BIOCHEMISTRY
including biophysical chemistry & molecular biology

ARTICLE

pubs.acs.org/biochemistry

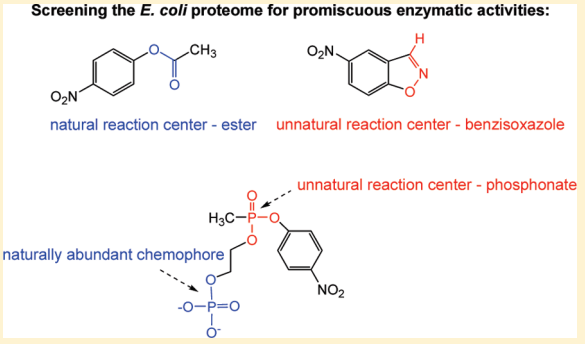
Role of Chemistry versus Substrate Binding in Recruiting Promiscuous Enzyme Functions

Olga Khersonsky,[†,§] Sergey Malitsky,[‡] Ilana Rogachev,[‡] and Dan S. Tawfik*[,†]

[†]Department of Biological Chemistry and [‡]Department of Plant Sciences, Weizmann Institute of Science, Rehovot 76100, Israel

Supporting Information

ABSTRACT: Two different scenarios for the recruitment of evolutionary starting points and their subsequent divergence to give new enzymes have been described. The coincidental, promiscuous starting activity may regard the same reaction chemistry on a new substrate (substrate ambiguity). Alternatively, substrate binding guides the recruitment of an enzyme whose reaction chemistry differs from that of the newly evolving one (catalytic promiscuity). While substrate ambiguity seems to underlie the divergence of most enzyme families, the relative levels of occurrence of these scenarios remain unknown. Screening the *Escherichia coli* proteome with a comparative series of xenobiotic substrates, we found that substrate ambiguity was, as anticipated, more frequent than reaction promiscuity. However, for at least one unnatural reaction (phosphonoesterase), a promiscuous enzyme was identified only when the substrate was decorated with the naturally abundant phosphate group. These findings support the prevailing hypothesis of chemistry-driven divergence but also suggest that recognition of familiar substrate motifs plays a role. In the absence of enzymes catalyzing the same chemistry, having a familiar, naturally occurring substrate motif (chemophore) such as phosphate may increase the likelihood of catalytic promiscuity. Chemophore anchoring may also find practical applications in identifying catalysts for unnatural reactions.



The hypothesis that latent, promiscuous functions served as starting points in the divergence of new enzyme families is now widely accepted.[1−5] However, at least two different scenarios for the recruitment of an evolutionary starting point and its subsequent divergence to give a new enzyme have been described.[3] Both scenarios begin with an existing genome−proteome encountering a new substrate, whose transformation into a given product is advantageous [either because substrate transformation renders it nontoxic (detoxification) or because its product provides a growth advantage]. In the chemistry-driven scenario, the organism happens to carry an enzyme that applies the same active-site chemistry to the new substrate to produce the desirable product (substrate ambiguity). The subsequent evolutionary process optimizes substrate binding (affinity and optimal alignment) to achieve a higher catalytic efficiency. In the substrate-driven scenario, the recruited enzyme's original substrate happens to share a key structural element with the new substrate. Substrate binding guides the recruitment of this enzyme to yield the desirable product, but the newly evolving chemistry differs from the original one (catalytic promiscuity). The catalytic chemistry and substrate binding are subsequently reshaped by mutation and selection.

Comparative analysis revealed that, in most cases, divergence is chemistry-driven, as key catalytic residues, and reaction intermediates or transition states, tend to be conserved throughout enzyme families and even within highly diverse superfamilies.[3,6] However, cases of substrate-driven divergence were also described.[3,7] Analysis by chemoinformatic tools demonstrated that most superfamilies exhibit conservation of certain substrate elements, or substructures dubbed here chemophores, such as phosphate, adenine, and sugar moieties.[8,9] Conserved substrate substructures have also been applied for the annotation of enzyme superfamilies.[10] Indeed, substrate binding and chemistry are inseparable with respect to the reactive group; esterase substrates all share an ester group. However, certain chemophores are overrepresented in substrates of enzyme superfamilies even though the catalyzed reactions do not necessarily involve these chemophores.[8,9,11]

Understanding the relative contributions of substrate binding versus chemistry may therefore promote our understating of enzyme evolution, and of ligand−protein interactions in general.[11] These issues also have biotechnological implications, because substrate ambiguity and catalytic promiscuity provide the basis of biocatalytic applications.[8,12] We therefore aimed to compare the relative contributions of shared chemistry (i.e., substrate ambiguity) versus shared substrate binding (i.e., catalytic promiscuity) in the recruitment of promiscuous enzyme functions, and the role of abundant chemophores in promoting the latter. To this end, we explored the relative levels of occurrence of these phenomena within the *Escherichia coli* genome−proteome. We applied the ASKA library (A Complete

Received: November 9, 2010
Revised: February 17, 2011
Published: February 18, 2011

ACS Publications © 2011 American Chemical Society

Set of *E. coli* K-12 ORF Archive) that encodes all *E. coli* proteins individually cloned in an overexpression vector,[13] as a model for the recruitment of latent, promiscuous enzyme functions when an organism's proteome encounters a new challenge.[14,15]

First, we screened the ASKA library with two synthetic, xenobiotic substrates. One substrate represented a naturally occurring reaction and the other an unnatural one. In the case of an unnatural reaction, recruitment of any enzyme would represent a case of catalytic promiscuity. In the case of a natural transformation, however, both catalytic promiscuity and substrate ambiguity are possible. The results of this screen can therefore indicate the relative abundance of catalytic promiscuity and substrate ambiguity in the *E. coli* proteome. Second, we tested how the probability of identifying enzymes that promiscuously catalyze an unnatural reaction is affected by decorating the unnatural substrate with abundant naturally occurring chemophores.

