Genome-wide Analysis of Substrate Specificities of the *Escherichia coli* Haloacid Dehalogenase-like Phosphatase Family*[S]

Ekaterina Kuznetsova[‡§¶], Michael Proudfoot[‡], Claudio F. Gonzalez[‡], Greg Brown[‡], Marina V. Omelchenko[||], Ivan Borozan[‡], Liran Carmel[||], Yuri I. Wolf[||], Hirotada Mori[**], Alexei V. Savchenko[‡§¶], Cheryl H. Arrowsmith[‡‡‡1], Eugene V. Koonin[||], Aled M. Edwards[‡§¶2], and Alexander F. Yakunin[‡¶3]

*From the ‡Banting and Best Department of Medical Research and the §Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L6, Canada, the ¶Ontario Center for Structural Proteomics, Ontario Cancer Institute, Toronto, Ontario M5G 2C4, Canada, the ||National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, the **Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0101, Japan, and ‡‡Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L6, Canada*

Haloacid dehalogenase (HAD)-like hydrolases are a vast superfamily of largely uncharacterized enzymes, with a few members shown to possess phosphatase, β-phosphoglucomutase, phosphonatase, and dehalogenase activities. Using a representative set of 80 phosphorylated substrates, we characterized the substrate specificities of 23 soluble HADs encoded in *the Escherichia coli* genome. We identified small molecule phosphatase activity in 21 HADs and β-phosphoglucomutase activity in one protein. The *E. coli* HAD phosphatases show high catalytic efficiency and affinity to a wide range of phosphorylated metabolites that are intermediates of various metabolic reactions. Rather than following the classical "one enzyme-one substrate" model, most of the *E. coli* HADs show remarkably broad and overlapping substrate spectra. At least 12 reactions catalyzed by HADs currently have no EC numbers assigned in Enzyme Nomenclature. Surprisingly, most HADs hydrolyzed small phosphodonors (acetyl phosphate, carbamoyl phosphate, and phosphoramidate), which also serve as substrates for autophosphorylation of the receiver domains of the two-component signal transduction systems. The physiological relevance of the phosphatase activity with the preferred substrate was validated *in vivo* for one of the HADs, YniC. Many of the secondary activities of HADs might have no immediate physiological function but could comprise a reservoir for evolution of novel phosphatases.

Most enzymes form families of paralogs whose members are related by sequence and catalyze similar reactions but have evolved specific biological functions. Comprehensive determination of the substrate specificities and selectivities of all metabolic enzymes in an organism is an essential step toward understanding the relationship between the proteome and the metabolome. By the most recent estimate, *Escherichia coli* possesses at least 1186 metabolic enzymes and 1005 metabolites (1). The most common functional group in the metabolome is phosphate; 35–40% of the metabolites contain a phosphate group (2). The pool of phosphorylated metabolites is controlled by the activity of diverse kinases and phosphatases, of which there are hundreds in the *E. coli* genome.

Haloacid dehalogenase (HAD)[4]-like hydrolases (3) represent the largest family of predicted small molecule phosphatases encoded in the genomes of bacteria, archaea, and eukaryotes, with 6,805 proteins in data bases. The great majority of these proteins have no known biochemical or biological function. In any individual genome, the number of HAD genes can range from 10 to 20 in different bacteria to 100 in humans and 115 in *Arabidopsis thaliana* (InterPro data base). HADs share little overall sequence similarity (15–30% identity), but they can be unequivocally identified by the presence of three short conserved sequence motifs (3) (supplemental Fig. 1). Most of the characterized HADs have phosphatase activity (CO–P bond hydrolysis), and several also catalyze dehalogenase (C–halogen bond hydrolysis), phosphonatase (C–P bond hydrolysis), and β-phosphoglucomutase (CO–P bond hydrolysis and intramolecular phosphoryl transfer) reactions (3, 4). The biochemically and structurally studied HADs include phosphoserine phosphatase SerB from *Methanococcus jannaschii* (5), phosphoglycolate phosphatase from *Thermoplasma acidophilum* (6), phosphonoacetaldehyde hydrolase from *Bacillus cereus* (7), β-phosphoglucomutase from *Lactococcus lactis* (8), haloacid dehalogenases from *Pseudomonas* sp. YL (9), and *Xanthobacter*

[4] The abbreviations used are: HAD, haloacid dehalogenase-like hydrolase; CHES, 2-(cyclohexylamino)ethanesulfonic acid; PLP, pyridoxal 5′-phosphate; PEP, phosphoenolpyruvate; MOPS, 4-morpholinepropanesulfonic acid; *p*NPP, *p*-nitrophenyl phosphate.

Supplemental Material can be found at:
http://www.jbc.org/cgi/content/full/M605449200/DC1

autotrophicus (10), and two *E. coli* phosphatases, YbiV and NagD (11, 12). However, the vast majority of HADs remains uncharacterized. Since these enzymes generally show little sequence similarity, the catalyzed reaction and, especially, the substrate specificity are hard to predict on the basis of sequence conservation and have to be determined experimentally.

The *E. coli* genome encodes five membrane-bound and 23 soluble HAD-like hydrolases, representing ~40% of the *E. coli* proteins with known or predicted small molecule phosphatase activity (1). None of the *E. coli* HADs is essential for bacterial growth (13). The physiological substrates have been experimentally identified for three soluble *E. coli* HADs, namely phosphoglycolate phosphatase Gph, 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase YrbI, and trehalose 6-phosphatase OtsB (14–16) Only three enzymes (Gph, YrbI, and NagD) have been characterized biochemically (12, 14, 15), and three-dimensional structures have been resolved for YbiV and NagD (11, 12).

Using a set of 80 representative phosphorylated metabolites, we characterized the substrate specificities of all 23 soluble *E. coli* HADs and found that they comprise a family of promiscuous phosphatases with overlapping substrate profiles and are capable of hydrolyzing a wide range of phosphorylated metabolites, including carbohydrates, nucleotides, organic acids, coenzymes, and small phosphodonors. Genetic analysis demonstrated that the activity of one of the HADs, YniC, toward its preferred substrate was biologically important. We further show that all *E. coli* HADs have phosphatase activity against small phosphate donors (acetyl phosphate, carbamoyl phosphate, phosphoramidate), which resembles the autophosphorylation reaction catalyzed by CheY fold receiver domains of the two-component regulatory systems. Together with the previously reported structural similarity, these results indicate that the HAD superfamily and the receiver domain originate from an ancestral low specificity phosphatase. Clustering of *E. coli* HADs on the basis of their phosphatase activities (k_{cat}/K_m) was incongruent with the sequence-based phylogeny. Thus, many of the secondary activities of HADs might be of no immediate functional importance but comprise a reservoir for evolution of phosphatases with novel specificities.

EXPERIMENTAL PROCEDURES

Gene Cloning and Protein Purification—For most HADs analyzed in this work, the genes were amplified by PCR from the *E. coli* DH5α genomic DNA and cloned into a modified pET15b (Novagen) as previously described (17). Several HADs (YbjI, YaeD, and YrbI) were cloned from the *E. coli* K12 W3110 genomic DNA into the archive vector pCA24N (Genobase database; available on the World Wide Web at ecoli.aist-nara.ac.jp). Purification of proteins for screening and biochemical characterization was performed as previously described (18). NagD was expressed in an insoluble form and was partially refolded from the inclusion bodies by a buffer exchange method using the Spin-Column Protein Folding Screen kit (SFC01-10) and the column PFC02 from ProFoldin Protein Folding Services according to the manufacturer's instructions (available on the World Wide Web at www.profoldin.com).

Enzymatic Screens and Assays—General phosphatase screens with *p*-nitrophenyl phosphate (*p*NPP) as substrate and natural substrate phosphatase screens with 80 phosphorylated compounds from Sigma (supplemental Table 1) were performed as previously described (18). Acetyl-phosphatase activity was assayed by measuring the acetyl-phosphate concentration using the hydroxylamine protocol of Lipmann and Tuttle (19). The production of fructose in enzymatic reactions was determined using an enzyme-coupled assay with fructose dehydrogenase (F5152; Sigma), essentially as previously described (20). This assay was adapted for 96-well microplates (200-μl reaction mixtures). Haloacid dehalogenase activity was determined spectrophotometrically by measuring the release of halide ions using the mercuric thiocyanide method (21). The assays were adapted to the 96-well format (125 μl) and contained 50 mM CHES buffer (pH 9.0), 10 mM substrate, and 5 μg of protein. Six compounds were used as substrates (*R*-chloropropionic acid, *S*-chloropropionic acid, bromoacetic acid, 4-chlorobenzoic acid, 2,2-dichloropropionic acid, and 2-bromopropionic acid), and L-2-haloacid dehalogenase from *Pseudomonas* sp. YL (22) was used as a positive control. Phosphonatase activity was assayed using phosphonoacetate as a substrate by measuring the release of inorganic phosphate using the Malachite Green reagent as previously described (23). β-Phosphoglucomutase activity was determined using a glucose-6-phosphate dehydrogenase-coupled assay and 1 mM β-glucose 1-phosphate as substrate (24).

For K_m and V_{max} determination, the phosphatase assays contained substrates at concentrations 0.005–2.0 mM. Kinetic parameters were determined by nonlinear curve fitting from the Lineweaver-Burk plot using the GraphPad Prism software (version 4.00 for Windows, GraphPad Software, San Diego, CA).

Mutagenesis and Growth Experiments—The *yniC* gene was deleted from the chromosome of the *E. coli* K-12 W3110 strain using a one-step inactivation method described by Datsenko and Wanner (25). The obtained ΔYniC strain contains an unmarked gene deletion, which was verified using PCR. The YniC-overexpressing strain was prepared by subcloning (PCR) of the wild-type *yniC* into the KpnI/HindIII sites of the arabinose-inducible vector pBAD-33 (26). The resulting plasmid (pKC1) and the empty (no insert) vector pBAD-33 (a control) were transformed into the wild-type W3110 strain. Cells were grown aerobically at 37 °C on the MOPS-buffered minimal medium containing 0.2% of succinate as a carbon source, and the expression of YniC was induced by the addition of 0.02% arabinose. The culture growth (A_{600}) was determined after 14 h of cultivation.

Bioinformatic Analyses—Hierarchical clustering of HADs (based on their substrate profiles) and substrates (based on their HAD spectra) was calculated using cosine correlations, and groups were clustered using the average method (R Foundation for Statistical Computing; available on the World Wide Web at www.R-project.org). For hierarchical clustering of proteins across HAD substrates, we consider each protein as a point in the *m*-dimensional space, where *m* designates the total number of HAD substrates. Each protein can then be represented as a vector A = (s1, . . . s*m*) of length |A| and unit (nor-

malized) vector A′(s1′ = s1/|A|, . . . s*m*′ = s*m*/|A|). We use the cosine angle between normalized vectors as a similarity measure between proteins across substrates. The protein profiles across substrates are clustered into groups using the average linkage method (R Foundation for Statistical Computing; available on the World Wide Web at www.R-project.org). For hierarchical clustering of substrates across proteins, we consider each protein as a point in the *n*-dimensional space, where *n* designates the total number of proteins. Each substrate can then be represented as a vector B = (p1, . . . p*n*) of length |B| and unit (normalized) vector B′(p1′ = p1/|A|, . . . p*n*′ = p*n*/|B|). We use the cosine angle between normalized vectors B′ as a similarity measure between substrates across proteins. The substrate profiles across proteins are clustered into groups using the average linkage method (R Foundation for Statistical Computing; available on the World Wide Web at www.R-project.org).

