Further, known active compounds are included in the top 1% of scored compounds for 9 of the 11 protein targets. It is once again worth noting that not all of the decoy compounds that score more favorably than these active compounds are necessarily inactive: they are simply presumed inactive for the purposes of this experiment, since they have typically not yet been tested for activity against these protein targets.

It must also be pointed out that screening against the ensemble of conformations was necessary, in most cases, to achieve this level of performance. While screening against an inhibitor-bound conformation did yield a top-ranked active compound in three cases, this approach did not identify a known active compound in the top 1% for *any* of the other eight protein targets (Table S2). Meanwhile, the use of a single pocket-optimized conformation found known active compounds in the top 1% for 4 of the 11 protein targets, but these were typically not ranked as highly as those obtained by screening using the other approaches (Table S3).

■ DISCUSSION

Identifying small-molecule inhibitors of protein–protein interactions remains a challenging task. Because these protein surfaces are typically not evolved to bind small molecules, they often lack the deep binding pockets to bind small molecules with high affinity.^{37,78} In the structurally characterized examples to date, some small-molecule inhibitors are found to bind using pre-existing pockets whereas others take advantage of transient pockets that are not observed in crystal structures of the unbound protein. This presents a natural problem for structure-based virtual screening approaches, since they require knowledge of the inhibitor-bound protein conformation (or at least something close to it): at protein interaction sites the unbound protein conformation may not bear sufficient resemblance to the inhibitor-bound conformation—in terms of available surface pockets—to allow successful virtual screening. Ligand-based methods for virtual screening are even less applicable, if very few known inhibitors are available.

In this study, we use biased simulations to predict low-energy pocket-containing conformations, starting from the unbound protein structures. We then use exemplars derived from these conformations as templates for rapidly screening very large compound libraries. By targeting protein conformations that would not otherwise be available without prior ligand-bound crystal structures, we enable discovery of potential inhibitors that would not be identified using other methodologies.

Comparison to Traditional Docking. For small-molecule inhibitors of protein interactions, the exemplar-based screening approach presented here offers two primary advantages over traditional, docking-based methods. The first advantage is accuracy: this is a notably challenging regime for traditional docking methods.⁷⁸ The screening benchmark we use in this study is intentionally difficult, since the TIMBAL data set is composed entirely of compounds that inhibit various protein–protein interactions and is thus devoid of decoy compounds that can be easily ruled out as unsuitable. When screening against a ligand-bound protein structure we find that exemplar screening performed with similar accuracy as DOCK 6 and AutoDock4. However, screening against a pocket-optimized conformation revealed the sensitivity of these methods to details of the protein structure, which in turn led to diminished performance. In contrast to this, overlaying ligands with exemplars strongly favors matching of the *overall* shape and chemical features, which in turn provides robustness to slight variations in the protein structure.

The second advantage, more dramatic than the first, is the fact that the exemplar-based screening approach is orders of magnitude faster than docking. Because large-scale virtual screens that involve docking are often constrained by computational requirements, compound libraries are typically filtered to remove redundancy while at the same time maximizing the diversity within the data set.^{79,80} However, small changes to a compound can have profound effects on binding affinity;^{81,82} as such, screening a complete library may yield hits from chemical series that may otherwise have been excluded based on the score of a single representative member included in the nonredundant set.

The speed advantage of exemplar-based screening approach relative to traditional docking approaches, moreover, also enables screening against multiple receptor conformations; this would not necessarily be feasible for a large compound library

using traditional docking, given the computational expense of screening against even a single receptor conformation.