EXPERIMENTAL PROCEDURES

Substrates. *p*-Nitrophenyl acetate (**1**) and benzisoxazole (**3**) were purchased from Sigma-Aldrich. Synthesis of other substrates is described in the Supporting Information.

ASKA Library. The ASKA archive clones are a set of plasmid clones containing all the predicted open reading frames (ORFs) of the *E. coli* K-12 genome. The ORFs of *E. coli* were cloned into the pCA24N vector with an N-terminal His tag, and the library was kindly provided by the *E. coli* project team from Nara Institute of Science and Technology (Nara, Japan).[13] Glycerol stocks of *E. coli* K-12 AG1 cells individually overexpressing each of the ~4300 different clones were provided in the format of 46 96-well plates and kept at −70 °C.

Screens. For the ASKA library screen, sets of six ASKA plates (as glycerol stocks) were pooled together into one plate, while the original well locations were retained. Plates 1−6, 7−12, 13−18, 19−24, 25−30, 31−36, 37−42, and 43−46 were unified, creating a pooled ASKA library in eight 96-well plates. The pooled glycerol stocks were inoculated into 2YT supplemented with 35 μg/mL chloramphenicol (100 μL) in 96-well plates and grown overnight at 37 °C. Overnight cultures (20 μL) were inoculated into 2YT supplemented with 35 μg/mL chloramphenicol (500 μL) in 96-deep well plates and grown to an OD_{600} of ~0.6. Overexpression was induced via addition of 1 mM IPTG; the cultures were grown for an additional 5 h and centrifuged, and the pellet was frozen overnight at −20 °C. The cells were lysed with lysis buffer [HEPES 50 mM (pH 7.25), 0.2% triton-100, 0.1 mg/mL lysozyme, ~1 unit benzonase], 250 μL/well, and the lysates were cleared by centrifugation. For the enzymatic assays, 100 μL of lysate was assayed with the various substrates dissolved in 50 mM HEPES (pH 7.25) by following the release of the phenol product in a PowerWave HT microtiter scanning spectrophotometer. Substrate concentrations and the monitoring wavelengths of the respective products are listed in Table S1 of the Supporting Information.

Deconvolution, Validation, and Retransformation. To deconvolute which of the six pooled clones within the wells found active in the initial library screen gave rise to the observed activity, the individual glycerol stocks of each of the six pooled clones were used for inoculation in 96-well plates. Growth, expression, and activity assays were performed as in the initial screen. In the case of 6-phosphobenzisoxazole, the cultures of the active wells were spread on LB agar plates containing 35 μg/mL chloramphenicol, and 16 colonies were randomly picked (3-fold oversampling), grown in liquid medium in 96-well plates, and screened to identify the individually active clones.

Following the deconvolution, overnight cultures of the active clones were plated on LB agar plates containing 35 μg/mL chloramphenicol. The enzymatic assays were repeated after the selected clones had been grown in duplicate (from two randomly picked colonies) as described above. Plasmid DNA was extracted from active clones, sequenced, and retransformed into *E. coli* BL21 DE3 cells, and the reaction rates were assayed as described above.

Characterization of Purified Enzyme Variants. 2YT medium (5 mL) supplemented with 35 μg/mL chloramphenicol was inoculated with a single colony. The culture was grown with shaking at 37 °C for ~15 h. 2YT medium (250 mL) supplemented with 35 μg/mL chloramphenicol was inoculated with the 5 mL overnight culture and grown at 37 °C with shaking to an OD_{600} of ~0.6. Overexpression was induced via addition of 1 mM IPTG; the cultures were grown for an additional 5 h and harvested, and the pellet was kept overnight at −20 °C. The cells were resuspended in 25 mL of lysis buffer (Table S2 of the Supporting Information) and lysed by sonication. The soluble fraction was loaded onto a Ni-NTA column (Qiagen) and washed with 10 and 20 mM imidazole, and the protein was eluted with 250 mM imidazole. This protocol yielded ≥90% pure protein as judged by sodium dodecyl sulfate−polyacrylamide gel electrophoresis. After dialysis against lysis buffer, protein concentrations were determined by absorption at 280 nm, and the samples were stored at 4 °C. For the enzymatic characterization, reactions were started via addition of 100 μL of substrate solutions to 100 μL of protein solutions (for specific conditions, see Table S2 of the Supporting Information). Product formation was monitored spectrophotometrically (Power HT microtiter scanning spectrophotometer), in 200 μL reaction volumes, using 96-well plates. The reported results are the average of at least two independent measurements.

RESULTS

Screening the ASKA Library. Given the ASKA library size (~4300 clones) and the need to screen it multiple times, we pooled the original library (46 96-well plates) into eight 96-well plates, such that each well combined six clones. Because of the large variability in growth rates (~25% of the ASKA clones show growth inhibition at various degrees[13]), not all potentially active variants were identified in our screens, although the identified ones were not systematically biased. Screening the ASKA library with promiscuous substrates involved detection of weak activities, and some of the screened activities also exhibited considerable background rates and nonspecific catalysis by bacterial lysate components. Given the noise and variability, clones identified in the initial screens were reassayed and their activity was validated. The complete screening procedure consisted of the following steps (see also the Supporting Information). (i) Initial screens were performed by growth of the pooled library in 96-well plates, addition of substrates to crude cell lysates, and monitoring product formation in a plate reader. (ii) For deconvolution, the six clones initially pooled in one well were taken from the original ASKA plates, grown, and screened individually as described above. (iii) For validation, active clones were regrown and reassayed in duplicate, and their identity was determined by DNA sequencing. (iv) For retransformation, plasmid DNA from active ASKA clones was isolated and retransformed into *E. coli*

