At the Periphery of the Amidohydrolase Superfamily: Bh0493 from *Bacillus halodurans* Catalyzes the Isomerization of D-Galacturonate to D-Tagaturonate[†,‡]

Tinh T. Nguyen,[§] Shoshana Brown,[∥] Alexander A. Fedorov,[⊥] Elena V. Fedorov,[⊥] Patricia C. Babbitt*[,∥,#] Steven C. Almo*[,⊥] and Frank M. Raushel*[,§]

Department of Chemistry, P.O. Box 30012, Texas A&M University, College Station, Texas 77842-3012, Department of Biopharmaceutical Sciences and Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, 1700 4th Street, San Francisco, California 94158-2550, and Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461

Received August 30, 2007; Revised Manuscript Received November 8, 2007

ABSTRACT: The amidohydrolase superfamily is a functionally diverse set of enzymes that catalyzes predominantly hydrolysis reactions involving sugars, nucleic acids, amino acids, and organophosphate esters. One of the most divergent members of this superfamily, uronate isomerase from *Escherichia coli*, catalyzes the isomerization of D-glucuronate to D-fructuronate and D-galacturonate to D-tagaturonate and is the only uronate isomerase in this organism. A gene encoding a putative uronate isomerase in *Bacillus halodurans* (Bh0705) was identified based on sequence similarity to uronate isomerases from other organisms. Kinetic evidence indicates that Bh0705 is relatively specific for the isomerization of D-glucuronate to D-fructuronate, confirming this functional assignment. Despite a low sequence identity to all other characterized uronate isomerases, phylogenetic and network-based analysis suggests that a second gene in this organism, Bh0493, is also a uronate isomerase, although it is an outlier in the group, with <20% sequence identity to any other characterized uronate isomerase from another species. The elucidation of the X-ray structure at a resolution of 2.0 Å confirms that Bh0493 is a member of the amidohydrolase superfamily with conserved residues common to other members of the uronate isomerase family. Functional characterization of this protein shows that unlike Bh0705, Bh0493 can utilize both D-glucuronate and D-galacturonate as substrates. In *B. halodurans*, Bh0705 is found in an operon for the metabolism of D-glucuronate, whereas Bh0493 is in an operon for the metabolism of D-galacturonate. These results provide the first identification of a uronate isomerase that operates in a pathway distinct from that for D-glucuronate. While most organisms that contain this pathway have only one gene for a uronate isomerase, sequence analysis and operon context show that five other organisms also appear to have two genes and one organism appears to have three genes for this activity.

The identification and functional assignment of enzymes encoded within completely sequenced genomes is a difficult and demanding task. According to some estimates, approximately 40% of newly sequenced genes have an unknown, uncertain, or incorrect functional assignment (*1*). The extent of the catalytic diversity inherent within metabolic and catabolic pathways will not be fully understood until the functional annotations of these newly sequenced genes have been addressed. The absence of a comprehensive understanding of the metabolic landscape has therefore motivated the development of new methodologies for the assignment of function for enzymes that catalyze reactions with ambiguous substrates. These emerging strategies include phylogenetic profiling (*2*, *3*), recognition of domain fusions and gene clusters (*4*−*8*), and other computational approaches (see ref *9* and references therein). Most of the commonly used techniques require sequence and/or structural comparisons to well-characterized homologues to make high quality functional predictions. However, these attempts at functional characterization can be problematic when the sequence or structural similarity is remote (*1*).

[†] This work was supported in part by the NIH (GM71790) and the Robert A. Welch Foundation (A-840). Tools used for visualization of protein networks were created by the UCSF Resource for Biocomputing, Visualization, and Informatics (http://www.rbvi.ucsf.edu), funded by NIH Grant P41 RR01081.

[‡] X-ray coordinates and structure factors for SeMet-substituted Bh0493 complexed with Zn and wild-type Bh0493 complexed with Zn have been deposited in the Protein Data Bank (PDB IDs 2Q6E and 2QO8).

* Corresponding authors. (P.C.B.) Tel.: (415) 476-3784; fax: (415) 514-4797; e-mail: babbitt@cgl.ucsf.edu. (S.C.A.) Tel.: (718) 430-2746; fax: (718) 430-8565; e-mail: almo@aecom.yu.edu. (F.M.R.) Tel.: (979) 845-3373; fax: (979) 845-9452; e-mail: raushel@tamu.edu.

[§] Texas A&M University.

[∥] Department of Biopharmaceutical Sciences, School of Pharmacy, University of California.

[⊥] Albert Einstein College of Medicine.

[#] Department of Pharmaceutical Chemistry, School of Pharmacy, University of California.

[1] Abbreviations: AHS, amidohydrolase superfamily; URI, uronate isomerase; ATCC, American Type Culture Collection; IPTG, isopropyl-thiogalactoside; HEPES, *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid; NCBI, National Center for Biotechnology Information; DTT, dithiothreitol; ICP-MS, inductively coupled plasma mass spectrometry; NagA, *N*-acetyl-D-glucosamine-6-phosphate deacetylase; ACMSD, α-amino-β-carboxymuconic-ε-semialdehyde decarboxylase; SFLD, Structure Function Linkage Database.