The distance between the catalytic efficiency (k_{cat}/K_m) profiles of 18 enzymes acting on 26 different substrates (see Table 2) was computed using the square distance approach (see supplemental materials and "Experimental Procedures"). A neighbor-joining tree was constructed from the distance matrix using the NEIGHBOR program of the PHYLIP package (28). Multiple alignments of amino acid sequences were constructed using the MUSCLE program (29) and optimized manually to ensure the correct alignment of known sequence motifs of the HAD superfamily (3). Sequence-based phylogenetic trees were constructed using the following methods: (i) neighbor-joining method, as implemented in the NEIGHBOR program of the PHYLIP package (28), (ii) least squares method as implemented in the FITCH program of the PHYLIP package (28), (iii) local maximum likelihood optimization of the least squares tree using the ProtML program of the MOLPHY package (30), (iv) quartet puzzling as implemented in the TREE-PUZZLE program (31), and (v) Markov chain Monte Carlo Bayesian estimation using the MrBayes program (32, 33).

RESULTS

The Enzymatic Activities of the E. coli HADs—HADs share relatively little overall sequence similarity (~15 to 30% identity), but they can be identified by the presence of three short conserved sequence motifs (3) (supplemental Fig. 1). 22 soluble HADs encoded by the *E. coli* genome were purified to homogeneity (supplemental Fig. 2) and screened for phosphatase, dehalogenase, phosphonatase, and β-phosphoglucomutase activities. 16 previously uncharacterized HADs showed detectable activity with *p*NPP, a general phosphatase substrate (18) (Table 1). All of these HADs required the presence of a divalent metal cation for activity ($Mg^{2+} > Co^{2+} > Mn^{2+} > Ni^{2+}$) and showed a broad range of activities and affinities to *p*NPP (V_{max} 0.03–4.95 μmol/min mg of protein; K_m = 0.92–17.8 mM). Several proteins, YcjU (HAD11), YigB (HAD7), OtsB (HAD17), YedP (HAD19), YaeD (HAD20), and HisB (HAD21), had no activity against *p*NPP. YcjU (HAD11) has been annotated in sequence data bases as a putative β-phosphoglucomutase, and indeed, we detected this activity using a specific assay with β-glucose-1-P as a substrate (3.81 ± 0.13 μmol/min/mg of protein). Low β-phosphoglucomutase activity (0.02–0.15 μmol/min/mg of

TABLE 1
Phosphatase activity of purified *E. coli* HADs against the model substrate *p*NPP and small phosphodonors

Phosphatase activity is shown in μmol/min/mg of protein (the mean values are shown; S.D. values for all substrates were 10.95 ± 2.99% of the mean value). Assays were performed in the presence of 2.5 mM Mg^{2+} and the following concentrations of substrates: 10 mM *p*NPP, 10 mM acetyl-P, 0.13 mM carbamoyl-P, or 0.13 mM imido-di-P. These substrate concentrations were saturating for most enzymes (except for carbamoyl-P, which showed no saturation with all tested proteins). ND, not detected.

HADs	*p*NPP	Acetyl-P	Carbamoyl-P	Imido-di-P
HAD1 (YniC)	0.60	11.1	0.23	0.01
HAD2 (YfbT)	1.13	28.4	0.92	0.08
HAD3 (YieH)	0.03	6.83	0.24	0.81
HAD4 (YihX)	0.22	6.72	0.10	0.44
HAD5 (YjjG)	4.95	38.5	0.13	3.14
HAD6 (YqaB)	0.84	8.08	1.55	0.49
HAD7 (YigB)	ND	0.50	0.07	0.25
HAD8 (YrfG)	0.60	2.06	0.05	0.02
HAD9 (SerB)	14.7	46.7	0.93	0.10
HAD10 (Gph)	0.12	10.2	1.06	1.90
HAD11 (YcjU)	ND	0.13	0.03	0.01
HAD12 (YbiV)	2.43	65.4	1.03	4.87
HAD13 (YidA)	0.70	28.2	3.88	21.3
HAD14 (YbhA)	0.06	6.59	0.27	2.41
HAD15 (YbjI)	0.29	3.98	0.12	0.35
HAD16 (YigL)	0.3	11.5	0.67	0.97
HAD17 (OtsB)	ND	ND	0.18	0.06
HAD18 (Cof)	0.11	0.33	0.04	0.05
HAD19 (YedP)	ND	ND	0.01	0.04
HAD20 (YaeD)	ND	0.26	0.02	3.28
HAD21 (HisB)	ND	ND	0.02	0.17
HAD22 (YrbI)	0.09	1.96	0.06	0.14

protein) was also detected in YihX (HAD4), YqaB (HAD6), YbhA (HAD14), Cof (HAD18), Gph (HAD10), and YaeD (HAD20). Screening of purified *E. coli* HADs for hydrolytic activity against phosphonoacetate (phosphonatase substrate) or several haloacid dehalogenase substrates (chloroacetate, bromoacetate, chloropropionate, dichloropropionate, bromopropionate, and chlorobenzoate) produced no positive results (data not shown). Thus, most of the soluble *E. coli* HADs are phosphatases.

Phosphatase Activity with Natural Substrates—In an effort to identify the physiological roles for the *E. coli* HADs, all purified proteins were screened for phosphatase activity (P_i release) against a set of 80 natural phosphatase substrates (supplemental Table 1) representing all main divisions of the *E. coli* phosphometabolome (nucleotides, phosphorylated carbohydrates, organic acids, and amino acids) (2). These screens did not detect activity for YcjU (HAD11), HisB (HAD21), and YedP (HAD19). As mentioned above, YcjU is a β-phosphoglucomutase, whereas HisB (predicted histidinol phosphatase) and YedP might be phosphatases that are highly specific to their respective substrates. The remaining 20 HADs showed significant phosphatase activity toward a variety of substrates, in many cases, at levels predicted to be of physiological relevance (Fig. 1). To keep the background low, the screens with natural substrates were performed with substrate concentrations of 0.1–0.2 mM, which were subsequently found to be nonsaturating for most HADs (Table 2). Therefore, the observed velocities do not represent the maximum activities of these enzymes. Nevertheless, these screens not only identified the positive substrates and revealed the broad substrate ranges for most HADs but also correctly identified their preferred substrates (as shown by the subsequent saturation experiments presented in Table 2). The broad substrate spectra of HADs could not be
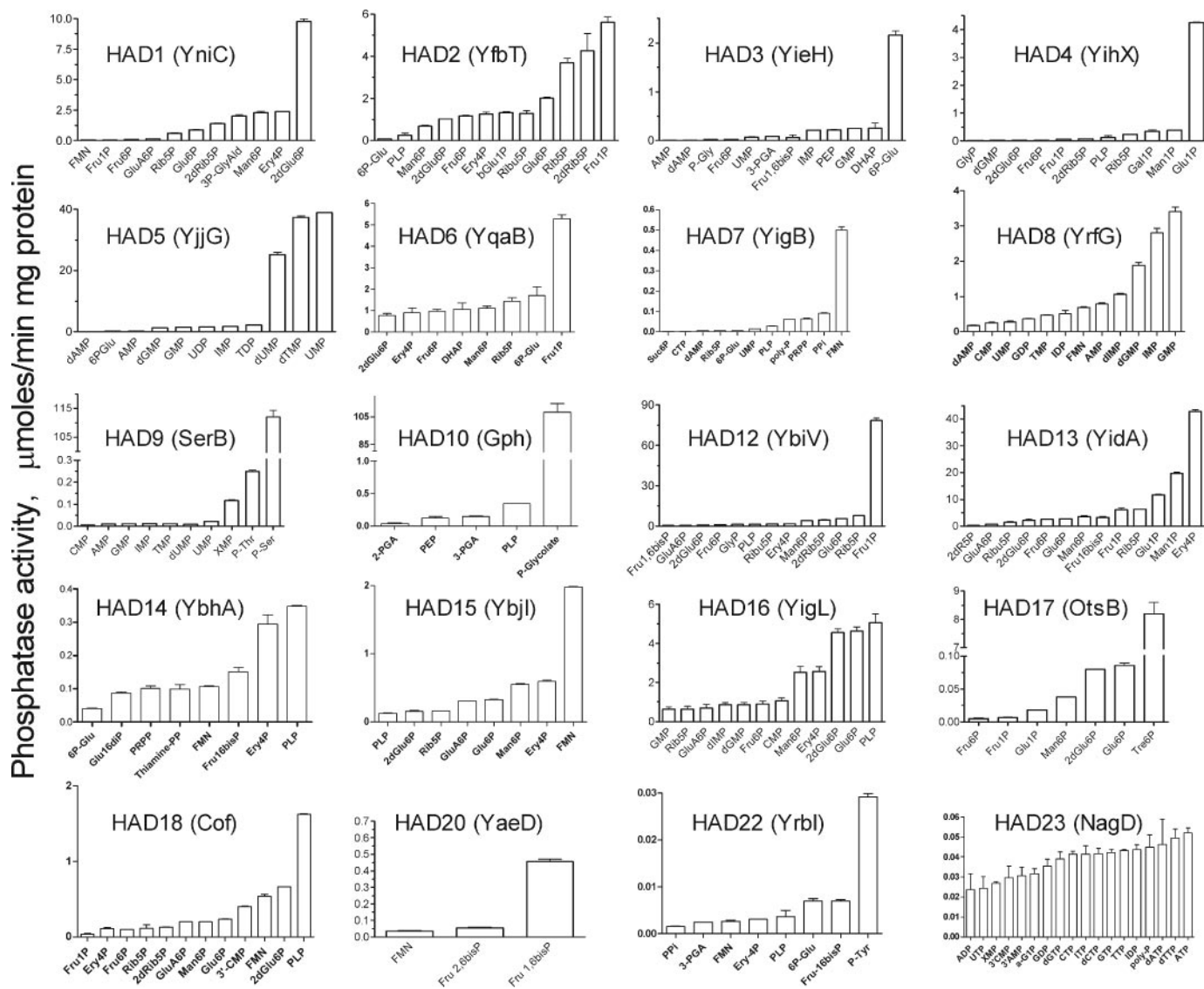
FIGURE 1. Substrate profiles of the *E. coli* HAD phosphatases. Reaction mixtures contained 0.25 mM natural substrate (0.15 mM for nucleoside triphosphates) and 1 μg of purified HAD. *2dGlu6P*, 2-deoxyglucose-6-P; *2dRib5P*, 2-deoxyribose-5-P; *DHAP*, dihydroxyacetone phosphate; *Ery4P*, erythrose-4-P; *Fru1P*, fructose 1-phosphate; *Fru1,6bisP*, fructose-1,6-bis-P; *Fru2,6bisP*, fructose-2,6-bis-P; *Fru6P*, fructose-6-P; *Gal1P*, galactose-1-P; *Glu1P*, α-glucose-1-P; *Glu6P*, glucose-6-P; *GluA6P*, glucosamine-6-P; *bGlu1P*, β-glucose-1-P; *Glu1,6bisP*, glucose-1,6-bis-P; *GlyP*, glycerophosphate; *2-PGA*, 2-phosphoglycerate; *3-PGA*, 3-phosphoglycerate; *Man1P*, mannose 1-P; *Man6P*, mannose 6-P; *6P-Glu*, 6-phosphogluconate; *PEP*, phosphoenolpyruvate; *3P-GlyAld*, 3-phosphoglyceraldehyde; *P-Gly*, phosphoglycolate; *PRPP*, phosphoribosyl pyrophosphate; *P-Thr*, phosphothreonine; *P-Tyr*, phosphotyrosine; *Poly-P*, polyphosphate; *PLP*, pyridoxal-5-P; *PPi*, inorganic pyrophosphate; *Rib5P*, ribose-5-P; *Ribu5P*, ribulose-5-P; *Suc6P*, sucrose-6-P.

accounted for by the presence of contaminating activities in purified preparations. Further extensive purification of HADs did not eliminate their activities toward different substrates, and alanine replacement mutagenesis of the catalytic Asp residues of two HADs, YniC and YidA, resulted in at least 99.9% abrogation of phosphatase activity.

