Diverse, High-Quality Test Set for the Validation of Protein−Ligand Docking Performance

Michael J. Hartshorn, Marcel L. Verdonk,* Gianni Chessari, Suzanne C. Brewerton, Wijnand T. M. Mooij, Paul N. Mortenson, and Christopher W. Murray

Astex Therapeutics, Ltd., 436 Cambridge Science Park, Milton Road, Cambridge CB4 0QA, United Kingdom

Received November 2, 2006

A procedure for analyzing and classifying publicly available crystal structures has been developed. It has been used to identify high-resolution protein−ligand complexes that can be assessed by reconstructing the electron density for the ligand using the deposited structure factors. The complexes have been clustered according to the protein sequences, and clusters have been discarded if they do not represent proteins thought to be of direct interest to the pharmaceutical or agrochemical industry. Rules have been used to exclude complexes containing non-drug-like ligands. One complex from each cluster has been selected where a structure of sufficient quality was available. The final Astex diverse set contains 85 diverse, relevant protein−ligand complexes, which have been prepared in a format suitable for docking and are to be made freely available to the entire research community (http://www.ccdc.cam.ac.uk). The performance of the docking program GOLD against the new set is assessed using a variety of protocols. Relatively unbiased protocols give success rates of approximately 80% for redocking into native structures, but it is possible to get success rates of over 90% with some protocols.

Introduction

Protein−ligand docking continues to be an area of intense interest to the pharmaceutical industry. Virtual screening is commonly used to generate hits against drug targets for which the structure is known, and docking is also heavily used in structure-based design projects to prioritize medicinal chemistry efforts.[1] Despite the success of docking methods, there is still a need to improve the performance of docking programs,[1,2] and one of the key steps toward this goal is to provide suitable validation sets to test methods.[3]

This paper is concerned with the construction of a new docking validation set, derived from the Protein Data Bank[4] (PDB), which it is hoped will prove useful in the assessment of docking methods. We decided that the requirements of such a validation set are as follows: (i) It is relevant to assess how useful the methods are to drug discovery (and agrochemical research). This means it should only include relevant protein targets containing drug-like complexes. (ii) It should contain a diverse list of protein−ligand complexes, with no particular target represented more than once, and a diversity of ligands with distinct molecular recognition types represented. This means that the set should be suitable to assess the general performance characteristics of a docking method. (iii) It should contain very high-quality experimental structures, and in particular, we would emphasize the availability of crystallographic structure factors so that the experimental binding mode of the ligands can be easily assessed. Ideally, the electron density (calculated from the structure factors) should account for all parts of the ligand. (iv) It should not contain complexes in which ligands make contact with protein atoms from multiple copies of crystallographically related subunits (unless the arrangement of the subunits is known to be physiologically relevant), because such contacts can sometimes affect the binding modes of ligands at the point of contact. (v) It should be sufficiently large so that improvements and differences in performance can be more reliably assigned to a real effect rather than random differences driven by the composition of the test set. (vi) It should contain relatively recent structures. This ensures the use of modern ligand geometry generation and refinement programs. Another reason for this restriction is to ensure that there is no overlap with the existing members of the CCDC/Astex test set (see below) so that the new test set could be used independently from older sets. (vii) It should be freely available to academics, software producers, and industrial companies so that the field as a whole could benefit from the time taken both to define the set and to prepare the ligands and proteins in a suitable form for docking.

A number of docking studies have been performed in the literature in which the validation has employed more than 50 protein−ligand complexes (i.e., the validation is consistent with requirement (v) above). The original GOLD test set consisted of 134 complexes.[5] These complexes later formed a large part of the 200 complexes used in the FlexX validation set[6] and the 154 complexes used in the testing of EUDOC.[7] Another substantial list of 284 complexes was used to test GLIDE,[8] and this was derived mainly from the FlexX validation set and complexes from the set used to derive the empirical scoring function Chemscore.[9] Similarly, the recent docking comparison by Chen et al.[10] uses 164 complexes and the testing of ConsDock by Paul et al.[11] uses 100 complexes, but in both cases, the vast majority of examples are derived from the GOLD and/or Chemscore sets. A criticism of all these validation sets is that despite being large in size they have been put together on a somewhat ad hoc basis, often leaning on the makeup of existing sets. In many ways, this is a good thing because it allows some comparison between different programs, but it means that any deficiencies in the sets are inherited from one performance assessment to another. Additionally, no formal attempt has been made to assess the diversity or relevance of the complexes included.

In 2004, the CCDC/Astex validation set was constructed.[12] This set contains 305 complexes incorporating the original 134 complexes of the GOLD set, 48 nonoverlapping complexes from the Chemscore set, and 123 new complexes. A substantial

* To whom correspondence should be addressed. Tel.: +44 1223 226206. Fax: +44 1223 226201. E-mail: m.verdonk@astex-therapeutics.com.

10.1021/jm061277y CCC: $37.00
Published on Web 01/19/2007

attempt was made to provide a clean subset of this list in which doubtful entries were removed on the basis of the following: (i) contacts between the ligand and crystallographically related chains; (ii) significant clashes between the protein and ligand; (iii) incorrect ligand representation; (iv) dubious ligand geometry; or (v) incongruity between ligand placement and electron density. Approximately 20% of the structures were considered doubtful and were removed using these criteria, and a further 6% of complexes were removed for diversity reasons, leaving a clean list of 224 compounds. The number of doubtful entries in the original list was worrying because many of these complexes had been routinely used in docking validations by different research groups. Also of some concern was the fact that structure factors had not been deposited in the PDB for about 75% of the clean list, so it was not possible to assess the ligand positioning versus the experimental electron density. In the construction of the Astex/CCDC set, there was no effort to restrict the list to only drug-like ligands or to proteins relevant to drug discovery.

Perola et al. have constructed a validation set that consists of 100 PDB entries and 50 complexes from their corporate database.[13] The authors have worked hard on ensuring that the complexes are relevant to drug discovery (requirement (i) above), and they have also considered diversity in the construction of the set, although it does contain multiple representations of the same protein. An amount equal to 44% of the publicly available complexes are kinases and 42% are proteases. The Perola set, therefore, does not match the diversity requirement that has been targeted in our work. Additionally, Perola et al. have not prioritized the availability of structure factors and electron density data in selecting entries. To illustrate the importance of this issue, Perola et al. identified 12 complexes for which none of the docking programs studied generated a good quality solution. Structure factors are available for three of these named complexes (1cet, 1nhu, and 1nhv). These three complexes would have been discarded in our work because we were not able to generate an electron density map that provides good support for the positioning of the entire ligand. A final potential problem with the Perola validation set is that there is no indication in the paper that the set is available to the research community.

From the foregoing discussion, it should be apparent that there is a great need for an entirely new validation set that is based on the six requirements outlined above. The present paper first describes the methodology used in the construction of such a set, which is based on a systematic analysis and classification of all the protein−ligand complexes in the PDB. In the next section, the composition of the set is discussed with regard to the representation of different protein classes, the presence of marketed drugs, and the availability of affinities. A key purpose of the validation set is to facilitate standardization of tests on docking programs, and so, the final part of the paper outlines the performance of the docking program, GOLD, on the validation set. Here, a number of different protocols are used that have been designed to mimic the type of protocols that have been employed in previous validation studies. The results illustrate that it is possible to get very different success rates depending on the protocol adopted and also underline how difficult it can be to compare the performance of different docking programs.[3] It is hoped that these results will be useful in benchmarking the performance of different programs against the new validation set in future studies.

Materials and Methods

Selection of Complexes. Sequence Analysis. The first stage in preparing the validation set was to perform a sequence analysis of the proteins contained within the PDB. The purpose of this analysis was to group together all of the amino acid chains that represent the same protein structure, for example, the cdk2 kinases, the HIV proteases, and so on. The analysis was performed using a local mirrored copy of the PDB as of June 27, 2006, which contained a total of 31 875 entries.

Each PDB file was processed to extract the amino acid sequence(s) contained within it. This was done by examining each alpha carbon (CA) atom and converting the residue name to a single letter amino acid code. Care was taken to handle residues with disordered CAs appropriately. The residue was only considered if the alternate location indicator was a space character or an A. New sequences were started whenever the chain name changed or a TER record was encountered. PDB files were ignored if they did not originate from experimental X-ray crystallography studies. This was defined by examining the value in the EXPDAT record of the PDB file.

A total of 62 727 amino acid chains were identified from the PDB. A Blast[14] database was constructed from these sequences, which was to be used for clustering the sequences into highly similar families. Each sequence from the set was compared against the rest of the database using a blast search with a p-value cutoff of 1.0 and with low complexity filtering turned off. This yielded a set of very similar sequences for each sequence in the input and it was empirically determined that a Blast sequence identity of >75% indicated that two sequences were of the same protein. Each set of matching sequences was assigned to a cluster. Once a sequence was part of a cluster, it was not available for inclusion in other clusters. This method robustly assigns the PDB sequences to clusters comprised of the same protein and handles the following difficulties that can occur with the sequences present in a PDB file: (i) A particular PDB file may contain chains of more than one protein. (ii) There may be loops missing because they could not be identified in the electron density of the structure. (iii) Different structures of the same protein may have used different constructs, so there may be N or C terminal extensions. (iv) There may be terminal extensions due to the presence of purification tags. (v) There may be a small number of mutations that were incorporated for crystallographic purposes.

This clustering procedure yielded a total of 9188 clusters, each corresponding to a distinct protein.

Ligand Analysis. To complete the characterization of protein/ligand complexes in the PDB, it is necessary to identify the chemical structures of ligands that are part of the PDB entry. Unfortunately, the PDB does not contain all of the information necessary to describe the valence structure of a general small molecule ligand (such as a drug or an inhibitor). Among other things, the following difficulties occur when trying to identify a small molecule structure present in a PDB file: (i) The bond orders of ligands are not specified in a standard way. (ii) The charge state of atoms in the ligand is not specified. (iii) Sometimes even identifying the element type of an atom is not trivial.

These problems greatly diminish the value of the PDB for storing protein/ligand complexes. A number of approaches have been devised for assigning the likely structure of a ligand from the information that is present in a PDB file. These range from manual curation methods, such as those used by Relibase,[15] to a number of geometric analyses that will generate a sensible valence structure for a bound ligand. To extract the ligand

structures, we have implemented our own algorithm based on the one described by Sayle[16] with a few extensions. The algorithm is briefly described here: (i) The PDB file is read and connectivity is assigned using standard bonding radii. (ii) Potential ligands are identified as groups of connected HETATM residues or connected components with less than 100 atoms. (iii) Elements are assigned from element codes stored in the ATOM record or from the ligand atom name using a set of rules. (iv) Hybridization of each atom center is assigned on the basis of the sum of the bond angles around the atom. (v) Where possible, five- and six-membered rings are made aromatic. (vi) A Kekulé structure is assigned by searching for the optimal arrangement of single and double bonds to aromatic atoms.

The aim of the ligand analysis is to identify a set of high-quality protein/ligand complexes, where the ligands are of therapeutic or agrochemical interest. Many protein structures in the PDB have cofactors or crystallization additives bound to them. These are of little interest for this work and it was decided to remove them completely from the analysis. This was done by drawing up a list of HET group identifiers that correspond to these ligands and ignoring all atoms that have identifiers in this list.

The starting list of identifiers was based on that described by Wang et al.[17,18] During the preparation of the ligand database, it was found that some additional HET groups occurred frequently in the PDB. The list of ignored HET group identifiers was extended by including all HET groups that occurred more than 20 times in the PDB (see Supporting Information).

The output from the ligand extraction phase is a series of MDL MOL files[19] that contain a sensible valence structure for a particular ligand. These ligands were converted into SMILES strings[20] using the Daylight program mol2smi.[21] Once in this form, the structures were stored in a searchable database. The structures were then annotated with physical properties that are necessary for classifying them as drug-like or not drug-like. These physical properties include molecular weight, ClogP[22] (calculated octanol/water partition coefficient), polar surface area, number of hydrogen bond donors, number of hydrogen bond acceptors, and heavy atom count.

Once the ligands had been stored and annotated, a number of filtering rules were devised to identify those that were drug-like. The filters were split into two groups: (i) Chemical pattern filters that remove any remaining ligands that slipped through the HET group filter. The chemical filters are described in Table 1. (ii) Physicochemical filters that remove molecules that do not have drug-like properties. These filters are also shown in Table 1.

PDB File Analysis. Clash/Symmetry Contacts. As described in the Introduction, a common problem with previous validation sets is that they contain structures with poor nonbonded interactions between the ligand and the active site, or that the ligand has interactions with a symmetry related molecule that is not physiologically relevant. Both of these problems introduce serious limitations on the usefulness of such complexes: (i) If the ligand has clashes with the active site, then the binding mode is probably incorrect. (ii) If the ligand is sandwiched between symmetry related protein molecules, then it is possible that these interactions influence the binding mode.

For these reasons, we have tried to identify all complexes that contain serious clashes between the protein structure and the ligand and those where the ligand forms any short-range nonbonded interactions with a symmetry copy of the protein or cofactors. Hydrogen atoms were ignored when determining clashes and symmetry interactions.