10.1021/bi7017738 CCC: $40.75 © 2008 American Chemical Society
Published on Web 01/03/2008

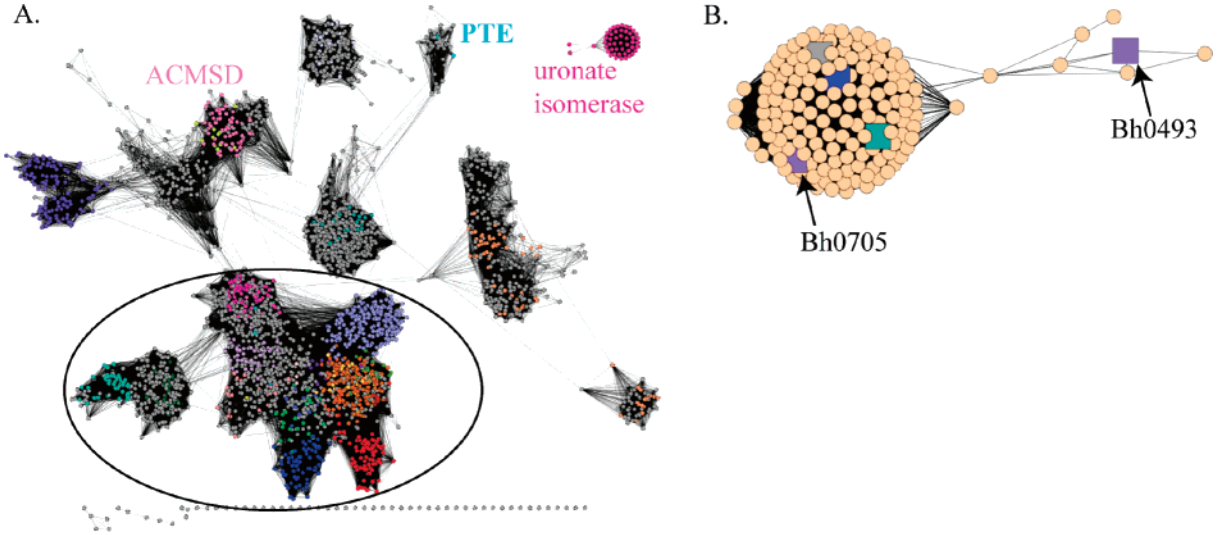


FIGURE 1: Network representation of the sequence relationships in the amidohydrolase superfamily. Each node in the network represents a single sequence in the uronate isomerase-like sequence set, and each edge represents the pairwise connection between two sequences with the most significant BLAST *E*-value (better than the cutoff) connecting the two sequences. Lengths of edges are not meaningful except that sequences in tightly clustered groups are relatively more similar to each other than sequences with few connections. (A) An 80% ID filtered set of amidohydrolase superfamily sequences from the SFLD showing connections with a BLAST *E*-value more significant than or equal to 10^{-4}. Sequences are colored by SFLD subgroup. Sequences with characterized functions originally identified as part of the amidohydrolase superfamily by Holm and Sander (*12*) are contained within the circle (except for phosphotriesterase, indicated with the PTE abbreviation). Additional groups discussed in the text are indicated with labels. (B) An unfiltered set of uronate isomerase-like sequences at a BLAST *E*-value more significant than or equal to 10^{-10}. Sequences that have been experimentally characterized as uronate isomerase and/or have an X-ray crystal structure are colored as follows: gray, Tm0064; lavender, Bh0705 and Bh0493; blue, Cc1490; and green, b3092.

Members of the amidohydrolase superfamily (AHS)[1] have been identified in every organism sequenced to date (*10*, *11*). This enzyme superfamily catalyzes predominantly hydrolytic reactions where a water molecule is activated by one or two divalent metal ions embedded within the active site (*11*). All of the enzymes within the AHS adopt a $(\beta/\alpha)_8$-barrel structural fold with an active site that resides at the C-terminal end of the β-barrel (*11*, *12*). One of the most divergent enzyme families in the AHS is uronate isomerase (URI). This enzyme family exhibits few statistically significant sequence links to any other member of the AHS (Figure 1), and structure-based comparisons are required to link this family to other families in the superfamily.