Phosphatase activity of the *E. coli* HADs toward natural substrates has neutral or slightly acidic pH optimums (pH 6.0–7.5) and is strictly dependent on the presence of a divalent metal cation (Mg^{2+}, Mn^{2+}, Co^{2+}, and Zn^{2+}). With Mg^{2+}, HADs usually showed higher activity, but they required higher concentrations of this metal for the saturation of phosphatase activity ($K_{m, Mg} = 0.05$–1.8 mM; $K_{m, Mn} = 0.006$–0.04 mM; $K_{m, Co} = 0.04$–0.84 mM) (supplemental Table 2). YniC (HAD1) was also found to be active in the presence of Zn^{2+} ($K_m = 0.10$–0.27

mM). The metal affinities of HADs (especially to Mg^{2+}; determined by activity) also varied 3–10 times depending on the substrate (supplemental Table 2). Metals also affected the kinetic parameters of the *E. coli* HADs with most substrates (Table 2). Typically, these enzymes showed a higher K_m (lower apparent substrate affinities) in the presence of Mg^{2+} and lower K_m with Mn^{2+} or Co^{2+}. Several enzymes (YniC (HAD1; mannose-6-P, 2-deoxyglucose-6-P), YfbT (HAD2; 2-deoxyribose-5-P), YbhA (HAD14; erythrose-4-P), YbjI (HAD15; mannose-6-P), and YigL (HAD16; 2-deoxyglucose-6-P)) showed no substrate saturation in the presence of Mg^{2+} but were saturated by these substrates in the presence of Mn^{2+} or Co^{2+} (Table 2). These results indicate that metals affect the binding of substrates to HADs, which seems to be a rate-limiting step for some HAD/substrate pairs in the presence of Mg^{2+}.
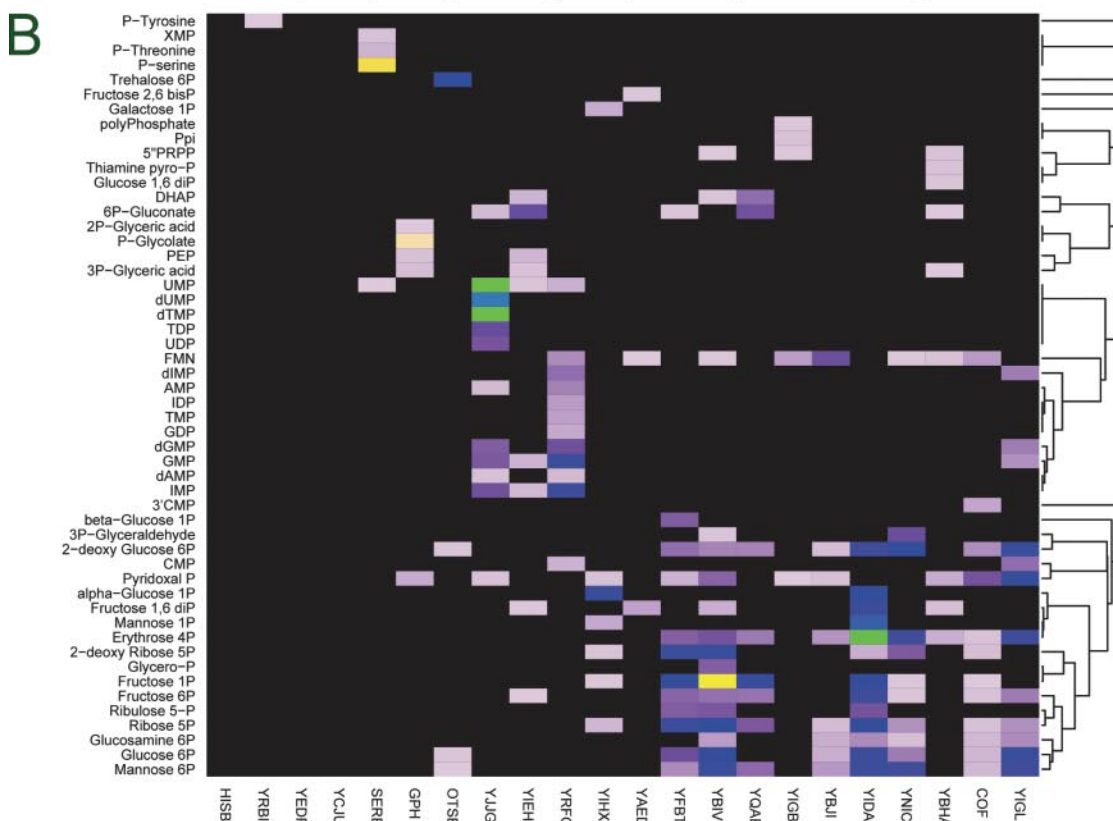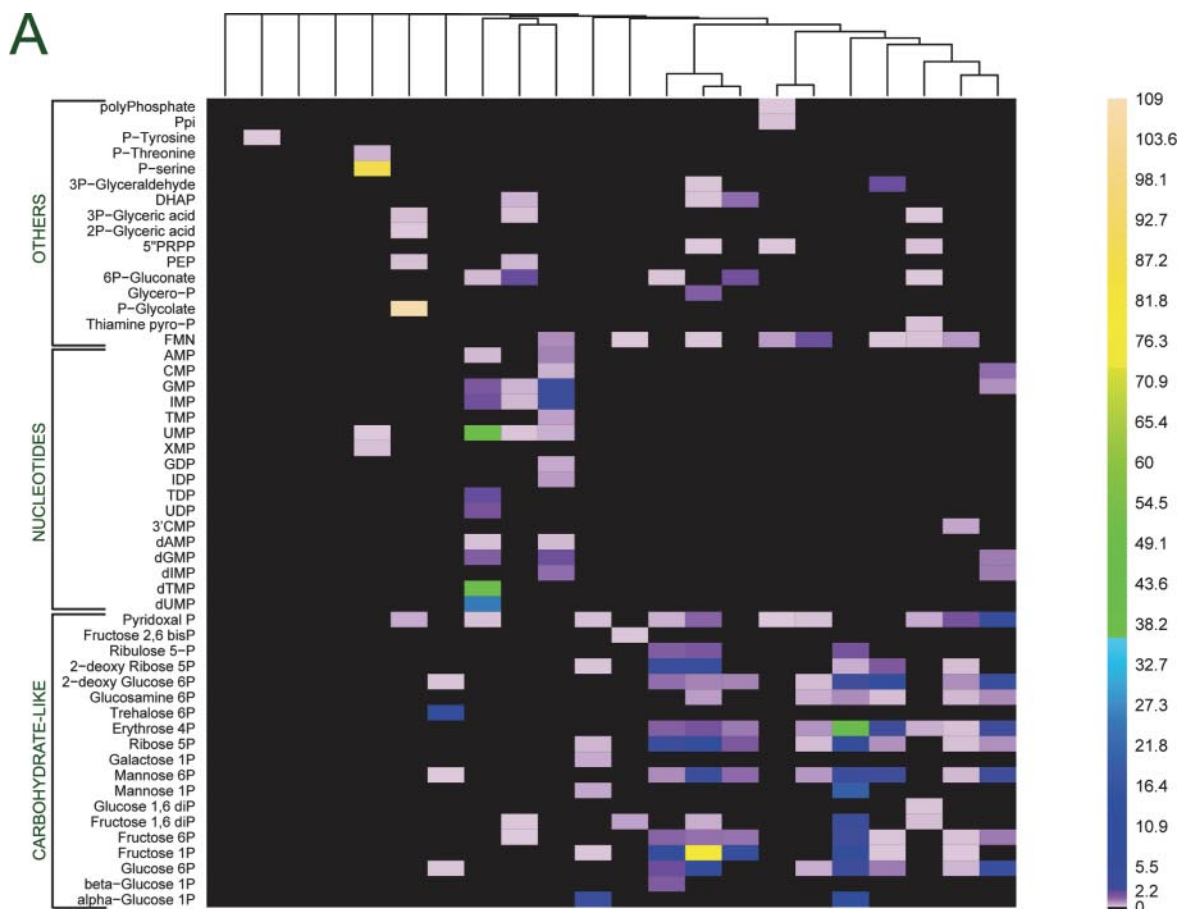
TABLE 2