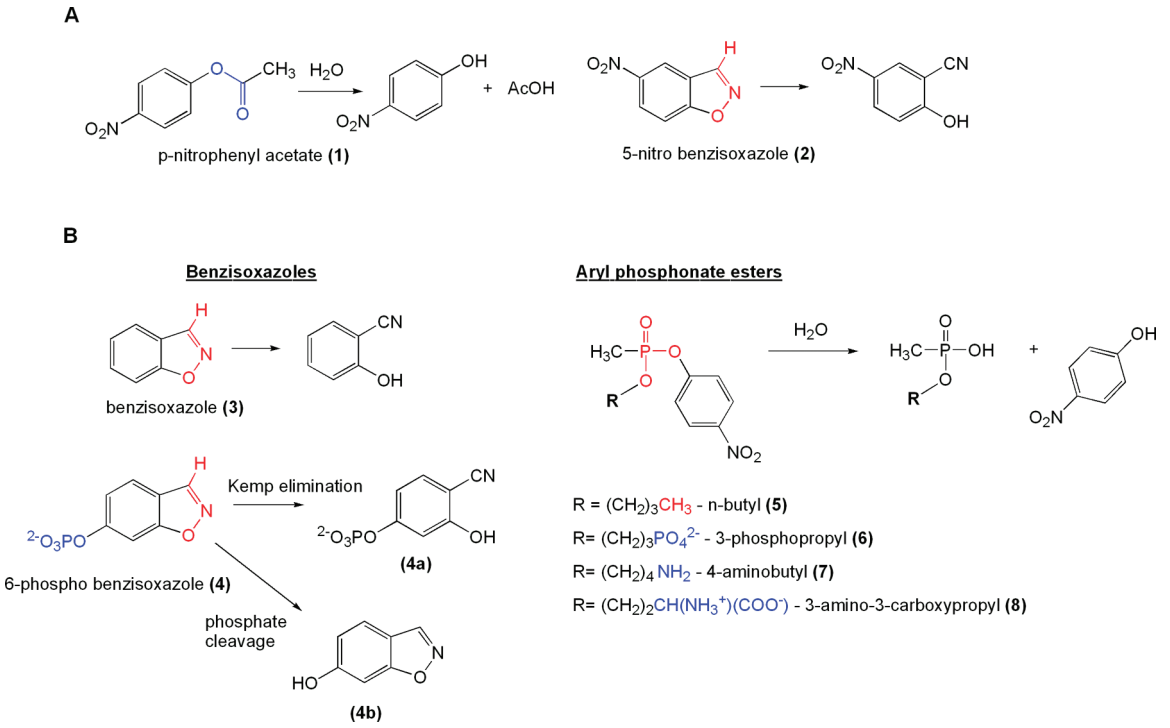


Figure 1. Substrates and reactions screened. Naturally occurring reaction centers and chemophores are colored blue and unnatural ones red. (A) Comparative substrates with natural (ester) and unnatural (benzisoxazole) reaction centers. (B) Substrates having unnatural reaction centers with and without naturally abundant chemophores.

cells. Several randomly chosen transformed colonies were grown, lysed, and assayed for activity. (v) For purified proteins, active variants were overexpressed, purified, and tested for enzymatic activity. Inhibition of the promiscuous activities by the natural substrates and/or cofactors of the enzyme was tested to confirm that the promiscuous activity occurs within the same active site.

Comparing Natural versus Unnatural Chemistries. To this end, we aimed to screen two synthetic, xenobiotic substrates with similar physicochemical properties, whereby transformation of one substrate represents a naturally occurring chemistry and that of the other substrate an unnatural one. Of various options explored, we chose *p*-nitrophenyl acetate (**1**) and 5-nitrobenzisoxazole (**2**) (Figure 1), which exhibit similar physical properties (e.g., size and hydrophobicity) and similar reactivity [the spontaneous decomposition rates are 3.8×10^{-5} min^{-1} for *p*-nitrophenyl acetate (pH 7.3)[16] and 6.96×10^{-5} min^{-1} for 5-nitrobenzisoxazole (pH 7.25)[17]]. Because the more reactive a substrate reaction is the higher the likelihood of promiscuity, inherent reactivity is a crucial factor that could strongly bias the results and thus limited our choice of substrates and reactions. *p*-Nitrophenyl acetate is an ester, and ester hydrolysis is a chemistry commonly catalyzed by natural esterases. The recruitment of an esterase catalyzing the hydrolysis of a xenobiotic ester such as *p*-nitrophenyl acetate is likely to proceed by the chemistry-driven scenario. The screen would therefore identify primarily cases of substrate ambiguity, although cases of catalytic promiscuity (i.e., nonhydrolytic enzymes that catalyze this ester's hydrolysis) may also occur. In contrast, the Kemp elimination of 5-nitrobenzisoxazole[18] is unknown in biological systems, and hence, no enzyme is likely to have evolved to catalyze this reaction. Therefore, any case of enzyme recruitment for the Kemp elimination would be a case of catalytic promiscuity.

In the initial screen (step i), ~5 times more wells were identified as being active with *p*-nitrophenyl acetate as with 5-nitrobenzisoxazole [>100 vs ~20 (Table 1)]. Following steps ii−iv, 11 clones with validated esterase activity were identified (Table 2). Seven known esterases, or other hydrolases, were identified, alongside two aldehyde dehydrogenases that exhibited promiscuous aryl-esterase activity [*feaB* and *aldH* (Figure 2A)]. The two remaining clones corresponded to nonenzymatic proteins, and although the observed esterase activity was associated with these plasmids, this activity probably resulted from alterations in the expression levels of endogenous esterases. For the unnatural chemistry, two enzymes with validated Kemp eliminase activity with 5-nitrobenzisoxazole were identified: purine nucleoside phosphorylase (*xapA*) and a predicted oxidoreductase (*ydbC*) (Table 2 and Figure 2B,C). As the latter can catalyze the breakdown of 5-nitrobenzisoxazole via oxidation rather than Kemp elimination, the identity of the elimination product was confirmed by liquid chromatography and mass spectrometry (LC−MS) (Figures S1−S3 of the Supporting Information).

