Information Decay in Molecular Docking Screens against Holo, Apo, and Modeled Conformations of Enzymes

Susan L. McGovern[†] and Brian K. Shoichet*[,‡]

Department of Molecular Pharmacology and Biological Chemistry, Northwestern University, 303 East Chicago Avenue, Chicago, Illinois 60611 and Department of Pharmaceutical Chemistry, University of California, San Francisco, Genentech Hall, 600 16th Street, San Francisco, California 94143-2240

Received January 21, 2003

Molecular docking uses the three-dimensional structure of a receptor to screen a small molecule database for potential ligands. The dependence of docking screens on the conformation of the binding site remains an open question. To evaluate the information loss that occurs as the active site conformation becomes less defined, a small molecule database was docked against the holo (ligand bound), apo, and homology modeled structures of 10 different enzyme binding sites. The holo and apo representations were crystallographic structures taken from the Protein Data Bank (PDB), and the homology-modeled structures were taken from the publicly available resource ModBase. The database docked was the MDL Drug Data Report (MDDR), a functionally annotated database of 95 000 small molecules that contained at least 35 ligands for each of the 10 systems. In all sites, at least 99% of the molecules in the MDDR were treated as nonbinding decoys. For each system, the holo, apo, and modeled structures were used to screen the MDDR, and the ability of each structure to enrich the known ligands for that system over random selection was evaluated. The best overall enrichment was produced by the holo structure in seven systems, the apo structure in two systems, and the modeled structure in one system. These results suggest that the performance of the docking calculation is affected by the particular representation of the receptor used in the screen, and that the holo structure is the one most likely to yield the best discrimination between known ligands and decoy molecules, but important exceptions to this rule also emerge from this study. Although each of the holo, apo, and modeled conformations led to enrichment of known ligands in all systems, the enrichment did not always rise to a level judged to be sufficient to justify the effort of a docking screen. Using a 20-fold enrichment of known ligands over random selection as a rough guideline for what might be enough to justify a docking screen, the holo conformation of the enzyme met this criterion in eight of 10 sites, whereas the apo conformation met this criterion in only two sites and the modeled conformation in three.

Introduction

Virtual screening fits libraries of candidate ligands into a model of a receptor binding site. When the receptor is represented at atomic resolution, virtual screening takes the form of molecular docking.[1−6] Like any screening technique, docking must distinguish potential ligands from a much larger selection of incompatible molecules (decoys). Docking programs attempt to make this distinction by sampling and evaluating multiple configurations of each database molecule for complementarity to the binding site, identifying a relatively small number of high-scoring "hits", which may then be tested experimentally.

Typically, docking calculations use an experimentally determined receptor structure. The number of receptor conformational states explored by these experimental structures is limited, and docking must often use a single conformation of a receptor, for instance a single ligand-bound (holo) conformation. A greater challenge is that many interesting targets have no experimental structure at all: of the 730 000 proteins with known sequence,[7] only about 5500[8] have had a structure determined. To bridge this gap, several methods, including comparative modeling, sequence threading, and de novo prediction, have been developed to calculate models of protein structure; these models are increasingly used directly in docking screens for novel ligands.[9−12] It is presently unclear how well the conformation of a receptor-binding site should be defined before it can be genuinely useful for docking. One might expect, for instance, that a receptor determined in a holo conformation may be a better target for inhibitor discovery, since more of the binding determinants will have been defined by conformational change on ligand binding. Conversely, a holo conformation might bias the docking screen, preventing the discovery of ligands much different from that particular ligand captured in the holo complex.

How does the quality of a docking screen decay with the quality of the binding site defined by the receptor structure? How "good" must a structure be before it can be used to screen for novel ligands? Are holo structures better than apo structures, and are both better than theoretically modeled structures? To investigate these questions, we chose 10 enzymes for which structures had been determined in both holo and apo forms, for

* To whom correspondence should be addressed: shoichet@cgl.ucsf.edu; tel: 415-514-4126, fax: 415-502-1411.
[†] Northwestern University.
[‡] University of California.

10.1021/jm0300330 CCC: $25.00
Published on Web 06/05/2003

Table 1. Enzyme Systems

	Holo			Apo		template		model		
enzyme[a]	ligand[b]	PDB	species	PDB	species	PDB	species	species	seq i.d., %[c]	seq i.d. in site, %[d]
DHFR	methotrexate	3dfr[47]	*L. casei*	6dfr[48]	*E. coli*	3dfr	*L. casei*	*P. aeruginosa*	34	64
PNP	immucillin-H	1b8o[49]	*B. taurus*	1pbn[18]	*B. taurus*	1b8o	*B. taurus*	*S. cerevisiae*	50	91
PARP	benzimidazole	1efy[21]	*G. gallus*	2paw[22]	*G. gallus*	1efy	*G. gallus*	*H. sapiens*	87	93
thrombin	Arg-aldehyde	1ba8[50]	*H. sapiens*	1hgt[51]	*H. sapiens*	1doj	*H. sapiens*	*O. cuniculus*	85	100
GART	folate analog	1c2t[52]	*E. coli*	1cdd[26]	*E. coli*	3gar	*E. coli*	*V. unguiculata*	28	45
SAHH	DHCeA[e]	1a7a[53]	*H. sapiens*	1b3r[54]	*R. norvegicus*	1b3r	*R. norvegicus*	*M. tuberculosis*	61	86
AR	tolrestat	1ah3[30]	*S. scrofa*	1ads[55]	*H. sapiens*	1c9w	*C. griseus*	*B. taurus*	69	59
AChE	huprine X	1e66[56]	*T. californica*	1ea5[56]	*T. californica*	1ea5	*T. californica*	*C. elegans*	36	82
TS	folate analog	2bbq[57]	*E. coli*	1bid[58]	*E. coli*	1qqq	*E. coli*	*M. tuberculosis*	66	88

[a] AChE, acetylcholinesterase; AR, aldose reductase; DHFR, dihydrofolate reductase; GART, glycinamide ribonucleotide transformylase; PARP, poly(ADP-ribose) polymerase; PNP, purine nucleoside phosphorylase; SAHH, S-adenosylhomocysteine hydrolase; TS, thymidylate synthase [b] Ligand in binding site used in docking experiment. See Supporting Information Table S2 for chemical structures. [c] Sequence identity between model and template, as supplied by ModBase.[35] [d] Sequence identity between model and template among residues within 5Å of ligand. [e] 2′-Hydroxy-3′-ketocyclopent-4′-enyladenine.

which homology modeled structures were available from a public database, and for which there were known ligands in a large, annotated database of small molecule structures. For each enzyme, 95 000 compounds of the MDL Drug Data Report (MDDR) database were docked against each of the three enzyme models: holo X-ray structure, apo X-ray structure, and modeled structure. The quality of the docking screen was judged based on the enrichment of the known ligands among the top-scoring hits compared to the other database molecules. For any given system, the known ligands constituted between 0.03% and 1% of the total database; the remaining database molecules were considered decoys.

We expected that the holo structures would give better enrichments than the apo structures, and that the apo structures would give better enrichments than the homology modeled structures. Although this expectation held up in more than half of the systems, we were often surprised.

Results

To investigate the effect of protein conformation on molecular docking, 10 enzyme systems were docked in three different conformations each: a crystallographically determined complexed (holo) conformation, a crystallographically determined uncomplexed (apo) conformation, and a homology modeled conformation of a species of the enzyme whose structure has not been experimentally determined (Table 1). Docking calculations were judged to be more or less successful based on their ability to enrich known ligands among the top-scoring docked molecules. Enrichment was evaluated using four criteria: the maximum "enrichment factor"[13] of the known ligands among the ranked hits (Table 2), the percentage of the docking-ranked MDDR database necessary to look through to find 25% of the known ligands (Table 2), overall profiles of enrichment factors and percentage of ligands found as a function of the percentage of the ranked database (Figures 1 and 2), and the geometry of the docked ligands compared to that observed crystallographically (Figures 3 and 4). The "enrichment factor" is defined as the number of known ligands found, at any given point in the docked-ranked list of compounds, divided by the number of ligands one would expect to find at random. For instance, if a system had 95 ligands in a database of 95 000 molecules, one would expect to find 1 ligand per bin of 1000 database molecules by random selection alone. If a docking screen ranked 10 ligands in the top-scoring 1000 molecules, this would represent a 10-fold improvement over random selection (a 10-fold enrichment factor) for this bin.

We have typically not attempted to recreate the crystallographic complex of the holo receptor and its cognate ligand by docking. For most of the systems

Table 2. Enrichment of Known Ligands by Each Enzyme Structure

enzyme[a]	% of db to find 25% of known ligands	max. enrichment factor	% of db where max. enrichment factor occurred
DHFR			
Holo*	2.01	28	0.40
Apo	3.92	14	0.10
Model	10.65	3	4.92
PNP − PO_4			
Holo	2.81	57	0.10
Apo	13.06	4	22.60
Model*	1.21	57	0.10
PNP + PO_4			
Holo*	0.40	85	0.10
Apo	N.D.	N.D.	N.D.
Model	0.90	43	0.40
PARP			
Holo*	2.81	10	3.82
Apo	3.82	9	4.92
Model	18.28	2	35.56
Thrombin			
Holo	6.63	19	0.10
Apo*	3.11	24	0.10
Model	3.52	20	0.10
GART			
Holo*	0.40	159	0.10
Apo	30.03	1	78.64
Model	6.83	4	9.44
SAHH			
Holo*	1.10	29	0.20
Apo	9.74	5	1.91
Model	20.49	3	4.82
AR			
Holo*	2.81	9	2.61
Apo	4.32	6	2.21
Model	10.65	4	3.31
AChE			
Holo*	6.33	28	0.30
Apo	9.04	7	0.90
Model	15.57	4	0.50
TS			
Holo	3.52	38	0.10
Apo*	2.11	27	0.30
Model	2.61	16	0.40

[a] * Protein conformation with the best overall enrichment of known ligands.