Kinetic parameters of the *E. coli* HADs with natural substrates

Protein and variable substrate	Metal[a]	K_m	Known intracellular concentrations of the variable substrates[b]	k_{cat}	k_{cat}/K_m
		mM	mM	s^{-1}	$M^{-1} s^{-1}$
HAD1 (YniC)					
2-Deoxyglucose-6-P	Mn^{2+}	0.61 ± 0.05		33 ± 0.8	5.4×10^4
Mannose-6-P	Zn^{2+}	4.7 ± 0.4		11 ± 0.4	2.2×10^3
2-Deoxyribose-5-P	Zn^{2+}	2.5 ± 0.1		9.4 ± 0.1	3.7×10^3
Ribose-5-P	Zn^{2+}	2.6 ± 0.3	0.15–0.63 (61)	2.8 ± 0.1	1.1×10^3
Glucose-6-P	Zn^{2+}	3.6 ± 0.5	0.18–2.0 (55)	25 ± 1.7	6.7×10^3
HAD2 (YfbT)					
Fructose-1-P	Mg^{2+}	1.0 ± 0.2		3.7 ± 0.24	3.5×10^3
Ribose-5-P	Mg^{2+}	2.2 ± 0.2	0.15–0.63 (61)	2.7 ± 0.09	1.2×10^3
Glucose-6-P	Co^{2+}	1.8 ± 0.2	0.18–2.0 (55)	13 ± 0.6	7.1×10^3
Fructose-6-P	Mg^{2+}	1.3 ± 0.2	0.54–0.91 (60)	3.1 ± 0.2	2.3×10^3
β-Glucose-1-P	Mn^{2+}	7.0 ± 0.3		7.5 ± 0.2	1.1×10^3
HAD3 (YieH)					
6-P-gluconate	Mg^{2+}	2.2 ± 0.2	0.1–15 (56)	16 ± 0.6	7.3×10^3
Phosphoenolpyruvate	Mn^{2+}	2.9 ± 0.7	0.15–8.6 (55, 60, 61)	0.75 ± 0.09	2.6×10^2
HAD4 (YihX)					
α-Glucose-1-P	Mg^{2+}	0.24 ± 0.02		1.4 ± 0.1	5.9×10^3
Fructose-1-P	Mn^{2+}	1.6 ± 0.1		0.10 ± 0.004	0.61×10^2
Acetyl-P	Mg^{2+}	3.6 ± 0.7	0.04–1.3 (66)	2.7 ± 0.2	0.73×10^3
Imido-diP	Mg^{2+}	0.028 ± 0.002		0.56 ± 0.02	20×10^3
HAD5 (YjjG)					
UMP	Mg^{2+}	2.4 ± 0.1		25 ± 0.5	1.0×10^4
UMP	Mn^{2+}	1.0 ± 0.0		4.4 ± 0.05	4.3×10^3
dTMP	Mg^{2+}	2.8 ± 0.2		31 ± 1.1	1.1×10^4
dTMP	Mn^{2+}	0.64 ± 0.03		16 ± 0.2	2.5×10^4
AMP	Mn^{2+}	1.8 ± 0.1	0.31–0.54 (55, 60)	1.3 ± 0.03	7.1×10^2
dGMP	Mg^{2+}	2.0 ± 0.19		2.0 ± 0.09	1.0×10^3
dGMP	Mn^{2+}	2.8 ± 0.32		2.1 ± 0.13	7.5×10^2
Fructose-6-P	Mn^{2+}	2.1 ± 0.19	0.54–0.91 (60)	0.18 ± 0.01	0.88×10^2
6-P-gluconate	Mg^{2+}	5.3 ± 0.49	0.1–15.0 (56)	0.88 ± 0.03	1.7×10^2
HAD6 (YqaB)					
Fructose-1-P	Mg^{2+}	1.7 ± 0.2		19.5 ± 1.1	2.0×10^4
6-P-gluconate	Mg^{2+}	3.9 ± 0.5	0.1–15.0 (56)	7.1 ± 0.3	1.8×10^3
HAD7 (YigB)					
FMN	Mg^{2+}	1.00 ± 0.1		9.4 ± 0.6	0.94×10^4
HAD8 (YrfG)					
GMP	Mg^{2+}	1.9 ± 0.2		5.5 ± 0.2	2.9×10^3
IMP	Mg^{2+}	1.2 ± 0.1		2.6 ± 0.1	2.3×10^3
HAD9 (SerB)					
Phosphoserine	Mg^{2+}	0.097 ± 0.006		66 ± 1	68×10^4
Acetyl-P	Mg^{2+}	6.2 ± 1.1	0.04–1.3 (66)	76 ± 6	12×10^3
Imido-di-P	Mg^{2+}	0.07 ± 0.007		0.11 ± 0.01	1.6×10^3
HAD10 (Gph)					
Acetyl-P	Mg^{2+}	8.9 ± 1.7	0.04–1.3 (66)	8.9 ± 0.8	1.0×10^3
Imido-di-P	Mg^{2+}	0.13 ± 0.02		21 ± 2	16×10^4
HAD12 (YbiV)					
Fructose-1-P	Mg^{2+}	1.4 ± 0.3		111 ± 6.4	8.0×10^4
Ribose-5-P	Mg^{2+}	2.4 ± 0.2	0.15–0.63 (61)	28 ± 0.9	1.2×10^4
Glucose-6-P	Mg^{2+}	3.1 ± 0.3	0.18–2.0 (55)	22 ± 0.7	6.9×10^3
Acetyl-P	Mg^{2+}	4.9 ± 0.7	0.04–1.3 (66)	101 ± 4	21×10^3
Imido-di-P	Mg^{2+}	0.12 ± 0.02		7.5 ± 0.6	62×10^3
HAD13 (YidA)					
Erythrose-4-P	Mg^{2+}	0.019 ± 0.001		19 ± 0.4	10×10^5
Mannose-1-P	Mg^{2+}	0.53 ± 0.03		21 ± 0.4	3.8×10^4
α-Glucose-1-P	Mg^{2+}	0.21 ± 0.03		20 ± 1.3	9.4×10^4
Ribose-5-P	Mg^{2+}	0.45 ± 0.05	0.15–0.63 (61)	9.2 ± 0.5	2.1×10^4
Fructose-1-P	Mg^{2+}	0.39 ± 0.03		11 ± 0.6	2.7×10^4
Fructose-6-P	Mg^{2+}	0.44 ± 0.07	0.54–0.91 (60)	5.4 ± 0.5	1.2×10^4
Glucose-6-P	Mg^{2+}	0.81 ± 0.09	0.18–2.0 (55)	11 ± 0.6	1.3×10^4
Acetyl-P	Mg^{2+}	3.9 ± 0.3	0.04–1.3 (66)	24 ± 1	6.2×10^3
Imido-di-P	Mg^{2+}	0.033 ± 0.002		14 ± 0	42×10^4
HAD14 (YbhA)					
PLP	Mg^{2+}	0.37 ± 0.05		1.0 ± 0.04	2.8×10^3
Fructose-1,6-bis-P	Mg^{2+}	2.4 ± 0.2	3.29–6.0 (55, 60, 61)	5.3 ± 0.1	2.2×10^3
HAD15 (YbjI)					
FMN	Mg^{2+}	2.3 ± 0.3		6.5 ± 0.4	2.8×10^3
Erythrose-4-P	Mg^{2+}	6.8 ± 1.3		3.3 ± 0.4	4.9×10^2
HAD16 (YigL)					
PLP	Mg^{2+}	1.5 ± 0.1		12 ± 0.5	7.7×10^3
2-Deoxyglucose-6-P	Mg^{2+}	7.5 ± 0.7		13 ± 0.5	1.8×10^3
β-Glucose-6-P	Mg^{2+}	5.9 ± 0.7		14 ± 0.7	2.4×10^3
HAD17 (OtsB)					
Trehalose-6-P	Mg^{2+}	0.61 ± 0.07	0.76 (62)	9.1 ± 0.5	1.5×10^4
HAD18 (Cof)					
PLP	Mg^{2+}	0.68 ± 0.07		0.58 ± 0.02	8.5×10^2
2-Deoxyglucose-6-P	Mg^{2+}	2.5 ± 0.3		0.50 ± 0.02	2.0×10^2
HAD20 (YaeD)					
Fructose-1,6-bis-P	Mg^{2+}	0.42 ± 0.06	3.3–6.0 (55, 60, 61)	0.19 ± 0.01	4.5×10^2

[a] The saturating metal concentrations used in the assays were 0.63–5.0 mM Mg^{2+}, 0.1–1.25 mM Mn^{2+}, 0.3–0.75 mM Co^{2+}, or 1 mM Zn^{2+}.

[b] Experimentally determined intracellular (*in vivo*) concentrations of the variable substrates. References are given in parentheses.

Analysis of the kinetic parameters of the *E. coli* HADs (Table 2) revealed that with saturating substrates, most HADs showed the same substrate preferences (k_{cat}/K_m) that were observed in substrate profiles (presented as phosphatase activities) with subsaturating substrate concentrations (Fig. 1). Moreover, for YfbT (HAD2), YjjG (HAD5), YrfG (HAD8), YbiV (HAD12), YbhA (HAD14), and YigL (HAD16), the level of substrate discrimination is quite low, given that k_{cat}/K_m values for several of their substrates agree within an order of magnitude. Comparison of the substrate profiles and kinetic parameters of various HADs also revealed that some phosphorylated metabolites can be hydrolyzed by several HADs with significant or even comparable catalytic efficiency. For example, ribose-5-P and glucose-6-P were substrates for YbiV (HAD12), YidA (HAD13), YniC (HAD1), and YfbT (HAD2), whereas fructose-1-P can be dephosphorylated by YfbT (HAD2), YihX (HAD4), YqaB (HAD6), YbiV (HAD12), and YidA (HAD13) (Table 2, Fig. 1). Although specificity is considered a hallmark of enzymatic activity and certain enzymes can be extraordinarily specific, there has been growing appreciation that substrate specificities are perhaps broader than is generally accepted, and many enzymes exhibit considerable catalytic and substrate promiscuity (34–36). Catalytic promiscuity (also called polyreactivity or moon-lighting activity) is defined as the ability of enzyme active sites to catalyze distinctly different chemical transformations (different types of bonds cleaved or formed or different catalytic mechanisms of bond making or breaking) (37, 38). Substrate promiscuity (also called substrate ambiguity or cross-reactivity) is defined as the ability of enzymes to catalyze one chemical transformation on several structurally related substrates. It is a generally accepted hypothesis that promiscuous activities serve as a starting point in the evolution of new enzymes and create the basis for the emergence of structurally and mechanistically related protein superfamilies (34, 35, 39–41). This theory is supported by the analysis of enzymes from enolase, amidohydrolase, thiyl radical, and crotonase superfamilies (40, 42, 43). Our work with the *E. coli* HADs demonstrated that these enzymes are metal-dependent small molecule phosphatases with various degrees of substrate promiscuity and overlapping substrate specificities.

Metabolites Hydrolyzed by the E. coli HADs—*E. coli* HADs hydrolyze a wide range of phosphorylated metabolites, including carbohydrates, nucleotides, organic acids, and coenzymes. Hierarchical clustering of the HAD phosphatase activities (observed in the screens) against 52 substrates showed that phosphorylated carbohydrates represent the major group of the HAD substrates (Fig. 2A). The most common substrates for these enzymes were fructose-1-P, glucose-6-P, mannose-6-P, 2-deoxyglucose-6-P, fructose-6-P, ribose-5-P, and erythrose-4-P (Fig. 2A). PLP and FMN also appear to be common substrates of HADs. In addition, hierarchical clustering recognized a group of three nucleotidases, YrfG (HAD8), YjjG (HAD5), and

YieH (HAD3). YrfG preferentially hydrolyzed purine nucleotides (GMP and IMP), and YjjG preferred pyrimidines (UMP, dUMP, and dTMP), whereas YieH hydrolyzed both purines and pyrimidines as secondary substrates. The fourth nucleotidase, NagD (HAD23) (which was only partially refolded and, therefore, was not included in the clustering), had an unusually broad substrate range and hydrolyzed deoxyribo- and ribonucleoside tri-, di-, and monophosphates, as well as polyphosphate and glucose-1-P (Fig. 1). Nucleotidase activity of *E. coli* NagD has been recently described by others, and a broad substrate range of this enzyme, with the highest catalytic efficiency toward nucleoside monophosphates, has been demonstrated (12).