Comparison to Side-Chain Mimicry. An emerging approach for designing of inhibitors of protein–protein interactions without docking entails mimicry of certain functional groups presented by the partner protein (or peptide). This approach has recently been demonstrated by Koes and colleagues: their approach identifies “anchor residues” important for a protein–protein interaction, then combines these into pharmacophores to search for compounds that mimic the three-dimensional geometry of these residues’ side chains.³⁸ Given the overabundance of certain side chains as anchor residues (phenylalanine, tyrosine, tryptophan, valine, and leucine), this approach lends itself naturally to construction of custom screening libraries (accessible through multi-component reaction chemistry) containing compounds mimicking these side chains.³⁹ Others have also used similar methods to extract specific side chains—or smaller functional groups—from protein complexes and defined pharmacophores based on these as templates for virtual screening.^{40–45}

Because their overlay with the protein’s natural binding partner is evident, the use of mimicry produces highly intuitive hit compounds. An important limitation, however, is that these hits represent a very limited subset of potentially active compounds. The interactions of the protein partner may not be optimal (given constraints of the protein geometry and/or lack of evolutionary pressure for tight binding), and this approach cannot be extended to conformations of the target protein other than that observed in the protein–protein complex. As noted earlier, many high affinity inhibitors acting at protein interaction sites take advantage of precisely these alternate conformations, because they contain surface pockets more suitable for small-molecule binding.³⁷

The exemplar-based screening we describe here is notably different from these approaches, in that it does not rely on mimicry of the protein partner, but instead builds a pharmacophore entirely from the structure of the target protein. Accordingly, it is not subject to the limitations described above: the resulting hits can fill the exemplar without carrying forward artifacts from the protein partner and can be applied equally well to other conformations of the target protein beyond simply the one used for protein binding.

Comparison to Other Receptor-Based Pharmacophore Methods. In the absence of small-molecule inhibitors (or natural substrates), others have also used features of the receptor to build pharmacophores;^{9,10,13,15,83–86} like these, the exemplars we describe here are essentially receptor-based pharmacophores. In general there are two types of approaches for building receptor-based pharmacophores: either by using the geometry of the receptor to generate the pharmacophore directly^{10,13} (as we have done in this study), or else by docking a series of “probe” compounds against the receptor to identify potential interaction sites at which these probes accumulate.^{9,15,83–86} For the most part, however, the use of receptor-based pharmacophores has thus far been confined to “traditional” drug target classes and not extended to protein interaction sites.

With the exception of the studies presented here, there has been—to our knowledge—only one other application of receptor-based pharmacophores for targeting a protein interaction site. In order to identify compounds inhibiting assembly of the heterodimeric HIV reverse transcriptase active complex, Grohmann and colleagues developed such an

approach.⁸⁶ To account for receptor flexibility, this group carried out a short molecular dynamics simulation, selected six snapshots, and then docked probe molecules against these conformations in order to identify potential interactions; this led to three unique pharmacophores. They then used these pharmacophores to carry out a virtual screen, and used the compounds that overlaid with their pharmacophores as a starting point for docking. The top-scoring compounds in these docking studies were advanced into a suite of biochemical assays, in which one of these compounds was found to exhibit (weak) activity.⁸⁶

The computational pipeline we describe here (Figure 1) differs from that of Grohmann and colleagues in several key respects. First, applying pocket optimization to generate a collection of receptor structures provides an efficient and effective means for sampling diverse low-energy conformations;^{46,56} in contrast, others have shown through molecular dynamics simulations of the same protein interaction sites that up to half of the corresponding pockets—which may be required for small-molecule binding—are very seldom sampled through (unbiased) molecular dynamics simulations.⁸⁷ The pocket optimization simulations have the additional advantage of requiring relatively modest computational resources, and they scale in a trivially parallel manner to yield a linear speedup with the number of available processors.

The second important difference results from the approach to building and applying the pharmacophore model. The computational pipeline described here does not include docking at any stage: the pharmacophore is built solely from the geometry of the receptor (without docking probes), and then models of the bound complexes are generated directly from the overlays to the pharmacophore (without docking the hit compounds). By avoiding explicitly protein–ligand docking at every stage, we enable rapid screening of large ensembles of receptor conformations, against very large compound libraries.

The approach of Grohmann and colleagues was immediately applied to HIV reverse transcriptase,⁸⁶ without examining the performance of this approach through any benchmarking studies. Moving forward, it will be interesting and exciting to parse the relative merits of each step in these computational pipelines and carefully determine the most effective method for identifying compounds that are potent and selective inhibitors of many other targets.

