

# Docking for fragment inhibitors of AmpC $\beta$ -lactamase

Denise G. Teotico<sup>1</sup>, Kerim Babaoglu<sup>1,2</sup>, Gabriel J. Rocklin, Rafaela S. Ferreira, Anthony M. Giannetti<sup>3</sup>, and Brian K. Shoichet<sup>4</sup>

Department of Pharmaceutical Chemistry, University of California, 1700 4th Street, MC 2550, San Francisco, CA 94158

Edited by Gregory A. Petsko, Brandeis University, Waltham, MA, and approved March 17, 2009 (received for review December 22, 2008)

Fragment screens for new ligands have had wide success, notwithstanding their constraint to libraries of 1,000–10,000 molecules. Larger libraries would be addressable were molecular docking reliable for fragment screens, but this has not been widely accepted. To investigate docking's ability to prioritize fragments, a library of >137,000 such molecules were docked against the structure of  $\beta$ -lactamase. Forty-eight fragments highly ranked by docking were acquired and tested; 23 had  $K_i$  values ranging from 0.7 to 9.2 mM. X-ray crystal structures of the enzyme-bound complexes were determined for 8 of the fragments. For 4, the correspondence between the predicted and experimental structures was high (RMSD between 1.2 and 1.4 Å), whereas for another 2, the fidelity was lower but retained most key interactions (RMSD 2.4–2.6 Å). Two of the 8 fragments adopted very different poses in the active site owing to enzyme conformational changes. The 48% hit rate of the fragment docking compares very favorably with "lead-like" docking and high-throughput screening against the same enzyme. To understand this, we investigated the occurrence of the fragment scaffolds among larger, lead-like molecules. Approximately 1% of commercially available fragments contain these inhibitors whereas only 10<sup>-7</sup>% of lead-like molecules do. This suggests that many more chemotypes and combinations of chemotypes are present among fragments than are available among lead-like molecules, contributing to the higher hit rates. The ability of docking to prioritize these fragments suggests that the technique can be used to exploit the better chemotype coverage that exists at the fragment level.

crystallography | drug design | hit rates | chemical space

Fragment-based screening has recently emerged as an important technique in early drug discovery (1–6). Typically constrained to libraries between 1,000 and 10,000 small molecules, it has nevertheless had hit rates of 5% or higher against many targets. Several of these have resisted ligand discovery by high-throughput screening (HTS), which exploits libraries containing 2–3 orders of magnitude more molecules than those typical for fragments. The high fragment hit rates have been attributed to 2 factors. First, fragments can complement subsites in the pocket without making the structural compromises necessary for larger, multifunctional molecules (7). Second, even a small fragment library typically explores many more chemotypes than is possible for libraries of larger molecules. The possible combinations of chemotypes rises exponentially with the size of the molecule (7), apparently increasing by a log order or more for every atom added in a reasonable size range (8, 9). Thus, one might well be more likely to find a good fragment from among a library of 1 thousand molecules than a good "drug-like" molecule from among a library of 1 million.

Still, the small size of fragment libraries leaves many accessible molecules untested. Even the most high-throughput fragment screen can now only address libraries of 10,000–20,000 compounds, whereas >250,000 fragments are commercially available (10). Although one might cover the 10<sup>6</sup> commercially available, rule-of-five-compliant (11) compounds with a well-designed library that is one-tenth that size, the collapse in chemical space with molecular size means that leaving out 90–99% of available fragments will leave out approximately the same ratio of che-

motypes. Thus, it would be interesting to interrogate much larger compound libraries than are currently feasible for the biophysical assays on which fragment screens now rely.

One method to do so is molecular docking. Docking screens compound libraries for molecules that physically complement protein-binding sites. The technique has been used to discover new ligands in the "lead-like" (typically  $\leq$ 350 Da) or drug-like ranges (typically  $\leq$ 500 Da) (12, 13); some of these new ligands have been determined in complex with their targets by crystallography in geometries consistent with the docking predictions (14–17). Docking has also been used to prioritize molecules for fragment-based screens for several targets (5, 18–20). Despite these successes, doubt remains about applying docking to fragments because of possible promiscuous binding modes, the lack of handles to fit fragments into the pocket, and biases in docking scoring functions (21). This skepticism has been buttressed by the lack of structures directly comparing docking fragment predictions with subsequent crystallographic results.

