JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

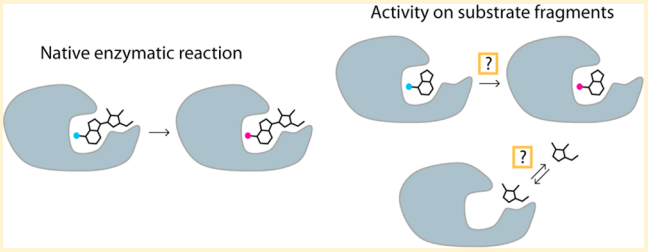
Substrate Deconstruction and the Nonadditivity of Enzyme Recognition

Sarah Barelier,[†] Jennifer A. Cummings,[‡] Alissa M. Rauwerdink,[‡] Daniel S. Hitchcock,[§] Jeremiah D. Farelli,[∥] Steven C. Almo,[⊥] Frank M. Raushel,*[,‡,§] Karen N. Allen,*[,∥] and Brian K. Shoichet*[,†,#]

[†]Department of Pharmaceutical Chemistry, University of California - San Francisco, 1700 Fourth Street, Byers Hall, San Francisco, California 94158, United States
[‡]Department of Chemistry, Texas A&M University, College Station, Texas 77843, United States
[§]Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas, United States
[∥]Department of Chemistry, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215-2521, United States
[⊥]Department of Biochemistry, Albert Einstein College of Medicine, New York, New York 10461, United States
[#]Faculty of Pharmacy, University of Toronto, Donnelly Centre Suite 604, 160 College Street, Toronto, Ontario, Canada, M5S 3E1

Supporting Information

ABSTRACT: Predicting substrates for enzymes of unknown function is a major postgenomic challenge. Substrate discovery, like inhibitor discovery, is constrained by our ability to explore chemotypes; it would be expanded by orders of magnitude if reactive sites could be probed with fragments rather than fully elaborated substrates, as is done for inhibitor discovery. To explore the feasibility of this approach, substrates of six enzymes from three different superfamilies were deconstructed into 41 overlapping fragments that were tested for activity or binding. Surprisingly, even those fragments containing the key reactive group had little activity, and most fragments did not bind measurably, until they captured most of the substrate features. Removing a single atom from a recognized substrate could often reduce catalytic recognition by 6 log-orders. To explore recognition at atomic resolution, the structures of three fragment complexes of the β-lactamase substrate cephalothin were determined by X-ray crystallography. Substrate discovery may be difficult to reduce to the fragment level, with implications for function discovery and for the tolerance of enzymes to metabolite promiscuity. Pragmatically, this study supports the development of libraries of fully elaborated metabolites as probes for enzyme function, which currently do not exist.



INTRODUCTION

While the number of protein sequences deposited in public databases continues to expand exponentially,[1] determining the function of the encoded proteins remains slow. Except where sequence identity to a protein of known function is high, the activity of a newly sequenced protein must be interrogated with candidate ligands or substrates. This can be done empirically, by screening for binding or substrate turnover[2−4] or by a mixture of computational prediction, for instance by docking molecular libraries[5−8] and subsequent experimental testing. Both approaches rely on screening libraries of small molecules, such as metabolites.[6] If the right metabolite, or a close analogue, is present in the library, it may be detected as substrate, whereas if it is not, either no activity will be assigned or it may be mis-assigned. In the latter case, more metabolites are needed in our screening libraries. However, the multiple chemotypes present in biological small molecules, and their exponential scaling when combined into more complex biological compounds, make full coverage of biorelevant chemical space difficult to ensure.

In drug discovery, the combinatorial explosion of chemotypes with molecular size has been addressed by screening libraries of fragment molecules.[9] Because fragments are smaller than druglike molecules (typically less than 17 non-hydrogen atoms), fragment chemical space is about 50 orders-of-magnitude smaller than druglike chemical space,[10] enabling fragment libraries to cover chemical space better than libraries of more complex molecules.[11] Individual fragment inhibitors usually present simple chemotypes that are only expanded out to fully elaborated molecules after initial hits are discovered; this has been a remarkably successful approach.[12−17]

A fragment-based strategy could be an attractive alternative to the full enumeration of metabolite space for substrate discovery. Not only would it cover potential substrate space far