The results described above indicate that the likelihood of recruiting promiscuous enzymes for xenobiotic substrates is higher for a naturally occurring chemical transformation such as ester hydrolysis compared to an unknown chemistry such as the Kemp elimination. Similar ratios of 3−5-fold were seen in the initial screen (~100 active wells with aryl ester **1** vs 20 wells with benzisoxazole **2**), as well as in individually validated variants: seven esterases/hydrolases exhibiting substrate ambiguity toward aryl ester **1** were identified. In comparison, we identified two enzymes exhibiting catalytic promiscuity toward aryl ester **1** and two enzymes that promiscuously catalyzed the elimination of benzisoxazole **2**. The activities observed in crude lysates were also much higher for cases of substrate ambiguity than for

Table 1. Summary of the ASKA Library Screens with Substrates 1−8

	esterase	Kemp elimination			phosphono-ester hydrolysis			
screening step	*p*-nitrophenyl acetate (**1**)	5-nitrobenzisoxazole (**2**)	benzisoxazole (**3**)	6-phosphobenzisoxazole (**4**)	*n*-butyl phosphopropyl (**5**)	3-aminobutyl (**6**)	4-carboxypropyl (**7**)	3-amino-3-(**8**)
(i) initial screen (no. of active wells)	>100[a]	22	2	24	11	15	9	9
(ii) deconvolution (no. of active clones)	18[a]	5	0	10	0	1	0	3
(iii) validation (no. of active clones with identified sequence)	11[a]	3	0	7	0	1	0	0
(iv) retransformation (no. of active clones verified in freshly transformed *E. coli*)	11[a]	2	0	4[b]	0	1	0	0

[a] Because of a large number of positive wells obtained in the screen with *p*-nitrophenyl acetate (>100), the 20 wells showing the highest rates were taken to deconvolution (step ii). These 20 wells gave rise to 11 clones with validated esterase activity. Two of these clones that exhibited the lowest activity were not enzymes, and one did not express an *E. coli* protein (Table 2). It is therefore likely that most of the significant, true hits, i.e., enzymes that catalyze *p*-nitrophenyl acetate hydrolysis, either promiscuously or via substrate ambiguity, were identified within the 20 most active wells. [b] Three of the four positive variants turned out to be phosphatases that catalyze hydrolysis of the 6-phosphate ester and have no Kemp eliminase activity. The fourth variant, *nadC*, lost activity with 6-phosphobenzisoxazole upon purification, and the origins of its activity (phosphatase or Kemp eliminase) could not be determined.