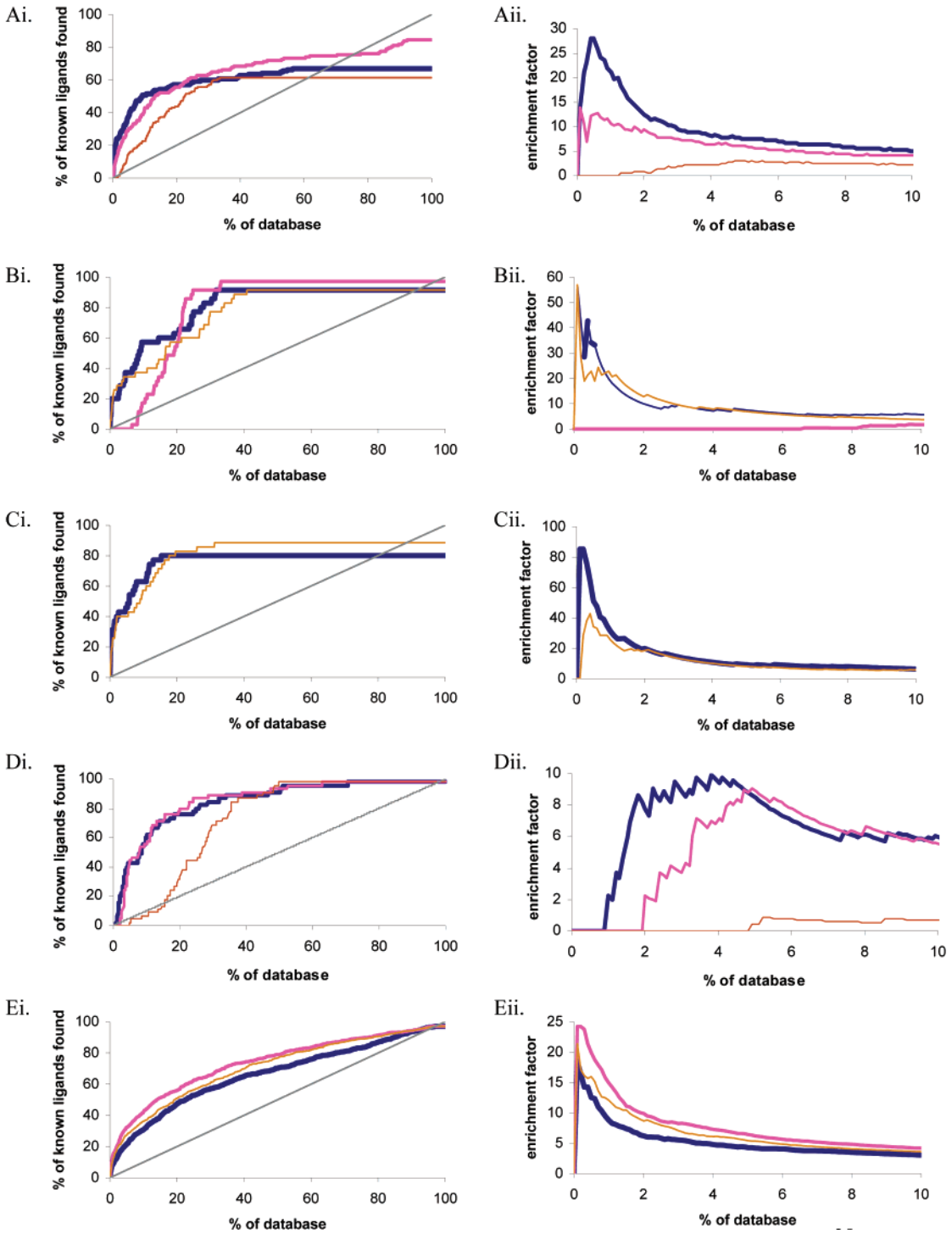

Figure 1. Enrichment of known ligands by holo (blue line), apo (magenta line), and modeled (orange line) conformations for five systems. A. DHFR, 142 known ligands. B. PNP without phosphate, 35 known ligands. C. PNP with phosphate, 35 known ligands. D. PARP, 45 known ligands. E. Thrombin, 699 known ligands. In each panel, subpanel *i* shows the cumulative percentage of known ligands found vs the percentage of rank-ordered database, and panel *ii* shows the enrichment of known ligands vs the percentage of rank-ordered database. The gray diagonal line in the left hand panels indicates the percentage of ligands one would expect to find by random selection.

studied here, there were only a handful of ligand-bound structures in the PDB, and the ligands in those structures were often absent from the MDDR or were covalently bound and therefore in a configuration inaccessible to our docking method. As a surrogate method to evaluate our approach, we have instead shown docked poses of ligands that are structurally similar to those from the crystallographic conformation and that have been experimentally shown to bind the target receptor, often at nanomolar concentrations (Supporting Information Table S2).

On the basis of these criteria, the best overall enrichment of known ligands was provided by the crystallographically determined holo conformation in seven systems, by the crystallographically determined apo conformation in two systems, and by the homology

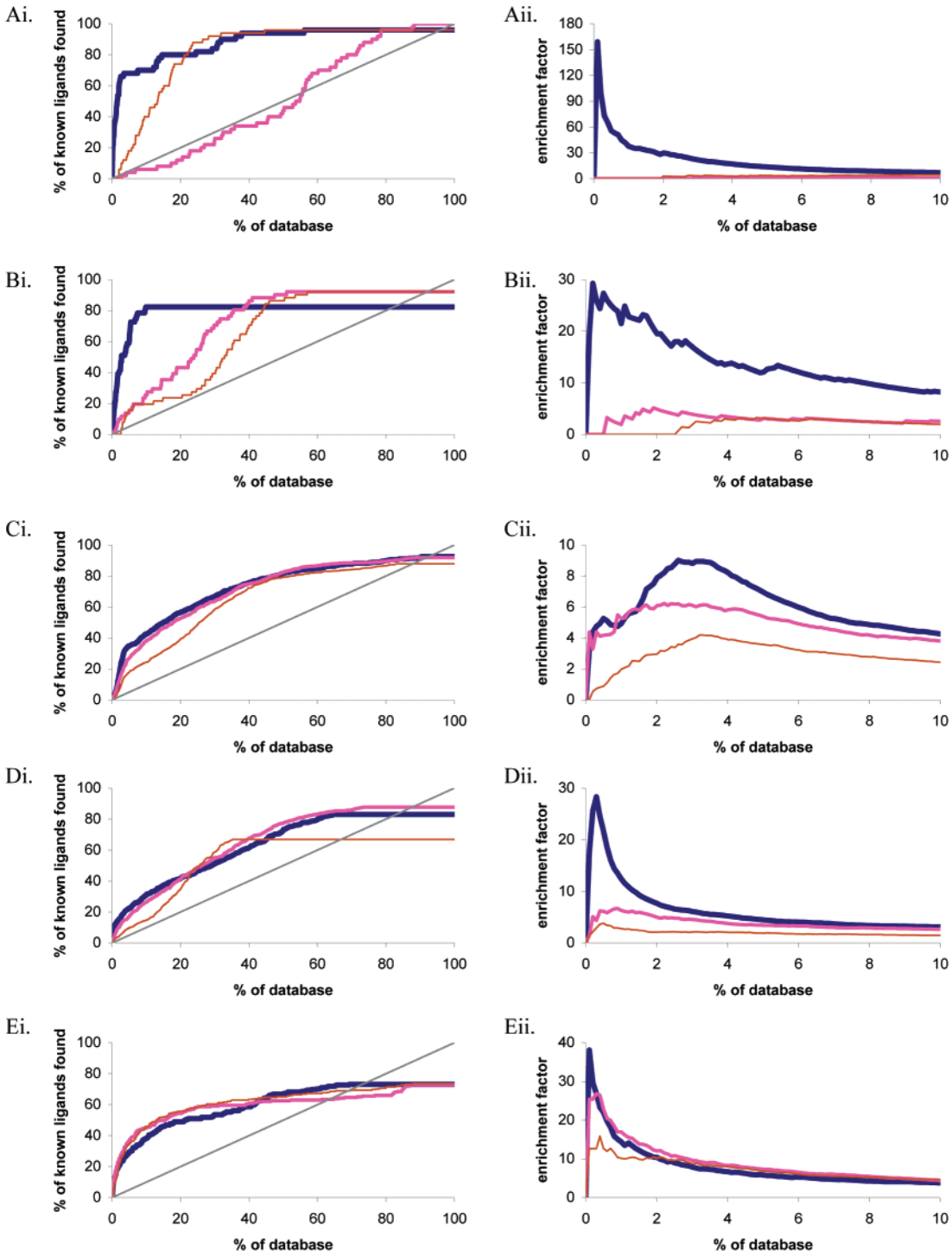

Figure 2. Enrichment of known ligands by holo (blue line), apo (magenta line), and modeled (orange line) conformations for five systems. A. GART, 50 known ligands. B. SAHH, 51 known ligands. C. AR, 908 known ligands. D. AChE, 680 known ligands. E. TS, 235 known ligands. In each panel, subpanel *i* shows the cumulative percentage of known ligands found vs the percentage of rank-ordered database and panel *ii* shows the enrichment of known ligands vs the percentage of rank-ordered database. The gray diagonal line in the left hand panels indicates the percentage of ligands one would expect to find by random selection.

modeled conformation in one system (Table 2). In subsequent sections we consider the performance of each individual system in detail.