Received: February 8, 2014
Published: May 2, 2014

   dx.doi.org/10.1021/ja501354q | J. Am. Chem. Soc. 2014, 136, 7374−7382

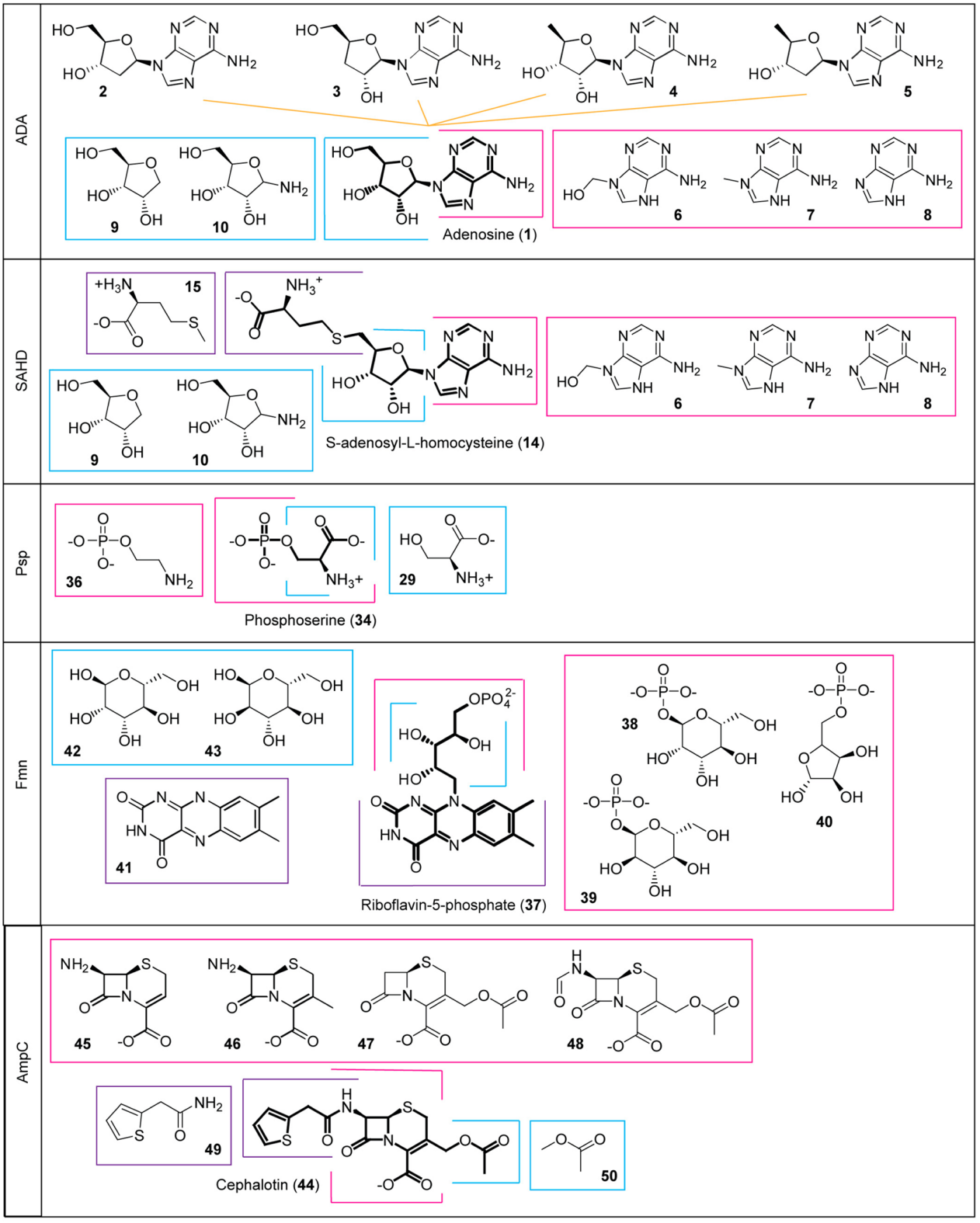


Figure 1. Multiple ways of fragmenting the substrates of adenosine deaminase (ADA); MTA/SAH deaminase (SAHD); phosphoserine phosphatase (PSP); flavin mononucleotide phosphatase (FMP); and AmpC β-lactamase (AmpC). Only one fragment was typically tested at a time. For isoaspartyl dipeptidase (IAD), recognition of the component amino-acid monomers was investigated.

more efficiently, but it would also increase the number of representative molecules that can be physically sourced; currently, many known metabolites and biogenic molecules are simply unavailable for testing. This is far less of a problem for fragments, where molecules containing core reactant groups are readily available; for instance, over 700,000 accessible fragments are cataloged in the ZINC database.[18] A key question is whether a substrate, stripped to the core reactive group on which the catalytic machinery of an enzyme acts, retains enough recognition elements to be an effective, or at least a detectable,

dx.doi.org/10.1021/ja501354q | J. Am. Chem. Soc. 2014, 136, 7374−7382

enzyme substrate. It could be that enzyme catalysis is so demanding that most of the atoms of the substrate must be engaged with the enzyme before catalysis will occur. Several lines of evidence support this view, including studies showing that fragmentation of cytidine into component fragments lowered the activity for cytidine deaminase by 4−9 orders-of-magnitude[19] and that fragmentation of a transition-state analogue of calf adenosine deaminase led to losses of up to 6 orders-of-magnitude in affinity.[20] Also, as shown by Jencks,[21−23] there is no reason why the binding energies of component fragments should sum up to the binding or catalytic activity of a full substrate, owing to the nonadditive, nonequilibrium effects of chemical connectivity. Conversely, other studies suggest that fragments can be built up additively for affinity and catalytic recognition. For instance, the well-studied enzyme chymotrypsin hydrolyzes a variety of substrates, including *p*-nitrophenyl acetate and *p*-nitroanilides.[24,25] These compounds only share a very reduced fragment at the reaction center (two common heavy atoms: C(=O)), which would argue in favor of a fragment-based approach. It has also been possible to deconstruct at least some substrates and transition-state analogues into component fragments. For example, the transition-state analogue immucillin-H, which inhibits purine nucleoside phosphorylase with a K_I of 28 pm, was deconstructed into purine and iminoribitol with a loss of only 30-fold in the multiplied affinities of the component fragments.[26] Crystallographic studies confirmed that the geometry of these fragments, when bound to the enzyme, recapitulates that of the entire immucillin-H.[27] Fragments of the thymidylate synthase substrate deoxyuridine monophosphate, 2′-deoxyuridine, and 2′,5′-dideoxyuridine, also conserve their binding mode when crystallized in the presence of phosphate, although smaller fragments are either poorly ordered in the active site or bind nonspecifically.[28] More recently, the substrate specificity of the mis-annotated enzyme Atu3266 from *Agrobacterium tumefaciens* C58 was determined, starting from a very weak fragment hit (k_{cat}/K_M = 4 M^{-1} s^{-1}) and resulting in a potent substrate (k_{cat}/K_M = 2.8 × 10^5 M^{-1} s^{-1}).[29] Similarly, for triosephosphate isomerase, the difference in activation barrier for the isomerization of whole substrate and substrate in pieces is large (6.6 kcal/mol) but product formation is still detectable.[30] Large increases in proteolytic activity also have been observed when long-chain substrates are hydrolyzed by pepsin and elastase.[31,32] Lastly, in addition to the successes in stepwise optimization of fragment inhibitors for drug discovery,[33−41] a fragment-based approach has been used to develop enzyme inhibitors, based on initial substrate turnover.[42,43] In short, there is evidence to both support and undermine the use of fragments for substrate discovery. The potential benefit of fragment screens against genomic targets, which may dramatically expand our ability to probe chemotype space, spurred us to explore this question further.

If it is true that fragments can be used as probe substrates, fragments of extant substrates should retain substantial turnover, assuming that they keep the reactive center intact. Here we explore this idea systematically on the substrates of six enzymes from three different enzyme superfamilies: the amidohydrolase superfamily members adenosine deaminase (ADA), 5-methylthioadenosine/*S*-adenosyl-homocysteine deaminase (SAHD), and isoaspartyl dipeptidase (IAD); the haloacid dehalogenase superfamily members phosphoserine phosphatase (PSP) and flavin mononucleotide phosphatase (FMP); and AmpC β-lactamase (AmpC). Substrates for each enzyme were deconstructed into fragments (Figure 1), which contained either the catalytic core of the full-length substrate, or "side-chains" of the substrate not directly involved in catalysis. The fragments were tested for enzyme turnover or enzyme inhibition, typically one at a time. In almost every case, the resulting fragments—several variants of which were tried for each enzyme—showed a drop of 10^6 or more in k_{cat}/K_M relative to the canonical substrate, and for most compounds binding could not be detected. These results have implications for our ability to extend the fragment approach, so successful for inhibitor discovery, to substrate discovery, and may illuminate the nature of the small-molecule environment against which enzymes have evolved.

RESULTS

Enzyme Assays. *Fragments of Adenosine As Substrates and Inhibitors of Adenosine Deaminase (ADA).* The canonical substrate of ADA, adenosine (**1**), was fragmented in stages, beginning with the removal of the hydroxyl groups at C2′, C3′, and C5′ of the molecule (**2**, **3**, **4**), and with the dideoxy derivative, 2′,5′-dideoxy adenosine (**5**). Subsequently, the entire ribose was removed to give 9-hydroxymethyladenine (**6**), 9-methyladenine (**7**), and adenine (**8**); all retained the core reactive purine amine on which the deamination reaction occurs. All fragments were tested as substrates for ADA (Table 1). The k_{cat}/K_M value for adenosine is 7.6 × 10^6 M^{-1} s^{-1}. While the 2′-deoxy derivative **2** retained most of that activity (k_{cat}/K_M 4.1 × 10^6 M^{-1} s^{-1}), the 3′-deoxy **3** lost over 1 order of magnitude of activity (k_{cat}/K_M 2 × 10^5 M^{-1} s^{-1}), and in the 5′-deoxy and dideoxy analogues **4** and **5** no activity could be detected. For the adenine derivatives **6**, **7**, and **8**, k_{cat}/K_m values were reduced by ∼5 orders of magnitude (Table 1).