Table 1. Additional Rules for Identifying Compounds That Are Likely To Be Drug-Like[a]

structure filters		physicochemical filters	
structure contains ligand	3434	exclude compounds with more than 10 rotatable bonds	1653
must contain oxygen or nitrogen	3400	exclude compounds with fewer than 10 heavy atoms	1356
must only contain H, C, N, O, F, P, S, Cl, and Br	2731	exclude compounds with MW outside range 100−600	1318
exclude common glycosylation structures	2692	exclude compounds that fail the rule of five[41]	1310
exclude simple saccharides	2477		
exclude tetrapeptides or bigger	2000		
exclude compounds containing $CH_2CH_2CH_2CH_2-$ linkers	1896		
exclude glycerol-like compounds	1821		

[a] Numbers of complexes surviving each filter are shown in adjacent column.

The clash checking was performed by reading the MOL file of each ligand and comparing it with the atoms in the original PDB file. Care was taken to ignore the atoms of the ligand that are present in the original PDB file. Atom Van-der-Waals radii were assigned using the rules of Tsai et al.,[23] and a clash was recorded if any pair of atoms was closer together than the sum of the radii minus 0.9 Å. We believe this to be a forgiving definition of a poor contact and it serves only to remove the very worst complexes.

To identify interactions with the symmetry environment, we used symmetry generation tools from our AutoSolve suite of programs.[24] A 10 Å shell of symmetry atoms was generated around all atoms present in the PDB file. Each ligand MOL file was compared against this shell of atoms, and if any contacts shorter than the sum of atom radii + 0.5 Å were found, a symmetry contact was recorded for that ligand.

Electron Density Generation. For each PDB file, we recorded whether or not structure factors are available in the PDB structure factor repository. Structure factors are essentially the primary experimental data for a crystal structure determination and can be used to calculate electron densities for all or part of a structure. We have found examination of electron densities to be extremely helpful in our structure-based drug design projects.[25] We decided to exclude structures for which we were not able to construct electron density maps.

Where structure factors were available, we constructed an OMIT map for each ligand using the procedure outlined in Mooij et al.[24] The OMIT map contains a close approximation to the original density observed for the ligand. This is similar in spirit to the approach adopted by Kleywegt for producing electron density maps in the Electron Density Server,[26] but we prefer to produce an F_o-F_c map rather than a $2F_o-F_c$ map. This facilitates the visualization of the ligand electron density because there is no electron density visible for correctly interpreted parts of the protein structure.

The OMIT maps were generated for all structures and then clipped to the region around the ligand atoms to minimize the storage requirements for the electron density maps. Although for the vast majority of complexes our software is able to recreate the ligand electron density maps, this was not possible

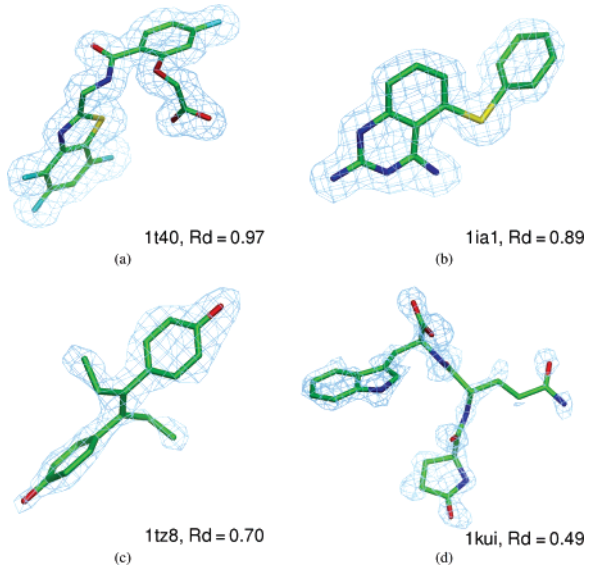

Figure 1. Complexes with a range of Rd coefficients. The first two complexes (1t40 and 1ia1) show the typically very high quality of the complexes that are included in the Astex Diverse Set. The 1tz8 complex has clear electron density for all atoms but shows the limit of what we considered for inclusion in the set. For entry 1kui, there is clear electron density for large parts of the ligand but there is no density for other parts. As a result, this last complex would not have passed our filtering criteria. All figures containing structures were prepared using AstexViewer.[28]

for a small number of complexes; the reasons for this have been described by Kleywegt in some detail.[26] Unfortunately, in these cases we had to exclude the structure from further examination.

Electron Density Scoring. The final analysis of the high-quality protein/ligand complexes in the PDB involved manual inspection of a large number of structures. To assist with this process we calculated a density correlation score (Rd) for each complex.[24] The density correlation measures how well the observed electron density correlates with the positions of the atoms in the ligand structure.

A perfect fit between electron density and ligand structure gives a correlation coefficient of 1.0. We found that structures with Rd > 0.9 were generally of very high quality. Structures with Rd < 0.7 were generally found not to be suitable for inclusion in the set. Lower correlation coefficients can result from weaker experimental support for the ligand positioning but very low coefficients can also indicate a technical issue for that specific structure. For example, errors in the specification of the cell dimensions or space group can lead to incorrect maps being generated. In the vast majority of cases, however, our map generation protocol yields high-quality maps. Figure 1 shows examples of ligand electron densities (F_o-F_c maps) with varying correlation coefficients.

Final Filtering. The previous sections have described the steps taken to process the PDB to produce clusters of related protein structures and chemical structures for the ligands that are bound to them. This section describes the process by which the final set of structures is identified before manually selecting the high-quality complexes of interest.

We are interested only in modern, high-quality protein structures, and for this reason, we limit the selection to structures determined after August 11, 2000 that have a resolution of 2.5 Å or better and for which structure factors are available. The limit on the date for the structures is chosen to reflect the fact that many improvements have been seen recently in methods for the determination of protein/ligand complexes, including the use of ligand structure generation programs such as Corina,[27] and improvements in the quality of crystallographic refinement procedures. Additionally, this date ensures that there is no overlap with structures from the Astex/CCDC validation set that was described previously.[12]

The final data set from this whole procedure is a set of clusters of protein structures. Each cluster contains only PDB structures that satisfy the conditions described above and that contain ligands that meet the criteria outlined in previous sections. To summarize, the remaining protein/ligand complexes meet the following criteria: (i) PDB file is deposited after 11-Aug-2000 (18 677 structures); (ii) PDB file resolution is 2.5 Å or better (12 650 structures); (iii) PDB file has structure factors deposited (9706 structures); (iv) the structure contains interesting ligands (3434 structures); (v) the structure contains drug-like ligand(s) (1310 structures); and (vi) the structure contains a ligand with no clashes with the binding site or interactions with symmetry units (836 structures)

Manual Selection. The automated filtering rules described above produced 427 occupied clusters, that is, modern, high-quality structures with drug-like ligands and electron density maps. Unfortunately, it is not possible to fully automate the final selection of the complexes. For example, it is not trivial to automatically detect which ligands are drugs and which targets are of relevance to drug discovery or agrochemistry and so on. Therefore, an interface was developed that presented the clusters in order of decreasing number of protein structure members.

Clicking on a cluster produces a list of the PDB structures that are in that cluster along with summary information about date of deposition, resolution, and the name of the protein as determined from the PDB file. Also the highest Rd for any of its ligands is shown. The list is sorted into decreasing order of Rd so that it is simple to identify the complexes with the best ligand density.

Selecting a structure in a particular cluster will show the ligands that are present in that PDB file. A chemical diagram for the ligand is shown along with the molecular weight, ClogP, the Rd, and the average B-factor and occupancy for atoms in the ligand. When a particular ligand is selected, the protein structure, the ligand structure, and the electron density map are displayed in a visualization window containing AstexViewer.[28] In this fashion, the candidate structures in each of the 427 clusters were examined, and where possible, one suitable representative for each cluster was selected.

While browsing the clusters, the header for each loaded PDB file is available. This makes it very easy to read some information about the protein/ligand complex and to judge if it is relevant to drug discovery. However, the final judgment on relevance is made when the primary literature for the complex has been consulted (see below). We would like to stress that at no point was a decision to include a complex biased by how well we believed a docking program would be able to dock the ligand. Each decision was based solely on the quality of the data and on the relevance of the complex. Hence, if there are any biases in the set, they merely reflect biases in the availability of relevant high-quality complexes in the PDB.

Composition of the Set. The initial manual selection resulted in candidate complexes, each representing one single protein cluster. The primary literature was then consulted for each complex and complexes were removed if there was no indication that the target was (potentially) relevant to the discovery of drugs or agrochemicals. Other reasons for excluding complexes at this stage included cases where the ligand had reacted during cocrystallization or soaking, where there were incomplete side

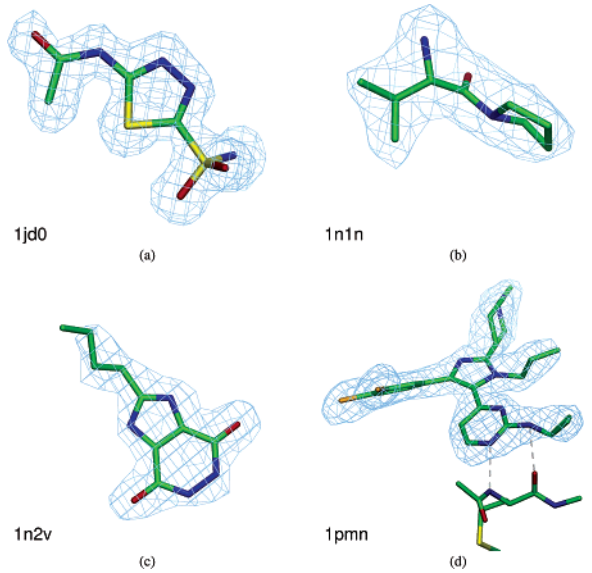

Figure 2. Examples of small structural issues that were deemed acceptable in the Astex Diverse Set. (a) In entry 1jd0, the ligand contains a cis amide bond that should be trans; it is difficult to tell from the electron density alone whether the amide should be cis or trans, but energetically the cis conformation is extremely unlikely. (b) In entry 1n1n, the nitrogen atom in the pyrrolidine ring should be planar, as it is part of an amide bond. (c) In entry 1n2v, there is no density for the terminal atom in the *n*-butyl chain of the ligand. (d) In entry 1pmn, the ligand amino-cyclopropyl group has not been placed in the density optimally; twisting the cyclopropyl would have resulted in a better fit with the density and would have created a much-improved geometry for the hydrogen bond formed between the ligand amine and the protein backbone carbonyl oxygen.

chains or missing loops in the vicinity of the binding site, or where the binding site contained mutated residues that were not relevant to drug discovery. Where possible, complexes that were removed at this stage were replaced with another acceptable complex from the same cluster. Complexes with minor structural issues were retained to keep the set at a reasonable size. A few examples of such minor structural issues are given in Figure 2.

This process resulted in 85 complexes that contain a drug-like ligand for which the electron density is essentially complete and unambiguous. Figure 3 shows a phylogenetic tree of the 85 complexes clustered based on global sequence identity of the proteins. It illustrates that, although the set contains several classes of targets, it is generally very diverse. In Figure 4a, the 85 targets are classified according to their relevance to drug or agrochemicals discovery. Nearly 90% of the proteins are direct targets for drug/agrochemicals discovery projects. Most of the remaining proteins are so-called "off-targets", which include cytochrome P450s and so on.

Chemical diagrams for the ligands in the set are shown in Figure 5, whereas Figure 4b shows the types of ligands that occur in the set. Of the 85 ligands in the set, 23 are approved drugs and an additional 6 are (or have been) in clinical trials. Where it is clear from the primary literature that the ligand originated from a drug discovery project, we have classified it as such; this class contains 35 ligands. The "substrates" group contains natural substrates, products, agonists, and so on, whereas the "substrate analogues" group contains close analogues of natural substrates for which it is not directly obvious from the primary literature if they originated from a drug discovery project. The "unassigned" group contains ligands for which we were unable to classify them in any of the other categories on the basis of the primary literature.

Figure 6 shows the distributions of the number of heavy atoms, of the number of rotatable bonds in the ligands, and of the crystallographic resolution of the 85 structures. It is clear that we have included fewer complexes with very small ligands and fewer complexes containing very large ligands compared to that of the CCDC/Astex validation set; in addition, we have included significantly fewer complexes with ligands containing a large number of rotatable bonds. Both these distributions are consistent with our aim to include only complexes containing drug-like ligands. No complexes were included in the Astex Diverse Set with resolution worse than 2.5 Å simply because of the resolution cutoff we applied. The new set contains a larger (but still relatively small) fraction of very high-resolution structures, probably due to the improvements in technology achieved in recent years. Additional distributions for various properties of the Astex Diverse Set can be found in the Supporting Information.

Where available in the primary literature, either directly or in one of the references, the potencies of the compounds were recorded and are listed in Table 2. Using these sources, we found potency data for 74 of the 85 complexes.

One of our initial concerns on looking through the list was the absence of some high profile drug targets for which structures are available. Such omissions in the list were not especially worrying when they came from families that were well represented in the set, such as serine/threonine/tyrosine kinases (example omissions: src, cdk4, mek), trypsin-like serine proteases (example omissions: thrombin, factor IX), metalloproteases (various MMPs omitted). Omissions that were of more concern were HCV protease, HCV polymerase, BACE/renin, elastases, PTP1B, and PPARγ/PPARδ. A detailed examination revealed that complexes involving these proteins had been discarded for a wide variety of reasons and often for multiple reasons per target or complex. The lack of availability of structure factors was the most common reason for omission, and this was particularly true for older drug targets such as PTP1B; the insistence on noncovalent inhibitors was important to the removal of several targets (e.g., HCV protease and elastase), whereas a number of more challenging drug targets were affected by the need to contain drug-like ligands (e.g., BACE, renin, HCV protease, and PPARγ/PPARδ). Poor experimental support for all atoms in the ligand (i.e., disorder) led to the removal of occasional complexes, as did the insistence on a high-resolution structure (e.g., for HCV polymerase and PPARγ/PPARδ). In conclusion, the examination of omissions revealed that targets and the associated complexes were excluded for a wide variety of valid reasons and that we can be safe in the assertion that there are only about 85 complexes from distinct targets that meet the strict criteria adopted in this paper.