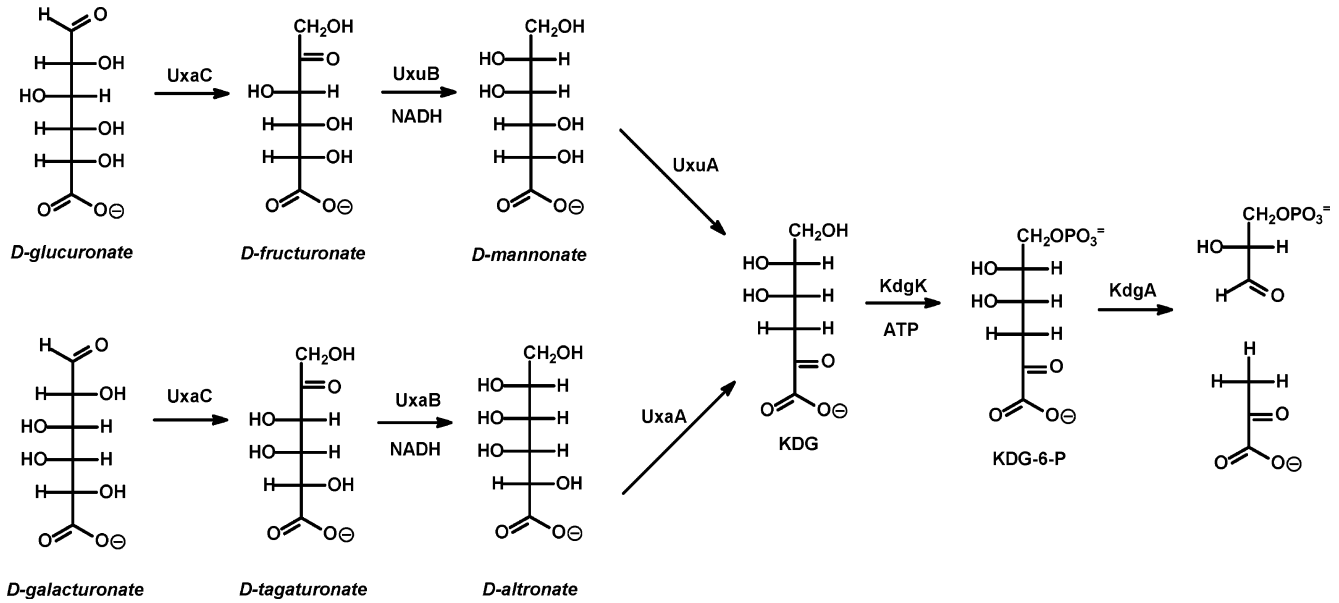
The URI from *Escherichia coli* (UxaC) catalyzes the isomerization of D-glucuronate to D-fructuronate and D-galacturonate to D-tagaturonate and represents the canonical URI found in the majority of organisms containing genes encoding this activity. These isomerization reactions represent the first step in the separate metabolic pathways for the utilization of D-glucuronate and D-galacturonate in bacteria. The isomerized products are subsequently reduced to D-mannonate and D-altronate by UxuB and UxaB, respectively, and then dehydrated by UxuA and UxaA to 2-keto-3-deoxy-D-gluconate. The entire metabolic pathway for the metabolism of these sugars in *E. coli* is summarized in Scheme 1. Consistent with its outlier status as compared to sequences of other families in the AHS, URI is the only known member of the AHS that apparently does not require the presence of a divalent cation for expression of catalytic activity (*13*). The enzyme from *E. coli* can bind up to one Zn per subunit, but this metal ion has little influence on the ability of this protein to catalyze the isomerization of D-glucuronate to D-fructuronate.

Our collective laboratories have begun to delineate the extent of the structural and functional diversity across the AHS of enzymes. One of the most divergent protein sequences identified to date from this investigation has emerged from the genome of the alkaliphilic bacterium *Bacillus halodurans* (gi: 15613056). This protein (Bh0493) was originally predicted to be a member of the AHS based upon a weak sequence similarity (<19%) with the structurally characterized uronate isomerase from *Thermotoga maritima* (Tm0064, gi: 15642839). Providing additional evidence for this assignment, the putative active site residues that originate from the ends of β-strands 1 and 8 are conserved in the *B. halodurans* enzyme. In this paper, we report the cloning, expression, structure, and functional characterization of Bh0493. The X-ray structure of Bh0493, determined with a single zinc ion in the active site, confirms that this enzyme is a member of the AHS. This enzyme will catalyze the isomerization of D-galacturonate to D-tagaturonate in a metabolic pathway that has diverged from that which performs the isomerization of D-glucuronate. We have also functionally characterized Bh0705 from *B. halodurans* (gi: 15613268), the amino acid sequence that more closely resembles that of a canonical uronate isomerase.

MATERIALS AND METHODS

Materials. D-Glucuronate and NADH were purchased from Sigma-Aldrich. D-Galacturonate was obtained from Acros. The inhibitors, D-arabinaric acid (**1**) and the monohydroxamate derivative of D-arabinaric acid (**2**), were synthesized as previously described (*13*). The DNA sequencing reactions were conducted by the Gene Technologies Laboratory at Texas A&M University. Determination of the metal content of the purified proteins was performed using ICP−MS. Gel filtration and anion exchange columns were purchased from GE HealthCare. Restriction enzymes and T4 DNA ligase

Scheme 1



were obtained from New England Biolabs. The pET30 vector was acquired from Novagen, and the Platinum Pfx DNA polymerase was purchased from Invitrogen. The Wizard Miniprep DNA purification kit was obtained from Promega, and the genomic DNA of *B. halodurans* was acquired from ATCC (BAA-125D). Mannonate dehydrogenase (UxuB) was purified from *E. coli* as previously described (*13*).