It may be that, while purines themselves are poor substrates, adding a purine and a ribose fragment simultaneously to the enzyme would enhance activity. To investigate this, adenine (**8**) was added to the enzyme which had been preincubated with 1-deoxy-D-ribose (**9**) (5 mM), but no change in rate was observed over that of adenine. Correspondingly, deamination of adenosine (**1**) was not inhibited by addition of either **9** or ribosylamine (**10**) at concentrations up to 6.5 mM, indicating that these two fragments do not bind detectably to the active site of the enzyme. To explore the specificity of ADA, cytosine (**11**) and guanine (**12**) were tested as substrates; neither showed detectable activity.

Fragments of S-Adenosyl-L-homocysteine As Substrates and Inhibitors of MTA-SAH Deaminase (SAHD). SAHD deaminates 5-methyl-thioadenosine (**13**), *S*-adenosyl-homocysteine (**14**) and adenosine (**1**) with k_{cat}/K_M values of 5 × 10^5 M^{-1} s^{-1}, 5.8 × 10^4 M^{-1} s^{-1}, and 9.2 × 10^3 M^{-1} s^{-1}, respectively (Table 2).[6] This enzyme is mechanistically related to ADA, but their sequence identity is low (22%); these two enzymes fall into two different clades of the amidohydrolase superfamily and are from different Clusters of Orthologous Groups.[6] Purine fragments that retain the reactive core of the larger substrates were tested for activity. No deamination could be detected for 9-hydroxymethyladenine (**6**), 9-methyladenine (**7**), and adenine (**8**) after incubating these compounds (0.1 mM) for 12 h with 0.001 mM enzyme (k_{cat}/K_M < 2.3 M^{-1} s^{-1}) (Table 2). To further probe binding, the ribose fragments 1-deoxy-D-ribose (**9**) and ribosyl amine (**10**), and methionine (**15**) were tested as inhibitors of SAHD using **13** as a substrate. At concentrations of 5.0 mM, **9**, **10**, and **15** did not inhibit the enzyme detectably (K_I > 5 mM). We note that the crystal

dx.doi.org/10.1021/ja501354q | J. Am. Chem. Soc. 2014, 136, 7374−7382

Table 1. Activity of Adenosine Deaminase (ADA) against Adenosine, Adenosine Fragments, Cytosine and Guanine

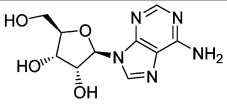
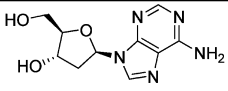
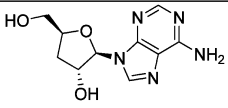
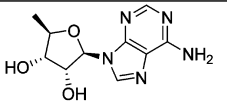
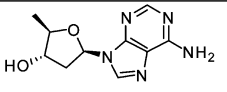
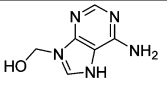
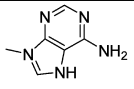
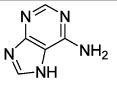
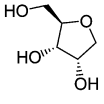
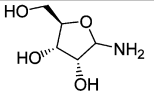
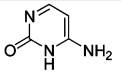
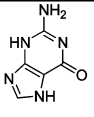
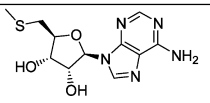
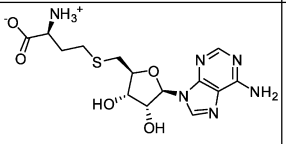
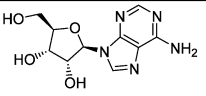
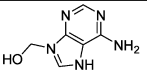
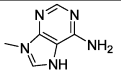
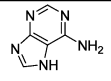
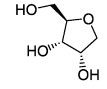
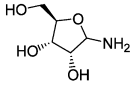
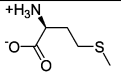
	Structure	Activity / Inhibition
1		$k_{cat}/K_M = 7.6 \times 10^6\ M^{-1}s^{-1}$
2		$k_{cat}/K_M = 4.1 \times 10^6\ M^{-1}s^{-1}$
3		$k_{cat}/K_M = 2 \times 10^5\ M^{-1}s^{-1}$
4		$k_{cat}/K_M < 0.01\ M^{-1}s^{-1}$
5		$k_{cat}/K_M < 0.01\ M^{-1}s^{-1}$
6		$k_{cat}/K_M = 11\ M^{-1}s^{-1}$
7		$k_{cat}/K_M = 52\ M^{-1}s^{-1}$
8		$k_{cat}/K_M = 30\ M^{-1}s^{-1}$
9		$K_I > 20$ mM
10		$K_I > 20$ mM
11		$k_{cat}/K_M < 3\ M^{-1}s^{-1}$
12		$k_{cat}/K_M < 3\ M^{-1}s^{-1}$

Table 2. Activity of MTA-SAH Deaminase (SAHD) against 5-Methyl-thioadenosine, *S*-Adenosyl-L-homocysteine, Adenosine and 5-Methyl-thioadenosine Fragments

	Structure	Activity / Inhibition
13		$k_{cat}/K_M = 5 \times 10^5\ M^{-1}s^{-1}$
14		$k_{cat}/K_M = 5.8 \times 10^4\ M^{-1}s^{-1}$
1		$k_{cat}/K_M = 9.2 \times 10^3\ M^{-1}s^{-1}$
6		$k_{cat}/K_M < 2.3\ M^{-1}s^{-1}$
7		$k_{cat}/K_M < 2.3\ M^{-1}s^{-1}$
8		$k_{cat}/K_M < 2.3\ M^{-1}s^{-1}$
9		$K_I > 5$ mM
10		$K_I > 5$ mM
15		$K_I > 5$ mM

structure of SAHD, determined in the presence of a high concentration of methionine (**15**) as an antioxidant, demonstrates that this fragment can bind to the protein in the absence of other substrates.[6] To investigate a possible synergistic role of **15**, the enzyme was preincubated with this fragment (up to 10 mM) before addition of **8**, but still no deamination was observed. Thus, for MTA-SAH deaminase, purine fragments of the canonical substrate, even though they retain the core adenine reactive center, lose over 5 orders of magnitude of activity relative to the natural substrate of the enzyme.

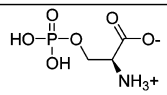
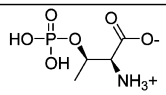
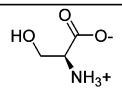
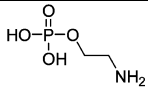
Amino Acids As Inhibitors of Isoaspartyl Dipeptidase (IAD). IAD hydrolyzes several β-aspartyl peptides, of which β-aspartyl-leucine (**16**) is the best recognized, with a k_{cat}/K_M of $10^5\ M^1\ s^{-1}$.[44,45] We were interested in determining if the enzyme would recognize any of the component amino acid monomers, even as inhibitors. Accordingly, IAD was assayed with 0.5 mM of the substrate **16** in the presence and absence of 2 mM concentrations of the following L-amino acids: leucine (**17**), alanine (**18**), glutamate (**19**), phenylalanine (**20**), glycine (**21**), histidine (**22**), isoleucine (**23**), lysine (**24**), methionine (**15**), asparagine (**25**), proline (**26**), glutamine (**27**), arginine (**28**), serine (**29**), threonine (**30**), valine (**31**), tryptophan (**32**), and tyrosine (**33**) (Table S1 Supporting Information [SI]). No measurable inhibition was observed for any of these amino acids (K_I < 4 mM).