We thus wanted to explore docking as a tool to prioritize fragments using AmpC  $\beta$ -lactamase as a model system. AmpC is well suited to addressing these questions because binding may be measured in a quantitative biochemical assay and biophysically by surface plasmon resonance (SPR), and high-resolution crystal structures can be readily obtained (22). Also, it is a protein that has been targeted by HTS and by docking (17) using the same library, and by docking alone (16). This allows us to compare the docking fragment hit rates with those from HTS and lead-like docking.

Here, we dock a library of 137,639 fragments against the crystal structure of AmpC  $\beta$ -lactamase and investigate the following questions. Can docking reliably prioritize fragments that inhibit in biochemical and biophysical assays? How do the hit rates compare with those from HTS, docking screens of lead-like molecules, or simply sampling fragments at random? What is the role of chemical space coverage in these different hit rates? How diverse are the fragments compared with the inhibitors discovered from the drug-like libraries? Are the docking predictions right for the right reasons—do experimental structures of the fragment inhibitors correspond to the docked poses?

## Results

**Docking Prioritization of Fragments.** A total of 137,639 compounds from the ZINC database were docked against a crystal structure

Author contributions: D.G.T., K.B., and B.K.S. designed research; D.G.T., K.B., G.J.R., R.S.F., and A.M.G. performed research; D.G.T., K.B., G.J.R., R.S.F., and A.M.G. analyzed data; and D.G.T. and B.K.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates have been deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank, www.rcsb.org (RCSB ID codes 3GV9, 3GV9, 3GRJ, 3GR2, 3GTC, 3GQZ, 3G5G).

<sup>1</sup>D.G.T. and K.B. contributed equally to this work.
























<sup>2</sup>Present address: Gilead Sciences, Inc., 333 Lakeside Drive, Foster City, CA 94404.

<sup>3</sup>Present address: Genentech, 1 DNA Way, South San Francisco, CA 94080.

<sup>4</sup>To whom correspondence should be addressed. E-mail: shoichet@cgl.ucsf.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0813029106/DCSupplemental](http://www.pnas.org/cgi/content/full/0813029106/DCSupplemental).

Table 1. Twenty-three fragment inhibitors identified biochemically and biophysically

Fragment structure	No.	Rank*	$K_i$ , $mM$	HAC <sup>†</sup>	Molecular mass, Da	LE <sup>‡</sup>
	1	270 <sup>§</sup>	2.0	17	236	0.21
	2	139 <sup>§</sup>	1.0	17	236	0.24
	3	274 <sup>§</sup>	3.0	14	193	0.24
	4	89 <sup>§</sup>	2.0	18	248	0.20
	5	210 <sup>§</sup>	6.7	18	246	0.16
	6	52 <sup>§</sup>	1.8	13	200	0.29
	7	76 <sup>§</sup>	3.5	17	234	0.20
	8	47 <sup>§</sup>	2.6	16	220	0.20
	9	18 <sup>§</sup>	0.7	18	244	0.24
	10	159 <sup>§</sup>	2.2	17	232	0.21
	11	193 <sup>§</sup>	6.0	13	219	0.23
	12	149 <sup>§</sup>	1.0	14	188	0.29
	13	1 <sup>¶</sup>	2.7	8	138	0.47
	14	15 <sup>¶</sup>	1.7/3.9 <sup>  </sup>	16	239	0.25
	15	373 <sup>¶</sup>	7.0	15	227	0.20
	16	22 <sup>¶</sup>	7.1	16	215	0.18
	17	34 <sup>¶</sup>	3.2	15	238	0.23
	18	27 <sup>¶</sup>	7.7/21.0 <sup>  </sup>	16	217	0.18
	19	87 <sup>¶</sup>	2.0/6.8 <sup>  </sup>	14	227	0.26
	20	159 <sup>¶</sup>	4.5	15	226	0.21
	21	162 <sup>¶</sup>	7.5	12	185	0.24
	22	48 <sup>¶</sup>	4.9/18.0 <sup>  </sup>	12	185	0.26
	23	39 <sup>¶</sup>	9.2	16	237	0.17

\*Docking rank.