Ensemble Docking. Already others have noted that docking against multiple receptor conformations (“ensemble docking”) can enhance performance relative to docking against a single static conformation.^{47–55} Indeed, our ensemble docking benchmark confirmed that using a collection of distinct conformations generated from an unbound structure, we obtained similar performance to that observed when using a protein structure solved in complex with an inhibitor (Figure 5). Further, when performing virtual screens across a diverse series of targets, screening against an ensemble of structures outperformed the analogous screens against an individual protein conformation, allowing discovery of more diverse active compounds (Figure 6, and Tables 1, S2, and S3).

Prior studies have sought to incorporate multiple conformations when using side chain mimicry^{42–44} or pharmacophores built from docking probe compounds.^{84–86} Typically, these approaches then consolidate pharmacophores into a small number of consensus templates for virtual screening. A potential risk inherent to this step is that the average location of a feature across several conformations may correspond to a

position that is not explicitly compatible with any of the individual conformations that contribute to this average. By instead creating and screening with multiple exemplars explicitly, we aim to identify complexes in which the ligand is truly optimal for a particular protein structure—rather than for an averaged protein conformation that may not be reflective of the individual members of the underlying ensemble.

Advantages of Pocket Optimization. These studies also provide another demonstration of the utility of pocket optimization to guide ligand discovery. Previously we showed that pocket optimization can be used to evaluate the druggability of a given protein surface though its propensity for forming pockets.⁴⁶ We also showed that the *shapes* of these pockets dictate the space of “allowed” ligands that can interact with a given surface and, thus, can dictate inhibitor selectivity.⁵⁶ This is further underscored through the results presented here, in which we demonstrate directly that features of these pockets can be used to identify complementary ligands, even in the absence of a ligand-bound structure to use as a starting point.

Putting these pieces together, we envision a unified pipeline for ligand discovery. Starting from the unbound protein structure, one could use pocket optimization to gauge druggability of the target surface. To ensure selectivity of the resulting inhibitors, one would then select only those pockets that are sampled by the desired protein target(s) and not by other members of this protein family. Exemplars built from these pockets would serve as templates for virtual screening, to identify inhibitors that recognize pockets unique to the protein target(s). Collectively, this approach offers the prospect of not only identifying compounds that bind to the protein target, but that do so with the desired selectivity.

Finally, we note that this approach may also enable “cryptic” allosteric sites to be addressed through novel ligands. In the absence of a binding partner at this site mimicry would not be possible, and the lack of a preformed binding pocket would render traditional docking unsuitable. In light of the multitude of such sites that potentially exist,⁸⁸ the exemplar screening method presented here may enable identification of new allosteric inhibitors that would be exceedingly challenging to discover using other existing methods.

■ ASSOCIATED CONTENT

S Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jcim.5b00572](https://doi.org/10.1021/acs.jcim.5b00572).

Supplementary methods, supplementary tables (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: johnk@ku.edu. Tel.: 785-864-8298.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Ragul Gowthaman and Andrea Bazzoli for providing benchmark sets, Sergey Lyskov for assistance with the ROSIE server, and Karen Khar for valuable discussions. We are grateful to OpenEye Scientific Software (Santa Fe, NM) for providing an academic license for the use of OMEGA, ROCS, and FastROCS. We are grateful to the developers of the DOCK and AutoDock software for providing academic licenses for the use of these programs. This work was supported by grants from the

National Institute of General Medical Sciences (1R01GM099959) and the National Science Foundation through XSEDE allocation MCB130049.