Fragments of Phosphoserine As Substrates and Inhibitors of Phosphoserine Phosphatase (PSP). PSP dephosphorylates phosphoserine (**34**) with a k_{cat}/K_M of $3.6 \times 10^5\ M^{-1}\ s^{-1}$ and phosphothreonine (**35**) with a k_{cat}/K_M of $3 \times 10^5\ M^{-1}\ s^{-1}$

dx.doi.org/10.1021/ja501354q | J. Am. Chem. Soc. 2014, 136, 7374−7382

(Table 3). Here too, the substrates are small to begin with, but a reasonable fragmentation was tested using the decarboxy

Table 3. Activity of Phosphoserine Phosphatase (PSP) against Phosphoserine, Phosphothreonine and Phosphoserine Fragments

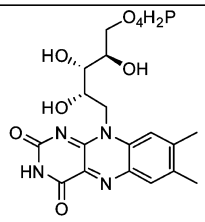
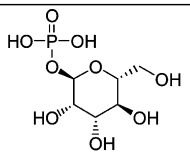
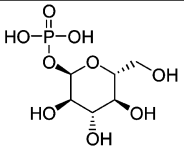
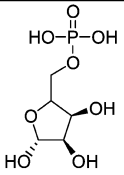
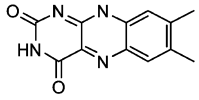
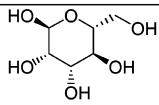
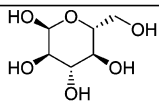
	Structure	Activity / Inhibition
34		$k_{cat}/K_M = 3.6 \times 10^5\ M^{-1}s^{-1}$
35		$k_{cat}/K_M = 3 \times 10^5\ M^{-1}s^{-1}$
29		$K_I = 0.5$ mM
36		$k_{cat}/K_M = 4.2\ M^{-1}s^{-1}$

analogue of phosphoserine, ethanolamine phosphate (**36**). This perturbation was sufficient to reduce activity on this molecule, relative to phosphoserine, by 5 orders of magnitude, to a k_{cat}/K_M of 4.2 M^{-1} s^{-1}. Meanwhile, serine (**29**) itself is a weak competitive inhibitor with a K_I of 0.5 mM.

Fragments of Flavin Mononucleotide As Substrates and Inhibitors of Flavin Mononucleotide Phosphatase (FMP). FMP catalyzes the dephosphorylation of flavin mononucleotide (**37**) with a k_{cat}/K_M of 2.4×10^4 M^{-1} s^{-1} (Table 4). Fragments where the flavin ring was removed, leaving sugar phosphate analogues, were tested as substrates. Mannose-1-phosphate (**38**), glucose-1-phosphate (**39**) and ribose-5-phosphate (**40**), though far less active as substrates than flavin mononucleotide, do retain substantial activity, with k_{cat}/K_M of 2×10^3 M^{-1} s^{-1}, 2.6×10^2 M^{-1} s^{-1}, and 2.1×10^1 M^{-1} s^{-1}, respectively. Meanwhile, the fragments lumichrome (**41**), mannose (**42**), and glucose (**43**) showed no measurable inhibition of FMP-catalyzed dephosphorylation of **38** (K_I values > 0.5, 20, and 20 mM, respectively). Unfortunately, investigating a possible synergistic effect of **41** on the activity of **38**, **39**, and **40** was not possible due to technical issues (fluorescence and background phosphate interference, see methods in SI).

Fragments of Cephalothin As Substrates and Inhibitors of AmpC β-Lactamase. AmpC catalyzes the hydrolysis of cephalothin (**44**) with a k_{cat}/K_M of 7×10^6 M^{-1} s^{-1} (Table 5). We began with fragments that retained the β-lactam ring, on which hydrolysis takes place, but lacked most of the side chains characteristic of β-lactam antibiotics. Starting with the minimal bicyclic cephalosporin ring system (**45**) and building up through **46** to cephalosporanic acid (**47**) itself, no measurable activity was observed. Only for 7-*N*-formyl-cephalosporanic acid (**48**), which almost entirely recapitulates cephalothin, is measurable activity achieved, and then it is almost entirely restored with a k_{cat}/K_M of 2.9×10^5 M^{-1} s^{-1} (Figure S1 in the SI). Thus, the addition of a single formyl moiety, distal to the catalytic center, restores 5 log-orders of activity to the cephalosporin family. Correspondingly, even binding for the

Table 4. Activities of Flavin Monophosphate Phosphatase (FMP) against Riboflavin-5-phosphate, Glucose-1-phosphate, Mannose-1-phosphate and Fragments

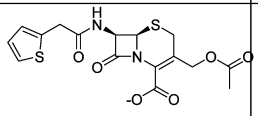
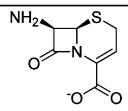
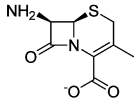
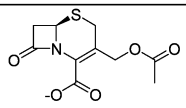
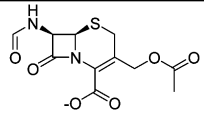
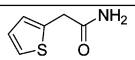
	Structure	Activity
37		$k_{cat}/K_M = 2.4 \times 10^4\ M^{-1}s^{-1}$
38		$k_{cat}/K_M = 2 \times 10^3\ M^{-1}s^{-1}$
39		$k_{cat}/K_M = 2.6 \times 10^2\ M^{-1}s^{-1}$
40		$k_{cat}/K_M = 2.1 \times 10^1\ M^{-1}s^{-1}$
41		$K_I > 0.5$ mM
42		$K_I > 20$ mM
43		$K_I > 20$ mM

smaller fragments **46** and **47** was difficult to detect, with K_I values of about 2 and 5 mM, respectively. Though the amidation of the N7 was sufficient to confer activity, the full thiophene acetamide fragment **49** did not detectably inhibit the enzyme (K_I > 20 mM), nor did the methyl acetate fragment **50**, representing the right-hand side of cephalothin (K_I > 20 mM).

X-ray Crystallography. The structures of three cephalothin fragments, **45**, **46**, and **48** in complex with AmpC were determined by X-ray crystallography, with resolutions ranging from 1.37 to 1.71 Å. The location of all ligands was unambiguous in $F_o - F_c$ difference electron density maps generated early in refinement (Figure S2 and Table S2 in the SI), allowing reliable model building of the enzyme−fragment complexes.

dx.doi.org/10.1021/ja501354q | J. Am. Chem. Soc. 2014, 136, 7374−7382

Table 5. Activity of AmpC β-Lactamase against Cephalotin and Its Fragments

	Structure	Activity
44		$k_{cat}/K_M = 7 \times 10^6\ M^{-1}s^{-1}$
45		$K_I > 2$ mM
46		$K_I \leq 2$ mM
47		$K_I \leq 5$ mM
48		$k_{cat}/K_M = 2.9 \times 10^5\ M^{-1}s^{-1}$
49		$K_I > 20$ mM
50		$K_I > 20$ mM

*N-Formyl 7-Aminocephalosporanic Acid (**48**) in Complex with AmpC.* Electron density for **48** in the AmpC active site is unambiguous in both monomers, with a clear covalent bond between Ser64 and the fragment substrate, as expected (Figure 2A and Figure S2 in the SI). The hydrolysis of the beta-lactam ring results in a carbonyl oxygen that interacts with S64 and A318 backbone nitrogens in the oxyanion hole of the enzyme.[46] This structure thus captures the stable acyl-enzyme intermediate step between the transition-state acylation and deacylation complexes.