Dihydrofolate Reductase (DHFR). DHFR is a key enzyme in folate biosynthesis and is inhibited by pteridines (such as methotrexate), pyrimidines (such as trimethoprim), and their analogues, of which 142 are annotated in the MDDR. The holo conformation yielded the best enrichment of the known ligands, with a maximum enrichment factor of 28-fold over random in the top 0.40% of the database (Figure 1Aii and Table 2). Overall, it was possible to fit 67% of the known ligands in the holo receptor conformation; 25% of these were identified in the top 2.01% of the database (Figure 1Ai). Pteridines and pyrimidines typically docked in a configuration similar to that of the crystallographic ligand, methotrexate (Figure 3Ai); known ligands in which the largest rigid fragment was bigger than a pteridine group did not fit in the site and were not scored.

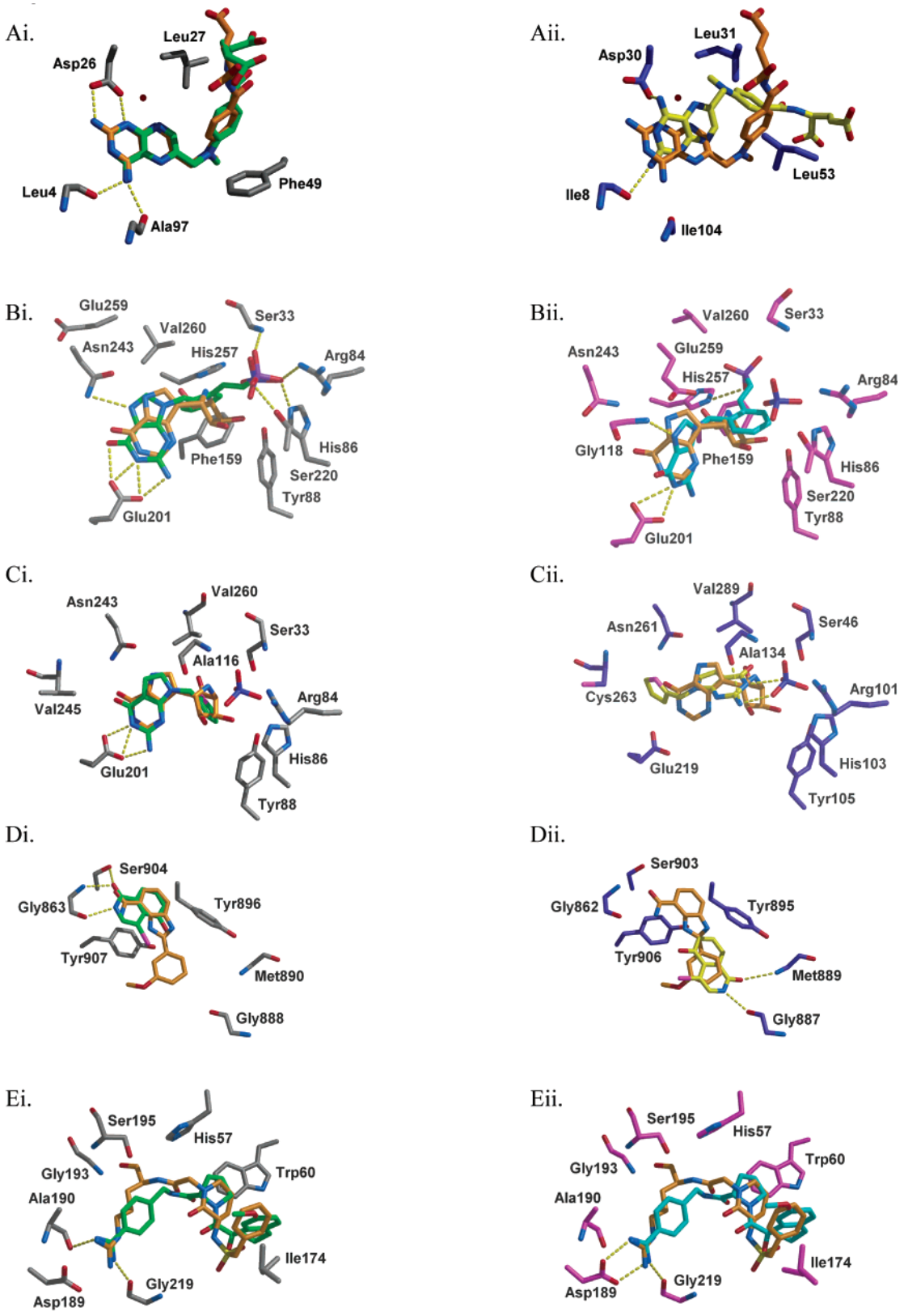

Figure 3. Docking predictions for five systems. A. DHFR. B. PNP without phosphate. C. PNP with phosphate. D. PARP. E. Thrombin. In each panel, subpanel *i* shows the predicted binding mode in the holo conformation and subpanel *ii* shows the predicted binding mode in either the apo or modeled conformation. In all panels, carbon atoms of the holo receptor are gray, carbon atoms of the crystallographic position of the ligand in the holo receptor are orange, carbon atoms of the docked ligand in the holo conformation are green, carbon atoms of the apo receptor are magenta, carbon atoms of the docked ligand in the apo conformation are cyan, carbon atoms of the modeled receptor are dark blue, carbon atoms of the docked ligand in the modeled conformation are yellow, oxygen atoms are red, and nitrogen atoms are blue. In panel C only, sulfur atoms are magenta. In panel D only, bromide atoms are magenta. Hydrogen bonds between the docked ligand and the receptor are illustrated with dashed yellow lines. See Supporting Information Table S2 for structures of the crystallographic and docked ligands.