■ REFERENCES

- (1) Ehrlich, P. *Beiträge zur experimentellen Pathologie und Chemothrapie*; Akademische Verlagsgesellschaft: Leipzig, 1909; p vii, 2l, 3–247.
- (2) Guner, O. F.; Bowen, J. P. Setting the record straight: the origin of the pharmacophore concept. *J. Chem. Inf. Model.* **2014**, *54*, 1269–1283.
- (3) Wermuth, C. G.; Ganellin, C. R.; Lindberg, P.; Mitscher, L. A. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* **1998**, *70*, 1129–1143.
- (4) Bohacek, R.; Boosalis, M. S.; McMardin, C.; Faller, D. V.; Perrine, S. P. Identification of novel small-molecule inducers of fetal hemoglobin using pharmacophore and ‘PSEUDO’ receptor models. *Chem. Biol. Drug Des.* **2006**, *67*, 318–328.
- (5) Cushman, M.; Insaf, S.; Ruell, J. A.; Schaeffer, C. A.; Rice, W. G. Synthesis of a cosalane analog with an extended polyanionic pharmacophore conferring enhanced potency as an anti-HIV agent. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 833–836.
- (6) Wender, P. A.; Koehler, K. F.; Sharkey, N. A.; Dell'Aquila, M. L.; Blumberg, P. M. Analysis of the phorbol ester pharmacophore on protein kinase C as a guide to the rational design of new classes of analogs. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 4214–4218.
- (7) Kier, L. B. Molecular orbital calculation of preferred conformations of acetylcholine, muscarine, and muscarone. *Mol. Pharmacol.* **1967**, *3*, 487–494.
- (8) Dassault Systèmes BIOVIA, *Discovery Studio Modeling Environment*, Release 4.5; Dassault Systèmes, San Diego, 2015.
- (9) Barillari, C.; Marcou, G.; Rognan, D. Hot-spots-guided receptor-based pharmacophores (HS-Pharm): a knowledge-based approach to identify ligand-anchoring atoms in protein cavities and prioritize structure-based pharmacophores. *J. Chem. Inf. Model.* **2008**, *48*, 1396–1410.
- (10) Ebalunode, J. O.; Ouyang, Z.; Liang, J.; Zheng, W. Novel approach to structure-based pharmacophore search using computational geometry and shape matching techniques. *J. Chem. Inf. Model.* **2008**, *48*, 889–901.
- (11) Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. The Conformational Parameter in Drug Design: The Active Analog Approach. In *Computer-Assisted Drug Design*; American Chemical Society: Washington, D.C., 1979; Vol. 112, Chapter 9, pp 205–226.
- (12) Hu, B.; Lill, M. A. Exploring the potential of protein-based pharmacophore models in ligand pose prediction and ranking. *J. Chem. Inf. Model.* **2013**, *53*, 1179–1190.
- (13) Lee, H. S.; Lee, C. S.; Kim, J. S.; Kim, D. H.; Choe, H. Improving virtual screening performance against conformational variations of receptors by shape matching with ligand binding pocket. *J. Chem. Inf. Model.* **2009**, *49*, 2419–2428.
- (14) Richmond, N. J.; Abrams, C. A.; Wolohan, P. R.; Abrahamian, E.; Willett, P.; Clark, R. D. GALAHAD: 1. pharmacophore identification by hypermolecular alignment of ligands in 3D. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 567–587.
- (15) Tintori, C.; Corradi, V.; Magnani, M.; Manetti, F.; Botta, M. Targets looking for drugs: a multistep computational protocol for the development of structure-based pharmacophores and their applications for hit discovery. *J. Chem. Inf. Model.* **2008**, *48*, 2166–2179.
- (16) 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*, 160–169.
- (17) Zhu, L. L.; Hou, T. J.; Chen, L. R.; Xu, X. J. 3D QSAR analyses of novel tyrosine kinase inhibitors based on pharmacophore alignment. *J. Chem. Inf. Model.* **2001**, *41*, 1032–1040.