*(6R,7R)-7-Amino-8-oxo-5-thia-1-azabicyclo[4.2.0]oct-2-ene-2-carboxylic Acid (**45**) in Complex with AmpC.* Although we could detect neither turnover of **45** nor binding to the enzyme, we were able to observe this compound in the hydrolyzed form in both monomers (Figure 2B and Figure S2 in the SI), via X-ray crystallography after a 10 min soak at 100 mM. The hydrolyzed form of **45** does not bind to the catalytic site, but rather to a distal subsite of the large binding site, interacting with Ser212, Tyr221, and Gly320. Two orientations of the ligand are observed, but in both the carboxylates hydrogen bond with the backbone nitrogens of Ser212 and Gly320, as previously observed for anionic ligands targeting this site.[46−51] It may be that this orientation represents an intermediate step in the exit of the product in AmpC—the contacting residues are well-conserved, even though Ser212 and Gly320 rarely interact with the better cephalosporin substrates of the enzyme.

*7-Amino-desacetoxycephalosporanic Acid (**46**) in Complex with AmpC.* While **46** was also inert to catalysis by AmpC, this compound did weakly inhibit the enzyme. In the crystal structure of the AmpC/**46** complex, **46** is bound at the surface of the protein, away from the active site, in its hydrolyzed form (Figure S2 in the SI). Intriguingly, the substrate cephalotin was previously observed to bind to the same site, also in the hydrolyzed form.[49] Neither the complex with **46** nor **45** supports a catalytically competent recognition of fragment substrates by β-lactamase, though both contain the core cephalosporin ring. The appearance of this hydrolysis product in the structure may reflect slow turnover by the enzyme, over the course of the crystallization experiment, but we cannot exclude the possibility that the hydrolysis product reflects a background reaction in solvent, to which β-lactams are prone.

DISCUSSION

A key observation from this study is that substrate fragments for five of the six enzymes lost essentially all measurable activity or binding, with a reduction of at least 5−6 orders of magnitude relative to that of the canonical substrate. Detectable activity was not regained until most of the substrate had been recapitulated, typically by adding back recognition elements far from the reactive center (Figure 3). Clipping only two or three heavy atoms, well removed from the reactive site, was sufficient to almost completely oblate enzymatic activity. For instance, removing a single hydroxyl from adenosine (**1**) to form 2′- or 3′-deoxyadenosine (**2** and **3**) reduces the activity of adenosine deaminase by 10-fold (Table 1), as does eliminating the thiophene ring from cephalothin for β-lactamase (Table 5). Removing one additional hydroxyl moiety from **2** to form 2′,5′-dideoxy adenosine (**5**) completely abolishes the adenosine deaminase activity, while removing the formyl group from **48** eliminates detectable AmpC activity. Molecules representing the core of the substrate, containing the key reactive group, and representing over 70% of its atoms and functionality, typically had little or no measurable activity against their respective enzymes (Figure 3). Moreover, most showed no detectable enzyme inhibition. Whereas fragment-based approaches are sometimes suitable for substrate discovery, as observed earlier for purine nucleoside phosphorylase[26,27] and Atu3266,[29] they seem unreliable as a general strategy.

The high sensitivity of substrate reactivity to small chemical insults seems to contrast with successful efforts to use fragment substrates as inhibitor leads[42,43] and with efforts to additively improve the affinity of inhibitory fragments. Peptide- and phosphate-bearing fragments have been used as substrate leads for proteases and phosphatases with great success.[52,53] Even in that work, however, substrate recognition was often decreased by 5 orders of magnitude compared to that of canonical substrates. This activity loss can be accepted when one knows the fundamental activity for which one is probing; it is harder to tolerate when seeking to identify the core enzyme activity from a screen of a diverse library of potential substrates.

It is challenging to reconcile the large effects on substrate recognition with those of inhibitors, where it is uncommon to find that removal of a single atom or small moiety reduces affinity by 7−9 kcal/mol, as was observed here. Not only did small changes in the substrate have pronounced effects, the residual pieces often did not measurably bind to the enzymes. At least partly, this may reflect the limits of detection over background, and more importantly the compounding effects of reduced substrate recognition on catalytic competence. Because substrate binding is usually weak, a loss of 3 kcal/mol in substrate affinity will not only reduce turnover by 100-fold (in the k_{cat}/K_M regime) but may also ramify through poor substrate

dx.doi.org/10.1021/ja501354q | J. Am. Chem. Soc. 2014, 136, 7374−7382

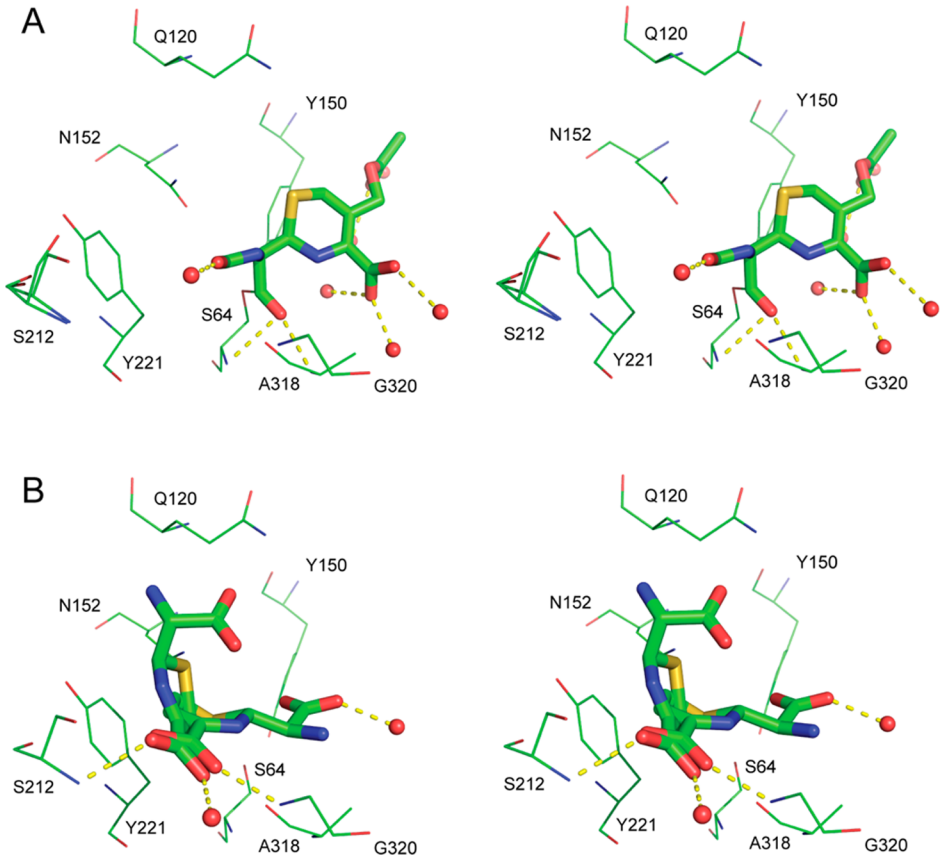


Figure 2. Stereoviews of key interactions within AmpC β-lactamase complexed structures. (A) The structure of AmpC in complex with **48** shows a covalent bond between Ser64 and the fragment substrate, (see also Figure S2 in the SI) and captures the stable acyl-enzyme intermediate step between the transition state acylation and deacylation complexes. (B) The structure of AmpC in complex with **45** shows the fragment in its product form, bound in two orientations to a distal subsite of the large overall binding site, interacting with Ser212, Tyr221, and Gly320.