Cloning and Overexpression of Bh0705 and Bh0493. The gene *uxa*C (Bh0705, gi: 15613268) from *B. halodurans* was amplified by PCR with two primers, 5′-GCAGGAGCATTAATGACGAATTTTCTTTCAGAAGACTTCCTCTTAATGA ACGAATACG-3′ containing an *Ase*I restriction site and 5′-GCGCAAGCTTTCATCATAGTCGAACAGTCTCCTTCATGTCG-3′ containing a *Hin*dIII restriction site. The *uxa*C gene was inserted into a pET30a (+) vector that had been digested with *Ase*I and *Hin*dIII. The protein was expressed and purified from BL21(DE3) star cells (Novagen) that had been transformed with the pET30a (+) plasmid containing the *uxa*C gene from *B. halodurans*. The gene for Bh0493 (gi: 15613056) was cloned with two primers, 5′-GCAGGAGCCATATGTCCATAAACAGTAGGGAAGTGTTACGGG-3′ and 5′-CGCGGAATTCTTATTACGTTTGCTGCTCCACCTTCACTGATGTGACG-3′ containing the restriction sites for *Nde*I and *Eco*RI, respectively. The gene product was ligated into a pET30a (+) vector and expressed in BL21(DE3) star cells. The same protocol was utilized to purify Bh0705 and Bh0493. A single colony of the cells containing the gene of interest was inoculated and incubated overnight in 5 mL of LB medium containing 50 μg/mL kanamycin at 37 °C. The overnight culture was used to inoculate 1 L of LB medium containing 50 μg/mL kanamycin. The cells were grown at 30 °C and induced with 0.4 mM IPTG when the cell density at A_{600} reached ~0.4−0.6. After 24 h, the cells were harvested at 4000*g*. The cell pellet was resuspended in 5 mL of 50 mM HEPES, pH 8.0 (buffer A) for every gram of cell paste and disrupted by sonication. The nucleic acids were removed by the dropwise addition of protamine sulfate to a final concentration of 2% w/v and subsequent centrifugation at 14 000*g*. The protein was precipitated between 50 and 70% saturation of ammonium sulfate, and the protein pellet was resuspended in a minimum amount of buffer A. The protein was loaded onto a pre-equilibrated Superdex 200 gel filtration column and eluted with buffer A. The fractions containing the protein of interest were pooled based on activity, loaded onto a Resource Q column, and eluted with a 0−30% gradient of 1 M NaCl in 20 mM HEPES, pH 8.0 (buffer B). The fractions containing the desired protein were pooled, and the purity of the protein was determined by SDS-gel electrophoresis. The purified enzyme was sterile-filtered and stored at 4 °C.

Cloning, Expression, and Purification of UxaB. The *uxa*B gene from *E. coli* (gi: 49176119) was amplified by PCR with two primers, 5′-GCAGGAGCCATATGAAAACACTAAATCGTCGCGATTTTCCCGGTGC-3′ and 5′-GCGCAAGCTTTTATTAGCACAACGGACGTACAGCTTCGCGCATCCCTTT TTCG-3′, containing *Nde*I and *Hin*dIII sites, respectively, and the resulting fragment was inserted into the *Nde*I and *Hin*dIII sites of the pET30 vector. The protein was expressed in the BL21(DE3) strain of *E. coli*. A single colony was used to inoculate a 5 mL overnight culture of LB medium containing 50 μg/mL kanamycin. The 5 mL overnight culture was inoculated into 1 L of LB containing the same concentration of kanamycin. The cells were grown at 30 °C, induced with 0.4 mM IPTG at an A_{600} of ~0.6, and incubated overnight. The cells were collected by centrifugation at 4000*g*, and the cell pellet was resuspended in buffer A. The cells were disrupted by sonication and centrifuged at 14 000*g*. The protein was soluble, and the supernatant solution was used directly in the enzymatic assays.

Sequence Collection. Two preliminary alignments of uronate isomerase-like sequences were created by aligning sequences similar to the structurally characterized *T. maritima* sequence (Tm0064) and the outlier *B. halodurans* sequence (Bh0493), respectively. Hidden Markov models were created from these alignments and used to search the NCBI protein database via the HMMSEARCH program. All hit sequences were then aligned and examined for conservation of residues conserved across the AHS (residues corre-

sponding to H26, H28, and D355 in Bh0493). Sequences that appeared to be fragments or did not conserve the three residues listed previously were removed from the sequence set.

Network Analysis for the Uronate Isomerase-like Sequence Set. BLAST (*14*) analysis was performed using the sequences in the uronate isomerase-like set as queries against the NCBI NR database at an *E*-value cutoff of 10^{-10}. Only hits in the uronate isomerase-like sequence set were used for further analysis. A Cytoscape (*15*) network was created from these BLAST results using an *E*-value cutoff of 10^{-10} (Figure 1). Connections between nodes are only shown if the *E*-value of the best BLAST hit between two sequences is at least as good as 10^{-10}. Tools used for visualization of protein networks were created by the UCSF Resource for Biocomputing, Visualization, and Informatics and are available from the Resource (http://www.rbvi.ucsf.edu). Each node in the network represents a single sequence in the uronate isomerase-like sequence set, and each edge represents the pairwise connection between two sequences with the most significant BLAST *E*-value (better than the cutoff) connecting the two sequences. Lengths of edges are not meaningful except that sequences in tightly clustered groups are relatively more similar to each other than sequences with few connections. The nodes were arranged using the yFiles organic layout provided with Cytoscape version 2.4. Sequences that have been experimentally characterized and/or have a solved X-ray crystal structure are colored as described in the legend to Figure 1.