- (18) Gao, Q.; Yang, L.; Zhu, Y. Pharmacophore based drug design approach as a practical process in drug discovery. *Curr. Comput.-Aided Drug Des.* **2010**, *6*, 37–49.
- (19) Palomer, A.; Cabre, F.; Pascual, J.; Campos, J.; Trujillo, M. A.; Entrena, A.; Gallo, M. A.; Garcia, L.; Mauleon, D.; Espinosa, A. Identification of novel cyclooxygenase-2 selective inhibitors using pharmacophore models. *J. Med. Chem.* **2002**, *45*, 1402–1411.
- (20) Kraemer, O.; Hazemann, I.; Podjarny, A. D.; Klebe, G. Virtual screening for inhibitors of human aldose reductase. *Proteins: Struct., Funct., Genet.* **2004**, *55*, 814–823.
- (21) Chen, J. Z.; Wang, J.; Xie, X. Q. GPCR structure-based virtual screening approach for CB2 antagonist search. *J. Chem. Inf. Model.* **2007**, *47*, 1626–1637.
- (22) Evers, A.; Klabunde, T. Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J. Med. Chem.* **2005**, *48*, 1088–1097.
- (23) Marriott, D. P.; Dougall, I. G.; Meghani, P.; Liu, Y. J.; Flower, D. R. Lead generation using pharmacophore mapping and three-dimensional database searching: application to muscarinic M(3) receptor antagonists. *J. Med. Chem.* **1999**, *42*, 3210–3216.
- (24) Ekins, S.; Johnston, J. S.; Bahadduri, P.; D’Souza, V. M.; Ray, A.; Chang, C.; Swaan, P. W. In vitro and pharmacophore-based discovery of novel hPEPT1 inhibitors. *Pharm. Res.* **2005**, *22*, 512–517.
- (25) Enyedy, I. J.; Sakamuri, S.; Zaman, W. A.; Johnson, K. M.; Wang, S. Pharmacophore-based discovery of substituted pyridines as novel dopamine transporter inhibitors. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 513–517.
- (26) Chang, C.; Ekins, S.; Bahadduri, P.; Swaan, P. W. Pharmacophore-based discovery of ligands for drug transporters. *Adv. Drug Delivery Rev.* **2006**, *58*, 1431–1450.
- (27) Fauman, E. B.; Rai, B. K.; Huang, E. S. Structure-based druggability assessment—identifying suitable targets for small molecule therapeutics. *Curr. Opin. Chem. Biol.* **2011**, *15*, 463–468.
- (28) Becker, J. W.; Rotonda, J.; Cryan, J. G.; Martin, M.; Parsons, W. H.; Sinclair, P. J.; Wiederrecht, G.; Wong, F. 32-Indolyl ether derivatives of ascomycin: three-dimensional structures of complexes with FK506-binding protein. *J. Med. Chem.* **1999**, *42*, 2798–2804.
- (29) Bruncko, M.; Oost, T. K.; Belli, B. A.; Ding, H.; Joseph, M. K.; Kunzer, A.; Martineau, D.; McClellan, W. J.; Mitten, M.; Ng, S. C.; Nimmer, P. M.; Oltersdorf, T.; Park, C. M.; Petros, A. M.; Shoemaker, A. R.; Song, X.; Wang, X.; Wendt, M. D.; Zhang, H.; Fesik, S. W.; Rosenberg, S. H.; Elmore, S. W. Studies leading to potent, dual inhibitors of Bcl-2 and Bcl-xL. *J. Med. Chem.* **2007**, *50*, 641–662.
- (30) Rush, T. S., 3rd; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* **2005**, *48*, 1489–1495.
- (31) Sun, H.; Stuckey, J. A.; Nikolovska-Coleska, Z.; Qin, D.; Meagher, J. L.; Qiu, S.; Lu, J.; Yang, C. Y.; Saito, N. G.; Wang, S. Structure-based design, synthesis, evaluation, and crystallographic studies of conformationally constrained Smac mimetics as inhibitors of the X-linked inhibitor of apoptosis protein (XIAP). *J. Med. Chem.* **2008**, *51*, 7169–7180.
- (32) Thanos, C. D.; Randal, M.; Wells, J. A. Potent small-molecule binding to a dynamic hot spot on IL-2. *J. Am. Chem. Soc.* **2003**, *125*, 15280–15281.
- (33) Vassilev, L. T.; Vu, B. T.; Graves, B.; Carvajal, D.; Podlaski, F.; Filipovic, Z.; Kong, N.; Kammlott, U.; Lukacs, C.; Klein, C.; Fotouhi, N.; Liu, E. A. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **2004**, *303*, 844–848.
- (34) Wang, Y.; Coulombe, R.; Cameron, D. R.; Thauvette, L.; Massariol, M. J.; Amon, L. M.; Fink, D.; Titolo, S.; Welchner, E.; Yoakim, C.; Archambault, J.; White, P. W. Crystal structure of the E2 transactivation domain of human papillomavirus type 11 bound to a protein interaction inhibitor. *J. Biol. Chem.* **2004**, *279*, 6976–6985.
- (35) Surade, S.; Blundell, T. L. Structural biology and drug discovery of difficult targets: the limits of ligandability. *Chem. Biol.* **2012**, *19*, 42–50.