<sup>†</sup> Heavy atom count — defined as the number of nonhydrogen atoms.<sup>‡</sup> (kcal/mol) per atom.<sup>§</sup> Ranking based on docking of 117,567 compounds.<sup>||</sup>  $K_i$  values calculated from SPR assays.<sup>¶</sup> Ranking based on docking of 47,997 compounds.

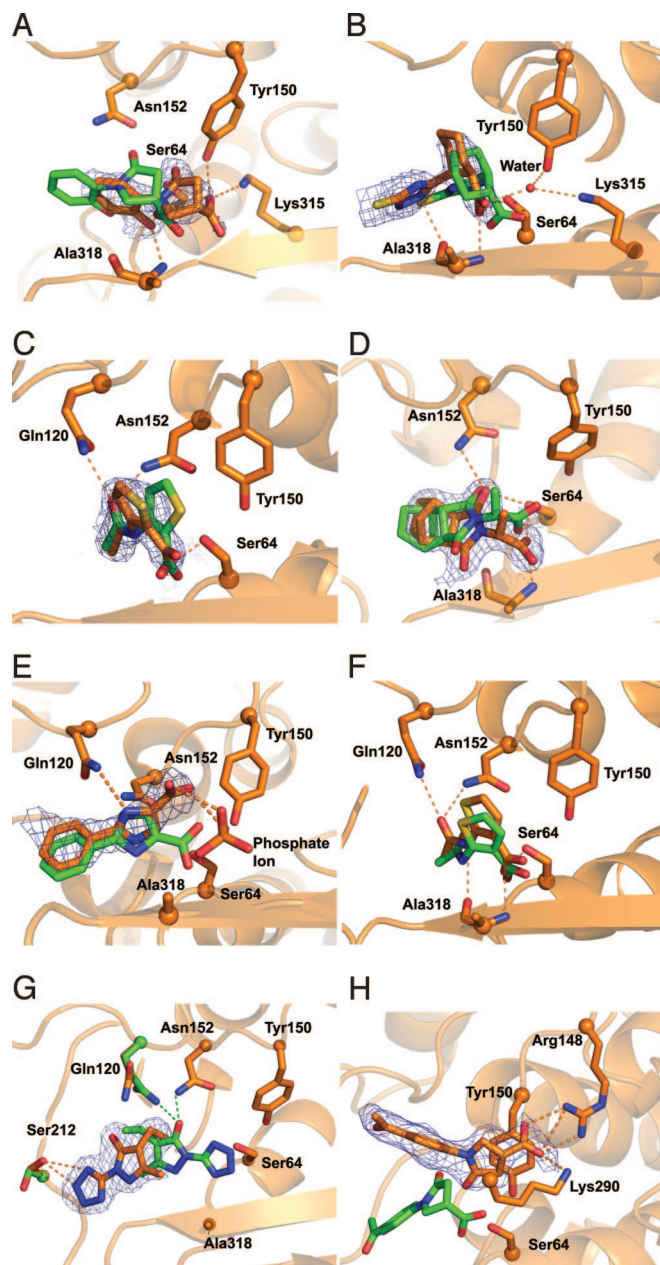
of AmpC. Two libraries were used for the docking calculations: the first developed by using the definition of fragments proposed by Carr and Rees (23) (47,997 compounds) and a second that removed restrictions on hydrogen bond acceptors and donors (89,642 new compounds plus 27,925 compounds overlapping with the original database). The total elapsed time for the calculation ranged from 5 h (47,997 compounds) to 32 h (117,567 compounds) on a cluster using between 4 (47,997 compounds) and 9 cpus (117,567 compounds). From among the 500 top-scoring molecules from each screen, 48 were purchased and tested in both enzyme and SPR assays (*Methods*).