Network Analysis for the Entire AHS. A custom database was created containing an 80% ID filtered set of sequences from the Structure Function Linkage Database (SFLD) (*16*). BLAST searches were then performed against the database at an *E*-value cutoff of 10^{-4}, using each sequence in the set as a query. Because of the large number of sequences, the time required to perform BLAST searches against the NCBI NR database, as stated previously, and remove false positives was prohibitive. Because this analysis was performed against this custom database that contains only sequences known to be members of the AHS, the generation of *E*-values is biased since the background model for calculating statistical significance is not random. Consequently, the *E*-values from this analysis cannot be directly compared to those for the network analysis of the uronate isomerase-like sequence set. A Cytoscape network was created based on the BLAST results, where each node represents a single sequence in the 80% ID filtered AHS sequence set and each edge represents the most significant BLAST *E*-value connecting the two sequences, as described for the network analysis for the URI-like sequence set. The nodes were arranged using the yFiles organic layout provided with Cytoscape version 2.4 and colored based on their SFLD subgroup assignment.

Sequence Alignments. The uronate isomerase sequence set was divided into three preliminary clusters, based on sequence similarity, and each cluster was separately aligned using Muscle (*17*). The alignments were then combined using Muscle's profile alignment mode. The alignment was further edited by hand to ensure that it conformed to a structural alignment (*18*, *19*) of 1J5S (for Tm0065), 2Q01 (for Cc1490), and 2QO8 (for Bh0493). Unaligned regions at the N− and C-termini were removed. The full sequence alignment is available in the Supporting Information. The filtered alignment provided in Figure 2 was created by extracting the sequences shown from the full alignment.

Tree Construction. A representative subset of uronate isomerase-like sequences was selected by filtering the entire sequence set to approximately 70% identity using the cd-hit program (*20*). A filtered version of the hand-edited alignment described in the previous section, containing only those sequences in the 70% identity set, was then created and used for tree construction with MrBayes (*21*, *22*).

Crystallization and Data Collection. Two different crystal forms were grown by vapor diffusion at room temperature: (i) selenomethionine (SeMet)-substituted wild-type Bh0493 complexed with Zn^{2+} and (ii) wild-type Bh0493 complexed with Zn^{2+}. The following crystallization conditions were used: (i) SeMet-substituted wild-type Bh0493: the protein solution contained SeMet-substituted Bh0493 (14.1 mg/mL) in 10 mM HEPES, pH 7.5, 150 mM NaCl, 10 mM methionine, 10% glycerol, 5 mM DTT, and 0.5 mM $ZnCl_2$, and the precipitant contained 25% PEG 3350, 0.1 M Tris pH 8.5, and 0.2 M NaCl. Crystals appeared in 3 days and exhibited diffraction consistent with the space group $P4_122$, with three molecules of Bh0493 per asymmetric unit. (ii) Wild-type Bh0493 complexed with Zn: the protein solution contained Bh0493 (16.9 mg/mL) in 20 mM Tris-HCl, pH 8.5, and 0.5 mM $ZnCl_2$, and the precipitant contained 45% polypropylene glycol and 0.1 M Bis-Tris, pH 6.5. Crystals appeared in 5 days and exhibited diffraction consistent with the space group $P1$, with 12 molecules of Bh0493 per asymmetric unit.

Prior to data collection, the crystals were transferred to cryoprotectant solutions composed of their mother liquors and 20% glycerol. After incubation for ~15 s, the crystals were flash-cooled in a nitrogen stream. A single-wavelength anomalous dispersion (SAD) data set for a crystal of SeMet-substituted wild-type Bh0493 with Zn was collected to 2.4 Å resolution at the NSLS X4A beamline (Brookhaven National Laboratory) on an ADSC CCD detector. The data set for wild-type Bh0493 with Zn was collected at the same beamline to 2.0 Å resolution. Intensities were integrated and scaled with DENZO and SCALEPACK (*23*).