- (36) Kuenemann, M. A.; Sperandio, O.; Labbe, C. M.; Lagorce, D.; Miteva, M. A.; Villoutreix, B. O. In silico design of low molecular weight protein-protein interaction inhibitors: Overall concept and recent advances. *Prog. Biophys. Mol. Biol.* **2015**, *119*, 20–32.
- (37) Wells, J. A.; McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007**, *450*, 1001–1009.
- (38) Koes, D. R.; Camacho, C. J. PocketQuery: protein-protein interaction inhibitor starting points from protein-protein interaction structure. *Nucleic Acids Res.* **2012**, *40*, W387–392.
- (39) Koes, D.; Khouri, K.; Huang, Y.; Wang, W.; Bista, M.; Popowicz, G. M.; Wolf, S.; Holak, T. A.; Domling, A.; Camacho, C. J. Enabling large-scale design, synthesis and validation of small molecule protein-protein antagonists. *PLoS One* **2012**, *7*, e32839.
- (40) Metz, A.; Schanda, J.; Grez, M.; Wichmann, C.; Gohlke, H. From determinants of RUNX1/ETO tetramerization to small-molecule protein-protein interaction inhibitors targeting acute myeloid leukemia. *J. Chem. Inf. Model.* **2013**, *53*, 2197–2202.
- (41) De Luca, L.; Barreca, M. L.; Ferro, S.; Christ, F.; Iraci, N.; Gitto, R.; Monforte, A. M.; Debysier, Z.; Chimirri, A. Pharmacophore-based discovery of small-molecule inhibitors of protein-protein interactions between HIV-1 integrase and cellular cofactor LEDGF/p75. *ChemMedChem* **2009**, *4*, 1311–1316.
- (42) Christ, F.; Voet, A.; Marchand, A.; Nicolet, S.; Desimmie, B. A.; Marchand, D.; Bardiot, D.; Van der Veken, N. J.; Van Remoortel, B.; Strelkov, S. V.; De Maeyer, M.; Chaltin, P.; Debysier, Z. Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. *Nat. Chem. Biol.* **2010**, *6*, 442–448.
- (43) Mustata, G.; Li, M.; Zevola, N.; Bakan, A.; Zhang, L.; Epperly, M.; Greenberger, J. S.; Yu, J.; Bahar, I. Development of small-molecule PUMA inhibitors for mitigating radiation-induced cell death. *Curr. Top. Med. Chem.* **2011**, *11*, 281–290.
- (44) Corradi, V.; Mancini, M.; Santucci, M. A.; Carlomagno, T.; Sanfelice, D.; Mori, M.; Vignaroli, G.; Falchi, F.; Manetti, F.; Radi, M.; Botta, M. Computational techniques are valuable tools for the discovery of protein-protein interaction inhibitors: the 14–3-3sigma case. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 6867–6871.
- (45) Voet, A.; Callewaert, L.; Ulens, T.; Vanderkelen, L.; Vanherreweghe, J. M.; Michiels, C. W.; De Maeyer, M. Structure based discovery of small molecule suppressors targeting bacterial lysozyme inhibitors. *Biochem. Biophys. Res. Commun.* **2011**, *405*, 527–532.
- (46) Johnson, D. K.; Karanicolas, J. Druggable protein interaction sites are more predisposed to surface pocket formation than the rest of the protein surface. *PLoS Comput. Biol.* **2013**, *9*, e1002951.
- (47) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*, 377–395.
- (48) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337*, 209–225.
- (49) Osguthorpe, D. J.; Sherman, W.; Hagler, A. T. Exploring protein flexibility: incorporating structural ensembles from crystal structures and simulation into virtual screening protocols. *J. Phys. Chem. B* **2012**, *116*, 6952–6959.
- (50) Huang, S. Y.; Zou, X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins: Struct., Funct., Genet.* **2007**, *66*, 399–421.
- (51) Cosconati, S.; Marinelli, L.; Di Leva, F. S.; La Pietra, V.; De Simone, A.; Mancini, F.; Andrisano, V.; Novellino, E.; Goodsell, D. S.; Olson, A. J. Protein flexibility in virtual screening: the BACE-1 case study. *J. Chem. Inf. Model.* **2012**, *52*, 2697–2704.
- (52) Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A. Representing receptor flexibility in ligand docking through relevant normal modes. *J. Am. Chem. Soc.* **2005**, *127*, 9632–9640.
- (53) Dixit, A.; Verkhivker, G. M. Integrating ligand-based and protein-centric virtual screening of kinase inhibitors using ensembles of multiple protein kinase genes and conformations. *J. Chem. Inf. Model.* **2012**, *52*, 2501–2515.