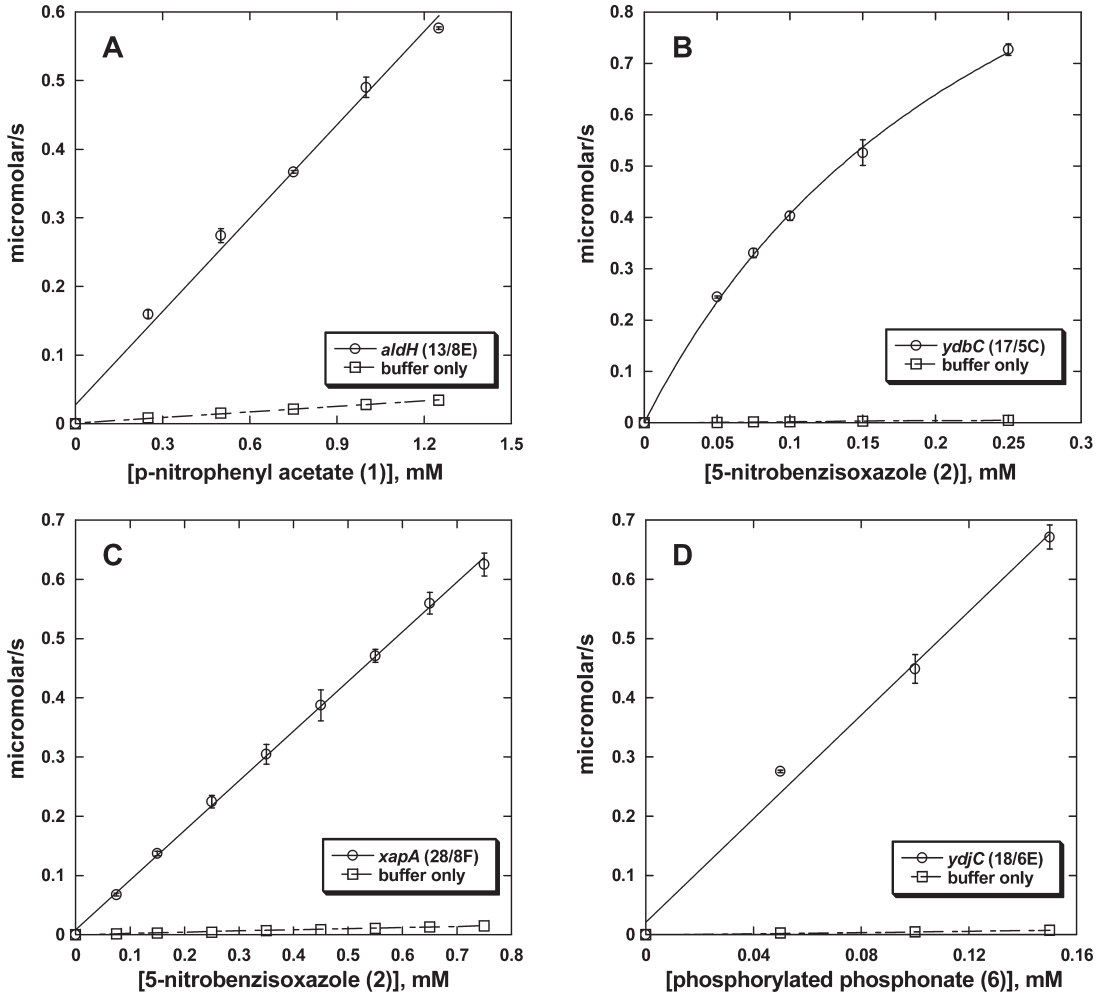


Figure 2. Activity plots for the promiscuous activities. (A) *aldH* (0.83 μM) with *p*-nitrophenyl acetate (**1**). (B) *ydbC* (13 μM) with 5-nitrobenzisoxazole (**2**). (C) *xapA* (0.83 μM) with 5-nitrobenzisoxazole (**2**). (D) *ydjC* (0.83 μM) with phosphorylated phosphonate (**6**). The data represent the average of two independent measurements, and the error bars represent the standard deviations. Catalytic parameters derived from these plots are listed in Table 2. Saturation was observed only in case of *ydbC*; for other proteins, k_{cat}/K_M values were derived by fitting the data to the linear regime of the Michaelis−Menten model.

Table 2. Identity and Activity of ASKA Clones Exhibiting Substrate Ambiguity and/or Promiscuous Activities

substrate	clones with verified activity (after validation and retransformation) (ASKA position, gene, and protein name)	activity in crude lysate after retransformation (μM product/min for 100 μL of lysate)	catalytic activity of purified enzymes, k_{cat}/K_M (M^{-1} s^{-1})[a]
p-nitrophenyl acetate (**1**)	5/7G, *aes*, acetyl esterase	3770 ± 430	2.05×10^4 [25]
	5/10B, *tesA*, thioesterase/protease	111 ± 8	hydrolysis of various *p*-nitrophenyl esters of amino-protected amino acids[26]
	30/10C, *yqiA*, predicted esterase	470 ± 30	hydrolysis of *p*-nitrophenyl butyrate[27]
	7/8A, *ybfF*, putative esterase	5.0 ± 0.4	hydrolysis of *p*-nitrophenyl butyrate[28]
	42/4D, *yjfP*, predicted hydrolase	13000 ± 400	hydrolysis of *p*-nitrophenyl butyrate[27]
	39/9D, *bioH*, carboxylesterase	675 ± 14	6.38×10^4 [29]
	13/11H, *feaB*, phenylacetaldehyde dehydrogenase	246 ± 2	5.6 nmol min^{-1} (mg of enzyme)$^{-1}$ [30]
	13/8E, *aldH*, γ-Glu γ-aminobutyraldehyde dehydrogenase	770 ± 67	502 ± 32[b]
	28/11E, *ypfH*, predicted esterase	1000 ± 80	hydrolysis of *p*-nitrophenyl butyrate[27]
	21/9E, part of expression vector	3.3 ± 0.4	not purified
	21/11G, *yecF*, putative inner membrane protein	3.28 ± 0.02	not purified
5-nitro-benzisoxazole (**2**)	17/5C, *ydbC*, predicted oxidoreductase	160 ± 33	k_{cat} = 0.114 ± 0.004 s^{-1} [c] K_M = 0.27 ± 0.02 mM k_{cat}/K_M = 430 ± 13 s^{-1} M^{-1}
	28/8F, *xapA*, purine nucleoside phosphorylase	6.9 ± 0.4	998 ± 32
6-phospho-benzisoxazole (**3**)	8/12D, *ybiV*, sugar phosphatase	16 ± 4	not purified; enzyme active as phosphatase[d]
	4/8C, *phoA*, bacterial alkaline phosphatase	230 ± 14	not purified; enzyme active as phosphatase
	43/1B, *ytjC*, putative phosphoglycerate mutase	33 ± 3	33 ± 2; enzyme active as phosphatase[e]
	2/1E, *nadC*, quinolinate phosphoribosyltransferase	18 ± 16 if it is phosphatase activity, 2.6 ± 2.3 if it is Kemp elimination	loss of activity with 6-phosphobenzisoxazole upon protein purification
3-phosphopropyl-*p*-nitrophenyl methyl phosphonate (**6**)	18/6E, *ydjC* (reannotated as *celG*), predicted cellobiose phosphatase	6.9 ± 0.2	486 ± 18[f]

[a] k_{cat}/K_M is provided when available, either from our measurements or from the literature. In some cases, the literature provides other parameters, which are cited as provided. [b] k_{cat}/K_M measured in the presence of 1 mM DTT and 4 mM NAD^+. In their absence, activity was >100-fold lower. [c] k_{cat}/K_M measured with 1.25 mM NAD^+; in its absence, activity was ~3-fold lower. [d] Phosphatase activity measured with 1 mM *p*-nitrophenyl phosphate and crude lysates (100 μL). [e] See Figure S5 of the Supporting Information. In addition, the purified *ytjC* dephosphorylated *p*-nitrophenyl phosphate [22 activity units (nanomoles of product per minute per milligram of protein), with 0.5 mM substrate]. [f] Activity with other phosphonates was also observed, albeit much lower than with the phosphorylated phosphonate (**6**): 115, 1.43, 1.37, and 2.06 activity units (nanomoles of product per minute per milligram of protein) for phosphorylated phosphonate (**6**), *n*-butyl phosphonate (**5**), aminobutyl phosphonate (**7**), and 3-amino-3-carboxypropyl phosphonate (**8**), respectively, with 0.15 mM substrates.

catalytic promiscuity. The activity levels are partial because differences in expression levels were not accounted for, yet, for example, the highest substrate ambiguity activity was ~13000 units, as opposed to ~770 units for the highest catalytically promiscuous activity (Table 2).