Dose–response curves were fit to inhibition numbers to obtain  $IC_{50}$  and  $K_i$  values [Table 1, [supporting information \(SI\) Figs. S1A and S2](#)]. Varying substrate and inhibitor concentrations allowed us to construct a Dixon Plot of 2 of the compounds, which confirmed competitive inhibition ([Fig. S1B](#)). Of the 48 compounds tested, 23 had  $K_i$  values of <10 mM, ranging from 700  $\mu$ M to 9.2 mM, with ligand efficiencies ranging from 0.16 to 0.47 (Table 1). Three inhibitors, compounds 6, 12, and 13, had ligand efficiencies of 0.29, 0.29, and 0.47 kcal/mol per atom. Dock rankings of the inhibitors ranged from 1 to 373 of 137,639 docked. This gives a 48% hit rate, defined as (number of actives/number of compounds experimentally tested)  $\times$  100. SPR experiments were conducted for 4 compounds, and for these biochemical  $K_i$  and biophysical  $K_d$  values correlated well (Table 1, [Fig. S3](#)).

To ensure that the high hit rate observed was not an artifact of screening at a high concentration, we tested a set of fragments chosen at random from the library. Of 20 random fragments tested, 1 inhibited detectably with a  $K_i$  of 3.1 mM. As it happened, this 1 random hit also scored well by docking (rank = 2,223, docking score =  $-60.3$  kcal/mol), ranking among the top 5% of all molecules docked.

Similarity analysis was performed to measure the diversity of the 23 fragment hits relative to each other and relative to that of 21 known lead-like inhibitors of AmpC ([Fig. S4](#)) (16, 17). By using FCFP-4 fingerprints (SciTegic Inc. and Accelrys Inc.) generated with Pipeline Pilot, the average pairwise Tanimoto coefficient within the fragment set was 0.22, and the lowest and highest Tanimoto coefficients were 0.13 and 0.26. Conversely, similarity coefficients within the lead-like sulfonamides and phthalimides previously discovered for AmpC (16, 17) fell between 0.21 and 0.55, with an average value of 0.43. Comparing the similarity between the 2 sets (i.e., fragments vs. lead-like), the minimum Tanimoto coefficient was 0.16, the maximum was 0.23, and the mean value was 0.19. Thus, the known lead-like inhibitors are internally less diverse than the fragments, and the fragments are structurally distinct from the known inhibitors.

**Comparing the Docking Predicted Pose to the Crystal Structure.** Of the 23 active fragments, 8 crystal structures were determined with resolutions from 1.5 Å to 2.5 Å (Table S1). The location of the inhibitors was unambiguous in the final  $2F_o - F_c$  electron density map ([Fig. 1 A–E, G, and H](#)). Fragment 22 ([Fig. 1 F](#)) was previously crystallized (24). To evaluate accuracy, the structure predicted by docking, directly from the initial calculation and without further refinement, was overlaid by using protein backbone and ligand atoms onto the crystallographic conformation. RMSD values were then calculated by using ligand atoms only. For fragments 21, 12, 1, and 22, the docking predictions corresponded closely to the X-ray results, with RMSD values of 1.2, 1.6, 1.4, and 1.3 Å, respectively ([Fig. 1 C–F](#)). For compounds 8 and 20, the RMSD values were 2.5 and 2.6 Å, making the correspondence substantially worse. Even here, both crystal structures retain most of the contacts with the catalytic residues predicted in the docking pose ([Fig. 1 A and B](#)). For fragment 8, the crystallographic orientation differs mainly by a translational shift. We note that this fragment had 2 configurations in the



**Fig. 1.** Overlay of docked pose (green) and crystallographic pose (orange) for 8 of the fragment inhibitors prioritized by docking. The compounds shown are: 8 ([A](#)), 20 ([B](#)), 21 ([C](#)), 1 ([D](#)), 12 ([E](#)), 22 ([F](#)), 3 ([G](#)), and 5 ([H](#)). The final  $2F_o - F_c$  maps contoured at  $1\sigma$  are shown for [A–E, G, and H](#). Compound 22, although discovered as part of the docking screen described here, was reported previously and no density is shown for it (24).

electron density, both modeled at 50% occupancy; here, we consider only that one closest to the docking prediction. For compound 20, the “warhead” carboxylate group interacts with both the catalytic serine and the oxyanion hole, as predicted in the docked pose ([Fig. 1 B](#)). The only substantial change is in the position of the triazole group, which results from favorable van der Waals interactions between the cyclohexyl ring and the protein.