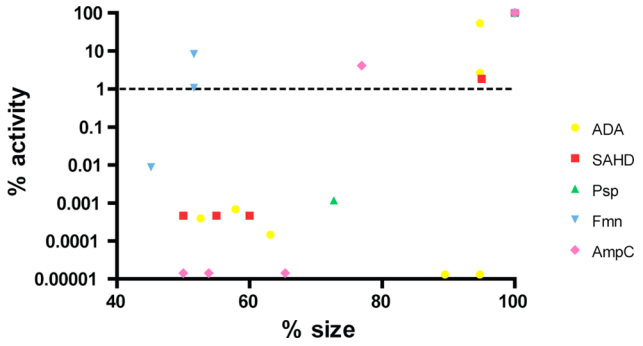


Figure 3. Plot of % compound activity (as compared to the entire substrate) as a function of size. Size is measured as the number of heavy atoms.

alignment to a lower k_{cat}. Together, this may reduce activity below the detection limit.

However, it also seems that, for many enzymes, there is a substantial nonadditivity to substrate recognition in catalysis. Thus, functional groups that contribute substantially to recognition when presented as part of the full substrate, are not recognized in substrate fragments. For inhibitors, conversely, recognition is often more additive. The idea of integrated, nonadditive recognition of substrates has precedence in experimental[20] and theoretical enzymology—returning to Jencks' idea that groups distal to the reactive center may provide crucial binding energy for catalytic recognition[21,22]—and is further highlighted in this study.

These results may have implications for promiscuity on the substrate side of enzyme–substrate interactions, providing a counterpoint to the promiscuity of enzyme recognition and evolution.[54,55] Enzyme promiscuity[56] and moonlighting activities[57] seem central to the evolution of new activities, and noncanonical substrates and ligands may be specifically recognized.[58] What this study suggests is that enzyme recognition of optimal substrates for any given reaction is mostly punctate, tolerating little diminution in substrate functional groups. Physically, this may reflect the integrated chemistry of small molecules, where each atom contributes to the electronic properties of the entire substrate. The sharp dependence on substrate structure may also reflect specificity constraints faced by enzymes confronted with multiple related metabolites, upon whose product multiple downstream enzymes depend. In such a system, it is possible to duplicate and subsequently mutate an enzyme without detrimental effect on other dependencies, exploiting intrinsic enzyme promiscuity. Conversely, changes in the substrate or the product could have cascading effects on the larger cellular system. Thus, while enzymes remain plastic to evolution overall, and may be promiscuous with unrelated substrates,[56,59] they may be selected for high fidelity around the structure of their cognate substrate, in which little change is tolerated; this idea has also been suggested for signaling molecules and their receptors.[60,61]

dx.doi.org/10.1021/ja501354q | J. Am. Chem. Soc. 2014, 136, 7374–7382

Several caveats merit mentioning. Most notably, some substrates do, in fact, tolerate fragmenting. That was observed here for flavin nucleotide phosphatase, and has been observed previously for purine nucleoside phosphorylase.[26,27] Also, one can always imagine larger substrates than the core molecules we have investigated. For instance, there are β-lactamase substrates that are substantially larger than cephalothin, and for these the removal of functional groups to yield a substrate just below the size of cephalothin only marginally affects their reactivity to β-lactamase.[62] However, the larger side chains of these molecules do not make strong interactions with β-lactamase,[63] and these may be seen as extraneous groups that are not germane to substrate recognition by the enzyme. The tolerance to diversity of substrate sizes may be more common for a detoxification enzyme like β-lactamase, which operates by itself and not in a pathway. Whereas we suspect that in most cases substrates do not tolerate fragmenting, and fragments will not be a pragmatic general strategy for function discovery, there will be cases where such a strategy may well be fruitful.[29] For instance, phosphoserine phosphatase only recognizes phosphoserine and phosphothreonine, whereas other phosphatases such as alkaline phosphatase are able to hydrolyze a large range of monophosphate substrates.[64] Such broad-specificity enzymes are expected to be less sensitive to small chemical changes in their substrates and to lend themselves more easily to substrate fragmentation. Finally, using substrate fragments as probes for new inhibitors[42,43] and fragments as leads for inhibitor discovery[48,65] is unaffected by this study; these approaches to inhibitor discovery continue to hold great promise.

These caveats should not obscure the main observations of this study. Most substrate functional groups may be essential for recognition and turnover by a cognate enzyme, and removal of even a small portion of the substrate, distal to the reactive center, can dramatically reduce the activity. Once deconstructed into smaller parts, the resulting fragments show little to no activity, and often do not measurably bind to the enzyme active site. Enzymes typically did not evolve for inhibitor binding as they did for substrates, and fidelity requirements appear much higher for the latter. This may reflect a need to engage a full substrate to drive reactivity, as distal functional groups can contribute critically to stabilization of a transition state.[21] The idea that the stringent specificity constraints on substrates, as opposed to inhibitors, reflects downstream affects of products on the cell is more speculative. Even here, one can imagine testable implications, as designed enzymes, not evolved against pathway constraints, may exhibit broader promiscuity against related substrates and fragments than do natural enzymes. Pragmatically, this study supports programs to synthesize and collect a relatively full repertoire of the core metabolites recognized by biology with which to interrogate protein function. This space, though far larger than one would prefer, is not, in the end, unbounded.

METHODS

Protein Expression and Purification. Isoaspartyl dipeptidase (IAD) and AmpC β-lactamase (AmpC) were expressed and purified as previously described.[45,66] For adenosine deaminase (ADA), MTA-SAH deaminase (SAHD), phosphoserine phosphatase (PSP) and flavin mononucleotide phosphatase (FMP) expression and purification, see methods in the SI.

Enzyme Assays. The hydrolysis rates and inhibition constants for adenosine deaminase and MTA-SAH deaminase activities were monitored spectrophotometrically in a direct assay. Inhibition of isoaspartyl dipeptidase was measured in a spectrophotometric assay coupling the formation of aspartate to the oxidation of NADH. The substrate hydrolysis rates and inhibition constants for phosphoserine phosphatase were measured spectrophotometrically using the Enzchek kit (Invitrogen). The hydrolysis rates and inhibition constants for flavin mononucleotide phosphatase were monitored using either the Enzchek kit or BIOMOL Green (Enzo Life Sciences). The hydrolysis of AmpC β-lactamase substrates and enzyme inhibition were monitored spectrophotometrically (methods in the SI).

Crystal Growth and Structure Determination. AmpC β-lactamase structures in complex with **45**, **46**, and **48** were obtained by soaking the crystals in the respective ligand solution. Structures were determined between 1.37 and 1.71 Å resolution, and the phases for all structures were determined by difference Fourier Methods (methods in the SI).

ASSOCIATED CONTENT

Supporting Information

Supplementary methods: protein expression and purification; enzymatic assays; crystal growth and structure determination. Supplentary tables: activity of iso-aspartyl dipeptidase against β-aspartyl-leucine in absence and in the presence of 18 L-amino acids; data collection and refinement statistics. Supplementary figures: activity of AmpC β-lactamase against *N*-formyl 7-amino cephalosporanic acid; stereoviews of the active site electron density for each AMPC complexed structure. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

drkallen@bu.edu (K.A.)
raushel@tamu.edu (F.R.)
bshoichet@gmail.com(B.S.)