Substrate Tagging with Naturally Occurring Chemophores. These results indicate that, as anticipated, catalytic promiscuity is a relatively rare event. Would, then, the probability of identifying an enzyme for an unknown chemistry increase if the substrate carries a naturally occurring chemophore? To test this hypothesis, we examined the results of a ligand-centric analysis of the occurrence of various chemophores in various enzyme superfamilies.[9] Some of the analyzed chemophores comprise a single bond only (e.g., C−N or C−O), yet others comprise a stand-alone group such as amine, carboxyl, or phosphate, each of which appears to be among the most abundant chemophores. In choosing which chemophores to test, we were also forced to consider the feasibility of substrate synthesis and background decomposition rates. The comparative substrates (Figure 1B) had to be chosen such that the attached chemophores, and the length of the linker connecting the chemophore and the reactive group, would not affect substrate reactivity due to intramolecular catalysis.

In view of the considerations mentioned above, two unnatural chemistries were explored: the Kemp elimination and phosphonate ester hydrolysis. For the Kemp elimination, benzisoxazole substrates with or without a phosphate chemophore were tested (Figure 1, substrates **3** and **4**, respectively), whereby the reactivity of both is essentially the same and ~100-fold lower than that of 5-nitrobenzisoxazole (**2**). Natural phosphonoesterases have been identified only in marine bacteria.[19] Serine hydrolases react with phosphomonoesters, but with no turnover. Few enzymes that catalyze phosphonoester hydrolysis are known, but these activities are either promiscuous or recently evolved toward pesticides or herbicides.[20] Methyl *p*-nitrophenyl phosphonoester substrates with no natural chemophore (**5**), or with phosphate (**6**), amine (**7**), or amine carboxylate (**8**) chemophores, were therefore tested. These chemophores did not exhibit measurable through-bond or intramolecular catalysis effects, and under the screening conditions, the background rates of these substrates differed by ≤2-fold.

Kemp Elimination Substrates with or without a Chemophore. While no activity was detected in the screen with

unsubstituted benzisoxazole **3**, four active clones were identified with phosphorylated benzisoxazole **4** (Tables 1 and 2). The native substrates of all four enzymes contain a phosphate group, and their native activities involve phosphate transfer or hydrolysis. As it turned out, phosphorylated benzisoxazole **4** not only presented an opportunity for catalytic promiscuity but could also react via the substrate ambiguity of phosphatases (Figure 1). The absorbance signals of these two reactions are partly overlapping at the applied screening wavelength [320 nm (Figure S4 of the Supporting Information)]. Indeed, two of the four active clones are known phosphatases [*ybiV* and *phoA* (Table 2)]. They exhibited phosphatase activity with this substrate as monitored at 290 nm (maximal absorbance for phosphatase activity), and with other aryl phosphates. The third clone, *ytjC*, is annotated as a putative phosphoglycerate mutase, yet we observed its phosphatase activity (Figure S5 of the Supporting Information). The fourth identified enzyme, quinolinate phosphoribosyl transferase (*nadC*), lost its promiscuous activity upon protein purification. The failure to identify Kemp eliminases relates primarily to the low reactivity of **3** and **4**, because promiscuous catalysts for the more reactive 5-nitrobenzisoxazole **2** were readily identified. However, detection difficulties in the elimination of **3** and **4** at the noisy wavelength of 320 nm (vs 380 nm for **2**) also play a role.

Phosphonate Esters without or with Various Chemophore Tags. In the screen of the ASKA library with phosphonoesters, one active enzyme (*ydjC*, a predicted cellobiose phosphatase) was found for phosphorylated phosphonate **6**. No active clones were obtained for the analogous phosphonoesters **5**, **7**, and **8** (Tables 1 and 2). Purified *ydjC* exhibited measurable activity with the phosphorylated substrate **6** [$k_{cat}/K_M \approx 500\ M^{-1}\ s^{-1}$ (Figure 2D)], and this activity was inhibited by phosphate ions (Table S2 of the Supporting Information). *ydjC* also exhibited very low yet measurable activity with the other phosphonates (**5**, **7**, and **8**) that do not carry a phosphate group (Table 2, footnote f). Despite its putative assignment, *ydjC* did not exhibit phosphatase activity with aryl phosphates. Further, removal of the phosphate group of **6** by alkaline phosphatase (as judged by TLC) had no effect on the rate of spontaneous hydrolysis of the *p*-nitrophenyl phosphonate ester. This rules out the possibility that *ydjC* dephosphorylates **6**, and the product breaks down rapidly because of intramolecular catalysis by the 3-hydroxy group.

Altogether, it appears that *ydjC*, whose natural substrate and reaction are unknown, exhibits catalytic phosphonoesterase promiscuity. However, in the absence of a phosphate group that anchors the substrate into its active site, this activity is very low and is certainly too low to be detected in crude lysates. Thus, although we were able to identify only one such case in the *E. coli* proteome, the identification of *ydjC* suggests that the likelihood of identifying an enzyme for an unknown chemistry (phosphonoesterase in this test case) may increase when the substrate contains an abundant naturally occurring chemophore such as phosphate.