- (54) Cheng, L. S.; Amaro, R. E.; Xu, D.; Li, W. W.; Arzberger, P. W.; McCammon, J. A. Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase. *J. Med. Chem.* **2008**, *51*, 3878–3894.
- (55) Ellingson, S. R.; Miao, Y.; Baudry, J.; Smith, J. C. Multi-conformer ensemble docking to difficult protein targets. *J. Phys. Chem. B* **2015**, *119*, 1026–1034.
- (56) Johnson, D. K.; Karanicolas, J. Selectivity by small-molecule inhibitors of protein interactions can be driven by protein surface fluctuations. *PLoS Comput. Biol.* **2015**, *11*, e1004081.
- (57) Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y. E.; Fleishman, S. J.; Corn, J. E.; Kim, D. E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popovic, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **2011**, *487*, 545–574.
- (58) *FastROCS*, 3.2.0.4; OpenEye Scientific Software, 2013.
- (59) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, *50*, 74–82.
- (60) *ROCS*, 3.2.0.3; OpenEye Scientific Software, 2013.
- (61) Lyskov, S.; Chou, F. C.; Conchuir, S. O.; Der, B. S.; Drew, K.; Kuroda, D.; Xu, J.; Weitzner, B. D.; Renfrew, P. D.; Sripakdeevong, P.; Borgo, B.; Havranek, J. J.; Kuhlman, B.; Kortemme, T.; Bonneau, R.; Gray, J. J.; Das, R. Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE). *PLoS One* **2013**, *8*, e63906.
- (62) Das, R.; Qian, B.; Raman, S.; Vernon, R.; Thompson, J.; Bradley, P.; Khare, S.; Tyka, M. D.; Bhat, D.; Chivian, D.; Kim, D. E.; Sheffler, W. H.; Malmstrom, L.; Wollacott, A. M.; Wang, C.; Andre, I.; Baker, D. Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home. *Proteins: Struct., Funct., Genet.* **2007**, *69* (Suppl 8), 118–128.
- (63) *Omega*, 2.5.1.4; OpenEye Scientific Software, 2013.
- (64) Hawkins, P. C.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572–584.
- (65) Higueruelo, A. P.; Jubb, H.; Blundell, T. L. TIMBAL v2: update of a database holding small molecules modulating protein-protein interactions. *Database* **2013**, *2013*, bat039.
- (66) Higueruelo, A. P.; Schreyer, A.; Bickerton, G. R.; Pitt, W. R.; Groom, C. R.; Blundell, T. L. Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database. *Chem. Biol. Drug Des.* **2009**, *74*, 457–467.
- (67) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Kruger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* **2014**, *42*, D1083–1090.
- (68) Kalliokoski, T.; Kramer, C.; Vulpetti, A.; Gedeck, P. Comparability of mixed IC₅₀(0) data - a statistical analysis. *PLoS One* **2013**, *8*, e61007.
- (69) Horvath, D.; Marcou, G.; Varnek, A. Do not hesitate to use Tversky-and other hints for successful active analogue searches with feature count descriptors. *J. Chem. Inf. Model.* **2013**, *53*, 1543–1562.
- (70) Gowthaman, R.; Miller, S. A.; Rogers, S.; Khowsathit, J.; Lan, L.; Bai, N.; Johnson, D. K.; Liu, C.; Xu, L.; Anbanandam, A.; Aube, J.; Roy, A.; Karanicolas, J. DARC: Mapping Surface Topography by Ray-Casting for Effective Virtual Screening at Protein Interaction Sites. *J. Med. Chem.* **2015**, DOI: 10.1021/acs.jmedchem.5b00150.