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Supported by U.S. NIH GM093342 (J.G.). We thank Nir London for reading this manuscript.

REFERENCES

(1) Magrane, M.; Consortium, U. *Database* **2011**, *2011*, bar009.
(2) Hobbs, M. E.; Vetting, M.; Williams, H. J.; Narindoshvili, T.; Kebodeaux, D. M.; Hillerich, B.; Seidel, R. D.; Almo, S. C.; Raushel, F. M. *Biochemistry* **2013**, *52*, 239.
(3) Lu, Z.; Dunaway-Mariano, D.; Allen, K. N. *Proteins* **2011**, *79*, 3099.
(4) Strushkevich, N.; Gilep, A. A.; Shen, L.; Arrowsmith, C. H.; Edwards, A. M.; Usanov, S. A.; Park, H. W. *Mol. Endocrinol.* **2013**, *27*, 315.
(5) Weiss, D. R.; Ahn, S.; Sassano, M. F.; Kleist, A.; Zhu, X.; Strachan, R.; Roth, B. L.; Lefkowitz, R. J.; Shoichet, B. K. *ACS Chem. Biol.* **2013**, *8*, 1018.
(6) Hermann, J. C.; Marti-Arbona, R.; Fedorov, A. A.; Fedorov, E.; Almo, S. C.; Shoichet, B. K.; Raushel, F. M. *Nature* **2007**, *448*, 775.
(7) Fan, H.; Hitchcock, D. S.; Seidel, R. D., 2nd; Hillerich, B.; Lin, H.; Almo, S. C.; Sali, A.; Shoichet, B. K.; Raushel, F. M. *J. Am. Chem. Soc.* **2013**, *135*, 795.
(8) Zhao, S.; Kumar, R.; Sakai, A.; Vetting, M. W.; Wood, B. M.; Brown, S.; Bonanno, J. B.; Hillerich, B. S.; Seidel, R. D.; Babbitt, P. C.; Almo, S. C.; Sweedler, J. V.; Gerlt, J. A.; Cronan, J. E.; Jacobson, M. P. *Nature* **2013**, *502*, 698.
(9) Erlanson, D. A. *Top. Curr. Chem.* **2012**, *317*, 1.
(10) Fink, T.; Reymond, J.-L. *J. Chem. Inf. Model.* **2007**, *47*, 342.
(11) Leach, A. R.; Hann, M. M. *Curr. Opin. Chem. Biol.* **2011**, *15*, 489.
(12) Silvestre, H. L.; Blundell, T. L.; Abell, C.; Ciulli, A. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 12984.