DISCUSSION

Our comparison of the frequency of promiscuously active enzymes in the proteome of *E. coli* suggests that, as anticipated, substrate ambiguity is more frequent than catalytic promiscuity. If the level of activity is also taken into consideration, the relative occurrences of substrate ambiguity versus catalytic promiscuity can become even higher. This trend was seen throughout our experiments: (1) in the comparison between aryl esterase and Kemp eliminase activities [**1** vs **2** (Figure 1A)] and (2) in the ~4-fold higher number of aryl esterases that are esterases and/or hydrolases (substrate ambiguity) compared to other enzymes that act via catalytic promiscuity (e.g., dehydrogenases). Finally, the screen with phosphobenzisoxazole **4** failed to identify cases of catalytic promiscuity (Kemp eliminase) but did identify at least three enzymes that act as phosphatases on this xenobiotic substrate. Our study examined a limited range of substrates and therefore provides only a quantitative conclusion regarding the prevalence of substrate ambiguity over catalytic promiscuity. Should other comparative substrates be chosen, different frequencies of "positive hits" are likely to be observed. However, the general trends observed here are likely to be retained.

The divergence of a new enzymatic function is therefore more likely to begin from an enzyme that catalyzes the same reaction rather than from an enzyme that binds the same or a similar substrate but catalyzes a different reaction. Given the huge diversity of catalytic chemistries within existing genomes, starting points with very similar or even identical chemistries are likely to be common. However, our results also suggest that in the absence of enzymes with the same chemistry, having a substrate "decorated" with an abundant, naturally occurring motif (chemophore), and a phosphate group in particular, increases the likelihood of finding a promiscuous catalyst. The chemophore serves as a common anchor that attracts the substrate to a larger fraction of active sites and allows its alignment with a suitable catalytic residue.

Although the screens described above identified only one case of promiscuity mediated by a shared substrate motif, or, chemophore, this hypothesis is also supported by the fact that certain chemophores are overrepresented in the substrates of certain enzyme superfamilies yet do not play a part in the reaction itself. Phosphate is possibly the most common example.[21] Indeed, a comparative study of enzymatic kinetic parameters indicated that the K_M values of small substrates (MW > 350 Da) tend to be rather limited ($\geq$0.1 mM). Tagging of such small substrates with commonly occurring groups decreases the K_M values. The most distinct example among large modifiers is coenzyme A (CoA). Among small to midsize modifying groups (or chemophores), only the phosphoryl group significantly decreases K_M values [the median ratio of K_M values for the same substrate with and without phosphate is 3.6 (A. Bar-Even, E. Noor, D. Davidi, Y. Savir, D. S. Tawfik, and R. Milo, manuscript submitted for publication)]. Thus, although a comparison of one hit (for the phosphate-tagged phosphonate **6**) to no hits [for the untagged (**5**) and for amino- and amino acid-tagged phosphonates (**7** and **8**, respectively)] bears no statistical significance, the fact that the phosphate-tagged substrate was catalyzed by an enzyme whose natural substrate is also phosphate-tagged may be more than a mere coincidence.

That the example we identified, *ydjC*, exhibited very low yet measurable activity with other phosphonates (untagged, or tagged with chemophores other than phosphate) further supports our hypothesis that a common natural chemophore, and phosphate in particular, can provide a powerful handle for newly directing substrates into enzyme active sites.

Chemophore anchoring may also find practical implications for biocatalysis, as the attachment of abundant chemophores such as phosphate to xenobiotic substrates may increase the chances of identifying catalysts for unnatural reactions. Chemophores were previously used in biocatalytic conversions with the aim of increasing the enzymatic activity with synthetic substrates.[22−24] However, the target enzymes were known in

advance. The added chemophores were also different for each case and were all unnatural. These chemophores probably exert specific effects on one particular target enzyme and substrate. The use of a natural chemophore, and of phosphate in particular, may comprise a more generic solution.

■ ASSOCIATED CONTENT

Supporting Information. Substrate synthesis details, ASKA library screening, protein purification procedures and results, and identification of reaction products by LC and LC−MS. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel. Phone: +972 8 934 3637. Fax: +972 8 934 4118. E-mail: tawfik@weizmann.ac.il.

Present Addresses

§Department of Biochemistry, University of Washington, Seattle, WA 98195.

Funding Sources

This work was supported by the Meil de Botton Aynsley, the Sasson & Marjorie Peress Philanthropic Fund, and the Adams Fellowship (Israel Academy of Science) to O.K. D.S.T. is the Nella and Leo Benoziyo Professor of Biochemistry.

■ ACKNOWLEDGMENT

We thank Svetlana Makiyevsky for help with organic synthesis, Kalia Bernath-Levin for help with ASKA screens, and Asaph Aharoni for help with LC−MS analyses. We thank the Nara Institute of Science and Technology for providing the ASKA library.

■ REFERENCES

(1) Jensen, R. A. (1974) Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol. 30*, 409–425.

(2) O'Brien, P. J., and Herschlag, D. (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol. 6*, R91–R105.

(3) Glasner, M. E., Gerlt, J. A., and Babbitt, P. C. (2006) Evolution of enzyme superfamilies. *Curr. Opin. Chem. Biol. 10*, 492–497.

(4) Khersonsky, O., and Tawfik, D. S. (2010) Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annu. Rev. Biochem. 79*, 471–505.

(5) Copley, S. D. (2010) in *Comprehensive Natural Products Chemistry* (Barton, D., and Meth-Cohn, O., Eds.) Elsevier, Oxford, U.K.

(6) Gerlt, J. A., and Babbitt, P. C. (1998) Mechanistically diverse enzyme superfamilies: The importance of chemistry in the evolution of catalysis. *Curr. Opin. Chem. Biol. 2*, 607–612.