In all these AmpC-fragment crystal structures, with 2 exceptions noted below, we typically observe hydrogen bonds between the ligand and the following residues: the catalytic Ser-64, one or both of the oxyanion hole nitrogens from Ala-318 and Ser-64,





is bias in the library, such as aminergic GPCR ligands (26, 27). Said another way, the chances of discovering interesting chemotypes for biological targets is many orders of magnitude higher when targeting molecules in the fragment weight range than even at slightly higher size ranges. These points have been made by others, and have been modeled by Hann and colleagues (7). The contribution of this work is to reduce general principles and models to specific quantification for a particular series of chemotypes actually found to inhibit an enzyme, and to show how structure-based docking can prioritize molecules from among a large set of available fragments that would be inaccessible by empirical approaches alone.

## Materials and Methods

**Docking.** Fragment-like molecules from the ZINC database (10) were docked into an apo AmpC structure (PDB ID code 1KE4) (28) by using DOCK3.5.54. Two screens of the ZINC fragments were conducted: once when the library contained 50,000 fragments and again after it had expanded to 137,000 fragments. For the first screen, compounds were filtered for the following properties:  $150 \leq$  molecular mass  $\leq 250$ Da,  $-2 \leq$  ClogP  $\leq 3$ , number of rotatable bonds  $\leq 3$ , number of H-bond donors  $\leq 2$ , and acceptors  $\leq 4$ . In the second screen, the restrictions on rotatable bonds and hydrogen bonding were removed. To prepare the protein-binding site, all water and ion molecules were removed except for Wat403 (in the first screen), and Wat466 and Wat565 (in the second) (SI Text).

For the docking calculations themselves, a sphere set was constructed based on 13 ligand-bound X-ray structures (17); these spheres determine how ligand orientations in the active site are calculated. Ligand-protein fit was calculated based van der Waals and electrostatic complementarity between enzyme and

ligand, corrected for ligand desolvation (see SI Text). After docking, the 500 top-ranking poses were visually inspected. In addition to simple rank, which for 137,000 compounds must be the primary criterion, compounds were prioritized for contacts to key catalytic residues (e.g., Ser-64 and the oxyanion hole), chemotype novelty, and commercial availability. This is our common practice in prosecuting docking screens, and these criteria are consistent with previous studies. Forty-eight high-scoring docking compounds were purchased and tested experimentally, as were 20 fragments selected at random (SI Text).

**Absorbance and SPR Assays.** For information on absorbance and SPR assays, see SI Text.

**Ligand Similarity and Substructure Search.** For information on ligand similarity and substructure search, See SI Text.

**Crystallography.** All AmpC/fragment structures were based on cocrystallization and structures determined by molecular replacement (see SI Text).

**Theoretical Number of Lead-Like Compounds.** Eq. 1 was used to determine how many lead-like compounds should theoretically contain the 23 active fragments as substructures. The GDB was used as a source of side chains to add to fragment scaffolds (8). Multiple boundary conditions were set to make the calculation feasible (see Results).

**ACKNOWLEDGMENTS.** We thank Yu Chen, Jerome Hert, Pascal Wassam, and Peter Kolb for discussions and assistance with chemoinformatics and Michael Mysinger, Sarah Boyce, and Yu Chen for reading the manuscript. This work was supported by National Institutes of Health Grants GM59957 and GM63813.