Setting Up Complexes. The ligand and protein were set up according to the rules laid out below. For the protein, the rules were only applied to atoms within 10 Å of the ligand.

Bond types for the ligands were taken directly from the primary literature. In cases where the primary literature indicated what the protonation state of the protein or ligand should be, these assignments were generally adopted. Otherwise, protonation states were assigned according to the following rules. Metal-coordinating groups in protein and ligand were protonated in accordance with what is typically observed in the Cambridge Structural Database (CSD).[29] For example, phenolic OH groups that coordinate metal ions are always deprotonated in the CSD, whereas aliphatic OH groups nearly always retain their proton when they coordinate to a metal ion. Hence, metal-coordinating tyrosine OH groups were deprotonated, but serine and threonine

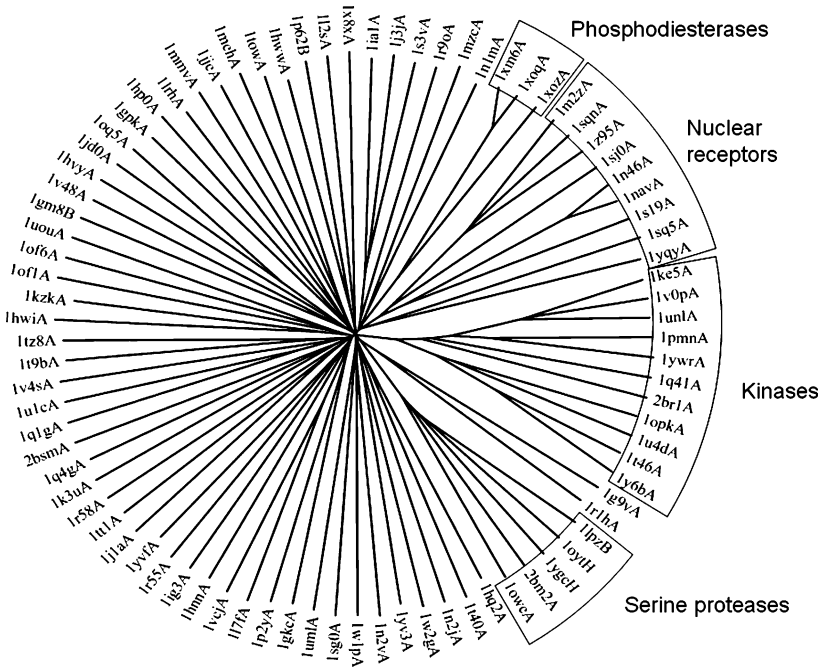

Figure 3. Phylogenetic tree of protein similarity in the Astex Diverse Set. The majority of structures in the set have almost no global sequence similarity. There are groups of structures from major protein families in the set. There are a total of 11 kinases, 9 nuclear receptors, 5 serine proteases, and 3 members of the phosphodiesterase family.

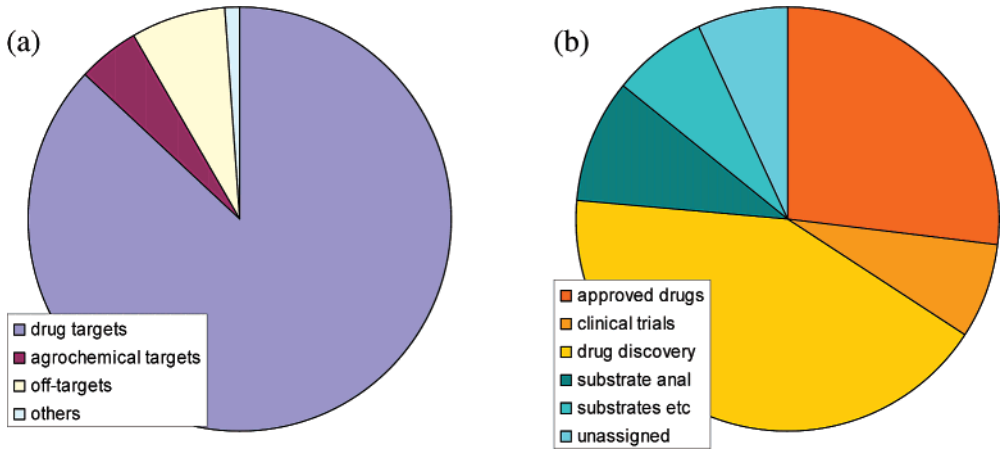

Figure 4. Protein classes (a) and ligand classes (b) in the Astex Diverse Set. The vast majority of proteins are drug targets or agrochemical targets, and most ligands are either approved drugs, compounds in clinical trials, or compounds originating from drug discovery projects.

OH groups that coordinate metals were left protonated. Protonation states of hydrogen-bonding groups were assigned according to our best judgment on their relative pK_as. Groups not involved in hydrogen bonds were protonated according to their pK_as.

Tautomers were assigned from the primary literature where they were described. Otherwise, tautomeric forms were chosen that were consistent with the metal-coordinating contacts and the hydrogen bonds. In cases where tautomers could not be derived from the interaction patterns formed by the functional groups involved, the tautomer that appeared most sensible was chosen.

Ligand hydrogen atoms were added using InsightII,[30] but for certain functional groups (all involving hydrogen atoms added to sp3 carbon atoms), this resulted in poor covalent bond angles. Hence, all hydrogens attached to sp3 carbon atoms were optimized in the MMFF force field[31] (keeping all other atoms fixed).

His, Asn, and Gln side chains can normally not be placed into the electron density map unambiguously. Hence, the crystallographer will usually have placed the side chain according to the hydrogen bond patterns. However, upon visual inspection, a few cases were identified where the side-chain placement was inconsistent with the hydrogen bond patterns. Hence, we searched (computationally) all 85 complexes for such inconsistencies, using the following protocol: (i) only His, Asn, or Gln side chains were checked with at least one side chain nitrogen or oxygen atom within 6 Å of a nonhydrogen ligand atom; (ii) water molecules were retained in this analysis; (iii) the number of hydrogen bonds formed was counted for all possible states of the side chain; for Asn and Gln side chains, these were simply the two rotameric states of the terminal carbamoyl group; for His side chains, in addition to the two rotameric states, two neutral uncharged states and the positively charged state were considered, resulting in a total of six states; (iv) the side chains that could achieve a higher number of hydrogen bonds by choosing an alternative rotamer were inspected visually; and (v) if visual inspection confirmed that indeed the hydrogen-bonding patterns of a side chain could be significantly improved by rotating it, then the side chain was flipped.

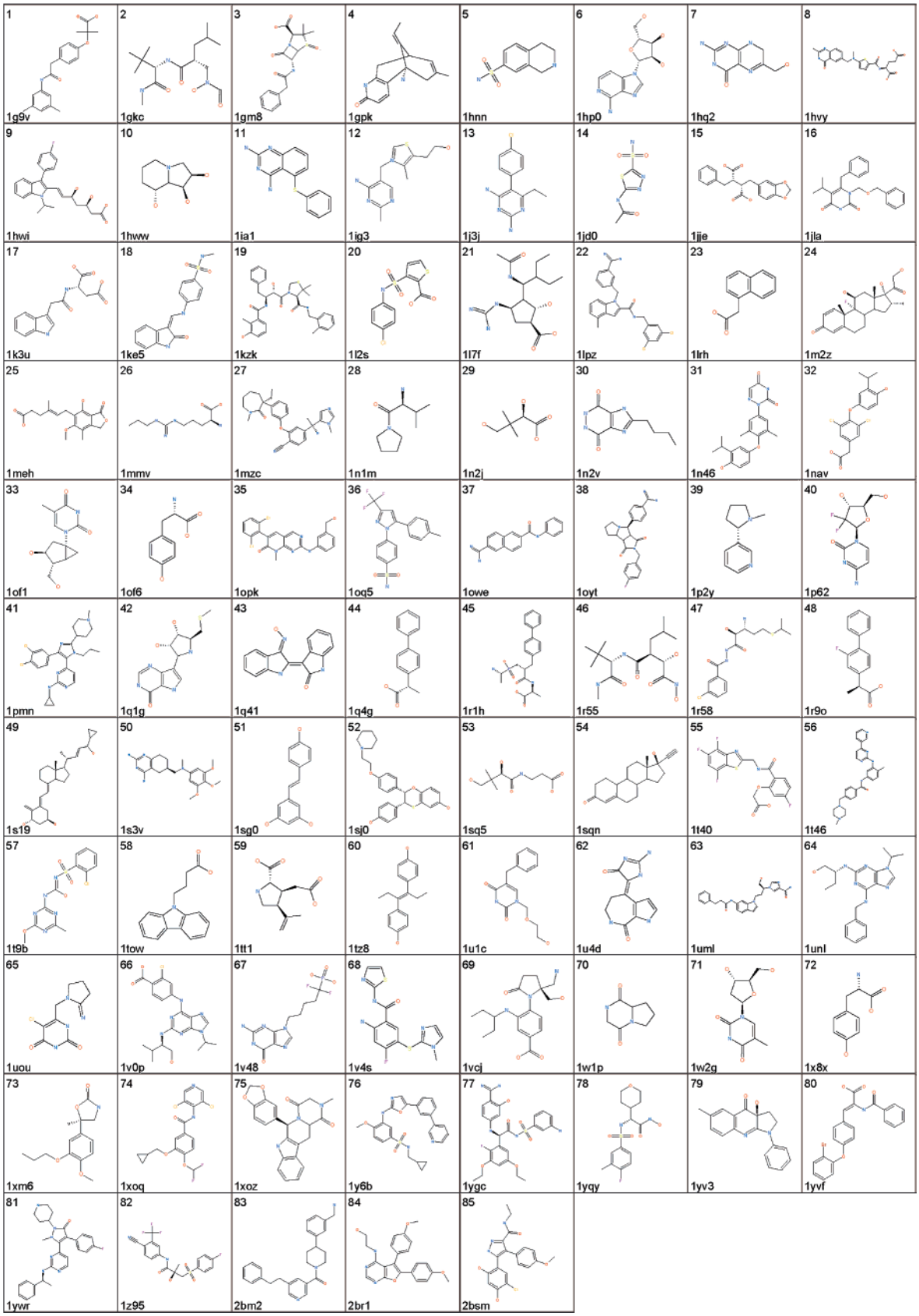

Figure 5. Chemical diagrams of the 85 ligands in the Astex Diverse Set.

In cases where protein side chains were disordered, the following approach was taken to select one of the conformers: (i) where possible, the conformer with the highest occupancy was selected; (ii) for side chains with equal occupancy, in cases where one of the conformers clashed with the ligand, the other conformer was selected; (iii) if no decision could be made based on rules (i) and (ii), the conformer listed first in the PDB file was selected.

All water molecules were removed from the complexes, except in cases where GOLD requires them to determine the

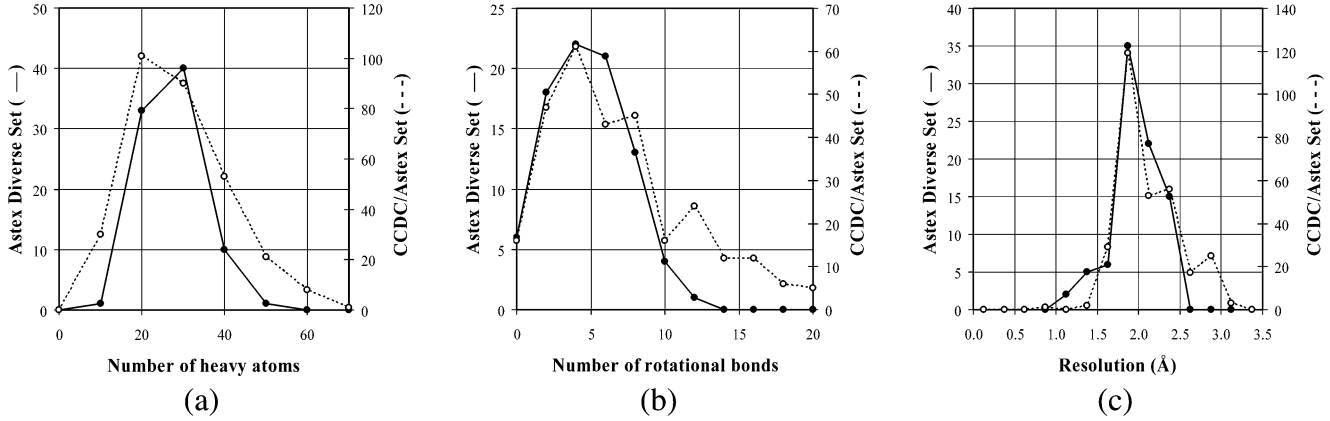

Figure 6. Distributions of heavy atom count (number of nonhydrogen atoms) in the ligand (a), number of ligand rotational bonds (b), and resolution of the crystal structures (c) for the CCDC/Astex Set (dotted lines) and the Astex Diverse Set (solid lines).

coordination geometry of a metal ion (this was the case for entries 1xm6 and 1xoq; in neither of these two cases do the water molecules interact directly with the ligand). Solvents and small ions like sulfates and phosphates were removed unless they mimic a cofactor. For example, in the thymidine kinase complex (1of1), the sulfate ion was retained because it mimics ATP. All cofactors were retained and were atom typed according to the rules outlined above.