(13) Scott, D. E.; Ehebauer, M. T.; Pukala, T.; Marsh, M.; Blundell, T. L.; Venkitaraman, A. R.; Abell, C.; Hyvonen, M. *ChemBioChem* **2013**, *14*, 332.
(14) Roughley, S.; Wright, L.; Brough, P.; Massey, A.; Hubbard, R. E. *Topics Curr. Chem.* **2012**, *317*, 61.
(15) Bollag, G.; Hirth, P.; Tsai, J.; Zhang, J.; Ibrahim, P. N.; Cho, H.; Spevak, W.; Zhang, C.; Zhang, Y.; Habets, G.; Burton, E. A.; Wong, B.; Tsang, G.; West, B. L.; Powell, B.; Shellooe, R.; Marimuthu, A.; Nguyen, H.; Zhang, K. Y. J.; Artis, D. R.; Schlessinger, J.; Su, F.; Higgins, B.; Iyer, R.; D'Andrea, K.; Koehler, A.; Stumm, M.; Lin, P. S.; Lee, R. J.; Grippo, J.; Puzanov, I.; Kim, K. B.; Ribas, A.; McArthur, G. A.; Sosman, J. A.; Chapman, P. B.; Flaherty, K. T.; Xu, X.; Nathanson, K. L.; Nolop, K. *Nature* **2010**, *467*, 596.
(16) Friberg, A.; Vigil, D.; Zhao, B.; Daniels, R. N.; Burke, J. P.; Garcia-Barrantes, P. M.; Camper, D.; Chauder, B. A.; Lee, T.; Olejniczak, E. T.; Fesik, S. W. *J. Med. Chem.* **2013**, *56*, 15.
(17) Talamas, F. X.; Ao-Ieong, G.; Brameld, K. A.; Chin, E.; de Vicente, J.; Dunn, J. P.; Ghate, M.; Giannetti, A. M.; Harris, S. F.; Labadie, S. S.; Leveque, V.; Li, J.; Lui, A. S.; McCaleb, K. L.; Najera, I.; Schoenfeld, R. C.; Wang, B.; Wong, A. *J. Med. Chem.* **2013**, *56*, 3115.
(18) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. *J. Chem. Inf. Model.* **2012**, *52*, 1757.
(19) Carlow, D.; Wolfenden, R. *Biochemistry* **1998**, *37*, 11873.
(20) Kati, W. M.; Acheson, S. A.; Wolfenden, R. *Biochemistry* **1992**, *31*, 7356.
(21) Jencks, W. P. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4046.
(22) Jencks, W. P. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1975**, *43*, 219.
(23) Page, M. I.; Jencks, W. P. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 1678.
(24) Hartley, B. S.; Kilby, B. A. *Biochem. J.* **1952**, *50*, 672.
(25) Christensen, U.; Ipsen, H. H. *Biochim. Biophys. Acta* **1979**, *569*, 177.
(26) Basso, L. A.; Santos, D. S.; Shi, W.; Furneaux, R. H.; Tyler, P. C.; Schramm, V. L.; Blanchard, J. S. *Biochemistry* **2001**, *40*, 8196.
(27) Shi, W. X.; Basso, L. A.; Santos, D. S.; Tyler, P. C.; Furneaux, R. H.; Blanchard, J. S.; Almo, S. C.; Schramm, V. L. *Biochemistry* **2001**, *40*, 8204.
(28) Stout, T. J.; Sage, C. R.; Stroud, R. M. *Structure* **1998**, *6*, 839.
(29) Ornelas, A.; Korczynska, M.; Ragumani, S.; Kumaran, D.; Narindoshvili, T.; Shoichet, B. K.; Swaminathan, S.; Raushel, F. M. *Biochemistry* **2013**, *52*, 228.
(30) Zhai, X.; Amyes, T. L.; Richard, J. P. *J. Am. Chem. Soc.* **2014**, *136*, 4145.
(31) Medzihradzky, K.; Voynick, I. M.; Medzihradszky-Schweiger, H.; Fruton, J. S. *Biochemistry* **1970**, *9*, 1154.
(32) Thompson, R. C.; Blout, E. R. *Proc. Natl. Acad. Sci. U.S.A.* **1970**, *67*, 1734.
(33) Schrader, F. C.; Glinca, S.; Sattler, J. M.; Dahse, H. M.; Afanador, G. A.; Prigge, S. T.; Lanzer, M.; Mueller, A. K.; Klebe, G.; Schlitzer, M. *ChemMedChem.* **2013**, *8*, 442.
(34) Egger, J.; Weckerle, C.; Cutting, B.; Schwardt, O.; Rabbani, S.; Lemme, K.; Ernst, B. *J. Am. Chem. Soc.* **2013**, *135*, 9820.
(35) Christopher, J. A.; Brown, J.; Dore, A. S.; Errey, J. C.; Koglin, M.; Marshall, F. H.; Myszka, D. G.; Rich, R. L.; Tate, C. G.; Tehan, B.; Warne, T.; Congreve, M. *J. Med. Chem.* **2013**, *56*, 3446.
(36) Brough, P. A.; Barril, X.; Borgognoni, J.; Chene, P.; Davies, N. G.; Davis, B.; Drysdale, M. J.; Dymock, B.; Eccles, S. A.; Garcia-Echeverria, C.; Fromont, C.; Hayes, A.; Hubbard, R. E.; Jordan, A. M.; Jensen, M. R.; Massey, A.; Merrett, A.; Padfield, A.; Parsons, R.; Radimerski, T.; Raynaud, F. I.; Robertson, A.; Roughley, S. D.; Schoepfer, J.; Simmonite, H.; Sharp, S. Y.; Surgenor, A.; Valenti, M.; Walls, S.; Webb, P.; Wood, M.; Workman, P.; Wright, L. *J. Med. Chem.* **2009**, *52*, 4794.
(37) Peat, T. S.; Rhodes, D. I.; Vandegraaff, N.; Le, G.; Smith, J. A.; Clark, L. J.; Jones, E. D.; Coates, J. A.; Thienthong, N.; Newman, J.; Dolezal, O.; Mulder, R.; Ryan, J. H.; Savage, G. P.; Francis, C. L.; Deadman, J. J. *PloS One* **2012**, *7*, e40147.
(38) Rega, M. F.; Wu, B.; Wei, J.; Zhang, Z.; Cellitti, J. F.; Pellecchia, M. *J. Med. Chem.* **2011**, *54*, 6000.
(39) Huth, J. R.; Park, C.; Petros, A. M.; Kunzer, A. R.; Wendt, M. D.; Wang, X.; Lynch, C. L.; Mack, J. C.; Swift, K. M.; Judge, R. A.; Chen, J.; Richardson, P. L.; Jin, S.; Tahir, S. K.; Matayoshi, E. D.; Dorwin, S. A.; Ladror, U. S.; Severin, J. M.; Walter, K. A.; Bartley, D. M.; Fesik, S. W.; Elmore, S. W.; Hajduk, P. J. *Chem. Biol. Drug Design* **2007**, *70*, 1.
(40) Jahnke, W.; Rondeau, J. M.; Cotesta, S.; Marzinzik, A.; Pelle, X.; Geiser, M.; Strauss, A.; Gotte, M.; Bitsch, F.; Hemmig, R.; Henry, C.; Lehmann, S.; Glickman, J. F.; Roddy, T. P.; Stout, S. J.; Green, J. R. *Nat. Chem. Biol.* **2010**, *6*, 660.
(41) Saalau-Bethell, S. M.; Woodhead, A. J.; Chessari, G.; Carr, M. G.; Coyle, J.; Graham, B.; Hiscock, S. D.; Murray, C. W.; Pathuri, P.; Rich, S. J.; Richardson, C. J.; Williams, P. A.; Jhoti, H. *Nat. Chem. Biol.* **2012**, *8*, 920.
(42) Soellner, M. B.; Rawls, K. A.; Grundner, C.; Alber, T.; Ellman, J. A. *J. Am. Chem. Soc.* **2007**, *129*, 9613.
(43) Baguley, T. D.; Xu, H. C.; Chatterjee, M.; Nairn, A. C.; Lombroso, P. J.; Ellman, J. A. *J. Med. Chem.* **2013**, *56*, 7636.
(44) Haley, E. E. *J. Biol. Chem.* **1968**, *243*, 5748.
(45) Marti-Arbona, R.; Fresquet, V.; Thoden, J. B.; Davis, M. L.; Holden, H. M.; Raushel, F. M. *Biochemistry* **2005**, *44*, 7115.
(46) Babaoglu, K.; Shoichet, B. K. *Nat. Chem. Biol.* **2006**, *2*, 720.
(47) Teotico, D. G.; Babaoglu, K.; Rocklin, G. J.; Ferreira, R. S.; Giannetti, A. M.; Shoichet, B. K. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 7455.
(48) Eidam, O.; Romagnoli, C.; Dalmasso, G.; Barelier, S.; Caselli, E.; Bonnet, R.; Shoichet, B. K.; Prati, F. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 17448.
(49) Beadle, B. M.; Trehan, I.; Focia, P. J.; Shoichet, B. K. *Structure* **2002**, *10*, 1025.
(50) Nukaga, M.; Kumar, S.; Nukaga, K.; Pratt, R. F.; Knox, J. R. *J. Biol. Chem.* **2004**, *279*, 9344.
(51) Powers, R. A.; Caselli, E.; Focia, P. J.; Prati, F.; Shoichet, B. K. *Biochemistry* **2001**, *40*, 9207.
(52) Rawls, K. A.; Lang, P. T.; Takeuchi, J.; Imamura, S.; Baguley, T. D.; Grundner, C.; Alber, T.; Ellman, J. A. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6851.
(53) Wood, W. J.; Patterson, A. W.; Tsuruoka, H.; Jain, R. K.; Ellman, J. A. *J. Am. Chem. Soc.* **2005**, *127*, 15521.
(54) Ben-David, M.; Wieczorek, G.; Elias, M.; Silman, I.; Sussman, J. L.; Tawfik, D. S. *J. Mol. Biol.* **2013**, *425*, 1028.
(55) Bar-Rogovsky, H.; Hugenmatter, A.; Tawfik, D. S. *J. Biol. Chem.* **2013**, *288*, 23914.
(56) Khersonsky, O.; Tawfik, D. S. *Annu. Rev. Biochem.* **2010**, *79*, 471.
(57) Huberts, D. H.; van der Klei, I. J. *Biochim. biophys. acta* **2010**, *1803*, 520.
(58) James, L. C.; Tawfik, D. S. *Protein Sci.* **2003**, *12*, 2183.
(59) Villiers, B. R.; Hollfelder, F. *ChemBioChem* **2009**, *10*, 671.
(60) Tomkins, G. M. *Science* **1975**, *189*, 760.
(61) Lin, H.; Sassano, M. F.; Roth, B. L.; Shoichet, B. K. *Nat. Methods* **2013**, *10*, 140.
(62) van Berkel, S. S.; Brem, J.; Rydzik, A. M.; Salimraj, R.; Cain, R.; Verma, A.; Owens, R. J.; Fishwick, C. W.; Spencer, J.; Schofield, C. J. *J. Med. Chem.* **2013**, *56*, 6945.
(63) Tremblay, L. W.; Xu, H.; Blanchard, J. S. *Biochemistry* **2010**, *49*, 9685.
(64) Coleman, J. E. *Annu. Rev. Biophys. Biomol. Structure* **1992**, *21*, 441.
(65) Geitmann, M.; Elinder, M.; Seeger, C.; Brandt, P.; de Esch, I. J.; Danielson, U. H. *J. Med. Chem.* **2011**, *54*, 699.
(66) Usher, K. C.; Blaszczak, L. C.; Weston, G. S.; Shoichet, B. K.; Remington, S. J. *Biochemistry* **1998**, *37*, 16082.