- Allen KN, et al. (1996). An experimental approach to mapping the binding surfaces of crystalline proteins. *J Phys Chem* 100:2605–2611.
- Ringe D, Mattos C (1999) Analysis of the binding surfaces of proteins. *Med Res Rev* 19:321–331.
- Mattos C, Ringe D (1996) Locating and characterizing binding sites on proteins. *Nat Biotechnol* 14:595–599.
- Verdonk ML, Hartshorn MJ (2004) Structure-guided fragment screening for lead discovery. *Curr Opin Drug Discov Dev* 7:404–410.
- Hubbard RE, Davis B, Chen I, Drysdale MJ (2007) The SeeDs approach: Integrating fragments into drug discovery. *Curr Top Med Chem* 7:1568–1581.
- Verlinde CL, Rudenko G, Hol WG (1992) In search of new lead compounds for trypanosomiasis drug design: A protein structure-based linked-fragment approach. *J Comput Aided Mol Des* 6:131–147.
- Hann MM, Leach AR, Harper G (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci* 41:856–864.
- Fink T, Raymond JL (2007) Virtual exploration of the chemical universe. *J Chem Inf Model* 47:342–353.
- Wester MJ, et al. (2008) Scaffold topologies. 2. Analysis of chemical databases. *J Chem Inf Model* 48:1311–1324.
- Irwin JJ, Shoichet BK (2005) ZINC—A free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45:177–182.
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Delivery Rev* 46:3–26.
- Oprea TI, Davis AM, Teague SJ, Leeson PD (2001) Is there a difference between leads and drugs? A historical perspective. *J Chem Inf Comput Sci* 41:1308–1315.
- Kellenberger E, et al. (2007) Identification of nonpeptide CCR5 receptor agonists by structure-based virtual screening. *J Med Chem* 50:1294–1303.
- Brenk R, et al. (2003) Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. *J Med Chem* 46:1133–1143.
- Rosenfeld RJ, et al. (2003) Automated docking of ligands to an artificial active site: Augmenting crystallographic analysis with computer modeling. *J Comput Aided Mol Des* 17:525–536.
- Powers RA, Morandi F, Shoichet BK (2002) Structure-based discovery of a novel, noncovalent inhibitor of AmpC beta-lactamase. *Structure (London)* 10:1013–1023.
- Babaoglu K, et al. (2008) Comprehensive mechanistic analysis of hits from high-throughput and docking screens against beta-lactamase. *J Med Chem* 51:2502–2511.
- Warner SL, et al. (2006) Identification of a lead small-molecule inhibitor of the Aurora kinases by using a structure-assisted, fragment-based approach. *Mol Cancer Ther* 5:1764–1773.
- Huang D, et al. (2005) Discovery of cell-permeable non-peptide inhibitors of beta-secretase by high-throughput docking and continuum electrostatics calculations. *J Med Chem* 48:5108–5111.
- Wyatt PG, et al. (2008) Identification of *N*-(4-piperidinyl)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design. *J Med Chem* 51:4986–4999.
- Hubbard RE, Chen I, Davis B (2007) Informatics and modeling challenges in fragment-based drug discovery. *Curr Opin Drug Discovery Dev* 10:289–297.
- Chen Y, Minasov G, Roth TA, Prati F, Shoichet BK (2006) The deacylation mechanism of AmpC beta-lactamase at ultrahigh resolution. *J Am Chem Soc* 128:2970–2976.
- Carr RA, Congreve M, Murray CW, Rees DC (2005) Fragment-based lead discovery: Leads by design. *Drug Discovery Today* 10:987–992.
- Babaoglu K, Shoichet BK (2006) Deconstructing fragment-based inhibitor discovery. *Nat Chem Biol* 2:720–723.
- Kramer B, Rarey M, Lengauer T (1999) Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* 37:228–241.
- Lowrie JF, Delisle RK, Hobbs DW, Diller DJ (2004) The different strategies for designing GPCR and kinase targeted libraries. *Comb Chem High Throughput Screen* 7:495–510.
- Reynolds CH, et al. (2001) Diversity and coverage of structural sublibraries selected using the SAGE and SCA algorithms. *J Chem Inf Comput Sci* 41:1470–1477.
- Powers RA, Shoichet BK (2002) Structure-based approach for binding site identification on AmpC beta-lactamase. *J Med Chem* 45:3222–3234.