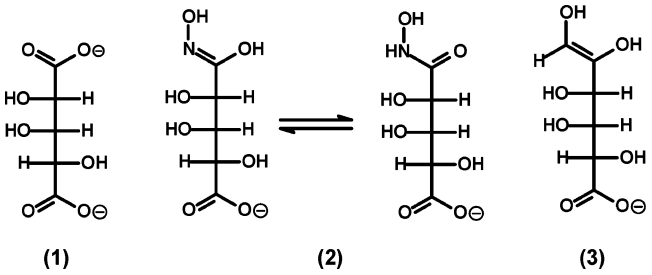
Structure Determination and Refinement. Initial attempts to determine the structure of Bh0493 by molecular replacement using the structure of uronate isomerase from *T. maritima* (PDB ID 1J5S) as a search model were unsuccessful. Instead, the structure of the SeMet-substituted wild-type Bh0493 complexed with Zn was determined by SAD with SOLVE (*24*); 48 of the 51 selenium sites were identified. These heavy atom sites were used to calculate initial phases that were improved by solvent flattening and NCS averaging with RESOLVE (*25*), yielding an interpretable map for three monomers in the asymmetric unit for space group $P4_122$. Iterative cycles of automatic rebuilding with ARP (*26*), manual rebuilding with TOM (*27*), and refinement with CNS (*28*) resulted in a model at 2.4 Å with R_{cryst} value of 0.231 and an R_{free} value of 0.246. The first residue and last 14 residues were disordered and were not included in the final model. The Zn^{2+} bound in the active site was clearly visible in the electron density maps for all three molecules in the asymmetric unit. Na^+ and Cl^-, located on the local 3-fold axis of the Bh0493 trimer, also exhibited good density.

```
TM0064 --HHMFLGEDYLLT-----------NR--AAVRLFNE-VKDLPIVDPHNHLDAKD-IV-E
 b3092 --MTPFMTEDFLLD-----------TE--FARRLYHDYAKDQPIFDYHCHLPPQQ-IA-E
BH0705 --MTNFLSEDFLLM-----------NE--YDRELYYTFAKNMPICDYHCHLSPQE-IW-E
CC1490 MARPLSFHEDRLFPSDP------ATRS--YARGLYAL-VKDLPIISPHGHTDPSW-FA-T
BH0493 -----------------MSINSRE---VLA-EKVKNA-VNNQPVTDMHTHL-F--SPNFG

TM0064 NKPWND------IWEVEGATDHYVWELMRRCGVS-E--EYI-TG-------SRS-NKEKW
 b3092 DYRFKN------LYDIWLKGDHYKWRAMRTNGVA-E--RLC-TG-------DAS-DREKF
BH0705 NKPFEN------MTKAWLGGDHYKWRAMRLNGVR-E--EFI-TG-------GAP-DKEKF
CC1490 NAPFQD------ATDLLLAPDHYLFRMLYSQGVS-L--DALKV-RSKAGVPDTD-PREAW
BH0493 ------EILLWDID-ELLT-YHYLVAEVMRWTDVSIEAF----W------AMS-KR-EQA

TM0064 LALAKV-F---PRFVGNPTYEWIHLDLWRRFNI-K-KVISEE-TAEEIWEETKKKL--PE
 b3092 DAWAAT-V---PHTIGNPLYHWTHLELRRPFGITG-KLLSPS-TADEIWNECNELLAQDN
BH0705 LAWAKT-V---PKTIGNPLYHWTHMELKTYFHF-H-QPLDET-NGENVWDACNRLLQQEA
CC1490 RVFASH-F---YLFRGTPSWVWLNHVFSQVFGF-T-EFLEAS-NADDYFDRITAALA-TD
BH0493 DLIWEELFIKRS-PV-SEACRGVLTCLQGLGLD--PA-----TR--DL-QVYREYFA-KK

TM0064 -MTPQKLLRD-MK---VEILCTTDDPVS-----T--LEHH--RKAKEA-V-E--GVTILP
 b3092 -FSARGIMQQ-MN---VKMVGTTDDPID-----S--LEHH--AEIAKD-GSF--TIKVLP
BH0705 -FTPRALIER-SN---VRAIGTTDDPTD-----S--LLYH--QKLQAD-DTF--HVKVIP
CC1490 AFRPRALFDR-FN---IETLATTEGPHE-----S--LQHH--AAIRESGW-G-G--HVIT
BH0493 -TS-EEQVDTVLQLANVSDVVMTND--PFDDNERISWL--EGK--Q------PD-SRFHA

TM0064 TWRPDRAMNVDKEG--WREYVEKMGERYGEDTST--L---D--GFLNALWKSHEHFKE-H
 b3092 SWRPDKAFNIEQAT--FNDYMAKLGEVSDTDIRR--F---A--DLQTALTKRLDHFAA-H
BH0705 TFRPDGALKIEQDS--FADWVAKLSDVTGESLDT--L---D--AFLHALKERLTFFDE-H
CC1490 AYRPDAVIDFEDER--SPRAFERFAETSGQDVYS--W---K--SYLEAHRLRRQAFID-A
BH0493 ALRLDPLLNE----YEQ--TKHRLRDW-GYKV-NDEWNEGSIQE-V---KRFLTDWIERM

TM0064 GCVASDHALL-EPSVYYV-DENRARAVHEKAFSGEKLTQDEINDY-KAFMMVQ-FGKMNQ
 b3092 GCKVSDHALD-VV-MFAEANEAELDSILARRLAGETLSEHEVAQF-KTAVLVF-LGAEYA
BH0705 GCRSSDHDMT-EV-PFVEVNEQEAQHIFRKRLANEGLTKVENEKY-KTFLMTW-LGKEYA
CC1490 GATSSDHGHP-TAATADL-SDVEAEALFNSLVKGD-VTPEKAELF-RAQMLTE-MAKMSL
BH0493 DPVYMAVSLPPTFSFP------------------------EESNRGRIIRDCLLPVAE

TM0064 ETNWVTQLHIGALRDYRDSLFKTL-GPDSGGDISTNFLRIAEGLRYFLNEF-DG-K---L
 b3092 RRGWVQQYHIGALRNNNLRQFKLL-GPDVGFDSIND-RPMAEELSKLLSKQ-NE-ENLLP
BH0705 ARGWVMQWHIGVMRNNNSRMLHKL-GPDTGFDSIGD-GQIAHATAKLLDLL-DK-QGALP
CC1490 DDGLVMQIHPGSHRNHNVGLLNSH-GRDKGADIPMR-TEYVDALKPLLTRLGNDPR---L
BH0493 KHNIPFAMMIGVKKRVH---P--ALG-D-AGDFVGK-ASM-DGVEHLLREY-P--N---N

TM0064 KIVLYVLDPTHL-PTISTIARAFPN------VYVGAPWWFNDSPFGMEMHLKYLASVDLL
 b3092 KTILYCLNPRDN-EVLGTMIGNFQGEGMPGKMQFGSGWWFNDQKDGMERQMTQLAQLGLL
BH0705 KTILYCVNPNAN-YILASMIGNFTESGVRGKVQFGSAWWFNDHIDGMRRQLTDLASVGLL
CC1490 SIILFTLDETTYSRELAPLAGHYPV------LKLGPSWWFHDSPEGMMRFREQVTETAGF
BH0493 KFLVTMLSRENQHE-LVVLARKFSN------LMIFGCWWFMNNPEIINEMTRMRMEMLG-

TM0064 YNLAGMVTDSRKLLSFGSRTEMFRRVLSNVVGEM-VEKGQIP--I-----KEA----R-E
 b3092 SRFVGMLTDSRSFLSY-TRHEYFRRILCQMIGRW-VEAGEAPADI-----NLL----G-E
BH0705 SNFIGMLTDSRSFLSY-PRHDYFRRILCQLIGSW-IKEGQLPPDM-----ERW----G-Q
CC1490 YNTVGFNDDTRAFLSIPARHDVARRVDSAFLARM-VAEHRMD--L-----VEA----E-E
BH0493 TSFIPQHSDARVLEQLIYKWHHSKSIIAEVLIDKYD--D-IL---QAGWEV--TEEEIKR

TM0064 LVKHVSYDGPKALF-F
 b3092 MVKNICFNNARD-Y-F
BH0705 IVQDICYNNVVD-Y-F
CC1490 LIVDLTYNLPKKAYKL
BH0493 DVADLFSRNFWRF-V-
```