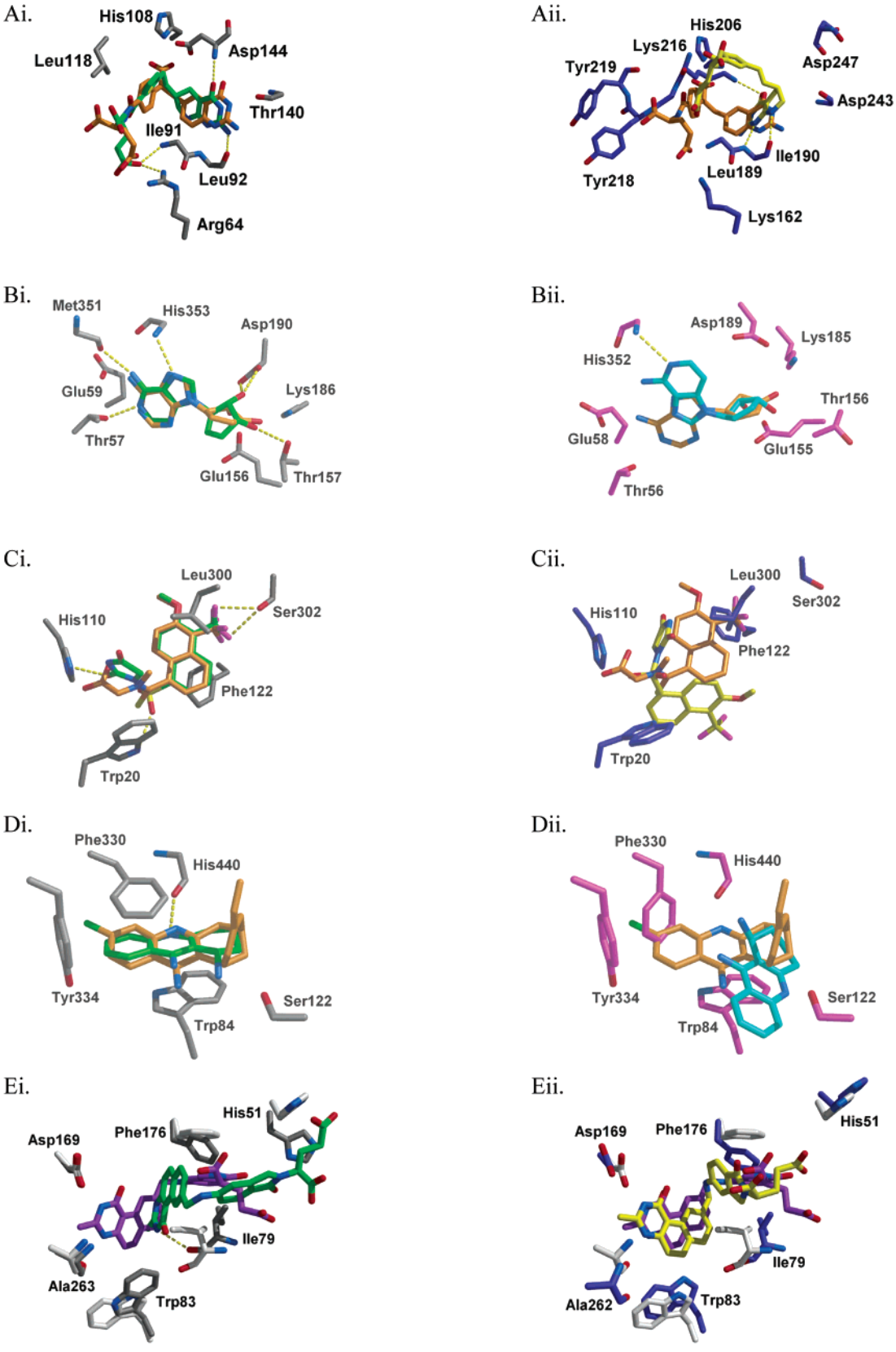

Figure 4. Docking predictions for five systems. A. GART. B. SAHH. C. AR. D. AChE. E. TS. In each panel, subpanel *i* shows the predicted binding mode in the holo conformation and subpanel *ii* shows the predicted binding mode in either the apo or modeled conformation. Atoms are colored as in Figure 3. In panel C only, fluoride atoms are magenta. In panel D only, chloride atoms on the crystallographic ligand are green. In panel E only, the second receptor depicted is from PDB structure 1syn with protein carbon atoms in white and ligand carbon atoms in purple; labels are given for the holo receptor used in the docking calculation (PDB structure 2bbq, carbon atoms in gray). See Supporting Information Table S2 for structures of the crystallographic and docked ligands.

The binding site in the apo conformation was larger and accommodated some ligands that were excluded from the holo conformation, allowing 84% of the known ligands to fit in the site. This site did not discriminate between decoys and known ligands as well as the holo conformation; its maximum enrichment was half that of the holo receptor (Table 2 and Figure 1Aii).

Although the holo structure of DHFR from *L. casei* was the template for the modeled conformation of DHFR from *P. aeruginosa* (root-mean square deviation (RMSD) of 0.25 Å over 149 Cα atoms), the binding geometries and enrichment of known ligands were worse in the modeled conformation. For instance, in the top-scoring configuration of methotrexate in this receptor, the pteridine group was flipped 180° relative to its position in the holo crystal structure (Figure 3Aii). This may be a result of two leucine side chains distal to the pteridine binding site: CD2 of Leu53 is 0.8 Å from the crystallographic position of the benzoic acid group of methotrexate, and CD1 of Leu31 is 1.9 Å from its glutamate tail (Figure 3Aii). These side chains effectively displace the benzoic acid group and force methotrexate to adopt another configuration. Side chains of the corresponding residues in the holo conformation of the *L. casei* enzyme, Phe49 and Leu27, are 3.9−4.3 Å from methotrexate and form a hydrophobic binding surface (Figure 3Ai). Although the modeled conformation identified 61% of the known ligands at a rate better than random selection (Figure 1Ai), the maximum enrichment of known ligands for this site was only 3-fold, occurring after the top 4.9% of the database.

Purine Nucleoside Phosphorylase (PNP) without Phosphate. PNP is a critical enzyme in the purine salvage pathway and has been the focus of extensive drug design efforts.[14−17] The active site of PNP binds both a nucleoside and a molecule of inorganic phosphate.[18] The 70 annotated PNP ligands in the MDDR are evenly divided between nucleotide analogues that contain a phosphate group and nucleoside analogues that do not contain a phosphate group. In our work, we assumed that the 35 PNP ligands bearing a phosphate moiety would bind the enzyme in the absence of inorganic phosphate; these ligands were counted as ligands for PNP devoid of phosphate, and their enrichment is described below. In separate calculations, we considered the other 35 PNP ligands that did not contain a phosphate group as ligands for PNP structures that incorporated a molecule of inorganic phosphate in the binding site; their enrichment is described in the subsequent section.

Both the holo and the modeled conformations of PNP identified 32 of the 35 phosphate-bearing PNP ligands in the MDDR, and both conformations yielded a maximum enrichment of 57-fold over random in the top 0.1% of the database (Figure 1B and Table 2). The modeled receptor concentrated the top 25% of known ligands in a smaller fraction of the database than the holo receptor (1.21% vs 2.81%); the model therefore produced the best overall enrichment (Table 2). Predicted binding modes for ligands were similar in both structures (Figure 3Bi and data not shown).

The apo conformation of PNP fares poorly in this test; although it eventually finds 34 of the 35 known ligands, the first ligand is not identified until the top 12.96% of the database (Figure 1Bi). The resulting enrichment is only 4-fold over random, occurring in the top 22.60% of the database (Table 2). Main chain and side chain conformational changes in the active site are likely responsible for this dismal performance. In the absence of ligand, the side chain of Glu259 encroaches on the purine binding site; when a nucleoside binds, Glu259 flips out of the site and is spatially replaced by Val260 (Figures 3Bi and 3Bii). Small molecules docked to the apo site therefore contorted themselves around Glu259; this often resulted in a displacement of the phosphate moiety from the well-defined phosphate binding pocket (Figure 3Bii). The assignment of the amide oxygen and nitrogen in the purine recognition residue Asn243 was also inverted in the apo structure relative to the holo structure, further contributing to the difficulty of docking ligands to this site.

PNP with Phosphate. Structures of phosphate-bound holo and modeled PNP were used to screen the MDDR; the PNP ligands that did not contain a phosphate moiety were considered ligands for this site. The holo PNP conformation produced the best enrichment of the 35 ligands in the MDDR, 85-fold in the top 0.1% of the database (Figure 1Cii and Table 2). These ligands were also highly concentrated at the top of the database; 25% were found in the top 0.40% of the database (Figure 1Ci and Table 2). Predicted binding modes of the known ligands typically captured many of the features observed in the holo structure (Figure 3Ci).

A few subtle differences in the modeled conformation resulted in a maximal enrichment half that of the holo structure, although this conformation also ranked the known ligands highly (25% in the top 0.90% of the database, Table 2). The active site in the modeled conformation was slightly larger than in the holo receptor, as reflected by the larger number of total molecules scored (44962 vs 27735, Supporting Information Table S1) in the modeled binding site. The larger volume appears to result from a few side chain differences. For instance, CG1 of Val245 in the holo structure forms one side of the purine binding pocket; the equivalent residue in the modeled structure, Cys263, has a hydrogen in that position (Figures 3Ci and 3Cii), thereby providing more space for ligand binding.

Poly(ADP-ribose) Polymerase (PARP). PARP is a nuclear protein involved in DNA repair; inhibition of PARP may increase the sensitivity of tumor cells to genotoxic chemotherapies.[19] Inhibitor design projects have typically focused on small molecules containing a carboxamide group;[20,21] this moiety or its analogues are found in most of the 45 annotated PARP ligands in the MDDR. In the docking screen, all three conformations of PARP identified 44 of these ligands. The holo conformation generated the best enrichment (10-fold over random) and scored 25% of the known ligands in the top 2.81% of the database (Figure 1D and Table 2). Many of the predicted configurations for these ligands included three hydrogen bonds with Gly863 and Ser904, previously shown to be important interactions for inhibitor binding in this site (Figure 3Di).[21,22] Ligand binding to the catalytic site of PARP induces few structural changes in the receptor; consequently, the

enrichment factors and docked poses of the ligands in the apo conformation were similar to those in the holo receptor (Figure 1D and Table 2).

Although the modeled structure of human PARP was based on the holo conformation of chicken PARP and shared 93% sequence identity in the binding site with its template (Table 1), it did not accelerate the identification of known ligands as well as the holo conformation (Figure 1Dii and Table 2). The best enrichment over random was 2-fold at 35.56% of the database; the 25% top scoring known ligands were distributed in the top 18.28% of the database (Figure 1Di and Table 2). Moreover, the docking modes predicted by this site did not reproduce the interactions with the residues equivalent to Gly863 and Ser904 (Figure 3Dii). Indeed, ligands were typically prohibited from occupying the critical part of the binding site because the side chain of Tyr906 drops down into the binding cavity in the modeled structure.

Thrombin. The serine protease thrombin controls the final step in the coagulation cascade, and the development of thrombin inhibitors has been the subject of considerable effort.[23] Much work has focused on the design of inhibitors directed toward the active site of thrombin; there are 699 such ligands in the MDDR. Surprisingly, the unbound conformation of thrombin identified the known ligands more rapidly than either the holo or modeled conformations, finding 25% of them in the top 3.11% of the database, with a maximum enrichment of 24-fold over random (Figure 1E and Table 2). The holo conformation did not identify known ligands as well, ranking the top 25% of them in the top 6.63% of the database with a maximum enrichment of 19-fold (Figure 1E and Table 2). This was unexpected because these binding sites are quite similar; most of the side chains do not move significantly upon ligand binding (RMSD of all heavy atoms within 5 Å of crystallographic ligand = 0.48 Å), but these small differences were sufficient to change the hydrogen bonding pattern of docked molecules. For the ligand illustrated in Figure 3E, its amidinium nitrogen atoms were placed 2.62 Å and 2.86 Å from the carboxylate oxygens of the specificity-determinant Asp189 in the apo conformation (Figure 3Eii); in the holo conformation, the nitrogen atoms were docked 3.40 Å and 3.63 Å away from this carboxylate group (Figure 3Ei). As a consequence of several such small changes, this ligand received a total DOCK score of −51.2 kcal/mol in the apo conformation and −38.3 kcal/mol in holo receptor; much of this difference was due to a better electrostatic score in the unbound conformation. Overall, the binding site in the apo conformation was slightly larger than in the holo conformation, allowing for more configurations of molecules to be scored; correspondingly, the run time was also longer (Supporting Information Table S1).