Results and Discussion

Standard Protocol. It is not the aim of this paper to test the docking performance of GOLD (or any other protein/ligand docking code) against the Astex Diverse Set, but rather to present and discuss this new set. However, it is the intention that researchers will use the set to improve their docking protocols and to compare the performance of their algorithms against other docking codes. Particularly in the light of this latter use of the new validation set, we believe it is good to stress again here that it is incredibly difficult to compare the performance of different docking tools in a fair way (see, e.g., Taylor et al.[32]). Docking programs can differ in their definition of the binding site, the way the complexes are preprocessed, the number of degrees of freedom considered, and so on, and this can have a significant effect on the success rates obtained. Therefore, to illustrate this, we present the performance of GOLD using several of these protocols, all of which have been used in the literature for evaluating other protein/ligand docking tools.

For all these validation runs, an in-house version of GOLD was used, the "default 1" settings described previously[33] were used for the genetic algorithm (GA), and the standard GOLD scoring function Goldscore[5] was used to drive the dockings. The results for the different protocols are shown in Table 3. Additional statistics on the docking runs can be found in the Supporting Information.

First, we tested the standard protocol that was used in previous validations of GOLD against a test set of protein/ligand complexes,[33] that is: (i) geometries for protein and ligand are taken directly from the PDB; (ii) binding sites are defined as all protein atoms within 6 Å of a nonhydrogen ligand atom; (iii) protein OH groups and NH_3 groups can rotate; and (iv) flexible ligand torsions can rotate and ring corners in ligands are allowed to flip to explore ring conformers. Using this standard protocol, GOLD predicts 81% of the complexes within 2.0 Å of the experimental binding mode. This is comparable to the GOLD success rate achieved against the "drug-like" subset of the CCDC/Astex validation set, for which, using the same protocols, Goldscore achieved a success rate of 79%.[33] Care should be taken not to overinterpret results obtained against different sets, as the sampling errors in validation sets of this size are 4−5%.[12]

Because in this validation set the structures themselves are not in doubt, docking failures are particularly interesting and analyzing them should help in the development of improved scoring functions. Although it is beyond the scope of this paper to produce an improved scoring function, it is interesting to highlight a few cases where GOLD fails to reproduce the X-ray binding mode and discuss possible reasons why this may be the case.

For a significant fraction of the complexes for which GOLD systematically struggles to dock the ligand correctly, water molecules play a key role in protein/ligand recognition. Two extreme examples are 1g9v and 1gm8, for which the ligand only forms one direct hydrogen bond with the protein but is involved in several water-mediated hydrogen bonds; in 1g9v, there are 11 water molecules in the direct vicinity of the ligand, and for 1gm8, there are 19 water molecules nearby. For 1hvy, the key recognition motif is reproduced almost exactly, but two acid groups that are solvent exposed (and together form hydrogen bonds to five water molecules) are misdocked. In the X-ray structure of 1xm6, the ligand does not interact with the zinc ion at the far end of the pocket, instead the ion is coordinated by two water molecules. As these water molecules are absent in the docking runs, GOLD predicts the ligand to coordinate to the zinc ion. The same happens for 1r9o, where in the X-ray structure a water molecule coordinates to the heme and the ligand acid group interacts with Arg108. In the absence of the water molecule, GOLD predicts the ligand acid group to coordinate to the heme. In addition, the standard version of Goldscore may over-reward the interaction between the ligand acid group and the iron atom in the heme. We recently reparameterized the Fe-coordination terms in Goldscore.[34] Two sets of new parameters were derived: one from Fe-coordination frequencies in the CSD and the other from Fe-coordination frequencies in the PDB. Interestingly, using either of these two new sets of parameters, the binding mode of this entry is predicted correctly. In real-life drug discovery applications, the important mediating water molecules would be included in the docking runs, giving the docking program a better chance of making correct predictions.

In 1yvf, the ligand contains a carboxylic acid group that is solvent exposed and forms no hydrogen bonds with the protein. The Goldscore function heavily rewards hydrogen bonds formed by charged groups and, hence, attempts to form hydrogen bonds

Table 2. Protein and Ligand Names for the 85 Complexes in the Set[a]

entry	target	compound	potency measure	potency (μM)	ref(s)
1g9v	deoxy hemoglobin	RSR-13			42
1gkc	matrix metalloprotease 9	reverse hydroxamate inhibitor			43
1gm8	penicillin G acylase	PGSO	K_m	16	44
1gpk	acetylcholinesterase	huperzine A	K_i	4.3	45
1hnn	phenylethanolamine *N*-methyltransferase	SK&F 29661	K_i	0.58	46
1hp0	purine specific nucleoside hydrolase	3-deaza-adenosine	K_i	0.2	47
1hq2	6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase	HP	K_d	0.17	48;49
1hvy	thymidylate synthase	tomudex	K_i	0.67	50;51
1hwi	HMG-CoA reductase	fluvastatin	IC_{50}	0.028	52
1hww	α-mannosidase II	swainsonine	IC_{50}	0.02	53
1ia1	dihydrofolate reductase	compound 3	IC_{50}	0.034	54
1ig3	thiamin pyrophosphokinase	thiamin or vitamin B1			55
1j3j	dihydrofolate reductase	pyrimethamine	K_i	0.0098	56
1jd0	carbonic anhydrase XII	acetazolamide	K_i	0.0057	57;58
1jje	metallo β-lactamase	compound 11	IC_{50}	0.0037	59
1jla	HIV-1 reverse transcriptase	TNK-651			60
1k3u	tryptophan synthase	*N*-[1*H*-indol−3-yl-acetyl]aspartic acid			61
1ke5	cyclin-dependent kinase 2	compound 98	IC_{50}	0.56	62
1kzk	HIV-1 protease	JE-2147 (also named AG1776 or KNI-764)	K_i	0.000 041	63
1l2s	β-lactamase	compound 1	K_i	26	64
1l7f	neuraminidase A	BCX-1812	IC_{50}	0.0008	65;66
1lpz	factor Xa	compound 41	K_i	0.025	67
1lrh	auxin-binding protein 1	1-naphthalene acetic acid			68
1m2z	glucocorticoid receptor	dexamethasone	K_d	0.06	69
1meh	inosine monophosphate dehydrogenase	mycophenolic acid			70
1mmv	neuronal nitric-oxide synthase	NG-propyl-L-arginine	K_i	0.057	71
1mzc	protein farnesyltransferase	compound 33a	IC_{50}	0.000 06	72
1n1m	dipeptidyl peptidase IV	valine-pyrrolidine	K_i	2	73
1n2j	pantothenate synthetase	pantoate			74
1n2v	tRNA-guanine transglycosylase	compound 6	K_i	83	75
1n46	thyroid hormone receptor β1	compound 3	K_i	0.000 03	76
1nav	thyroid hormone receptor α1	compound 15	IC_{50}	0.025	77
1of1	thymidine kinase	(*S*)-MCT	K_m	4.1	78
1of6	DAHP synthase	tyrosine	K_i	0.9	79
1opk	c-Abl tyrosine kinase	PD166326	IC_{50}	0.000 15	80
1oq5	carbonic anhydrase II	celecoxib	IC_{50}	0.0021	81
1owe	urokinase	compound 6	K_i	0.631	82
1oyt	thrombin	compound 4	K_i	0.057	83
1p2y	cytochrome P450cam	nicotine	K_s	10	84
1p62	deoxycytidine kinase	gemcitabine	K_m	22	85
1pmn	c-Jun terminal kinase 3	compound 1	IC_{50}	0.007	86
1q1g	purine nucleoside phosphorylase	MT-ImmH	K_d	0.0027	87
1q41	glycogen synthase kinase 3β	indirubin-3′-monoxime	IC_{50}	0.022	88
1q4g	prostaglandin H2 synthase 1	α-methyl-4-biphenylacetic acid	K_i	0.13	89;90
1r1h	neprilysin	compound 1	K_i	0.0012	91;92
1r55	ADAM33	marimastat	K_i	0.16	93;94
1r58	methionine aminopeptidase 2	A357300	IC_{50}	0.11	95
1r9o	cytochrome P450 2C9	flurbiprofen	K_s	9.6	96
1s19	vitamin D nuclear receptor	calcipotriol	EC_{50}	0.0017	97
1s3v	dihydrofolate reductase	compound 2	IC_{50}	0.038	98;99
1sg0	quinone reductase 2	resveratrol	K_d	0.034	100
1sj0	estrogen receptor α	compound 4-D	IC_{50}	0.0008	101
1sq5	pantothenate kinase	pantothenate	K_m	41	102;103
1sqn	progesterone receptor	norethindrone	K_d	0.0004	104
1t40	aldose reductase	IDD552	IC_{50}	0.011	105
1t46	c-kit tyrosine kinase	gleevec	IC_{50}	0.413	106;107
1t9b	acetohydroxyacid synthase	chlorsulfuron	K_i	0.127	108;109
1tow	adipocyte fatty acid-binding protein	compound 1	IC_{50}	0.57	110
1tt1	glutamate receptor 6	kainate	K_i	64.6	111
1tz8	transthyretin	diethylstilbestrol			112
1u1c	uridine phosphorylase	5-benzyl-acyclouridine	K_i	4.3	113
1u4d	activated Cdc42 kinase 1	debromohymenialdisine			114
1uml	adenosine deaminase	compound 4c	K_i	0.03	115
1unl	cyclin-dependent kinase 5	(*R*)-roscovitine	IC_{50}	0.2	116
1uou	thymidine phosphorylase	TPI	IC_{50}	0.02	117;118
1v0p	protein kinase 5	purvalanol B	IC_{50}	0.13	119;120
1v48	purine nucleoside phosphorylase	DFPP-G	K_i	0.0069	121
1v4s	glucokinase	compound A	K_m	1000	122
1vcj	neuraminidase B	BANA207	IC_{50}	26	123

Table 2. Continued

entry	target	compound	potency measure	potency (μM)	ref(s)
1w1p	chitinase B	cyclo-(gly-L-pro)	IC_{50}	5000	124
1w2g	thymidylate kinase	deoxythymidine	K_i	27	125;126
1x8x	tyrosyl-tRNA synthetase	L-tyrosine			127
1xm6	phosphodiesterase 4B	(*R*)-mesopram	IC_{50}	0.42	128
1xoq	phosphodiesterase 4D	roflumilast	IC_{50}	0.000 68	128
1xoz	phosphodiesterase 5A	tadalafil	IC_{50}	0.0012	128
1y6b	vascular endothelial growth factor receptor 2	compound 46	IC_{50}	0.038	129
1ygc	factor VIIa	G17905	K_i	0.000 35	130
1yqy	lethal factor	hydroxamate LFI	K_i	0.024	131
1yv3	myosin II	blebbistatin	IC_{50}	4.9	132
1yvf	NS5B polymerase	compound 59	IC_{50}	0.1	133
1ywr	p38 kinase	compound 15d	IC_{50}	0.032	134
1z95	androgen receptor	*R*-bicalutamide	K_i	0.076	135
2bm2	βII tryptase	compound 16b	K_i	0.015	136
2br1	Chk1	compound 1	IC_{50}	15.4	137
2bsm	heat shock protein 90	compound 11 (VER49009)	IC_{50}	0.14	138

[a] Where available, the potency of the compound against the target is given.

Table 3. Docking Performance for the Protocols Tested in This Study[a]

standard (6Å)	80.5 (0.5)
smaller site (4Å)	86.5 (0.4)
larger site (10Å)	80.4 (0.5)
preoptimized polar Hs	86.9 (0.3)
X-ray waters present	98.6 (0.1)
Corina ligand geometry	75.2 (0.4)

[a] The success rates listed are the percentage of complexes for which the top-ranked GOLD solution is within 2 Å of the experimental binding mode, averaged over 20 runs. Errors in the mean are given in parentheses.

between the ligand acid group and the hydrogen bond donors on the protein. As a result, this entry is systematically docked incorrectly. One of the oxygen atoms in the carboxylic acid group of the ligand in 1tow is involved in hydrogen bonds to both the hydroxyl group of Tyr128 and the Arg126 side chain. The second oxygen atom hydrogen bonds to two water molecules. Goldscore again rewards the formation of hydrogen bonds between charged groups and lets the ligand acid group hydrogen bond to Arg126 via both its oxygen atoms, resulting in an incorrect docking (see Figure 7a). In 1sq5 GOLD completely misdocks the ligand. Instead of letting the ligand acid group hydrogen bond with Tyr1240 and Asn1282, it flips the ligand around in the binding site, so that the ligand acid group hydrogen bonds to the backbone NH groups of Thr1126 and Asp1127. In doing so, the ligand is predicted to bind deeper in the pocket, resulting in a more favorable Van-der-Waals term. However, the ligand acid group is now lodged between two protein acid groups, which clearly is not very favorable.