FIGURE 2: Protein sequence alignment of representative uronate isomerase-like sequences from *T. maritima* (Tm0064), *E. coli* (b3092), *B. halodurans* (Bh0705 and Bh0493), and *C. crescentus* (Cc1490). The three sequence groups discussed in the text and shown in Figure 3 are indicated by the colors on the sequence headers (green, group 1; orange, group 2; and purple, group 3). Residues within the active site that are positioned such that they may influence substrate binding and/or catalysis are colored. Red indicates positions conserved across the entire uronate isomerase sequence set (exception—one of the tryptophan residues in the WWF motif is not conserved in gi: 7531233), and blue indicates positions that are completely conserved within a group but not across the entire sequence set. β-Strands for Bh0493, mentioned in the text, are indicated with gray shading.

The structure of wild-type Bh0493 complexed with Zn was determined by molecular replacement with PHASER (*29*), using the SeMet-substituted Bh0493 as the search model. Iterative cycles of automatic rebuilding with ARP, manual rebuilding with TOM, and refinement with CNS were performed. The model was refined at 2.0 Å with an R_{cryst} value of 0.226 and an R_{free} value of 0.239. Each of the 12 molecules in the asymmetric unit has Zn^{2+} bound in the

Scheme 2



active site, and additionally, each of the four trimers in the asymmetric unit has a Zn^{2+} ion located on their local 3-fold axis.

Enzyme Activity and Substrate Specificity. The isomerase activities of Bh0705 and Bh0493 were measured with D-glucuronate and D-galacturonate. The enzymatic reactions with D-glucuronate and D-galacturonate were coupled to the reduction of D-fructuronate and D-tagaturonate with the appropriate dehydrogenase. These assays monitor the oxidation of NADH to NAD^+, and the change in absorbance was measured spectrophotometrically at 340 nm. The standard assays contained 50 mM HEPES, pH 8.0, 0.2 mM NADH, ~2 μM either D-mannonate dehydrogenase or D-tagaturonate dehydrogenase, and various substrate concentrations in a final volume of 250 μL (*13*).