The modeled conformation of thrombin from *O. cuniculus* was built on an unbound conformation of human thrombin. It performed better than the holo structure but slightly worse than the apo structure, ranking 25% of the known ligands in top 3.52% of the database with a maximum enrichment of 20-fold at the top 0.10% of the database (Figure 1E and Table 2). Predicted docking modes were typically similar to those seen with the apo conformation (data not shown).

Glycinamide Ribonucleotide Transformylase (GART). GART uses a folate cofactor to catalyze the first step of de novo purine biosynthesis.[24] Because tumor cells depend more on de novo purine biosynthesis than do normal cells, which typically use salvage pathways, GART has been extensively explored as an anticancer target. Folate-based inhibitors of GART have been shown to inhibit the growth of tumor cells;[25] the MDDR contains 50 annotated GART ligands that are folate derivatives. The holo conformation of GART produced the highest enrichment of known ligands among all systems studied in this project, 159-fold over random in the top 0.1% of the database (Figure 2Aii and Table 2). This receptor ranked many of the 50 known ligands highly, placing 25% of them in the top 0.40% of the database (Figure 2Ai and Table 2). The predicted docking modes of the known ligands reproduce many of the same interactions with GART as 10-formyl-5,8,-10-trideazafolate, the crystallographic ligand (Figure 4Ai).

The apo conformation of GART identified known ligands at a rate similar to that of random selection (Figure 2Ai). This appears to result primarily from the position of the loop of residues 142−144, which adopts a closed conformation in the monomer selected for docking.[26] The residues in this loop, including the catalytically important Asp144, partially occupy the active site and force the ligands to dock outside of the folate recognition site (not shown).

The modeled structure GART from *V. unguiculata* identifies ligands at a rate much higher than random selection (Figure 2Ai), but for the wrong reason: in all cases, the folate moiety is docked backward into the binding site, and the remainder of the ligand wraps around a loop of the receptor that infringes on the benzoic acid binding site, especially the extended side chain of Lys216 (Figure 4Aii). This extra loop is absent from the holo and apo conformations of GART that were docked. The peculiar placement of this loop may reflect the low sequence identity between the model and the template on which it was built (28% overall, 45% in the binding site, Table 1), because the template structure does not have a loop in a similar location (not shown).

S-Adenosylhomocysteine Hydrolase (SAHH). SAHH is a key regulatory enzyme in several methyltransferase pathways, and inhibitors of this enzyme have demonstrated antiviral, antiarthritic, and antiparasitic properties.[27−29] Many of the known SAHH inhibitors are adenosine derivatives; 51 such molecules are included in the MDDR. These were most rapidly identified by the holo conformation of the enzyme; 25% were found within in the top 1.10% of the database with a maximum enrichment of 29-fold over random (Figure 2B and Table 2). Notably, the ligands did not score well; 26 of the 38 known ligands scored had unfavorable DOCK scores, primarily dominated by high (poor) van der Waals scores. This stems from the fairly small binding site defined by the crystallographic ligand 2′-hydroxy-3′-ketocyclopent-4′-enyladenine (DHCeA) (Figure 4Bi and Supporting Information Table S2); many of the SAHH ligands in the MDDR contain at least one more heavy atom substituent than DHCeA. Despite these poor scores, the smaller site served as a steric filter to remove many decoys, as only 11% (Supporting

Information Table S1) of the total molecules in the database fit into the site and were scored at all. Predicted binding modes for the scored ligands were similar to that of the crystallographic ligand (Figure 4Bi).

In the unbound conformation of SAHH, the cofactor domain is further from the catalytic domain, creating a more open binding pocket. This larger site allowed 91% (Supporting Information Table S1) of the total molecules in the MDDR to be scored. Docked configurations of the ligands typically exploited the additional space; for instance, the purine moiety of most known ligands was flipped 180° relative to that of the crystallographic ligand (Figure 4Bii). Consequently, the site did not discriminate between known ligands and decoys as well as the holo conformation; the top scoring 25% of the ligands were found in the top 9.74% of the database with a maximum enrichment of 5-fold over random (Figure 2B and Table 2).

The apo conformation of rat SAHH was used as the template for the modeled conformation of *M. tuberculosis* SAHH, so the modeled structure also has an open binding site with the cofactor domain positioned far from the catalytic domain. Despite the good backbone alignment of the apo and modeled structures (RMSD = 0.70 Å over 387 Cα atoms), small side chain differences in the modeled binding site led to poorer binding geometries in the modeled structure than in the apo structure (not shown) and ranking of known ligands at a rate similar to random selection (25% of known ligands in the top 20.49% of the database, Table 2 and Figure 2Bi).

Aldose Reductase (AR). AR reduces glucose to sorbitol, and the accumulation of sorbitol is thought to be responsible for various complications associated with chronic diabetes. Therefore, inhibition of AR is of great interest, and many different classes of ligands have been explored, including spirohydantoins and several acetic acid derivatives.[30] There are 908 diverse AR ligands contained in the MDDR; enrichment of these ligands was the greatest in the holo conformation: 9-fold over random at 2.61% of the database (Figure 2Cii and Table 2). This mediocre performance is partly a consequence of the significant flexibility of the AR binding site. The tolrestat-bound conformation was used in this study; many of the top-scoring known ligands structurally resembled tolerestat (Figure 4Ci and Supporting Information Table S1). This is likely a consequence of the particular conformations of Phe122, Leu300, and Thr111 in this structure (Figure 4Ci).[30]

The apo conformation did slightly worse than the holo structure with a maxiumum enrichment of 6-fold over random at 2.21% of the database. Again, the conformational flexibility of the site seems to be the main cause of the poor enrichment. The modeled receptor showed the worst enhancement of the known ligands: 4-fold over random at 3.31% of the database. Interestingly, the modeled conformation was built using a structure of a different enzyme, aldehyde reductase (instead of aldose reductase), as a template; this is likely responsible for the lower sequence identity of the model and template sequences in the binding site than over the entire protein length (59% vs 69%, Table 1).

Acetylcholinesterase (AChE). Alzheimer's disease has been associated with a decrease in acetylcholine levels; consequently, inhibition of acetylcholinesterase has been pursued as a pharmaceutical approach to slow the progress of this disease. Several AChE inhibitors have been identified, including the clinically used drugs tacrine and rivastigmine. The MDDR contains 680 compounds annotated as AChE inhibitors; the holo AChE receptor yielded the highest maximum enrichment of these ligands: 28-fold over random at 0.30% of the database; 25% of these ligands were found in the top 6.33% of the database (Figure 2D and Table 2). The docked positions of many of the known ligands typically involved stacking interactions between Phe330 and Trp84 (Figure 4Di). In the apo conformation, the side chain of Phe330 is perpendicular to its position in the holo conformation; this abolished the hydrophobic stacking site and forced ligands to dock in an alternate position (Figure 4Dii). The corresponding enrichment produced by the unbound conformation was less than that of the bound conformation, 7-fold at 2.21% of the database (Figure 2Dii and Table 2).

Although the structure of AChE from *C. elegans* was modeled on that of the apo conformation (Table 1), the model did not enrich the known ligands as rapidly as the apo receptor (Figure 2Di). This appears to result from a residue substitution in the active site: Ser122 in the apo receptor is spatially replaced by a tyrosine in the model. Because the backbone of Ser122 and the tyrosine are aligned, the bulky tyrosine side chain extends into the binding site and infringes on the carbobicyclic binding site (Figure 4Dii), again causing ligands to adopt a docking mode different from the crystallographic ligand. As in the apo receptor, the modeled conformation of the residue equivalent to Phe330 prevents it from stacking with docked ligands. The smaller size of the binding site in the model also reduced the total number of molecules scored and reduced the run time to 20% that of the apo conformation (Supporting Information Table S1).

Thymidylate Synthase (TS). TS catalyzes the last committed step in de novo thymidylate biosynthesis and is a well-studied target for anticancer drug design. The enzyme recognizes folate analogues in its cofactor binding site, 235 of which are annotated in the MDDR. The holo conformation of TS produced the highest enrichment factor, 38-fold at the top 0.10% of the database, but it identified the known TS ligands at a slower overall rate than either the apo or modeled conformations (Figure 2Ei). This may result from the diversity of TS ligands in the MDDR: the ligand in the holo conformation was smaller than many of the known TS ligands in the database. Consequently, the holo conformation yielded better enrichment of ligands most similar to its native ligand, but it was unable to correctly dock larger, bulkier molecules. Figure 4Ei illustrates this problem for BW1843U89, a 90 pM inhibitor of TS[31] that has a benzoquinazoline ring system followed by an isoindolinyl moiety. To accommodate these large groups, several TS residues, including Ile79, Phe176, and Trp83,[31] rotate significantly from their position in complex with CB3727-polyglu, the liganded conformation used for this docking study. This latter conformation was therefore unable to correctly dock BW1843U89 in its crystallographic position (Figure 4Ei).