Other Protocols. Binding Site Definition. An obvious factor affecting the extent of the search problem is the size of the binding site presented to the docking program. In real-life applications, binding sites are probably mostly defined manually, hence incorporating knowledge about the target and so on, but for large-scale validations, normally a fixed definition is used to define which atoms form part of the binding site. These definitions vary between research groups and it can be quite difficult to reproduce the exact protocols used (often authors do not report the definitions used). In our standard protocol, we include all protein atoms that are within 6 Å of any nonhydrogen ligand atom in the X-ray structure. To assess the effect of the size of the binding site presented to the docking tool, we tried two additional cutoff values: 4 Å and 10 Å. It is clear from Table 3 that reducing the size of the pocket increases the success rates significantly. Increasing the size of the binding site to 10 Å does not have a negative effect of the performance for these search settings, but it does at faster settings: using the "default 4" GA settings, the success rate using a 6 Å pocket is 78%, whereas this drops to 71% when a 10 Å pocket is used (see Supporting Information).

Preoptimizing Polar Hydrogens. By default, GOLD rotates protein OH groups on serines, threonines, and tyrosines and NH_3 groups on lysines. Other docking programs do not always have this functionality and require polar hydrogens to be preoptimized in some way. This is usually done by optimizing them in the presence of the ligand (in its X-ray binding mode). Although it is clearly very useful to be able to orient polar hydrogens automatically, docking against a protein with fixed pregenerated positions for the polar hydrogen atoms can actually be very helpful in cases where it is known how a flexible OH or NH_3 group interacts with the ligand. However, keeping the polar hydrogen atoms fixed in these preoptimized positions reduces the search space and the potential for false minima. To investigate the effect this has on the success rates obtained, we made a small change to GOLD that allows us to keep the polar hydrogens fixed. The initial positions for the polar hydrogens were generated by running GOLD in "local optimization" mode,[33] with simplex optimization switched off. It is clear from Table 3 that using the preoptimized polar hydrogens for the protein yields significantly improved success rates. One of the complexes that GOLD now predicts correctly is 1tow (see Figure 7b). When polar hydrogen atoms are optimized automatically, GOLD predicts both oxygen atoms of the ligand carboxylate group to H-bond with Arg126 (see above), orienting the Tyr128 hydroxyl away from the ligand. However, preoptimizing the polar hydrogen atoms of the complex orients the Tyr128 OH hydrogen atom toward the ligand carboxylate, making the incorrect binding mode described above less favorable and resulting in a correctly predicted binding mode.

The improvement in success rates serves as a caveat to adversely judging performance of docking programs that optimize polar hydrogens (e.g., GOLD) versus other programs for which the polar hydrogens have been optimized prior to docking. A further concern is that the minimization protocol adopted in many validations (i.e., local optimization of the protein prior to docking in the presence of the X-ray conformation of the ligand) can lead to additional biases toward the correct ligand conformation beyond simple optimization of the polar hydrogens. Unfortunately, it would be a substantial amount of work for us to perform a local optimization with a scoring function similar to the one used during the GOLD docking, so

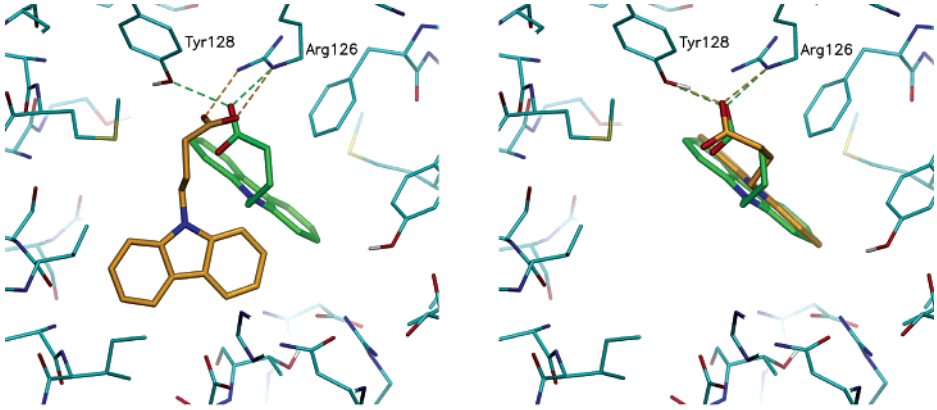

Figure 7. Top-ranked GOLD solution for entry 1tow when (a) GOLD is run in standard mode (i.e., optimizing polar hydrogen atoms automatically) or when (b) the polar hydrogen atoms are preoptimized on the X-ray complex and then kept fixed during the docking run. The X-ray binding mode of the ligand is shown in green, whereas the binding modes predicted by GOLD are shown in amber. When protein polar hydrogen atoms are optimized automatically during the docking, GOLD incorrectly predicts the binding mode of the ligand (a). However, when the polar hydrogen atoms are preoptimized on the protein−ligand complex and then kept fixed during the docking, the correct binding mode is predicted (b).

it is difficult to quantify what effect preoptimization of the entire complex might have on docking success rates.

Including X-ray Water Molecules. Sometimes water molecules play key roles in protein/ligand recognition, and in such cases, it is important to include them.[35,36] However, some authors have suggested including all X-ray water molecules in a validation study.[37,38] This can severely restrict the size of the pocket, particularly in high-resolution structures, by essentially leaving a template of the ligand's shape. Although we do not advocate using such a protocol in a validation exercise, we did include it here to assess the affect it has on docking success rates. Hence, we merged all X-ray water molecules within 6 Å of any non-H ligand atom into the protein and preoptimized the orientation of all polar hydrogen atoms and water molecules, using the "local optimization" mode, with simplex optimization switched off. Polar hydrogens and water molecule orientations were then kept fixed in these optimized orientations during the docking runs. As is clear from Table 3, this has a dramatic effect on the success rates. GOLD now correctly reproduces the binding mode of nearly every complex. We believe the main reason behind this massive improvement is that the search space is drastically reduced by the presence of the water molecules and often almost exclusively limited to the correct binding mode. Hence, we do not believe indiscriminately including all X-ray water molecules is generally helpful to drug discovery projects. However, incorporating key water molecules during the docking can be useful. We recently reported on the development and validation of an extension to GOLD that allows water molecules to rotate and to toggle "on" and "off" during the docking.[36] We will illustrate here how such a method can help improve binding mode predictions, using the 1xm6 entry as an example. If no water molecules are included, GOLD erroneously predicts the ligand to coordinate to the zinc ion (see Figure 8a). However, if we include the water molecule that coordinates the zinc ion and dock the ligand using the water-mediation option, GOLD automatically predicts this water molecule to be switched "on" (i.e., present), and the ligand binding mode is predicted correctly (see Figure 8b).

Ligand Geometry. Most authors (including ourselves), when validating their docking algorithms, take the geometry of the ligand directly from the PDB file, possibly optimizing it as part of the complex before docking (see above). However, in a drug discovery setup, the ligand geometry will often be created automatically by a three-dimensional (3D) conformer generator

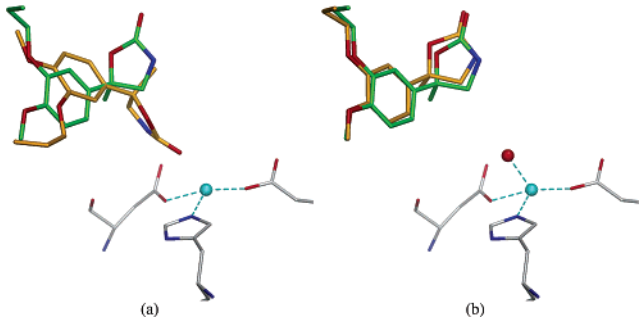

Figure 8. Top-ranked GOLD solution for entry 1xm6 when (a) no water molecules are included in the binding site and GOLD is run in standard mode or when (b) the water molecule coordinating the zinc ion is included in the binding site and allowed to rotate and to toggle "on" and "off" during the docking. The X-ray binding mode of the ligand is shown in green, whereas the binding modes predicted by GOLD are shown in amber. Without the water molecule, GOLD incorrectly predicts the ligand to coordinate to the zinc ion in the binding site. If, however, the water molecule is included, it is predicted to be present and the correct binding mode is produced for the ligand.

like Corina,[27] and this can have an effect on the performance of a docking tool.[39] Therefore, it is probably useful to also report docking success rates for docking ligand geometries generated automatically from the SMILES strings of the ligands. Here, we used Corina to generate 3D input geometries (with correct tautomers and charge states) and docked those using our standard protocol. When these geometries are used, it results in a 5% drop-off in success rates (see Table 3). The reasons why using Corina geometries introduces extra failures are variable, and we intend to investigate these in more detail in a future study. Here we will discuss the 1sj0 entry as an example. The ligand in this entry has a central 2,3-dihydro-benzo[1,4]-oxathiine ring system, with two substituents on the oxathiine ring: an equatorial 3-[4-hydroxyphenyl] and an axial 2-[4-(2-(1-piperdinyl)ethoxy)phenyl] substituent (see Figure 9). The input geometry generated by Corina has the larger substituent equatorially and the smaller substituent axially (i.e., the opposite configuration to the one observed in the X-ray complex). Hence, GOLD is relying on its ring-flipping mechanism to produce the correct conformation of the ligand. For a ring corner to be flipped by GOLD, the four ring atoms adjacent to the flipping corner need to be approximately in the same plane, otherwise the transformations used by GOLD to generate the alternative

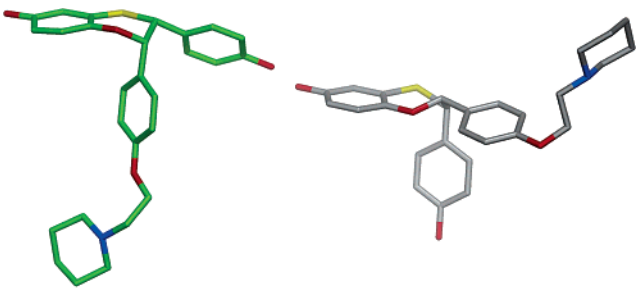

Figure 9. Conformation of the ligand in entry 1sj0, as observed in the X-ray structure (a) and as generated by Corina (b). In the X-ray geometry, the larger substituent on the oxathiine ring is axial and the smaller substituent is equatorial, whereas in the Corina geometry, this is the other way around. As the GOLD ring-flipping algorithm only flips the larger substituent, it cannot access the bioactive conformation from the Corina geometry.

conformer will distort the ring bond angles and distances.[40] The four atoms adjacent to the 2-substituent are roughly in the same plane and are flipped by GOLD, but the four atoms adjacent to the 3-substituent are not in the same plane and, as a result, GOLD does not flip this corner. Hence, GOLD cannot produce the X-ray conformation of the ligand as, in order to do so, both corners need to be flipped. As a result, GOLD systematically fails to dock this compound correctly. GOLD can be forced to flip both corners by increasing the tolerance on the ring distortions. This does produce the correct solution (rmsd = 0.8 Å) from the Corina geometry, but the oxathiine ring is quite distorted.

Future Work. The Astex Diverse Set is primarily aimed at allowing users to have a high-quality set to test algorithms and drive the production of new or improved scoring functions, and this is what we ourselves intend to use the test set for in future work. It is also interesting to ask whether the validation set can be usefully extended. In our view, it would be difficult to increase the size of the set substantially without compromising the criteria used in its construction, although one viable approach would be to allow multiple ligands for each protein. This would be comparatively easy to do because our interface allows the examination of each protein in turn to see whether multiple ligands match the specified criteria.

Another, perhaps more serious, weakness of the set is that it can only be used in native docking validation experiments, that is, docking a ligand back into the protein conformation derived from the experimentally observed conformation with that ligand. This introduces a clear bias in docking validations because the effects of subtle or pronounced induced fit are not tested. For this reason, we intend to use the clustering analysis to define a cross-docking validation set based on the 85 protein targets identified in this work, and this set will also include relevant apo structures for the targets.

Conclusions

We have produced a new validation set for docking applications. The Astex Diverse Set meets the six criteria we identified as important in the generation of a high-quality diverse set. The set only includes complexes for which we have successfully generated electron densities and have confirmed that the electron density is supportive of the deposited experimental geometry for the entire ligand. The 85 complexes in the set are all from different drug discovery or agrochemical targets, and the associated ligands all meet drug-like criteria; 23 of the ligands are approved drugs and an additional six are currently in clinical trials. The set has no overlap with the previous CCDC/Astex set and has been made freely available to the scientific community (http://www.ccdc.cam.ac.uk).

Our approach has involved clustering together all the high-quality complexes for a particular protein and allowing easy visualization and manual selection of high-quality complexes. In future work we intend to construct a cross-docking test set based upon similar principles.

We have shown that docking success rates are heavily dependent on the protocols used to set up the protein and ligand. The size of the binding site directly affects the size of the search problem, and higher success rates are achieved using smaller definitions of the binding sites. Preoptimizing protein polar hydrogens also improves docking performance significantly. GOLD optimizes the protein polar hydrogens during the docking process, and this is clearly a very useful feature. However, researchers may have information on the orientation of some polar hydrogens, and in those cases, it would be useful to be able to fix these groups, a feature currently not available in the publicly available version of GOLD. When we add all X-ray water molecules within 6 Å of the ligand, GOLD predicts the binding modes of nearly all complexes correctly. We do not believe this is a good test of the performance of a docking program, as many of the water molecules are likely to be ligand specific. Using Corina-generated geometries for the ligands reduces the success rates by about 5%. All these results highlight that it is very difficult to compare the performance of different docking programs when the protocols to prepare the protein and ligand differ. We suggest that researchers compare their results to the protocol in Table 3 that is closest to the protocol they use in their investigations.