Hierarchical clustering of 52 metabolites based on their ability to serve as substrates for the HADs separated these compounds into distinct groups of structurally related molecules (Fig. 2B). The dendrogram in Fig. 2B, which was based solely on the similarity of the enzyme profiles, grouped together hexoses (mannose-6-P, glucose-6-P, and glucosamine-6-P), pentoses (ribose-5-P and ribulose-5-P), purine nucleotides (GMP, dGMP, dAMP, and IMP), pyrimidine nucleotides (UMP, dUMP, dTMP, TDP, and UDP), pyrophosphate-containing metabolites (thiamine pyrophosphate, 5-phosphoribosyl-l-pyrophosphate, PP_i, and polyphosphate), and amino acids (P-serine and P-threonine). This indicates that structurally related substrates are recognized by similar sets of HADs, demonstrating a direct relationship between the chemical structure of HAD substrates and their biological activity (HAD spectra).

The metabolites hydrolyzed by the *E. coli* HADs are intermediates of various metabolic pathways and reactions (glycolysis, pentose phosphate pathway, gluconeogenesis, and intermediary sugar and nucleotide metabolism). YbiV (HAD12), YfbT (HAD2), YqaB (HAD6), and YidA (HAD13) showed high phosphatase activity with fructose-1-P (Table 2, Fig. 1). In *E. coli*, this metabolite is produced by the PEP-dependent phosphotransferase system PtsI, PtsH, FruA, and FruB during transport of extracellular fructose across the cytoplasmic membrane (44, 45). At the next step, 1-phosphofructokinase FruK (EC 2.7.1.56) phosphorylates fructose-1-P, yielding the glycolytic intermediate fructose 1,6-bis-P (46). Analysis of the kinetic parameters of four HADs demonstrated that YidA (HAD13) had the lowest K_m (highest affinity) to this substrate ($K_m = 0.39 \pm 0.03$ mM), whereas YbiV (HAD12) had the highest catalytic activity and efficiency (Table 2). YrfG (HAD8) preferentially hydrolyzed purine nucleotides (GMP, IMP), whereas YjjG (HAD5) preferred pyrimidines (UMP, dUMP, and dTMP). The observed substrate affinities of YrfG and YjjG ($K_m = 0.64$–2.44 mM; Table 2) were within the range reported for other characterized nucleotidases (0.01–22 mM; BRENDA data base). NagD (HAD23) preferentially hydrolyzed purine nucleotides (IMP, UMP, and GMP) (12) (Fig. 1). For YieH (HAD3), Cof (HAD18), and YigL (HAD16) nucleotides were identified as secondary substrates (Fig. 1). Thus, like eukaryotes, *E. coli* has multiple

FIGURE 2. **Hierarchical clustering of 22 *E. coli* HADs (*A*) and 52 HAD substrates (*B*) by using the cosine similarity measurement.** *A*, the horizontal *x* axis clusters HADs into groups based on the similarity of their substrate profiles (phosphatase activities from Fig. 1) using normalized activity levels (see details under "Experimental Procedures"). *B*, the vertical *y* axis clusters substrates based on the similarity of HAD profiles between substrates using normalized activity levels. For both graphs, the similarities between objects were calculated using cosine correlations, and the groups were clustered using the average method. The activity levels (μmol/min/mg of protein) in both figures were color-coded as shown in *A* (*black* indicates zero activity).

E. coli Haloacid Dehalogenase-like Phosphatases

intracellular nucleotidases with different substrate specificities, namely YrfG, YjjG, NagD, Cof, YieH, and the previously characterized SurE and YfbR (47). These *E. coli* nucleotidases can dephosphorylate a wide range of 2′-, 3′-, and 5′-nucleotides and, together with nucleoside kinases (Gsk, Tdk, and Udk), can establish regulatory substrate cycles controlling the nucleotide levels in the bacterial cell, similar to those described in mammals (48).

Several HADs catalyzed the dephosphorylation of two important coenzymes, pyridoxal phosphate (PLP) and FMN. PLP was the preferred substrate for *E. coli* Cof (HAD18) and YbhA (HAD14), and several other HADs (YbiV (HAD12), YihX (HAD4), YbjI (HAD15), and Gph (HAD10)) showed significant phosphatase activity with this molecule (0.2–1.44 μmol of P$_i$/min/mg of protein) (Table 2, Fig. 1). In mammalian cells, the intracellular level of PLP is regulated, to a large extent, by the PLP phosphatase (49–51), which is also an enzyme of the HAD superfamily and has K_m and $V_{\max}$ for this substrate similar to those determined here for *E. coli* HADs.

Four *E. coli* HADs (YbjI (HAD15), YigB (HAD7), Cof (HAD18), and YrfG (HAD8)) were capable of dephosphorylating FMN, an activity without EC number assigned (Table 2, Fig. 1). The KEGG data base indicates that in *E. coli*, three acid phosphatases are involved in the dephosphorylation of FMN: AppA, SurE, and AphA. However, two of them (AppA and AphA) are periplasmic nonspecific phosphatases and, therefore, not likely to be involved in the control of the intracellular FMN level, whereas no FMN dephosphorylating activity has been found in SurE in our recent work (47). We have found that FMN was a preferred substrate for two HADs, YbjI (HAD15) and YigB (HAD7) (Table 2, Fig. 1), and they represent more likely candidates for the FMN dephosphorylation function in *E. coli* cells. Generally, we suggest that the activity of several HADs could be important in the regulation of the intracellular levels of PLP and FMN in *E. coli*.

We identified four HADs (YaeD (HAD20), YbhA (HAD14), YidA (HAD13), and YbiV (HAD12)) with fructose 1,6-bis-P dephosphorylation activity (Table 2, Fig. 1). Fructose 1,6-bisphosphatase (EC 3.1.3.11) catalyzes the second irreversible step in gluconeogenesis producing fructose-6-P that is further converted into glucose-6-P by phosphoglucose isomerase Pgi (52). Two fructose 1,6-bisphosphatases have been characterized in *E. coli*, namely Fbp and GlpX, both members of the inositol polyphosphate 1-phosphatase-like superfamily (53, 54). Like Fbp and GlpX, YaeD (HAD20) and YbhA (HAD14) hydrolyzed this metabolite to yield fructose-6-P (data not shown). By contrast, YidA (HAD13) and YbiV (HAD12) yielded fructose as the final product of fructose 1,6-bis-P hydrolysis (data not shown), which is compatible with their ability to dephosphorylate both fructose 1- and 6-phosphates (Fig. 1). YaeD (HAD20) showed higher affinity to fructose 1,6-bis-P than YbhA (HAD14), whereas the latter had higher activity and catalytic efficiency (k_{cat}/K_m) with this substrate (Table 2). Given that the reported intracellular concentration of fructose 1,6-bis-P in *E. coli* (3–6 mM) (55–57) is significantly higher than the K_m of YaeD (HAD20) and YbhA (HAD14) (Table 2), the phosphatase activity of both proteins could affect the *in vivo* pool of this crucial metabolite.

Three *E. coli* HADs showed high phosphatase activity with α-D-glucose-1-P (YidA (HAD13) and YihX (HAD4)) or β-D-glucose-1-P (YfbT (HAD2)) (Table 2, Fig. 1). These proteins were strictly specific to one anomer of glucose-1-P (α or β) and did not hydrolyze the other anomer. In *E. coli*, the α-form is produced from glycogen by glycogen phosphorylase GlgP or in other reactions of pentose, glucuronate, or nucleotide sugar metabolism, whereas β-D-glucose-1-P might be produced by yet unidentified maltose or trehalose phosphorylases (58). Both anomers can be used in various biosynthetic reactions or channeled to glycolysis or the Entner-Doudoroff pathway through conversion into the respective α- and β-glucose 6-phosphates by specific phosphoglucomutases, α-phosphoglucomutase Pgm (EcoCyc data base) or β-phosphoglucomutase YcjU (this work). *E. coli* also makes the α-glucose 1-phosphatase Agp, a periplasmic enzyme that acts primarily as a glucose scavenger (59). Although, in the KEGG data base, this periplasmic phosphatase is annotated as an enzyme responsible for intracellular dephosphorylation of α-glucose-1-P, our results indicate that both YihX (HAD4) and YidA (HAD13) have at least 2 times lower K_m (higher affinity) to this substrate, and therefore, these enzymes are likely to be the principal phosphatases involved in the intracellular metabolism of α-glucose-1-P in *E. coli*. YfbT (HAD2) is the first phosphatase found to selectively hydrolyze β-D-glucose-1-P and showing no activity with the α-form of this substrate.

Finally, many HADs demonstrated phosphatase activities against various intermediates of two central metabolic pathways, glycolysis (gluconeogenesis) and the pentose phosphate pathway (glucose-6-P, fructose-6-P, dihydroxyacetone phosphate, 3-phosphoglyceraldehyde, 3-phosphoglycerate, 2-phosphoglycerate, PEP, 6-P-gluconate, erythrose-4-P, ribose-5-P, ribulose-5-P) (Fig. 1). In several cases, *E. coli* HADs showed high affinities with these substrates (Table 2), comparable with their known intracellular concentrations. The K_m values of YidA (HAD13) and YfbT (HAD2) for glucose-6-P, fructose-6-P, and ribose-5-P (0.44–2.23 mM; Table 2) were comparable with the known intracellular concentrations of these metabolites (0.18–2.0 mM glucose-6-P, 0.54–0.91 mM fructose-6-P, and 0.15–0.63 mM ribose-5-P) in *E. coli* (55, 60, 61). The K_m of YieH (HAD3) to PEP (2.89 mM) and of OtsB (HAD17) to trehalose-6-P (0.61 mM) were also within the range of the known physiological concentrations of these metabolites (0.15–8.63 mM for PEP and 0.76 mM for trehalose-6-P) (55, 60–62). Therefore, the *in vivo* activities of HADs could potentially modulate the metabolite pools and metabolic fluxes through a broad variety of pathways.

HADs Hydrolyze Acetyl Phosphate and Other Small Phosphodonors—The promiscuity of the HAD phosphatase family raises interesting questions regarding the evolution of the activities and the emergence of new activities from ancestral ones. One clue into the evolution of the HAD family comes from structural analysis of phosphoserine phosphatase from *M. jannaschii*, which showed significant structural similarity to CheY, a receiver domain/response regulator of the two-component signal transduction system CheA-CheY (63). In two-component systems, the receiver domains of response regulator proteins are either phosphorylated by their cognate sensor his-

tidine kinases or autophosphorylate using acetyl phosphate (acetyl-P), carbamoyl phosphate (carbamoyl-P), or phosphoramidate as phosphodonors (64, 65). The structural similarity and the inferred common ancestry of response regulators and the HADs prompted us to test if any of the *E. coli* HADs were able to hydrolyze small phosphodonors. All purified *E. coli* HADs, including even YcjU (HAD11), a β-phosphoglucomutase, showed various levels of Mg^{2+}-dependent activity toward these substrates (imidodiphosphate was used as a phosphoramidate substrate) (Table 1). Most proteins showed higher activities (≥ 1 μmol/min/mg of protein) with acetyl-P and lower activities with carbamoyl-P or imido-di-P, but YidA (HAD13), YbiV (HAD12), and Gph (HAD10) had high activity with all three substrates (Table 1). To investigate whether these small phosphodonors are specific substrates for HAD phosphatases, we purified seven non-HAD phosphatases from *E. coli* (CCA, YfdR, YcdX, SurE, CysQ, GlpX, and YfbR) and eight HADs from other organisms (Cof, YigL, and YrfG from *Salmonella typhimurium*; PA0006, PA0065, and PA3172 from *Pseudomonas aeruginosa*; RPA3639 and RPA4337 from *Rhodopseudomonas palustris*) and assayed them for phosphatase activity against these compounds. Although all of these proteins showed high phosphatase activity against *p*NPP, only HAD phosphatases were active toward small phosphodonors. Thus, the phosphohydrolase activity toward acetyl-P, carbamoyl-P, and imido-di-P is a specific property of the HADs.