The binding sites in the apo and modeled TS conformations were more spacious and could accommodate larger ligands than the holo receptor. Consequently, molecules such as BW1843U89 were correctly docked into these receptors (Figure 4Eii). Overall, the apo and modeled receptors identified 25% of the known ligands in the top 2.11% and 2.61% of the database, respectively. The larger size of these binding pockets also allowed for sampling of more orientations and conformations than the holo receptor, leading to longer run times (Supporting Information Table S1).

Discussion

This survey of 10 docking systems is an initial evaluation of information decay as a consequence of receptor structure in molecular docking. On the basis of the enrichment of known ligands from a database containing at least 99% decoy molecules, the success of the docking calculation typically decayed as the quality of the receptor structure, as judged by preorganization for ligand binding, decreased from holo to apo to modeled structures. This trend corresponded with the decreasing ability of each structure to capture the features of the active site that were important for ligand recognition.

Detailed examination of where the docking screens failed illustrates some of the limitations of each type of structure. For instance, holo structures can be overspecialized: the holo TS structure recognized ligands similar to the one captured in the holo complex (folate analogues) but not ligands that were markedly different (e.g., BW1843489). Apo conformations, such as the apo GART structure, may not resemble the ligand-bound conformation of the protein. Modeled structures, even those with a high sequence identity to their template structures, can have poorly placed side chains in the active site (e.g., modeled PARP).

Rigid protein docking typically requires the selection of a single receptor conformation. Which conformation is the best choice? Almost all of the structures studied here identified known ligands at a rate better than random selection, suggesting that screening with any type of structure is better than nothing. But to be useful for ligand discovery, docking must discriminate between true ligands and decoys at a high enough rate such that at least some ligands are ranked in the top 0.2−0.5% of the scored database. In a typical virtual screen, one might visually inspect only the top 500 scoring molecules (the top 0.5% of the database used here) to select a subset of molecules for experimental testing. If we imagine that on the order of 0.1% of a diverse database might have activity for a given target, then to have more than a handful of hits in the top 0.5% of a database of the size used here would require an enrichment factor of around 20-fold, which is typically considered a good improvement over random selection.[32] With an enrichment of 20-fold, one could expect 10 hits in the top 500 compounds, such that there is a 2% chance that a given compound is active. In an academic lab, one might reasonably test 20 to 50 compounds; even with a 20-fold enrichment, the chance that all 20 compounds are inactive is still 67%; for 50 compounds tested, the chance that all are inactive is 36%. If 500 compounds are tested, as might be done in an industrial setting, the chance that all 500 are inactive is less than 0.005%.

In the systems studied here, enrichment of 20-fold over random within the top 0.5% or better of the database was provided by eight holo structures (including thrombin), two apo structures, and three modeled structures (Table 2). This suggests that, at least in this initial survey of docking systems, meaningful enrichment of ligands is most likely to occur with crystallographic holo structures and less likely with either apo or modeled structures. Additionally, not all of the receptor structures produced good predictions of binding modes; indeed, some structures achieved a remarkable enrichment of known ligands despite docking these molecules incorrectly into the active site (for instance, modeled GART).

Given the widespread use of modeled structures in molecular docking,[9−12] it is worth considering their performance in more detail. The modeled structures almost always identified known ligands at a rate better than random selection (Figures 1 and 2). Models that shared a high sequence identity with their template structures often performed better, but this was not always true. In three systems in which the sequence identity between the model and template in the binding site was greater than 80%, the modeled receptor showed the poorest discrimination between known ligands and decoys among all of the modeled conformations studied, occasionally succeeding only slightly better random selection. This was often a consequence of changes in the geometry of the binding site,[33] including a side chain rotation (Phe330 in AChE, Tyr906 in PARP), a single residue substitution (tyrosine for serine in AChE), and a domain movement (SAHH). In two of those systems, SAHH and AChE, the modeled receptor used an unbound structure as a template; the results for these models would probably have been better if a complexed template had been used. The PARP model was built on the same crystallographic holo conformation that was docked (Table 1), but even in this case the model fared more poorly than both of the experimentally determined structures (Figure 1D).

There are a few caveats to note. First, we assume that all molecules annotated in the MDDR as ligands of a given receptor are actual inhibitors of that receptor. Not all of the molecules in the MDDR have been experimentally shown to bind their purported targets. Conversely, we assume that only those molecules in the MDDR explicitly annotated as ligands for a particular target will bind that target. These factors should lead to some errors in each system, yet they should have little effect on the relative performance of the receptor structures.

Second, we have assumed that all ligands of an enzyme from one species will bind a homologous enzyme from a different species. This supposition will certainly not be true for all ligands of all enzymes, but conservation of active site geometries suggests that it is a reasonable assumption, especially as the shared sequence identity between homologues increases (Table 1).

These caveats should not obscure the three main points to emerge from this study. First, the performance of a molecular docking screen depends on the particular conformation of the receptor structure used in the calculation. Second, when the possible receptor confor-

mations include holo, apo, and modeled structures, the crystallographically determined holo conformation is the one most likely to yield meaningful enrichment of known ligands from a database containing mostly decoy molecules. Finally, exceptions to this rule can occur when the holo structure is overspecialized; in these cases, the apo or the modeled conformation may provide better discrimination between true ligands and decoys.

Experimental Section

System Selection. Nine distinct enzyme systems were selected for molecular docking. Each system had at least 35 ligands in the MDDR, a crystallographic complexed (holo) structure in the PDB,[34] a crystallographic uncomplexed (apo) structure in the PDB, and a homology modeled structure in the publicly available database ModBase[35] (Table 1). Models were used as supplied by ModBase to avoid introducing our own bias into the system. Models built on holo templates were used for DHFR, PNP, and PARP. Holo models were unavailable for the other systems, so models built on apo templates were used. All models selected had reliable fold and model assignments with a model score[36] of 1.00 and a sequence alignment *E*-value[37] $\leq 2 \times 10^{-53}$. ModBase was used as the source of the modeled structures because it was a publicly available database; it may be that models from other programs would have performed better. We note that MODELLER, the program that generated the models in ModBase, has consistently performed well, and the models used appeared to be free of gross errors in the binding site.

Database Preparation. The 2000.2 version of the MDDR was filtered[38] to remove molecules containing phosphene; saturated terminal alkyl chains of *n*-heptane or greater; percholorate; more than seven fluorines; more than six chlorines, bromines or iodines; acid halide; carbazide; acid anhydride; peroxide; isocyanate; isothiocyanate; phosphorane; phosphate or sulfur halide; carbodiimide; or cyanohydrin; leaving 95,579 unique molecules in the database. Protons were added using Omega 0.9 (OpenEye Scientific Software, Santa Fe, NM); to account for multiple protonation states and multiple conformations of unsaturated rings, there were a total of 133 068 molecular entries in the database. Conformations of these molecules were generated using Omega 0.9 and stored in a multiconformer database. Details on database preparation will be described elsewhere; briefly, multiple conformations were calculated for each database molecule so that the largest rigid fragment for any given molecule was always in the same position. This defined a frame of reference for the other conformations calculated for the molecule, as described previously for docking ligand conformational ensembles.[39] The database used here extends this ligand ensemble by a fully hierarchical organization of ligand conformational information, reducing redundancy throughout the "tree" of ligand conformations and not just in the rigid fragment alone as in the previous implementation.[39] Partial atomic charges, solvation energies,[13] and van der Waals parameters[40] for the ligands were calculated as previously described.

Molecular Docking. To place all three structures in the same frame of reference, the crystallographic apo and homology modeled structures were aligned onto the crystallographic holo structure by matching Cα backbone atoms of well-defined secondary structural elements including helices, β-sheets, and some loops. All Cα RMSDs for matching onto the crystallographic holo receptor were <1.0 Å except for matching the AChE model onto 1e66 (Cα RMSD = 2.08 Å). SAHH structures were aligned using only the Cα atoms of the catalytic domain. This alignment had no influence on scoring of the docked ligands; it also simplified the comparison of docked and crystallographic ligand positions. After alignment, protons were added to the receptors with SYBYL (Tripos, St. Louis, MO).

Each receptor was then prepared for docking in a similar manner. A grid-based excluded volume map was calculated using DISTMAP.[41] CHEMGRID[40] was used to calculate an AMBER-based[40] van der Waals potential for the receptor. DelPhi[42] was used to calculate an electrostatic potential for the receptor with an internal dielectric of 2 and an external dielectric of 78. To approximate the effect of ligand binding, the effective dielectric of the binding site was reduced by identifying the volume occupied by ligand atoms as a low dielectric region. Several low dielectric atom-sized elements were placed inside the binding site; their positions were obtained from the positions of the ligand atoms in the crystallographic holo structure and (for PNP, GART, PARP, AR, DHFR, and SAHH) from SPHGEN.[43] The same low dielectric volume elements were used in all three protein conformations for each system. These elements did not bear any charge and were only used to lower the dielectric of the binding site; they did not affect excluded volume or van der Waals calculations.

The holo and apo crystal structures for some of the proteins contained additional ligands (cofactors or substrates) in addition to the ligand in the binding site targeted for docking. When present, these additional ligands were included in the excluded volume map, electrostatic potential map, and van der Waals potential map for the receptor. For the modeled conformation, the coordinates of additional ligands were obtained from the following PDB structures: DHFR, NADPH from 3dfr; PNP, PO_4 from 1b8o; GART, β-GAR from 1c2t; SAHH, NAD^+ from 1b3r; AR, $NADP^+$ from 1c9w; and TS, 2′-deoxyuridine-5′-monophosphate from 1bid.