- (71) Willett, P. Similarity-based approaches to virtual screening. *Biochem. Soc. Trans.* **2003**, *31*, 603–606.
- (72) Brozell, S. R.; Mukherjee, S.; Baliaus, T. E.; Roe, D. R.; Case, D. A.; Rizzo, R. C. Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 749–773.
- (73) Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA* **2009**, *15*, 1219–1230.
- (74) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.
- (75) McGovern, S. L.; Shoichet, B. K. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **2003**, *46*, 2895–2907.
- (76) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **2012**, *52*, 1757–1768.
- (77) Bazzoli, A.; Kelow, S. P.; Karanicolas, J. Enhancements to the Rosetta energy function enable improved identification of small molecules that inhibit protein-protein interactions. *PLoS One* **2015**, *10*, e0140359.
- (78) Gowthaman, R.; Deeds, E. J.; Karanicolas, J. Structural properties of non-traditional drug targets present new challenges for virtual screening. *J. Chem. Inf. Model.* **2013**, *53*, 2073–2081.
- (79) Cosconati, S.; Forli, S.; Perryman, A. L.; Harris, R.; Goodsell, D. S.; Olson, A. J. Virtual Screening with AutoDock: Theory and Practice. *Expert Opin. Drug Discovery* **2010**, *5*, 597–607.
- (80) Cerqueira, N. M.; Gesto, D.; Oliveira, E. F.; Santos-Martins, D.; Bras, N. F.; Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Receptor-based virtual screening protocol for drug discovery. *Arch. Biochem. Biophys.* **2015**, *582*, 56–67.
- (81) Schonherr, H.; Cernak, T. Profound methyl effects in drug discovery and a call for new C-H methylation reactions. *Angew. Chem., Int. Ed.* **2013**, *52*, 12256–12267.
- (82) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity? *J. Med. Chem.* **2002**, *45*, 4350–4358.
- (83) Ajay, D.; Sobhia, M. E. Simplified receptor based pharmacophore approach to retrieve potent PTP-LAR inhibitors using apoenzyme. *Curr. Comput.-Aided Drug Des.* **2011**, *7*, 159–172.
- (84) Carlson, H. A.; Mastukawa, 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*, 2100–2114.
- (85) Damm, K. L.; Carlson, H. A. Exploring experimental sources of multiple protein conformations in structure-based drug design. *J. Am. Chem. Soc.* **2007**, *129*, 8225–8235.
- (86) Grohmann, D.; Corradi, V.; Elbasyouny, M.; Baude, A.; Horenkamp, F.; Laufer, S. D.; Manetti, F.; Bott, M.; Restle, T. Small molecule inhibitors targeting HIV-1 reverse transcriptase dimerization. *ChemBioChem* **2008**, *9*, 916–922.
- (87) Eyrisch, S.; Helms, V. Transient pockets on protein surfaces involved in protein-protein interaction. *J. Med. Chem.* **2007**, *50*, 3457–3464.
- (88) Bowman, G. R.; Geissler, P. L. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 11681–11686.
- (89) 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.