With acetyl-P or imido-di-P as substrates, YidA (HAD13), YihX (HAD4), YbiV (HAD12), SerB (HAD9), and Gph (HAD10) demonstrated classical Michaelis-Menten kinetics with higher activity toward acetyl-P (except for Gph) and lower K_m to imido-di-P (Table 2). With carbamoyl-P, all tested HADs (YidA, YbiV, YqaB, and Gph) showed no saturation, indicating that substrate binding was the limiting step in the hydrolysis of this substrate. The K_m values of the *E. coli* HADs with acetyl-P (3.64–8.93 mM) were substantially higher than the reported intracellular concentrations of acetyl-P (0.04–1.3 mM) (66). Therefore, the physiological role of the acetyl phosphatase activity of these enzymes remains uncertain. The ability of the *E. coli* HADs to hydrolyze imido-di-P (a phosphoramidate-like substrate, N–P bond hydrolysis) is a previously unknown catalytic activity of these enzymes. This reaction yields free NH_4^+ (data not shown), indicating that HADs also hydrolyze phosphoramidate, the immediate product of imido-di-P hydrolysis. Very recently, the first microbial phosphoramidase, the *E. coli* HinT, was identified and characterized, but its physiological function remains unknown (67). Eight more *E. coli* enzymes can now be annotated with this activity: YieH (HAD3), YjjG (HAD5), Gph (HAD10), YbiV (HAD12), YidA (HAD13), YbhA (HAD14), YigL (HAD16), and YaeD (HAD20).

In Vivo Validation of an in Vitro Activity—To determine whether any of the *in vitro* activities observed in our experiments represent the corresponding biologically relevant reactions, we performed genetic and physiological experiments for a representative *in vitro* phosphatase activity against 2-deoxyglucose-6-P observed in several *E. coli* HADs. 2-deoxyglucose-6-P is a toxic analogue of glucose-6-P, and its intracellular concentration can be very high (up to 100 mM) (68). YniC (HAD1) showed higher catalytic efficiency (k_{cat}/K_m) and affinity ($K_m =$
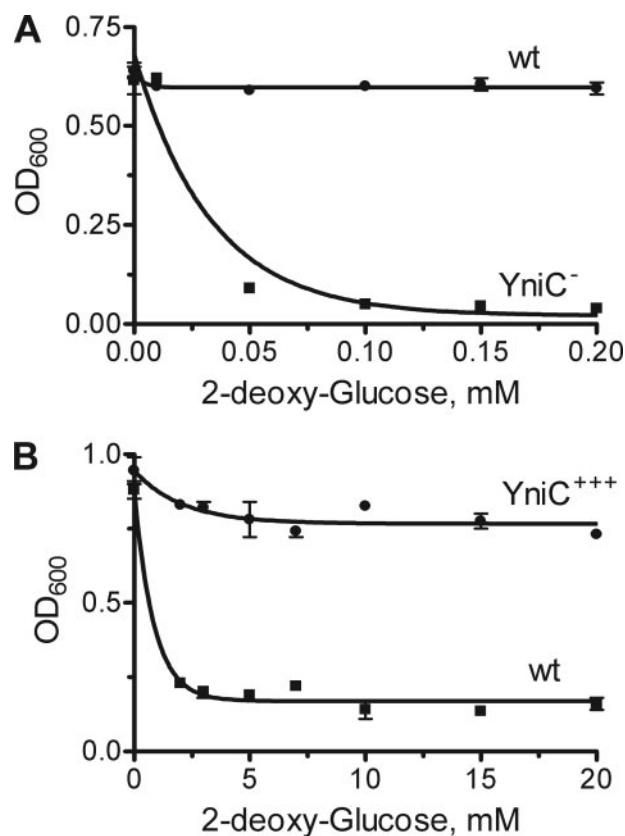

FIGURE 3. **Effect of *yniC* deletion (*A*) or overexpression (*B*) on the growth of the *E. coli* strains.** The wild-type (*wt*), YniC deletion (*YniC$^-$*), or YniC overexpressing (*YniC^{+++}*) strains were grown under aerobic conditions on a minimal medium with 0.2% succinate, 0.02% arabinose (an inducer), and various concentrations of 2-deoxyglucose (0–20 mM). 25-ml culture tubes containing 5 ml of medium were inoculated with 0.1 ml of the overnight culture and incubated in a shaker (200 strokes/min) at 37 °C. After 14 h, the growth was determined by measuring the absorbance at 600 nm.

0.61 mM) to this substrate than YigL (HAD16) and Cof (HAD18) (Table 2), suggesting that *in vivo* YniC might be the principal phosphatase responsible for the hydrolysis of this compound. It has been shown previously that 2-deoxyglucose is taken up by *E. coli* and phosphorylated to yield 2-deoxyglucose-6-P (69). The *E. coli yniC* deletion mutant was much more sensitive to the presence of 2-deoxyglucose in the growth medium ($IC_{50} = 0.02$ mM) than the wild-type strain ($IC_{50} = 0.59$ mM) (Fig. 3). By contrast, the YniC-overproducing strain grew well even in the presence of 20 mM 2-deoxyglucose (Fig. 3). These results indicate that the 2-deoxyglucose 6-phosphatase activity of YniC (HAD1) plays an important role in the resistance of *E. coli* to 2-deoxyglucose and suggest that the *in vitro* activities of HADs toward their preferred and at least some of the secondary substrates reflect their functionally relevant *in vivo* activities.

Substrate Specificities, Sequence Similarity, and Evolutionary Relationships among HADs—The availability of substrate profiles for the full complement of soluble *E. coli* HADs and the kinetic parameters for most of them allowed us to examine the relationship between protein evolution within a large family of enzymes and evolution of their substrate preferences. To this end, we first constructed sequence-based phylogenetic trees of *E. coli* HADs using five methods, namely neighbor-joining,
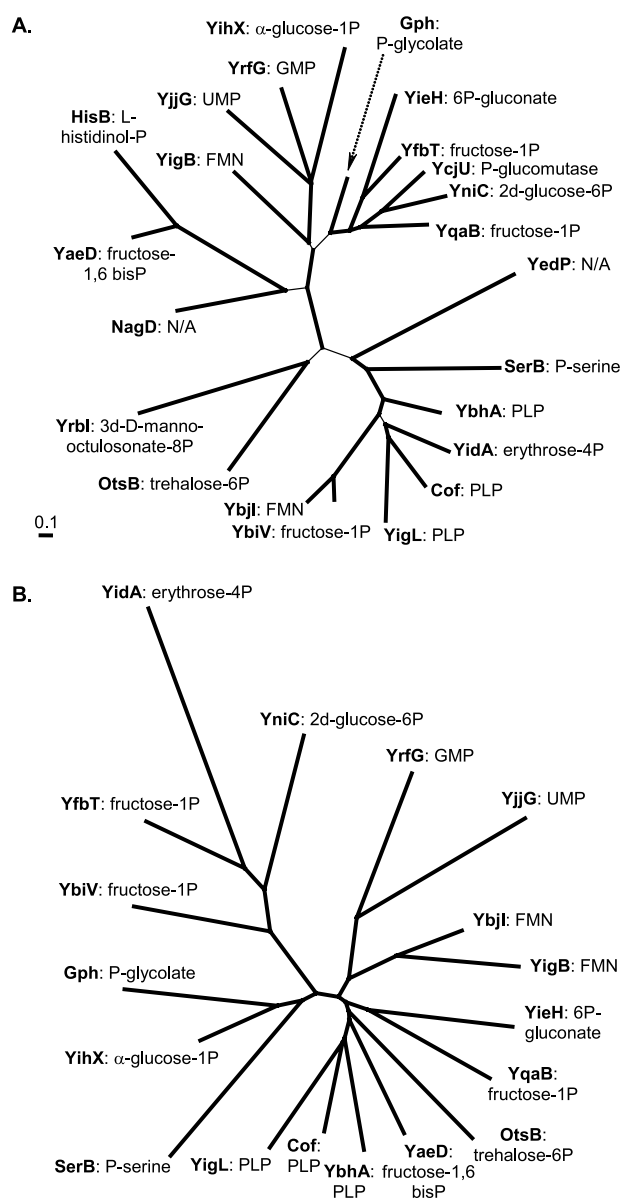
A.

B.

FIGURE 4. **Sequence-based (*A*) and activity-based (*B*) phylogenetic trees of the *E. coli* HADs.** For each protein, the preferred substrates are shown. *A*, a sequence-based phylogenetic tree created using the maximum likelihood (local optimization of minimum evolution) method. Branches that are common between the phylogenetic trees reconstructed with different methods (see "Experimental Procedures") are indicated with *thick lines*. *B*, a neighbor-joining tree constructed from the matrix of quadratic distances between the catalytic efficiencies (k_{cat}/K_m) of HADs (Table 2).

least squares, maximum likelihood obtained by local optimization of a minimum evolution tree, quartet puzzling, and Bayesian inference (see "Experimental Procedures"). These methods yielded very similar phylogenetic trees; the maximum likelihood tree is shown on Fig. 4*A*. Most of the HADs fell into four well supported clusters of paralogs, which can be designated the YieH (HAD3), YbhA (HAD14), HisB, and YihX (HAD4) subfamilies after their representative members, and several HADs (*e.g.* OtsB or YrbI) whose phylogenetic affinities were less certain (Fig. 4*A*). Notably, however, each of the subfamilies includes at least one sugar phosphatase, suggesting that this specificity might be ancestral in the HAD superfamily.

In addition to the traditional, sequence-based phylogenetic tree (Fig. 4*A*), we constructed a neighbor-joining tree on the basis of the matrix of quadratic distances between the catalytic efficiencies (k_{cat}/K_m) of the HADs (Fig. 4*B*). In general, although the sequence-based trees reconstructed using different methods show remarkable agreement, the topology of the activity-based tree showed no congruence with that of the sequence-based phylogenetic trees (Fig. 4, compare *A* and *B*). This lack of compatibility of the tree topologies suggests that many of the secondary substrate specificities could be of lesser functional importance and might evolve (largely) neutrally.