Ligand atoms from the holo structure served as receptor matching positions (spheres) to dock database molecules in the site. These positions defined the orientations sampled by the ligand in the site, using the DOCK matching algorithm in which sets of receptor site positions are matched against sets of ligand atom positions.[43] The spheres were individually labeled[44] based on their hydrogen bonding properties and charges. For each system, the same set of spheres was used for the holo, apo, and modeled conformations.

The Northwestern version of DOCK3.5 was used to flexibly dock the ligands in the MDDR into the active site of each receptor, based on a hierarchical method of sampling ligand conformations.[45] To sample ligand orientations, bin sizes varied for each system but the same parameters were applied to the holo, apo, and modeled conformation of each protein. Over all systems, ligand and receptor bins were set to 0.4−0.6 Å and overlap bins were set to 0.2−0.4 Å; the distance tolerance for matching ligand atoms to receptor matching sites in all cases was 1.2 Å. Each ligand configuration was sampled for steric fit; those passing the steric filter were scored for combined electrostatic and van der Waals complementarity. Each energy score was adjusted by an electrostatic and an apolar desolvation term calculated for each ligand by the program AMSOL, as described.[13] This precalculated desolvation penalty from AMSOL calculates the cost of moving the ligand from water to a dielectric of 2. This desolvation cost was based on the orientation adopted by the ligand in the binding site according to how buried any given ligand atom was by the receptor. Ligands atoms within 4.5 Å of 12 receptor atoms were considered fully desolvated; contact with fewer receptor atoms resulted in a proportionally lower desolvation penalty. This contact-based algorithm is admittedly crude and under development, but previous work suggests that it yields more reasonable scores than either full ligand desolvation or no ligand desolvation at all. The best-scoring conformation for any given ligand orientation was then minimized with 20 steps of simplex rigid-body minimization.[46] Docking statistics, including overall run times, are listed in Supporting Information Table S1.

Ligand Identification. All molecules annotated in the MDDR as an inhibitor of each protein system studied were initially counted as ligands for the protein; covalent inhibitors were removed from each set to obtain the final set of "known ligands" for each protein. The PNP ligands were divided into two sets: ligands bearing a phosphate group were used as ligands for the "PNP without PO_4" system, and ligands not

bearing a phosphate group were used as ligands for the "PNP with PO_4" system.

For each docking calculation, all molecules in the MDDR that could be fit in the site were ranked by their energy score. For molecules represented in multiple protonation states in the database, only the top-scoring version of each molecule was kept. This rank-ordered list was then divided into bins of 96 molecules each (approximately 0.1% of the unique molecules in the MDDR). The ability of DOCK to identify a subset of known ligands from all ligands in the MDDR was evaluated in two ways: first, the cumulative percentage of known molecules found in each bin of 96 molecules and second, the enrichment factor over random selection, calculated as described.[13]

Acknowledgment. We thank D. Lorber and J. Irwin for technical assistance and thoughtful comments; J. Horn, P. Focia, and T. Roth for reading this manuscript; MDL Inc. (San Leandro, CA) for providing the MDDR database and ISIS software; and Open Eye Scientific Software (Santa Fe, NM) for providing Omega. This work was funded by NIH GM59957 (to B.K.S.). S.L.M. was partly supported by NIH T32-GM08061 (K. Mayo, PI) and is a Presidential Scholar of Northwestern University.

Abbreviations

AChE, acetylcholinesterase; AR, aldose reductase; DHFR, dihydrofolate reductase; β-GAR, β-glycinamide ribonucleotide; GART, glycinamide ribonucleotide transformylase; MDDR, MDL Drug Data Report; NAD(P)(H), nicotinamide adenine dinucleotide (phosphate), (reduced); PARP, poly(ADP-ribose) polymerase; PNP, purine nucleoside phosphorylase; RMSD, root-mean square deviation; SAHH, *S*-adenosylhomocysteine hydrolase; TS, thymidylate synthase.

Supporting Information Available: Docking statistics and chemical structures of the crystallographic ligands and representative docked ligands in Figures 3 and 4. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Schnecke, V.; Swanson, C. A.; Getzoff, E. D.; Tainer, J. A.; Kuhn, L. A. Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. *Proteins* **1998**, *33*, 74−87.
(2) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100−5109.
(3) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.
(4) Gschwend, D. A.; Good, A. C.; Kuntz, I. D. Molecular docking towards drug discovery. *J. Mol. Recognit.* **1996**, *9*, 175−186.
(5) Hoffmann, D.; Kramer, B.; Washio, T.; Steinmetzer, T.; Rarey, M. et al. Two-stage method for protein−ligand docking. *J. Med. Chem.* **1999**, *42*, 4422−4433.
(6) Abagyan, R.; Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **2001**, *5*, 375−382.
(7) http://us.expasy.org/sprot/.
(8) Wang, G.; Dunbrack, R. L. http://www.fccc.edu/research/labs/dunbrack/pisces/.
(9) Enyedy, I. J.; Ling, Y.; Nacro, K.; Tomita, Y.; Wu, X. et al. Discovery of small-molecule inhibitors of Bcl-2 through structure-based computer screening. *J. Med. Chem.* **2001**, *44*, 4313−4324.
(10) Honma, T.; Hayashi, K.; Aoyama, T.; Hashimoto, N.; Machida, T. et al. Structure-based generation of a new class of potent Cdk4 inhibitors: new de novo design strategy and library design. *J. Med. Chem.* **2001**, *44*, 4615−4627.
(11) Schapira, M.; Raaka, B. M.; Samuels, H. H.; Abagyan, R. In silico discovery of novel Retinoic Acid Receptor agonist structures. *BMC Struct. Biol.* **2001**, *1*, 1.
(12) Zuccotto, F.; Zvelebil, M.; Brun, R.; Chowdhury, S. F.; Di Lucrezia, R. et al. Novel inhibitors of *Trypanosoma cruzi* dihydrofolate reductase. *Eur. J. Med. Chem.* **2001**, *36*, 395−405.
(13) Wei, B. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W.; Shoichet, B. K. A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol.* **2002**, *322*, 339−355.
(14) Montgomery, J. A.; Niwas, S.; Rose, J. D.; Secrist, J. A., 3rd; Babu, Y. S. et al. Structure-based design of inhibitors of purine nucleoside phosphorylase. 1. 9-(arylmethyl) derivatives of 9-deazaguanine. *J. Med. Chem.* **1993**, *36*, 55−69.
(15) Halazy, S.; Eggenspiller, A.; Ehrhard, A.; Danzin, C. Phosphonate derivatives of N^9-benzylguanine: A new class of potent purine nucleoside phosphorylase inhibitors. *Bioorg. Med. Chem. Lett.* **1992**, *2*, 407−410.
(16) Andricopulo, A. D.; Yunes, R. A. Structure−activity relationships for a collection of structurally diverse inhibitors of purine nucleoside phosphorylase. *Chem. Pharm. Bull. (Tokyo)* **2001**, *49*, 10−17.
(17) Farutin, V.; Masterson, L.; Andricopulo, A. D.; Cheng, J.; Riley, B. et al. Structure−activity relationships for a class of inhibitors of purine nucleoside phosphorylase. *J. Med. Chem.* **1999**, *42*, 2422−2431.
(18) Mao, C.; Cook, W. J.; Zhou, M.; Federov, A. A.; Almo, S. C. et al. Calf spleen purine nucleoside phosphorylase complexed with substrates and substrate analogues. *Biochemistry* **1998**, *37*, 7135−7146.
(19) Griffin, R. J.; Curtin, N. J.; Newell, D. R.; Golding, B. T.; Durkacz, B. W. et al. The role of inhibitors of poly(ADP-ribose) polymerase as resistance- modifying agents in cancer therapy. *Biochimie* **1995**, *77*, 408−422.
(20) Suto, M. J.; Turner, W. R.; Arundel-Suto, C. M.; Werbel, L. M.; Sebolt-Leopold, J. S. Dihydroisoquinolinones: the design and synthesis of a new series of potent inhibitors of poly(ADP-ribose) polymerase. *Anticancer Drug Des.* **1991**, *6*, 107−117.
(21) White, A. W.; Almassy, R.; Calvert, A. H.; Curtin, N. J.; Griffin, R. J. et al. Resistance-modifying agents. 9. Synthesis and biological properties of benzimidazole inhibitors of the DNA repair enzyme poly(ADP-ribose) polymerase. *J. Med. Chem.* **2000**, *43*, 4084−4097.
(22) Ruf, A.; de Murcia, G.; Schulz, G. E. Inhibitor and NAD+ binding to poly(ADP-ribose) polymerase as derived from crystal structures and homology modeling. *Biochemistry* **1998**, *37*, 3893−3900.
(23) Balasubramanian, B. N., Ed. *Advances in the Design and Development of Thrombin Inhibitors*; 1995; pp 999−1007.
(24) Dev, I. K.; Harvey, R. J. N10−Formyltetrahydrofolate is the formyl donor for glycinamide ribotide transformylase in *Escherichia coli*. *J. Biol. Chem.* **1978**, *253*, 4242−4244.
(25) Beardsley, G. P.; Moroson, B. A.; Taylor, E. C.; Moran, R. G. A new folate antimetabolite, 5,10-dideaza-5,6,7,8-tetrahydrofolate is a potent inhibitor of de novo purine synthesis. *J. Biol. Chem.* **1989**, *264*, 328−333.
(26) Almassy, R. J.; Janson, C. A.; Kan, C. C.; Hostomska, Z. Structures of apo and complexed *Escherichia coli* glycinamide ribonucleotide transformylase. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6114−6118.
(27) Henderson, D. M.; Hanson, S.; Allen, T.; Wilson, K.; Coulter-Karis, D. E. et al. Cloning of the gene encoding Leishmania donovani S-adenosylhomocysteine hydrolase, a potential target for antiparasitic chemotherapy. *Mol. Biochem. Parasitol.* **1992**, *53*, 169−183.
(28) Wolos, J. A.; Frondorf, K. A.; Esser, R. E. Immunosuppression mediated by an inhibitor of S-adenosyl-L-homocysteine hydrolase. Prevention and treatment of collagen-induced arthritis. *J. Immunol.* **1993**, *151*, 526−534.
(29) Wnuk, S. F.; Yuan, C. S.; Borchardt, R. T.; Balzarini, J.; De Clercq, E. et al. Nucleic acid related compounds. 84. Synthesis of 6′-(*E* and *Z*)- halohomovinyl derivatives of adenosine, inactivation of *S*-adenosyl-L- homocysteine hydrolase, and correlation of anticancer and antiviral potencies with enzyme inhibition. *J. Med. Chem.* **1994**, *37*, 3579−3587.
(30) Urzhumtsev, A.; Tete-Favier, F.; Mitschler, A.; Barbanton, J.; Barth, P. et al. A 'specificity' pocket inferred from the crystal structures of the complexes of aldose reductase with the pharmaceutically important inhibitors tolrestat and sorbinil. *Structure* **1997**, *5*, 601−612.
(31) Stout, T. J.; Stroud, R. M. The complex of the anti-cancer therapeutic, BW1843U89, with thymidylate synthase at 2.0 A resolution: implications for a new mode of inhibition. *Structure* **1996**, *4*, 67−77.
(32) Kick, E. K.; Roe, D. C.; Skillman, A. G.; Liu, G.; Ewing, T. J. et al. Structure-based design and combinatorial chemistry yield low nanomolar inhibitors of cathepsin D. *Chem. Biol.* **1997**, *4*, 297−307.
(33) Fradera, X.; de la Cruz, X.; Silva, C. H. T. P.; Gelpi, J. L.; Luque, F. J. et al. Ligand-induced changes in the binding sites of proteins. *Bioinformatics* **2002**, *18*, 939−948.