Although it has not been the aim of this current paper to improve the existing GOLD scoring functions, inspection of docking failures has identified a few potential areas where GOLD and the Goldscore function may be improved. For example, Goldscore heavily rewards hydrogen bonds involving charged groups. This has led to incorrect dockings in several cases where a charged ligand group is either solvent exposed or forming hydrogen bonds with neutral protein groups. Also, the ring-flipping algorithm in GOLD can make certain ring conformations inaccessible, which can be problematic if ligand geometries are generated automatically from SMILES strings. We intend to use the new validation set to investigate and address these and potential other issues in a future study.

Acknowledgment. The authors thank Simon Bowden, Jason Cole, John Liebeschuetz, Willem Nissink, Richard Sykes, and Robin Taylor at the Cambridge Crystallographic Data Centre for general discussions concerning the set and for enabling the Astex Diverse Set to be made freely available via their website.

Supporting Information Available: The following additional information is available: (i) a list of the HET groups excluded in this study; (ii) docking results for all protocols using different GA settings for both the Goldscore and the Chemscore function; (iii) run times for the standard protocol using different GA settings for both the Goldscore and the Chemscore function; (iv) a breakdown of how often each complex is predicted correctly using the standard protocol; and (v) distributions for various properties of the complexes in the Astex Diverse Set. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935−949.

(2) Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862−865.
(3) Cole, J. C.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Comparing protein−ligand docking programs is difficult. *Proteins* **2005**, *60*, 325−332.
(4) 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*, 235−242.
(5) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(6) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein−ligand docking. *Proteins* **1999**, *37*, 228−241.
(7) Pang, Y. P.; Perola, E.; Xu, K.; Prendergast, F. G. EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J. Comput. Chem.* **2001**, *22*, 1750−1771.
(8) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(9) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions. 1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.
(10) Chen, H.; Lyne, P. D.; Giordanetto, F.; Lovell, T.; Li, J. On evaluating molecular-docking methods for pose prediction and enrichment factors. *J. Chem. Inf. Model.* **2006**, *46*, 401−415.
(11) Paul, N.; Rognan, D. ConsDock: A new program for the consensus analysis of protein−ligand interactions. *Proteins* **2002**, *47*, 521−533.
(12) Nissink, J. W. M.; Murray, C. W.; Hartshorn, M. J.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A new test set for validating predictions of protein−ligand interaction. *Proteins* **2002**, *49*, 457−471.
(13) Perola, E.; Walters, W. P.; Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **2004**, *56*, 235−249.
(14) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.
(15) Hendlich, M.; Bergner, A.; Gunther, J.; Klebe, G. Relibase: design and development of a database for comprehensive analysis of protein−ligand interactions. *J. Mol. Biol.* **2003**, *326*, 607−620.
(16) Sayle, R. PDB: Cruft to content (Perception of molecular connectivity from 3D coordinates); Daylight user meeting MUG01, 2001 (http://www.daylight.com/meetings/mug01/Sayle/m4xbondage.html).
(17) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.* **2004**, *47*, 2977−2980.
(18) Wang, R.; Fang, X.; Lu, Y.; Yang, C. Y.; Wang, S. The PDBbind database: methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111−4119.
(19) Elsevier MDL, 2440 Camino Ramon, Suite 300, San Ramon, CA 94583, 2006.
(20) Weininger, D. SMILES a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1998**, *28*, 31−36.
(21) Daylight Chemical Information Systems, Inc., Aliso Vieja, CA; www.daylight.com, 2006.
(22) BioByte Corp. 201 W. 4th St., #204, Claremont, CA 91711−4707, 2006.
(23) Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* **1999**, *290*, 253−266.
(24) Mooij, W. T.; Hartshorn, M. J.; Tickle, I. J.; Sharff, A. J.; Verdonk, M. L.; Jhoti, H. Automated protein−ligand crystallography for structure-based drug design. *ChemMedChem* **2006**, *1*, 827−838.
(25) Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H. Fragment-based lead discovery using X-ray crystallography. *J. Med. Chem.* **2005**, *48*, 403−413.
(26) Kleywegt, G. J.; Harris, M. R.; Zou, J. Y.; Taylor, T. C.; Wahlby, A.; Jones, T. A. The Uppsala Electron-Density Server. *Acta Crystallogr., Sect. D* **2004**, *60*, 2240−2249.
(27) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537−547.
(28) Hartshorn, M. J. AstexViewer: a visualisation aid for structure-based drug design. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 871−881.
(29) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B* **2002**, *58*, 380−388.
(30) Accelrys, Inc., InsightII, San Diego: Accelrys, Inc., 2005.
(31) Halgren, T. A. The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules and vdW parameters. *J. Am. Chem. Soc.* **1992**, *114*, 7827−7843.
(32) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein−small molecule docking methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151−166.
(33) Verdonk, M. L.; Cole, J. C.; Hartshorn, M.; Murray, C. W.; Taylor, R. D. Improved protein−ligand docking using GOLD. *Proteins* **2003**, *52*, 609−623.
(34) Kirton, S. B.; Murray, C. W.; Verdonk, M. L.; Taylor, R. D. Prediction of binding modes for ligands in the cytochromes P450 and other heme-containing proteins. *Proteins* **2005**, *58*, 836−844.
(35) Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: Placing discrete water molecules during protein−ligand docking predictions. *Proteins* **1999**, *34*, 17−28.
(36) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein−ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504−6515.
(37) Goto, J.; Kataoka, R.; Hirayama, N. Ph4Dock: pharmacophore-based protein−ligand docking. *J. Med. Chem.* **2004**, *47*, 6804−6811.
(38) De Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N. P. E. Binding mode prediction of cytochrome P450 and thymidine kinase protein−ligand complexes by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2005**, *48*, 2308−2318.
(39) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004**, *47*, 45−55.
(40) Glen, R. C.; Payne, A. W. A genetic algorithm for the automated generation of molecules within constraints. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 181−202.
(41) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.
(42) Safo, M. K.; Moure, C. M.; Burnett, J. C.; Joshi, G. S.; Abraham, D. J. High-resolution crystal structure of deoxy hemoglobin complexed with a potent allosteric effector. *Protein Sci.* **2001**, *10*, 951−957.
(43) Rowsell, S.; Hawtin, P.; Minshull, C. A.; Jepson, H.; Brockbank, S. M.; Barratt, D. G.; Slater, A. M.; McPheat, W. L.; Waterson, D.; Henney, A. M.; Pauptit, R. A. Crystal structure of human MMP9 in complex with a reverse hydroxamate inhibitor. *J. Mol. Biol.* **2002**, *319*, 173−181.
(44) McVey, C. E.; Walsh, M. A.; Dodson, G. G.; Wilson, K. S.; Brannigan, J. A. Crystal structures of penicillin acylase enzyme−substrate complexes: structural insights into the catalytic mechanism. *J. Mol. Biol.* **2001**, *313*, 139−150.
(45) Dvir, H.; Jiang, H. L.; Wong, D. M.; Harel, M.; Chetrit, M.; He, X. C.; Jin, G. Y.; Yu, G. L.; Tang, X. C.; Silman, I.; Bai, D. L.; Sussman, J. L. X-ray structures of Torpedo californica acetylcholinesterase complexed with (+)-huperzine A and (−)-huperzine B: structural evidence for an active site rearrangement. *Biochemistry* **2002**, *41*, 10810−10818.
(46) Martin, J. L.; Begun, J.; McLeish, M. J.; Caine, J. M.; Grunewald, G. L. Getting the adrenaline going: crystal structure of the adrenaline-synthesizing enzyme PNMT. *Structure* **2001**, *9*, 977−985.
(47) Versees, W.; Decanniere, K.; Pelle, R.; Depoorter, J.; Brosens, E.; Parkin, D. W.; Steyaert, J. Structure and function of a novel purine specific nucleoside hydrolase from *Trypanosoma vivax*. *J. Mol. Biol.* **2001**, *307*, 1363−1379.
(48) Blaszczyk, J.; Li, Y.; Shi, G.; Yan, H.; Ji, X. Dynamic roles of arginine residues 82 and 92 of *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: crystallographic studies. *Biochemistry* **2003**, *42*, 1573−1580.
(49) Li, Y.; Blaszczyk, J.; Wu, Y.; Shi, G.; Ji, X.; Yan, H. Is the critical role of loop 3 of Escherichia coli 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase in catalysis due to loop-3 residues arginine-84 and tryptophan-89? Site-directed mutagenesis, biochemical, and crystallographic studies. *Biochemistry* **2005**, *44*, 8590−8599.
(50) Phan, J.; Koli, S.; Minor, W.; Dunlap, R. B.; Berger, S. H.; Lebioda, L. Human thymidylate synthase is in the closed conformation when complexed with dUMP and raltitrexed, an antifolate drug. *Biochemistry* **2001**, *40*, 1897−1902.