The present work might offer insights into even more ancient history of the HAD hydrolases and their evolutionary progenitors. We showed that small intracellular phosphodonors (acetyl-P, carbamoyl-P, imido-di-P) are universal substrates for all *E. coli* HADs (as well as for HADs from other organisms) but not for any of the other phosphatases we were able to test (Table 1). This functional similarity meshes well with the monophyly of the HADs and the receiver domain superfamily, which together form a distinct branch in the evolutionary tree of the Rossmann fold domains (70), because the receiver domains can autophosphorylate using the same small phosphodonors. Conceivably, the common ancestor of the HADs and the receiver domains was a broad specificity phosphatase that could use as substrates, at least, small phosphodonors and various phosphorylated sugars (Fig. 5). Subsequently, after duplication, the evolving HAD and receiver domain families apparently retained different subsets of the multiple functions of the ancestor (71, 72) (subfunctionalization), which, in some cases, was followed by neofunctionalization (*i.e.* emergence of new specificities in some of the paralogs) (73). Thus, the receiver domains retained the ability to hydrolyze small phosphodonors and evolved the activity against another N–P bond that occurs in the phosphohistidine of phosphorylated histidine kinases (Fig. 5). In contrast, the HADs expanded, through a series of duplications, the repertoire of phosphatase activities against a broad variety of metabolites, as demonstrated by their broad substrate profiles (Fig. 5). Conceivably, the more specialized phosphatases like OtsB, SerB, or Gph were locked into their narrow specificities at a later stage of evolution. Dehalogenase, phosphonatase, and protein phosphatase activities observed in several HADs from other organisms (7, 9, 10, 74), but not found in *E. coli*, represent further substrate specialization of the HADs. The present findings on the substrate profiles of the HAD superfamily phosphatases emphasize the enormous and still largely unexplored versatility of enzymes that can evolve within the same structural scaffold through duplication, subfunctionalization, and neofunctionalization (75).

DISCUSSION

Our results demonstrate that HADs represent the largest family of small molecule phosphatases in *E. coli*. 12 phosphatase activities catalyzed by the *E. coli* HADs have no EC number assigned (supplemental Table 3). Two HADs are involved in lipopolysaccharide biosynthesis: YrbI (3-deoxy-D-manno-octulosonate 8-phosphatase) and YaeD (GmhB, DD-heptose 1,7-bisphosphatase) (15, 76). The phosphoglycolate phosphatase Gph has been recently shown to be involved in the degradation
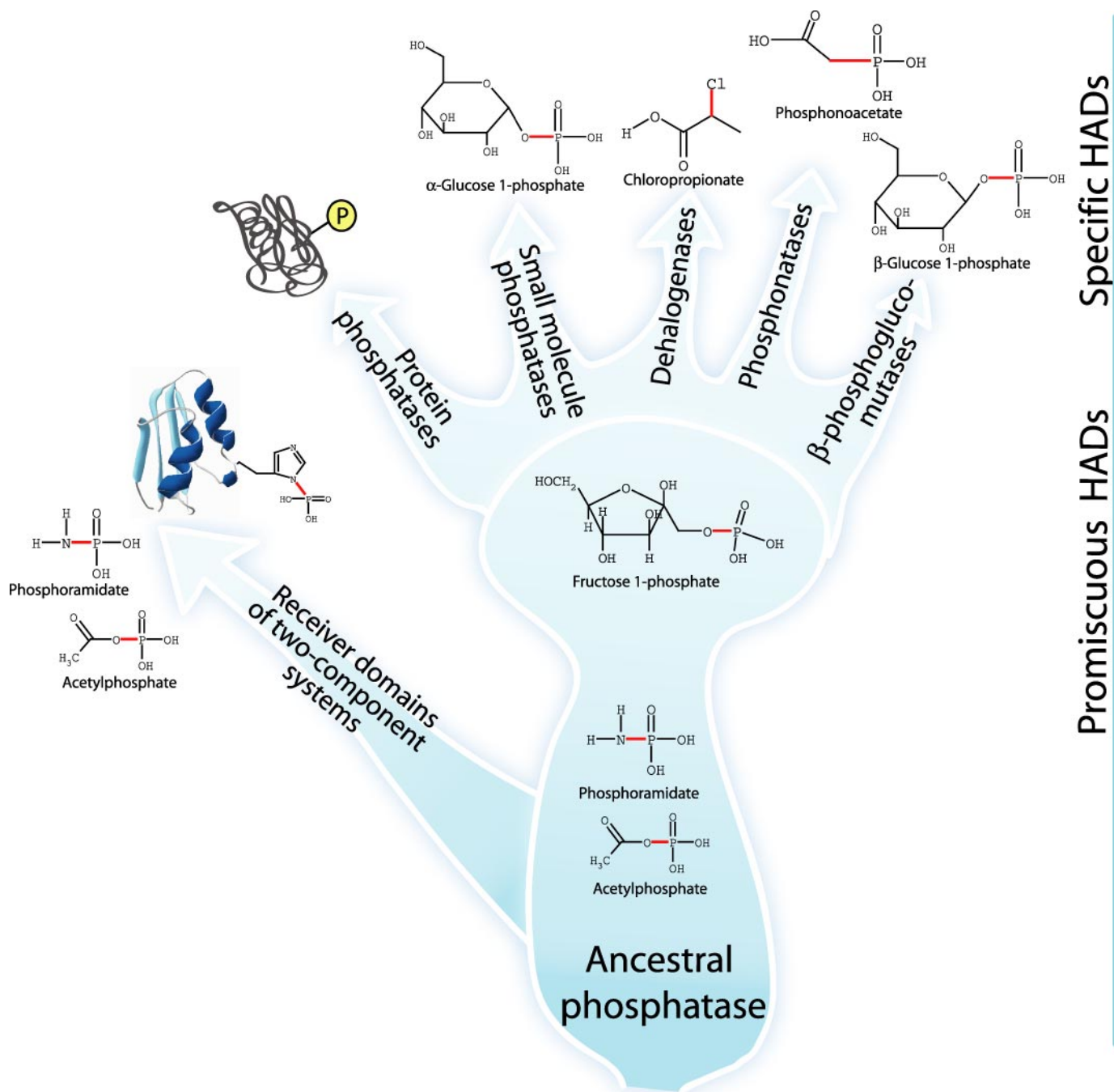
FIGURE 5. **Proposed evolutionary history of the modern HADs and receiver domains of two-component regulatory systems.** Hydrolyzable bonds in the HAD substrates are shown in *red*.

of 2-phosphoglycolate produced during DNA repair (14). Database annotations suggested that SerB (HAD9) is a phosphoserine phosphatase involved in the biosynthesis of serine, whereas OtsB (HAD17) is a trehalose 6-phosphatase that plays an important role in the resistance of *E. coli* to various shocks (heat, cold, osmotic, oxygen) (1, 77). The present work directly shows that these proteins, indeed, possess the respective phosphatase activities (Fig. 1, Table 2). Additionally, we demonstrate that 2-deoxyglucose 6-phosphatase activity of YniC (HAD1) plays an important role in the resistance of *E. coli* cells to 2-deoxyglucose (Fig. 3). Similar detoxification functions can be suggested for other promiscuous *E. coli* HADs (YfbT, YbiV,

YidA, YjjG, YihX, and YigL). Recently, the human HAD phosphatase MDP-1 has been proposed to be involved in glycation repair to free proteins from the glycation products derived from direct nonenzymatic glycosylation (glycation) of proteins by aldose phosphates (glucose-6-P, ribose-5-P, and erythrose-4-P) (78). This enzyme dephosphorylates lysozyme glycated with glucose-6-P, converting it to a substrate for fructosamine-3-kinase, the next enzyme of the glycation repair cascade (78). Nonenzymatic glycation of proteins and chromosomal DNA has been demonstrated in *E. coli* (79, 80). Therefore, we can suggest that several *E. coli* HAD phosphatases (YniC (HAD1), YfbT (HAD2), YbiV (HAD12), YidA (HAD13), and YigL

(HAD16)) might be also involved in glycation repair by direct dephosphorylation of phosphoglycated proteins or DNA or by preventing the intracellular concentrations of the phosphorylated aldoses from reaching deleterious levels.

One of the principal findings of this work is that nearly all HADs can hydrolyze a wide range of phosphorylated molecules present in bacterial cells and that some of these reactions are predicted to be of physiological significance, based on known intracellular substrate concentrations and the enzyme properties that we determined. The recent, complementary work by Thornton and co-workers (27) on ligand selectivity and competition between enzymes *in silico* suggests that cellular metabolites also have significant promiscuity in the recognition of enzymes. Thus, the specificity of the enzyme-substrate recognition *in vivo* is, in general, achieved by a combination of enzyme and substrate selectivity. The broad occurrence of natural substrate promiscuity in different classes of enzymes (phosphatases, acetyltransferases, kinases, epimerases, esterases, and amidohydrolases) strongly suggests that this feature is important *in vivo*. Thus, systems biologists must take this property of enzyme/substrate interactions into account when modeling cellular metabolic pathways and networks.

A fundamental question is whether the secondary activities of HADs (and other promiscuous enzymes) are biologically relevant and maintained by selection or represent functionally neutral variations. The lack of congruence between the phylogenetic tree of the HADs and the clustering of their activities (k_{cat}/K_m or substrate profiles) (Fig. 4) seems to be most compatible with the latter possibility. However, even if most of the secondary specificities are not directly maintained by selection, the breadth of the substrate spectra of HADs might be crucial in evolutionary terms, since they supply the pool from which new specificities could evolve. The ability of HADs to hydrolyze different types of bonds and their substrate promiscuity described here might have contributed to the evolution of new enzymatic specificities involved in the regulation of bacterial metabolism and environmental response.

Note Added in Proof—While this manuscript was under review, Burroughs *et al.* (Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D., and Aravind, L. (2006) *J. Mol. Biol.* **361**, 1003–1034) reported the results of the sequence- and structure-based analysis of the entire HAD superfamily. These authors presented a comprehensive classification of the HAD superfamily based on their sequences, structures, phyletic distribution patterns, and phylogenetic tree analysis. Consistent with our model, the authors suggested that the diversification of the substrate specificities of the primordial HAD hydrolases provided the raw evolutionary material for the development of new enzymes. Additionally, we want to mention two more works, which also pointed out the structural similarity between HADs and the receiver domain of two-component regulatory systems (Ridder, I. S., and Dijkstra, B. W. (1999) *Biochem. J.* **339**, 223–226; Meng, E. C., Polacco, B. J., and Babbitt, P. C. (2004) *Proteins* **55**, 962–976).