(34) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N. et al. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(35) Pieper, U.; Eswar, N.; Stuart, A. C.; Ilyin, V. A.; Sali, A. MODBASE, a database of annotated comparative protein structure models. *Nucleic Acids Res.* **2002**, *30*, 255−259.
(36) Melo, F.; Sanchez, R.; Sali, A. Statistical potentials for fold assessment. *Protein Sci.* **2002**, *11*, 430−448.
(37) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z. et al. Gapped BLAST and PSI−BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.
(38) Hann, M.; Hudson, B.; Lewell, X.; Lifely, R.; Miller, L. et al. Strategic pooling of compounds for high-throughput screening. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 897−902.
(39) Lorber, D. M.; Shoichet, B. K. Flexible ligand docking using conformational ensembles. *Protein Sci.* **1998**, *7*, 938−950.
(40) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **1992**, *13*, 505−524.
(41) Shoichet, B. K.; Bodian, D. L.; Kuntz, I. D. Molecular docking using shape descriptors. *J. Comput. Chem.* **1992**, *13*, 380−397.
(42) Gilson, M. K.; Honig, B. H. Calculation of electrostatic potentials in an enzyme active site. *Nature* **1987**, *330*, 84−86.
(43) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.
(44) Shoichet, B. K.; Kuntz, I. D. Matching chemistry and shape in molecular docking. *Protein Eng.* **1993**, *6*, 723−732.
(45) Lorber, D. M.; Udo, M. K.; Shoichet, B. K. Protein−protein docking with multiple residue conformations and residue substitutions. *Protein Sci.* **2002**, *11*, 1393−1408.
(46) Gschwend, D. A.; Kuntz, I. D. Orientational sampling and rigid-body minimization in molecular docking revisited: on-the-fly optimization and degeneracy removal. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 123−132.
(47) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. Crystal structures of *Escherichia coli* and Lactobacillus casei dihydrofolate reductase refined at 1.7 A resolution. I. General features and binding of methotrexate. *J. Biol. Chem.* **1982**, *257*, 13650−13662.
(48) Bystroff, C.; Oatley, S. J.; Kraut, J. Crystal structures of *Escherichia coli* dihydrofolate reductase: the NADP+ holoenzyme and the folate. NADP+ ternary complex. Substrate binding and a model for the transition state. *Biochemistry* **1990**, *29*, 3263−3277.
(49) Fedorov, A.; Shi, W.; Kicska, G.; Fedorov, E.; Tyler, P. C. et al. Transition state structure of purine nucleoside phosphorylase and principles of atomic motion in enzymatic catalysis. *Biochemistry* **2001**, *40*, 853−860.
(50) Kapples, K. J.; Shutske, G. M.; Bores, G. M.; Huger, F. P. Synthesis and in vitro acetylcholinesterase inhibitory activity of some 1-substituted analogues of velnacrine. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 2789−2792.
(51) Skrzypczak-Jankun, E.; Carperos, V. E.; Ravichandran, K. G.; Tulinsky, A.; Westbrook, M. et al. Structure of the hirugen and hirulog 1 complexes of alpha-thrombin. *J. Mol. Biol.* **1991**, *221*, 1379−1393.
(52) Greasley, S. E.; Yamashita, M. M.; Cai, H.; Benkovic, S. J.; Boger, D. L. et al. New insights into inhibitor design from the crystal structure and NMR studies of *Escherichia coli* GAR transformylase in complex with beta-GAR and 10-formyl-5, 8, 10-trideazafolic acid. *Biochemistry* **1999**, *38*, 16783−16793.
(53) Turner, M. A.; Yuan, C. S.; Borchardt, R. T.; Hershfield, M. S.; Smith, G. D. et al. Structure determination of selenomethionyl S-adenosylhomocysteine hydrolase using data at a single wavelength. *Nat. Struct. Biol.* **1998**, *5*, 369−376.
(54) Hu, Y.; Komoto, J.; Huang, Y.; Gomi, T.; Ogawa, H. et al. Crystal structure of S-adenosylhomocysteine hydrolase from rat liver. *Biochemistry* **1999**, *38*, 8323−8333.
(55) Wilson, D. K.; Bohren, K. M.; Gabbay, K. H.; Quiocho, F. A. An unlikely sugar substrate site in the 1.65 A structure of the human aldose reductase holoenzyme implicated in diabetic complications. *Science* **1992**, *257*, 81−84.
(56) Dvir, H.; Wong, D. M.; Harel, M.; Barril, X.; Orozco, M. et al. 3D structure of *Torpedo californica* acetylcholinesterase complexed with huprine X at 2.1 A resolution: kinetic and molecular dynamic correlates. *Biochemistry* **2002**, *41*, 2970−2981.
(57) Kamb, A.; Finer-Moore, J.; Calvert, A. H.; Stroud, R. M. Structural basis for recognition of polyglutamyl folates by thymidylate synthase. *Biochemistry* **1992**, *31*, 9883−9890.
(58) Stout, T. J.; Sage, C. R.; Stroud, R. M. The additivity of substrate fragments in enzyme-ligand binding. *Structure* **1998**, *6*, 839−848.
(59) Gangjee, A.; Yu, J.; McGuire, J. J.; Cody, V.; Galitsky, N. et al. Design, synthesis, and X-ray crystal structure of a potent dual inhibitor of thymidylate synthase and dihydrofolate reductase as an antitumor agent. *J. Med. Chem.* **2000**, *43*, 3837−3851.
(60) Li, J.-H.; Tays, K. L.; Zhang, J. Oxo-substituted compounds, process of making, and compositions and methods for inhibiting PARP activity. In PCT Intl. Appl., 1999.
(61) Karlsson, O.; Linschoten, M.; Nystrom, J.-E. New amidino derivatives and their use as thrombin inhibitors. PCT Intl. Appl., 1998.
(62) Habeck, L. L.; Leitner, T. A.; Shackelford, K. A.; Gossett, L. S.; Schultz, R. M. et al. A novel class of monoglutamated antifolates exhibits tight-binding inhibition of human glycinamide ribonucleotide formyltransferase and potent activity against solid tumors. *Cancer Res.* **1994**, *54*, 1021−1026.
(63) Narayanan, S. R.; Keller, B. T.; Borcherding, D. R.; Scholtz, S. A.; Borchardt, R. T. 9-(*trans*-2′,*trans*-3′-Dihydroxycyclopent-4′-enyl) derivatives of adenine and 3-deazaadenine: potent inhibitors of bovine liver *S*-adenosylhomocysteine hydrolase. *J. Med. Chem.* **1988**, *31*, 500−503.
(64) Malamas, M. S.; Sestanj, K.; Millen, J. Naphthalenylsulfonylhydantoins as aldose reductase inhibitors. *Eur. J. Med. Chem.* **1991**, *26*, 369−374.

JM0300330