(51) Rutenber, E. E.; Stroud, R. M. Binding of the anticancer drug ZD1694 to *E. coli* thymidylate synthase: assessing specificity and affinity. *Structure* **1996**, *4*, 1317−1324.
(52) Istvan, E. S.; Deisenhofer, J. Structural mechanism for statin inhibition of HMG-CoA reductase. *Science* **2001**, *292*, 1160−1164.
(53) van den Elsen, J. M.; Kuntz, D. A.; Rose, D. R. Structure of Golgi alpha-mannosidase II: a target for inhibition of growth and metastasis of cancer cells. *EMBO J.* **2001**, *20*, 3008−3017.
(54) Whitlow, M.; Howard, A. J.; Stewart, D.; Hardman, K. D.; Chan, J. H.; Baccanari, D. P.; Tansik, R. L.; Hong, J. S.; Kuyper, L. F. X-ray crystal structures of *Candida albicans* dihydrofolate reductase: high resolution ternary complexes in which the dihydronicotinamide moiety of NADPH is displaced by an inhibitor. *J. Med. Chem.* **2001**, *44*, 2928−2932.
(55) Timm, D. E.; Liu, J.; Baker, L. J.; Harris, R. A. Crystal structure of thiamin pyrophosphokinase. *J. Mol. Biol.* **2001**, *310*, 195−204.
(56) Yuvaniyama, J.; Chitnumsub, P.; Kamchonwongpaisan, S.; Vanichtanankul, J.; Sirawaraporn, W.; Taylor, P.; Walkinshaw, M. D.; Yuthavong, Y. Insights into antifolate resistance from malarial DHFR-TS structures. *Nat. Struct. Biol.* **2003**, *10*, 357−365.
(57) Vullo, D.; Innocenti, A.; Nishimori, I.; Pastorek, J.; Scozzafava, A.; Pastorekova, S.; Supuran, C. T. Carbonic anhydrase inhibitors. Inhibition of the transmembrane isozyme XII with sulfonamides—a new target for the design of antitumor and antiglaucoma drugs? *Bioorg. Med. Chem. Lett.* **2005**, *15*, 963−969.
(58) Whittington, D. A.; Waheed, A.; Ulmasov, B.; Shah, G. N.; Grubb, J. H.; Sly, W. S.; Christianson, D. W. Crystal structure of the dimeric extracellular domain of human carbonic anhydrase XII, a bitopic membrane protein overexpressed in certain cancer tumor cells. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 9545−9550.
(59) Toney, J. H.; Hammond, G. G.; Fitzgerald, P. M.; Sharma, N.; Balkovec, J. M.; Rouen, G. P.; Olson, S. H.; Hammond, M. L.; Greenlee, M. L.; Gao, Y. D. Succinic acids as potent inhibitors of plasmid-borne IMP-1 metallo-beta-lactamase. *J. Biol. Chem.* **2001**, *276*, 31913−31918.
(60) Ren, J.; Nichols, C.; Bird, L.; Chamberlain, P.; Weaver, K.; Short, S.; Stuart, D. I.; Stammers, D. K. Structural mechanisms of drug resistance for mutations at codons 181 and 188 in HIV-1 reverse transcriptase and the improved resilience of second generation non-nucleoside inhibitors. *J. Mol. Biol.* **2001**, *312*, 795−805.
(61) Weyand, M.; Schlichting, I.; Marabotti, A.; Mozzarelli, A. Crystal structures of a new class of allosteric effectors complexed to tryptophan synthase. *J. Biol. Chem.* **2002**, *277*, 10647−10652.
(62) Bramson, H. N.; Corona, J.; Davis, S. T.; Dickerson, S. H.; Edelstein, M.; Frye, S. V.; Gampe, R. T., Jr.; Harris, P. A.; Hassell, A.; Holmes, W. D.; Hunter, R. N.; Lackey, K. E.; Lovejoy, B.; Luzzio, M. J.; Montana, V.; Rocque, W. J.; Rusnak, D.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): design, synthesis, enzymatic activities, and X-ray crystallographic analysis. *J. Med. Chem.* **2001**, *44*, 4339−4358.
(63) Reiling, K. K.; Endres, N. F.; Dauber, D. S.; Craik, C. S.; Stroud, R. M. Anisotropic dynamics of the JE-2147-HIV protease complex: drug resistance and thermodynamic binding mode examined in a 1.09 Å structure. *Biochemistry* **2002**, *41*, 4582−4594.
(64) Powers, R. A.; Morandi, F.; Shoichet, B. K. Structure-based discovery of a novel, noncovalent inhibitor of AmpC beta-lactamase. *Structure* **2002**, *10*, 1013−1023.
(65) Babu, Y. S.; Chand, P.; Bantia, S.; Kotian, P.; Dehghani, A.; El Kattan, Y.; Lin, T. H.; Hutchison, T. L.; Elliott, A. J.; Parker, C. D.; Ananth, S. L.; Horn, L. L.; Laver, G. W.; Montgomery, J. A. BCX-1812 (RWJ-270201): discovery of a novel, highly potent, orally active, and selective influenza neuraminidase inhibitor through structure-based drug design. *J. Med. Chem.* **2000**, *43*, 3482−3486.
(66) Smith, B. J.; McKimm-Breshkin, J. L.; McDonald, M.; Fernley, R. T.; Varghese, J. N.; Colman, P. M. Structural studies of the resistance of influenza virus neuramindase to inhibitors. *J. Med. Chem.* **2002**, *45*, 2207−2212.
(67) Matter, H.; Defossa, E.; Heinelt, U.; Blohm, P. M.; Schneider, D.; Muller, A.; Herok, S.; Schreuder, H.; Liesum, A.; Brachvogel, V.; Lonze, P.; Walser, A.; Al Obeidi, F.; Wildgoose, P. Design and quantitative structure−activity relationship of 3-amidinobenzyl-1*H*-indole-2-carboxamides as potent, nonchiral, and selective inhibitors of blood coagulation factor Xa. *J. Med. Chem.* **2002**, *45*, 2749−2769.
(68) Woo, E. J.; Marshall, J.; Bauly, J.; Chen, J. G.; Venis, M.; Napier, R. M.; Pickersgill, R. W. Crystal structure of auxin-binding protein 1 in complex with auxin. *EMBO J.* **2002**, *21*, 2877−2885.
(69) Bledsoe, R. K.; Montana, V. G.; Stanley, T. B.; Delves, C. J.; Apolito, C. J.; McKee, D. D.; Consler, T. G.; Parks, D. J.; Stewart, E. L.; Willson, T. M.; Lambert, M. H.; Moore, J. T.; Pearce, K. H.; Xu, H. E. Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* **2002**, *110*, 93−105.
(70) Prosise, G. L.; Luecke, H. Crystal structures of Tritrichomonasfoetus inosine monophosphate dehydrogenase in complex with substrate, cofactor and analogs: a structural basis for the random-in ordered-out kinetic mechanism. *J. Mol. Biol.* **2003**, *326*, 517−527.
(71) Bretscher, L. E.; Li, H.; Poulos, T. L.; Griffith, O. W. Structural characterization and kinetics of nitric-oxide synthase inhibition by novel N5-(iminoalkyl)- and N5-(iminoalkenyl)-ornithines. *J. Biol. Chem.* **2003**, *278*, 46789−46797.
(72) deSolms, S. J.; Ciccarone, T. M.; MacTough, S. C.; Shaw, A. W.; Buser, C. A.; Ellis-Hutchings, M.; Fernandes, C.; Hamilton, K. A.; Huber, H. E.; Kohl, N. E.; Lobell, R. B.; Robinson, R. G.; Tsou, N. N.; Walsh, E. S.; Graham, S. L.; Beese, L. S.; Taylor, J. S. Dual protein farnesyltransferase-geranylgeranyltransferase-I inhibitors as potential cancer chemotherapeutic agents. *J. Med. Chem.* **2003**, *46*, 2973−2984.
(73) Rasmussen, H. B.; Branner, S.; Wiberg, F. C.; Wagtmann, N. Crystal structure of human dipeptidyl peptidase IV/CD26 in complex with a substrate analog. *Nat. Struct. Biol.* **2003**, *10*, 19−25.
(74) Wang, S.; Eisenberg, D. Crystal structures of a pantothenate synthetase from *M. tuberculosis* and its complexes with substrates and a reaction intermediate. *Protein Sci.* **2003**, *12*, 1097−1108.
(75) Brenk, R.; Naerum, L.; Gradler, U.; Gerber, H. D.; Garcia, G. A.; Reuter, K.; Stubbs, M. T.; Klebe, G. Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. *J. Med. Chem.* **2003**, *46*, 1133−1143.
(76) Dow, R. L.; Schneider, S. R.; Paight, E. S.; Hank, R. F.; Chiang, P.; Cornelius, P.; Lee, E.; Newsome, W. P.; Swick, A. G.; Spitzer, J.; Hargrove, D. M.; Patterson, T. A.; Pandit, J.; Chrunyk, B. A.; LeMotte, P. K.; Danley, D. E.; Rosner, M. H.; Ammirati, M. J.; Simons, S. P.; Schulte, G. K.; Tate, B. F.; DaSilva-Jardine, P. Discovery of a novel series of 6-azauracil-based thyroid hormone receptor ligands: potent, TR beta subtype-selective thyromimetics. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 379−382.
(77) Ye, L.; Li, Y. L.; Mellstrom, K.; Mellin, C.; Bladh, L. G.; Koehler, K.; Garg, N.; Garcia Collazo, A. M.; Litten, C.; Husman, B.; Persson, K.; Ljunggren, J.; Grover, G.; Sleph, P. G.; George, R.; Malm, J. Thyroid receptor ligands. 1. Agonist ligands selective for the thyroid receptor beta1. *J. Med. Chem.* **2003**, *46*, 1580−1588.
(78) Schelling, P.; Claus, M. T.; Johner, R.; Marquez, V. E.; Schulz, G. E.; Scapozza, L. Biochemical and structural characterization of (South)-methanocarbathymidine that specifically inhibits growth of herpes simplex virus type 1 thymidine kinase-transduced osteosarcoma cells. *J. Biol. Chem.* **2004**, *279*, 32832−32838.
(79) Hartmann, M.; Schneider, T. R.; Pfeil, A.; Heinrich, G.; Lipscomb, W. N.; Braus, G. H. Evolution of feedback-inhibited beta/alpha barrel isoenzymes by gene duplication and a single mutation. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 862−867.
(80) Nagar, B.; Hantschel, O.; Young, M. A.; Scheffzek, K.; Veach, D.; Bornmann, W.; Clarkson, B.; Superti-Furga, G.; Kuriyan, J. Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* **2003**, *112*, 859−871.
(81) Weber, A.; Casini, A.; Heine, A.; Kuhn, D.; Supuran, C. T.; Scozzafava, A.; Klebe, G. Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: new pharmacological opportunities due to related binding site recognition. *J. Med. Chem* **2004**, *47*, 550−557.
(82) Wendt, M. D.; Rockway, T. W.; Geyer, A.; McClellan, W.; Weitzberg, M.; Zhao, X.; Mantei, R.; Nienaber, V. L.; Stewart, K.; Klinghofer, V.; Giranda, V. L. Identification of novel binding interactions in the development of potent, selective 2-naphthamidine inhibitors of urokinase. Synthesis, structural analysis, and SAR of *N*-phenyl amide 6-substitution. *J. Med. Chem.* **2004**, *47*, 303−324.
(83) Olsen, J. A.; Banner, D. W.; Seiler, P.; Obst, S. U.; D'Arcy, A.; Stihle, M.; Muller, K.; Diederich, F. A fluorine scan of thrombin inhibitors to map the fluorophilicity/fluorophobicity of an enzyme active site: evidence for C−F···C═O interactions. *Angew. Chem., Int. Ed.* **2003**, *42*, 2507−2511.
(84) Strickler, M.; Goldstein, B. M.; Maxfield, K.; Shireman, L.; Kim, G.; Matteson, D. S.; Jones, J. P. Crystallographic studies on the complex behavior of nicotine binding to P450cam (CYP101). *Biochemistry* **2003**, *42*, 11943−11950.
(85) Sabini, E.; Ort, S.; Monnerjahn, C.; Konrad, M.; Lavie, A. Structure of human dCK suggests strategies to improve anticancer and antiviral therapy. *Nat. Struct. Biol.* **2003**, *10*, 513−519.