(7) Teichmann, S. A., Rison, S. C., Thornton, J. M., Riley, M., Gough, J., and Chothia, C. (2001) The evolution and structural anatomy of the small molecule metabolic pathways in *Escherichia coli*. *J. Mol. Biol. 311*, 693–708.

(8) Nobeli, I., Favia, A. D., and Thornton, J. M. (2009) Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol. 27*, 157–167.

(9) Nobeli, I., Spriggs, R. V., George, R. A., and Thornton, J. M. (2005) A ligand-centric analysis of the diversity and evolution of protein-ligand relationships in *E. coli*. *J. Mol. Biol. 347*, 415–436.

(10) Chiang, R. A., Sali, A., and Babbitt, P. C. (2008) Evolutionarily conserved substrate substructures for automated annotation of enzyme superfamilies. *PLoS Comput. Biol. 4*, e1000142.

(11) Ji, H. F., Kong, D. X., Shen, L., Chen, L. L., Ma, B. G., and Zhang, H. Y. (2007) Distribution patterns of small-molecule ligands in the protein universe and implications for origin of life and drug discovery. *Genome Biol. 8*, R176.1–R176.13.

(12) Bornscheuer, U. T., and Kazlauskas, R. J. (2004) Catalytic promiscuity in biocatalysis: Using old enzymes to form new bonds and follow new pathways. *Angew. Chem., Int. Ed. 43*, 6032–6040.

(13) Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2005) Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): Unique resources for biological research. *DNA Res. 12*, 291–299.

(14) Soo, V. W., Hanson-Manful, P., and Patrick, W. M. (2011) Artificial gene amplification reveals an abundance of promiscuous resistance determinants in Escherichia coli.. *Proc. Natl. Acad. Sci. U.S.A. 108*, 1199–1200.

(15) Desai, K. K., and Miller, B. G. (2010) Recruitment of genes and enzymes conferring resistance to the nonnatural toxin bromoacetate. *Proc. Natl. Acad. Sci. U.S.A. 107*, 17968–17973.

(16) Wei, Y., and Hecht, M. H. (2004) Enzyme-like proteins from an unselected library of designed amino acid sequences. *Protein Eng., Des. Sel. 17*, 67–75.

(17) Rothlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S., and Baker, D. (2008) Kemp elimination catalysts by computational enzyme design. *Nature 453*, 190–194.

(18) Casey, M. L., Kemp, D. S., Paul, K. C., and Cox, D. D. (1973) The physical organic chemistry of benzisoxazoles. I. The mechanism of the base-catalyzed decomposition of benzisoxazoles. *J. Org. Chem. 38*, 2294–2301.

(19) White, A. E. (2009) New insights into bacterial acquisition of phosphorus in the surface ocean. *Proc. Natl. Acad. Sci. U.S.A. 106*, 21013–21014.

(20) van Loo, B., Jonas, S., Babtie, A. C., Benjdia, A., Berteau, O., Hyvonen, M., and Hollfelder, F. (2010) An efficient, multiply promiscuous hydrolase in the alkaline phosphatase superfamily. *Proc. Natl. Acad. Sci. U.S.A. 107*, 2740–2745.

(21) Srinivasan, V., and Morowitz, H. J. (2009) Analysis of the intermediary metabolism of a reductive chemoautotroph. *Biol. Bull. 217*, 222–232.

(22) Olivo, H. F., and Hemenway, M. S. (1999) Total Synthesis of (±)-Epibatidine Using a Biocatalytic Approach. *J. Org. Chem. 64*, 8968–8969.

(23) de Raadt, A., Griengl, H., and Weber, H. (2001) The concept of docking and protecting groups in biohydroxylation. *Chem.—Eur. J. 7*, 27–31.

(24) Savile, C. K., and Kazlauskas, R. J. (2006) The 3-(3-pyridine)-propionyl anchor group for protease-catalyzed resolutions: p-Toluenesulfinamide and sterically hindered secondary alcohols. *Adv. Synth. Catal. 348*, 1183–1192.

(25) Kanaya, S., Koyanagi, T., and Kanaya, E. (1998) An esterase from *Escherichia coli* with a sequence similarity to hormone-sensitive lipase. *Biochem. J. 332* (Pt 1), 75–80.

(26) Pacaud, M., Sibilli, L., and Le Bras, G. (1976) Protease I from *Eschrechia coli*. *Eur. J. Biochem. 69*, 141–151.

(27) Kuznetsova, E., Proudfoot, M., Sanders, S. A., Reinking, J., Savchenko, A., Arrowsmith, C. H., Edwards, A. M., and Yakunin, A. F. (2005) Enzyme genomics: Application of general enzymatic screens to discover new enzymes. *FEMS Microbiol. Rev. 29*, 263–279.

(28) Kim, J. H., Jang, K. S., Yang, Y. H., Kim, Y. G., Lee, J. H., Oh, M. K., Kim, B. G., and Lee, C. S. (2008) Rapid functional identification of putative genes based on the combined in vitro protein synthesis with mass spectrometry: A tool for functional genomics. *Anal. Biochem. 375*, 11–17.

(29) Sanishvili, R., Yakunin, A. F., Laskowski, R. A., Skarina, T., Evdokimova, E., Doherty-Kirby, A., Lajoie, G. A., Thornton, J. M., Arrowsmith, C. H., Savchenko, A., Joachimiak, A., and Edwards, A. M. (2003) Integrating structure, bioinformatics, and enzymology to discover function: BioH, a new carboxylesterase from *Escherichia coli*. *J. Biol. Chem. 278*, 26039–26045.

(30) Ferrandez, A., Prieto, M. A., Garcia, J. L., and Diaz, E. (1997) Molecular characterization of PadA, a phenylacetaldehyde dehydrogenase from *Escherichia coli*. *FEBS Lett. 406*, 23–27.