REFERENCES

1. Keseler, I. M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I. T., Peralta-Gil, M., and Karp, P. D. (2005) *Nucleic Acids Res.* **33**, D334–D337
2. Nobeli, I., Ponstingl, H., Krissinel, E. B., and Thornton, J. M. (2003) *J. Mol. Biol.* **334**, 697–719
3. Koonin, E. V., and Tatusov, R. L. (1994) *J. Mol. Biol.* **244**, 125–132
4. Allen, K. N., and Dunaway-Mariano, D. (2004) *Trends Biochem. Sci.* **29**, 495–503
5. Wang, W., Kim, R., Jancarik, J., Yokota, H., and Kim, S. H. (2001) *Structure (Camb.)* **9**, 65–71
6. Kim, Y., Yakunin, A. F., Kuznetsova, E., Xu, X., Pennycooke, M., Gu, J., Cheung, F., Proudfoot, M., Arrowsmith, C. H., Joachimiak, A., Edwards, A. M., and Christendat, D. (2004) *J. Biol. Chem.* **279**, 517–526
7. Morais, M. C., Zhang, W., Baker, A. S., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2000) *Biochemistry* **39**, 10385–10396
8. Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2002) *Biochemistry* **41**, 8351–8359
9. Hisano, T., Hata, Y., Fujii, T., Liu, J. Q., Kurihara, T., Esaki, N., and Soda, K. (1996) *J. Biol. Chem.* **271**, 20322–20330
10. Ridder, I. S., Rozeboom, H. J., Kalk, K. H., Janssen, D. B., and Dijkstra, B. W. (1997) *J. Biol. Chem.* **272**, 33015–33022
11. Roberts, A., Lee, S. Y., McCullagh, E., Silversmith, R. E., and Wemmer, D. E. (2005) *Proteins* **58**, 790–801
12. Tremblay, L. W., Dunaway-Mariano, D., and Allen, K. N. (2006) *Biochemistry* **45**, 1183–1193
13. Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balazsi, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I., Gelfand, M. S., Bhattacharya, A., Kapatral, V., D'Souza, M., Baev, M. V., Grechkin, Y., Mseeh, F., Fonstein, M. Y., Overbeek, R., Barabasi, A. L., Oltvai, Z. N., and Osterman, A. L. (2003) *J. Bacteriol.* **185**, 5673–5684
14. Pellicer, T. M., Felisa Nunez, M., Aguilar, J., Badia, J., and Baldoma, L. (2003) *J. Bacteriol.* **185**, 5815–5821
15. Wu, J., and Woodard, R. W. (2003) *J. Biol. Chem.* **278**, 18117–18123
16. Strom, A. R., and Kaasen, I. (1993) *Mol. Microbiol.* **8**, 205–210
17. Zhang, R. G., Skarina, T., Katz, J. E., Beasley, S., Khachatryan, A., Vyas, S., Arrowsmith, C. H., Clarke, S., Edwards, A., Joachimiak, A., and Savchenko, A. (2001) *Structure (Camb.)* **9**, 1095–1106
18. Kuznetsova, E., Proudfoot, M., Sanders, S. A., Reinking, J., Savchenko, A., Arrowsmith, C. H., Edwards, A. M., and Yakunin, A. F. (2005) *FEMS Microbiol. Rev.* **29**, 263–279
19. Lipmann, F., and Tuttle, L. C. (1950) *Biochim. Biophys. Acta* **4**, 301–309
20. Ameyama, M., Shinagawa, E., Matsushita, K., and Adachi, O. (1981) *J. Bacteriol.* **145**, 814–823
21. Iwasaki, I., Utsumi, S., Hagino, K., and Ozawa, T. (1956) *Bull Chem. Soc. Jpn.* **29**, 860–864
22. Kurihara, T., Liu, J. Q., Nardi-Dei, V., Koshikawa, H., Esaki, N., and Soda, K. (1995) *J Biochem. (Tokyo)* **117**, 1317–1322
23. Baykov, A. A., Evtushenko, O. A., and Avaeva, S. M. (1988) *Anal. Biochem.* **171**, 266–270
24. Zhang, G., Dai, J., Wang, L., Dunaway-Mariano, D., Tremblay, L. W., and Allen, K. N. (2005) *Biochemistry* **44**, 9404–9416
25. Datsenko, K. A., and Wanner, B. L. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–6645
26. Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995) *J. Bacteriol.* **177**, 4121–4130
27. Macchiarulo, A., Nobeli, I., and Thornton, J. M. (2004) *Nat. Biotechnol.* **22**, 1039–1045
28. Felsenstein, J. (1996) *Methods Enzymol.* **266**, 418–427
29. Edgar, R. C. (2004) *Nucleic Acids Res.* **32**, 1792–1797
30. Adachi, J., and Hasegawa, M. (1992) *Comput. Sci. Monogr.* **27**, 1–77
31. Schmidt, H. A., Strimmer, K., Vingron, M., and von Haeseler, A. (2002) *Bioinformatics* **18**, 502–504
32. Huelsenbeck, J. P., and Ronquist, F. (2001) *Bioinformatics* **17**, 754–755
33. Ronquist, F., and Huelsenbeck, J. P. (2003) *Bioinformatics* **19**, 1572–1574
34. Jensen, R. A. (1976) *Annu. Rev. Microbiol* **30**, 409–425
35. O'Brien, P. J., and Herschlag, D. (1999) *Chem. Biol.* **6**, R91–R105

36. Copley, S. D. (2003) *Curr. Opin. Chem. Biol.* **7,** 265–272
37. James, L. C., and Tawfik, D. S. (2001) *Protein Sci.* **10,** 2600–2607
38. Bornscheuer, U. T., and Kazlauskas, R. J. (2004) *Angew Chem. Int. Ed. Engl.* **43,** 6032–6040
39. Petsko, G. A., Kenyon, G. L., Gerlt, J. A., Ringe, D., and Kozarich, J. W. (1993) *Trends Biochem. Sci.* **18,** 372–376
40. Gerlt, J. A., and Babbitt, P. C. (2001) *Annu. Rev. Biochem.* **70,** 209–246
41. Aharoni, A., Gaidukov, L., Khersonsky, O., Mc, Q. G. S., Roodveldt, C., and Tawfik, D. S. (2005) *Nat. Genet.* **37,** 73–76
42. Bartlett, G. J., Borkakoti, N., and Thornton, J. M. (2003) *J. Mol. Biol.* **331,** 829–860
43. Roodveldt, C., and Tawfik, D. S. (2005) *Biochemistry* **44,** 12728–12736
44. Postma, P. W., Lengeler, J. W., and Jacobson, G. R. (1993) *Microbiol. Rev.* **57,** 543–594
45. Charbit, A., Reizer, J., and Saier, M. H., Jr. (1996) *J. Biol. Chem.* **271,** 9997–10003
46. Buschmeier, B., Hengstenberg, W., and Deutscher, J. (1985) *FEMS Microbiol. Lett.* **29,** 231–235
47. Proudfoot, M., Kuznetsova, E., Brown, G., Rao, N. N., Kitagawa, M., Mori, H., Savchenko, A., and Yakunin, A. F. (2004) *J. Biol. Chem.* **279,** 54687–54694
48. Bianchi, V., and Spychala, J. (2003) *J. Biol. Chem.* **278,** 46195–46198
49. Lumeng, L., and Li, T. K. (1975) *J. Biol. Chem.* **250,** 8126–8131
50. Jang, Y. M., Kim, D. W., Kang, T. C., Won, M. H., Baek, N. I., Moon, B. J., Choi, S. Y., and Kwon, O. S. (2003) *J. Biol. Chem.* **278,** 50040–50046
51. Fonda, M. L. (1992) *J. Biol. Chem.* **267,** 15978–15983
52. Froman, B. E., Tait, R. C., and Gottlieb, L. D. (1989) *Mol. Gen. Genet.* **217,** 126–131
53. Babul, J., and Guixe, V. (1983) *Arch. Biochem. Biophys.* **225,** 944–949
54. Donahue, J. L., Bownas, J. L., Niehaus, W. G., and Larson, T. J. (2000) *J. Bacteriol.* **182,** 5624–5627
55. Lowry, O. H., Carter, J., Ward, J. B., and Glaser, L. (1971) *J. Biol. Chem.* **246,** 6511–6521
56. de Silva, A. O., and Fraenkel, D. G. (1979) *J. Biol. Chem.* **254,** 10237–10242
57. Buchholz, A., Takors, R., and Wandrey, C. (2001) *Anal. Biochem.* **295,** 129–137
58. Linn, E. C. C. (1996) in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (Neidhardt, F. C., ed) 2nd Ed., pp. 307–342, American Society for Microbiology Press, Washington, D. C.
59. Pradel, E., and Boquet, P. L. (1988) *J. Bacteriol.* **170,** 4916–4923
60. Bhattacharya, M., Fuhrman, L., Ingram, A., Nickerson, K. W., and Conway, T. (1995) *Anal. Biochem.* **232,** 98–106
61. Buchholz, A., Hurlebaus, J., Wandrey, C., and Takors, R. (2002) *Biomol. Eng.* **19,** 5–15
62. Rimmele, M., and Boos, W. (1994) *J. Bacteriol.* **176,** 5654–5664
63. Cho, H., Wang, W., Kim, R., Yokota, H., Damo, S., Kim, S. H., Wemmer, D., Kustu, S., and Yan, D. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 8525–8530
64. Lukat, G. S., McCleary, W. R., Stock, A. M., and Stock, J. B. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 718–722
65. Silversmith, R. E., Appleby, J. L., and Bourret, R. B. (1997) *Biochemistry* **36,** 14965–14974
66. McCleary, W. R., and Stock, J. B. (1994) *J. Biol. Chem.* **269,** 31567–31572
67. Chou, T. F., Bieganowski, P., Shilinski, K., Cheng, J., Brenner, C., and Wagner, C. R. (2005) *J. Biol. Chem.* **280,** 15356–15361
68. Thompson, J. (1987) *FEMS Microbiol. Rev.* **46,** 221–231
69. Dietz, G. W., and Heppel, L. A. (1971) *J. Biol. Chem.* **246,** 2881–2884
70. Aravind, L., Mazumder, R., Vasudevan, S., and Koonin, E. V. (2002) *Curr. Opin. Struct. Biol.* **12,** 392–399
71. Lynch, M., and Force, A. (2000) *Genetics* **154,** 459–473
72. Massingham, T., Davies, L. J., and Lio, P. (2001) *BioEssays* **23,** 873–876
73. He, X., and Zhang, J. (2005) *Genetics* **169,** 1157–1164
74. Gohla, A., Birkenfeld, J., and Bokoch, G. M. (2005) *Nat. Cell Biol.* **7,** 21–29
75. Anantharaman, V., Aravind, L., and Koonin, E. V. (2003) *Curr. Opin. Chem. Biol.* **7,** 12–20
76. Kneidinger, B., Marolda, C., Graninger, M., Zamyatina, A., McArthur, F., Kosma, P., Valvano, M. A., and Messner, P. (2002) *J. Bacteriol.* **184,** 363–369
77. Kandror, O., DeLeon, A., and Goldberg, A. L. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 9727–9732
78. Fortpied, J., Maliekal, P., Vertommen, D., and Van Schaftingen, E. (2006) *J. Biol. Chem.* **281,** 18378–18385
79. Mironova, R., Niwa, T., Hayashi, H., Dimitrova, R., and Ivanov, I. (2001) *Mol. Microbiol.* **39,** 1061–1068
80. Mironova, R., Niwa, T., Handzhiyski, Y., Sredovska, A., and Ivanov, I. (2005) *Mol. Microbiol.* **55,** 1801–1811