(86) Scapin, G.; Patel, S. B.; Lisnock, J.; Becker, J. W.; LoGrasso, P. V. The structure of JNK3 in complex with small molecule inhibitors: structural basis for potency and selectivity. *Chem. Biol.* **2003**, *10*, 705−712.
(87) Shi, W.; Ting, L. M.; Kicska, G. A.; Lewandowicz, A.; Tyler, P. C.; Evans, G. B.; Furneaux, R. H.; Kim, K.; Almo, S. C.; Schramm, V. L. Plasmodium falciparum purine nucleoside phosphorylase: crystal structures, immucillin inhibitors, and dual catalytic function. *J. Biol. Chem.* **2004**, *279*, 18103−18106.
(88) Bertrand, J. A.; Thieffine, S.; Vulpetti, A.; Cristiani, C.; Valsasina, B.; Knapp, S.; Kalisz, H. M.; Flocco, M. Structural characterization of the GSK-3beta active site using selective and nonselective ATP-mimetic inhibitors. *J. Mol. Biol.* **2003**, *333*, 393−407.
(89) Gupta, K.; Kaub, C. J.; Carey, K. N.; Casillas, E. G.; Selinsky, B. S.; Loll, P. J. Manipulation of kinetic profiles in 2-aryl propionic acid cyclooxygenase inhibitors. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 667−671.
(90) Gupta, K.; Selinsky, B. S.; Kaub, C. J.; Katz, A. K.; Loll, P. J. The 2.0 Å resolution crystal structure of prostaglandin H2 synthase-1: structural insights into an unusual peroxidase. *J. Mol. Biol.* **2004**, *335*, 503−518.
(91) Chen, H.; Noble, F.; Mothe, A.; Meudal, H.; Coric, P.; Danascimento, S.; Roques, B. P.; George, P.; Fournie-Zaluski, M. C. Phosphinic derivatives as new dual enkephalin-degrading enzyme inhibitors: synthesis, biological properties, and antinociceptive activities. *J. Med. Chem.* **2000**, *43*, 1398−1408.
(92) Oefner, C.; Roques, B. P.; Fournie-Zaluski, M. C.; Dale, G. E. Structural analysis of neprilysin with various specific and potent inhibitors. *Acta Crystallogr., Sect. D* **2004**, *60*, 392−396.
(93) Orth, P.; Reichert, P.; Wang, W.; Prosise, W. W.; Yarosh-Tomaine, T.; Hammond, G.; Ingram, R. N.; Xiao, L.; Mirza, U. A.; Zou, J.; Strickland, C.; Taremi, S. S.; Le, H. V.; Madison, V. Crystal structure of the catalytic domain of human ADAM33. *J. Mol. Biol.* **2004**, *335*, 129−137.
(94) Zou, J.; Zhu, F.; Liu, J.; Wang, W.; Zhang, R.; Garlisi, C. G.; Liu, Y. H.; Wang, S.; Shah, H.; Wan, Y.; Umland, S. P. Catalytic activity of human ADAM33. *J. Biol. Chem.* **2004**, *279*, 9818−9830.
(95) Sheppard, G. S.; Wang, J.; Kawai, M.; BaMaung, N. Y.; Craig, R. A.; Erickson, S. A.; Lynch, L.; Patel, J.; Yang, F.; Searle, X. B.; Lou, P.; Park, C.; Kim, K. H.; Henkin, J.; Lesniewski, R. 3-Amino-2-hydroxyamides and related compounds as inhibitors of methionine aminopeptidase-2. *Bioorg. Med. Chem Lett.* **2004**, *14*, 865−868.
(96) Wester, M. R.; Yano, J. K.; Schoch, G. A.; Yang, C.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0 Å resolution. *J. Biol. Chem.* **2004**, *279*, 35630−35637.
(97) Tocchini-Valentini, G.; Rochel, N.; Wurtz, J. M.; Moras, D. Crystal structures of the vitamin D nuclear receptor liganded with the vitamin D side chain analogues calcipotriol and seocalcitol, receptor agonists of clinical importance. Insights into a structural basis for the switching of calcipotriol to a receptor antagonist by further side chain modification. *J. Med. Chem.* **2004**, *47*, 1956−1961.
(98) Cody, V.; Luft, J. R.; Pangborn, W.; Gangjee, A.; Queener, S. F. Structure determination of tetrahydroquinazoline antifolates in complex with human and *Pneumocystis carinii* dihydrofolate reductase: correlations between enzyme selectivity and stereochemistry. *Acta Crystallogr., Sect. D* **2004**, *60*, 646−655.
(99) Gangjee, A.; Zaveri, N.; Kothare, M.; Queener, S. F. Nonclassical 2,4-diamino-6-(aminomethyl)-5,6,7,8-tetrahydroquinazoline antifolates: synthesis and biological activities. *J. Med. Chem.* **1995**, *38*, 3660−3668.
(100) Buryanovskyy, L.; Fu, Y.; Boyd, M.; Ma, Y.; Hsieh, T. C.; Wu, J. M.; Zhang, Z. Crystal structure of quinone reductase 2 in complex with resveratrol. *Biochemistry* **2004**, *43*, 11417−11426.
(101) Kim, S.; Wu, J. Y.; Birzin, E. T.; Frisch, K.; Chan, W.; Pai, L. Y.; Yang, Y. T.; Mosley, R. T.; Fitzgerald, P. M.; Sharma, N.; Dahllund, J.; Thorsell, A. G.; DiNinno, F.; Rohrer, S. P.; Schaeffer, J. M.; Hammond, M. L. Estrogen receptor ligands. II. Discovery of benzoxathiins as potent, selective estrogen receptor alpha modulators. *J. Med. Chem.* **2004**, *47*, 2171−2175.
(102) Ivey, R. A.; Zhang, Y. M.; Virga, K. G.; Hevener, K.; Lee, R. E.; Rock, C. O.; Jackowski, S.; Park, H. W. The structure of the pantothenate kinase−ADP−pantothenate ternary complex reveals the relationship between the binding sites for substrate, allosteric regulator, and antimetabolites. *J. Biol. Chem.* **2004**, *279*, 35622−35629.
(103) Song, W. J.; Jackowski, S. Kinetics and regulation of pantothenate kinase from *Escherichia coli*. *J. Biol. Chem.* **1994**, *269*, 27051−27058.
(104) Madauss, K. P.; Deng, S. J.; Austin, R. J.; Lambert, M. H.; McLay, I.; Pritchard, J.; Short, S. A.; Stewart, E. L.; Uings, I. J.; Williams, S. P. Progesterone receptor ligand binding pocket flexibility: crystal structures of the norethindrone and mometasone furoate complexes. *J. Med. Chem.* **2004**, *47*, 3381−3387.
(105) Ruiz, F.; Hazemann, I.; Mitschler, A.; Joachimiak, A.; Schneider, T.; Karplus, M.; Podjarny, A. The crystallographic structure of the aldose reductase−IDD552 complex shows direct proton donation from tyrosine 48. *Acta Crystallogr., Sect. D* **2004**, *60*, 1347−1354.
(106) Manley, P. W.; Cowan-Jacob, S. W.; Buchdunger, E.; Fabbro, D.; Fendrich, G.; Furet, P.; Meyer, T.; Zimmermann, J. Imatinib: a selective tyrosine kinase inhibitor. *Eur. J. Cancer* **2002**, *38* (Suppl 5), S19−S27.
(107) Mol, C. D.; Dougan, D. R.; Schneider, T. R.; Skene, R. J.; Kraus, M. L.; Scheibe, D. N.; Snell, G. P.; Zou, H.; Sang, B. C.; Wilson, K. P. Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *J. Biol. Chem.* **2004**, *279*, 31655−31663.
(108) Duggleby, R. G.; Pang, S. S.; Yu, H.; Guddat, L. W. Systematic characterization of mutations in yeast acetohydroxyacid synthase. Interpretation of herbicide-resistance data. *Eur. J. Biochem.* **2003**, *270*, 2895−2904.
(109) McCourt, J. A.; Pang, S. S.; Guddat, L. W.; Duggleby, R. G. Elucidating the specificity of binding of sulfonylurea herbicides to acetohydroxyacid synthase. *Biochemistry* **2005**, *44*, 2330−2338.
(110) Lehmann, F.; Haile, S.; Axen, E.; Medina, C.; Uppenberg, J.; Svensson, S.; Lundback, T.; Rondahl, L.; Barf, T. Discovery of inhibitors of human adipocyte fatty acid-binding protein, a potential type 2 diabetes target. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4445−4448.
(111) Mayer, M. L. Crystal structures of the GluR5 and GluR6 ligand binding cores: molecular mechanisms underlying kainate receptor selectivity. *Neuron* **2005**, *45*, 539−552.
(112) Morais-de-Sa, E.; Pereira, P. J.; Saraiva, M. J.; Damas, A. M. The crystal structure of transthyretin in complex with diethylstilbestrol: a promising template for the design of amyloid inhibitors. *J. Biol. Chem.* **2004**, *279*, 53483−53490.
(113) Bu, W.; Settembre, E. C.; el Kouni, M. H.; Ealick, S. E. Structural basis for inhibition of *Escherichia coli* uridine phosphorylase by 5-substituted acyclouridines. *Acta Crystallogr., Sect. D* **2005**, *61*, 863−872.
(114) Steinbacher, S.; Kaiser, J.; Wungsintaweekul, J.; Hecht, S.; Eisenreich, W.; Gerhardt, S.; Bacher, A.; Rohdich, F. Structure of 2*C*-methyl-D-erythritol-2,4-cyclodiphosphate synthase involved in mevalonate-independent biosynthesis of isoprenoids. *J. Mol. Biol.* **2002**, *316*, 79−88.
(115) Terasaka, T.; Kinoshita, T.; Kuno, M.; Seki, N.; Tanaka, K.; Nakanishi, I. Structure-based design, synthesis, and structure−activity relationship studies of novel non-nucleoside adenosine deaminase inhibitors. *J. Med. Chem.* **2004**, *47*, 3730−3743.
(116) Mapelli, M.; Massimiliano, L.; Crovace, C.; Seeliger, M. A.; Tsai, L. H.; Meijer, L.; Musacchio, A. Mechanism of CDK5/p25 binding by CDK inhibitors. *J. Med. Chem.* **2005**, *48*, 671−679.
(117) Cole, C.; Reigan, P.; Gbaj, A.; Edwards, P. N.; Douglas, K. T.; Stratford, I. J.; Freeman, S.; Jaffar, M. Potential tumor-selective nitroimidazolylmethyluracil prodrug derivatives: inhibitors of the angiogenic enzyme thymidine phosphorylase. *J. Med. Chem.* **2003**, *46*, 207−209.
(118) Norman, R. A.; Barry, S. T.; Bate, M.; Breed, J.; Colls, J. G.; Ernill, R. J.; Luke, R. W.; Minshull, C. A.; McAlister, M. S.; McCall, E. J.; McMiken, H. H.; Paterson, D. S.; Timms, D.; Tucker, J. A.; Pauptit, R. A. Crystal structure of human thymidine phosphorylase in complex with a small molecule inhibitor. *Structure* **2004**, *12*, 75−84.
(119) Holton, S.; Merckx, A.; Burgess, D.; Doerig, C.; Noble, M.; Endicott, J. Structures of *P. falciparum* PfPK5 test the CDK regulation paradigm and suggest mechanisms of small molecule inhibition. *Structure* **2003**, *11*, 1329−1337.
(120) Le Roch, K.; Sestier, C.; Dorin, D.; Waters, N.; Kappes, B.; Chakrabarti, D.; Meijer, L.; Doerig, C. Activation of a *Plasmodium falciparum* cdc2-related kinase by heterologous p25 and cyclin H. Functional characterization of a *P. falciparum* cyclin homologue. *J. Biol. Chem.* **2000**, *275*, 8952−8958.
(121) Luic, M.; Koellner, G.; Yokomatsu, T.; Shibuya, S.; Bzowska, A. Calf spleen purine-nucleoside phosphorylase: crystal structure of the binary complex with a potent multisubstrate analogue inhibitor. *Acta Crystallogr., Sect. D* **2004**, *60*, 1417−1424.
(122) Kamata, K.; Mitsuya, M.; Nishimura, T.; Eiki, J.; Nagata, Y. Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase. *Structure* **2004**, *12*, 429−438.
(123) Lommer, B. S.; Ali, S. M.; Bajpai, S. N.; Brouillette, W. J.; Air, G. M.; Luo, M. A benzoic acid inhibitor induces a novel conformational change in the active site of Influenza B virus neuraminidase. *Acta Crystallogr., Sect. D* **2004**, *60*, 1017−1023.

(124) Houston, D. R.; Synstad, B.; Eijsink, V. G.; Stark, M. J.; Eggleston, I. M.; van Aalten, D. M. Structure-based exploration of cyclic dipeptide chitinase inhibitors. *J. Med. Chem.* **2004**, *47*, 5713−5720.
(125) Fioravanti, E.; Adam, V.; Munier-Lehmann, H.; Bourgeois, D. The crystal structure of *Mycobacterium tuberculosis* thymidylate kinase in complex with 3′-azidodeoxythymidine monophosphate suggests a mechanism for competitive inhibition. *Biochemistry* **2005**, *44*, 130−137.
(126) Vanheusden, V.; Van Rompaey, P.; Munier-Lehmann, H.; Pochet, S.; Herdewijn, P.; Van Calenbergh, S. Thymidine and thymidine-5′-*O*-monophosphate analogues as inhibitors of *Mycobacterium tuberculosis* thymidylate kinase. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3045−3048.
(127) Kobayashi, T.; Takimura, T.; Sekine, R.; Kelly, V. P.; Kamata, K.; Sakamoto, K.; Nishimura, S.; Yokoyama, S. Structural snapshots of the KMSKS loop rearrangement for amino acid activation by bacterial tyrosyl-tRNA synthetase. *J. Mol. Biol.* **2005**, *346*, 105−117.
(128) Card, G. L.; England, B. P.; Suzuki, Y.; Fong, D.; Powell, B.; Lee, B.; Luu, C.; Tabrizizad, M.; Gillette, S.; Ibrahim, P. N.; Artis, D. R.; Bollag, G.; Milburn, M. V.; Kim, S. H.; Schlessinger, J.; Zhang, K. Y. Structural basis for the activity of drugs that inhibit phosphodiesterases. *Structure* **2004**, *12*, 2233−2247.
(129) Harris, P. A.; Cheung, M.; Hunter, R. N., III; Brown, M. L.; Veal, J. M.; Nolte, R. T.; Wang, L.; Liu, W.; Crosby, R. M.; Johnson, J. H.; Epperly, A. H.; Kumar, R.; Luttrell, D. K.; Stafford, J. A. Discovery and evaluation of 2-anilino-5-aryloxazoles as a novel class of VEGFR2 kinase inhibitors. *J. Med. Chem.* **2005**, *48*, 1610−1619.
(130) Olivero, A. G.; Eigenbrot, C.; Goldsmith, R.; Robarge, K.; Artis, D. R.; Flygare, J.; Rawson, T.; Sutherlin, D. P.; Kadkhodayan, S.; Beresini, M.; Elliott, L. O.; DeGuzman, G. G.; Banner, D. W.; Ultsch, M.; Marzec, U.; Hanson, S. R.; Refino, C.; Bunting, S.; Kirchhofer, D. A selective, slow binding inhibitor of factor VIIa binds to a nonstandard active site conformation and attenuates thrombus formation in vivo. *J. Biol. Chem.* **2005**, *280*, 9160−9169.
(131) Shoop, W. L.; Xiong, Y.; Wiltsie, J.; Woods, A.; Guo, J.; Pivnichny, J. V.; Felcetto, T.; Michael, B. F.; Bansal, A.; Cummings, R. T.; Cunningham, B. R.; Friedlander, A. M.; Douglas, C. M.; Patel, S. B.; Wisniewski, D.; Scapin, G.; Salowe, S. P.; Zaller, D. M.; Chapman, K. T.; Scolnick, E. M.; Schmatz, D. M.; Bartizal, K.; MacCoss, M.; Hermes, J. D. Anthrax lethal factor inhibition. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7958−7963.
(132) Allingham, J. S.; Smith, R.; Rayment, I. The structural basis of blebbistatin inhibition and specificity for myosin II. *Nat. Struct. Mol. Biol.* **2005**, *12*, 378−379.
(133) Pfefferkorn, J. A.; Greene, M. L.; Nugent, R. A.; Gross, R. J.; Mitchell, M. A.; Finzel, B. C.; Harris, M. S.; Wells, P. A.; Shelly, J. A.; Anstadt, R. A.; Kilkuskie, R. E.; Kopta, L. A.; Schwende, F. J. Inhibitors of HCV NS5B polymerase. Part 1: Evaluation of the southern region of (2*Z*)-2-(benzoylamino)-3-(5-phenyl-2-furyl)acrylic acid. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2481−2486.
(134) Golebiowski, A.; Townes, J. A.; Laufersweiler, M. J.; Brugel, T. A.; Clark, M. P.; Clark, C. M.; Djung, J. F.; Laughlin, S. K.; Sabat, M. P.; Bookland, R. G.; VanRens, J. C.; De, B.; Hsieh, L. C.; Janusz, M. J.; Walter, R. L.; Webster, M. E.; Mekel, M. J. The development of monocyclic pyrazolone based cytokine synthesis inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2285−2289.
(135) Bohl, C. E.; Gao, W.; Miller, D. D.; Bell, C. E.; Dalton, J. T. Structural basis for antagonism and resistance of bicalutamide in prostate cancer. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6201−6206.
(136) Levell, J.; Astles, P.; Eastwood, P.; Cairns, J.; Houille, O.; Aldous, S.; Merriman, G.; Whiteley, B.; Pribish, J.; Czekaj, M.; Liang, G.; Maignan, S.; Guilloteau, J. P.; Dupuy, A.; Davidson, J.; Harrison, T.; Morley, A.; Watson, S.; Fenton, G.; McCarthy, C.; Romano, J.; Mathew, R.; Engers, D.; Gardyan, M.; Sides, K.; Kwong, J.; Tsay, J.; Rebello, S.; Shen, L.; Wang, J.; Luo, Y.; Giardino, O.; Lim, H. K.; Smith, K.; Pauls, H. Structure-based design of 4-(3-aminomethylphenyl)piperidinyl-1-amides: novel, potent, selective, and orally bioavailable inhibitors of betaII tryptase. *Bioorg. Med. Chem.* **2005**, *13*, 2859−2872.
(137) Foloppe, N.; Fisher, L. M.; Howes, R.; Kierstan, P.; Potter, A.; Robertson, A. G.; Surgenor, A. E. Structure-based design of novel Chk1 inhibitors: insights into hydrogen bonding and protein−ligand affinity. *J. Med. Chem.* **2005**, *48*, 4332−4345.
(138) Dymock, B. W.; Barril, X.; Brough, P. A.; Cansfield, J. E.; Massey, A.; McDonald, E.; Hubbard, R. E.; Surgenor, A.; Roughley, S. D.; Webb, P.; Workman, P.; Wright, L.; Drysdale, M. J. Novel, potent small-molecule inhibitors of the molecular chaperone Hsp90 discovered through structure-based design. *J. Med. Chem.* **2005**, *48*, 4212−4215.

JM061277Y