Inhibition Studies. The inhibition of the reactions catalyzed by Bh0705 and Bh0493 was quantified using two compounds that have previously been shown to inhibit the isomerization of D-glucuronate by uronate isomerase from *E. coli*. D-Arabinaric acid (**1**) and (2*S*,3*R*,4*S*)-4-(hydroxycarbamoyl)-2,3,4-trihydroxybutanoate (**2**) were tested as inhibitors using standard assay conditions with D-glucuronate as the substrate and D-mannonate dehydrogenase as the coupling enzyme. The structures of these compounds are presented in Scheme 2.

Data Analysis. The kinetic parameters, k_{cat} and k_{cat}/K_m, for the enzyme Bh0705 were determined by fitting the initial velocity data to eq 1, where v is the initial velocity, E_t is the total enzyme concentration, k_{cat} is the turnover number, A is the substrate concentration, and K_a is the Michaelis constant. The double-reciprocal plots for Bh0493 were nonlinear (concave down) and fitted to eq 2 (*30*). In this equation, k_1 is the maximum velocity at low substrate concentration, and the sum of k_1 and k_2 is the maximum velocity at high substrate concentration. The apparent Michaelis constants at low and high substrate concentrations are K_1 and K_2, respectively. Competitive inhibition patterns by the compounds that mimic the putative *cis*-enediol intermediate were fit to eq 3, where I is the inhibitor concentration and K_i is the slope inhibition constant

$$v/E_t = k_{cat}[A]/(K_a + [A]) \quad (1)$$

$$v/E_t = (k_1[A]/(K_1 + [A]) + (k_2[A])/(K_2 + [A])) \quad (2)$$

$$v/E_t = (k_{cat}[A])/(K_a(1 + (I/K_i)) + [A]) \quad (3)$$

RESULTS

Bioinformatics. As of April 2007, approximately 150 uronate isomerase-like sequences were identified in the NCBI NR database. Except for two archaeal sequences, all are from bacteria, where they play roles in metabolism of D-glucuronic and D-galacturonic acids. As shown in Figure 3, these sequences cluster into three groups. Group 1 contains one of the two functionally characterized uronate isomerases from *B. halodurans* (Bh0705) and that from *E. coli* (not included in the tree). Group 2 contains the structurally characterized uronate isomerase from *Caulobacter crescentus*. Group 3 contains the second functionally characterized uronate isomerase from *B. halodurans* (Bh0493).

Purification and Properties of Bh0705. The gene for the enzyme that encodes for Bh0705 was expressed in BL21-(DE3) cells. The protein Bh0705 was sparingly soluble after the cells were disrupted by sonication. Approximately 14 mg of purified protein was obtained from 14 g of wet cell paste. The metal content of the purified protein was determined to contain 0.9 equiv of Zn^{2+} per subunit using ICP−MS. SDS-PAGE analysis revealed the presence of a single band at approximately 55 kDa for Bh0705. This value is in agreement with the reported gene sequence.

Characterization of Bh0705. The catalytic activity for Bh0705 was determined with D-glucuronate and D-galacturonate as substrates. This enzyme was found to isomerize both of these compounds to their respective ketoacid sugar products, and the kinetic constants are presented in Table 1. The protein Bh0705 was found to be more specific for the isomerization of D-glucuronate as compared to D-galacturonate since the ratio of k_{cat}/K_m for these two compounds is approximately 100. The two active site directed inhibitors, D-arabinaric acid (**1**) and the hydroxamate derivative (**2**), that were designed to mimic the proposed *cis*-enediol intermediate (**3**) were found to be potent competitive inhibitors for the isomerization of D-glucuronate. The inhibition constants are provided in Table 2.

Purification and Properties of Bh0493. The plasmid containing the gene encoding protein Bh0493 was transformed into BL21(DE3) cells, and the protein was expressed after induction with IPTG. The protein was soluble, and significant amounts of protein were obtained and purified to homogeneity. Using ICP−MS, the protein was found to contain 0.5 equiv of Zn^{2+} per subunit. SDS-PAGE indicated that the subunit molecular weight is approximately 49 kDa, which corresponds to the calculated molecular weight of Bh0493 based on the DNA sequence.

Characterization of Bh0493. The enzymatic activity of Bh093 was tested using D-glucuronate and D-galacturonate as substrates. The enzyme catalyzes the isomerization of D-glucuronate to D-fructuronate and D-galacturonate to D-tagaturonate. The double-reciprocal plots for both substrates were biphasic (concave downward), and thus, the initial velocity data were fitted to eq 2. The double-reciprocal plots for the isomerization of D-glucuronate and D-galacturonate are shown in Figure 4. The kinetic constants at high and low concentrations of substrate are provided in Table 1. The structural mimics of the proposed *cis*-enediol intermediate were found to be competitive inhibitors for the isomerization of D-glucuronate when the assays were conducted at low substrate concentrations. The kinetic constants are presented in Table 2.

Structure of Bh0493. The structures of two crystal forms of Bh0493 were determined and refined. A tetragonal crystal form was determined to a resolution of 2.4 Å and a triclinic crystal form at a resolution of 2.0 Å (Table 3). Both crystal

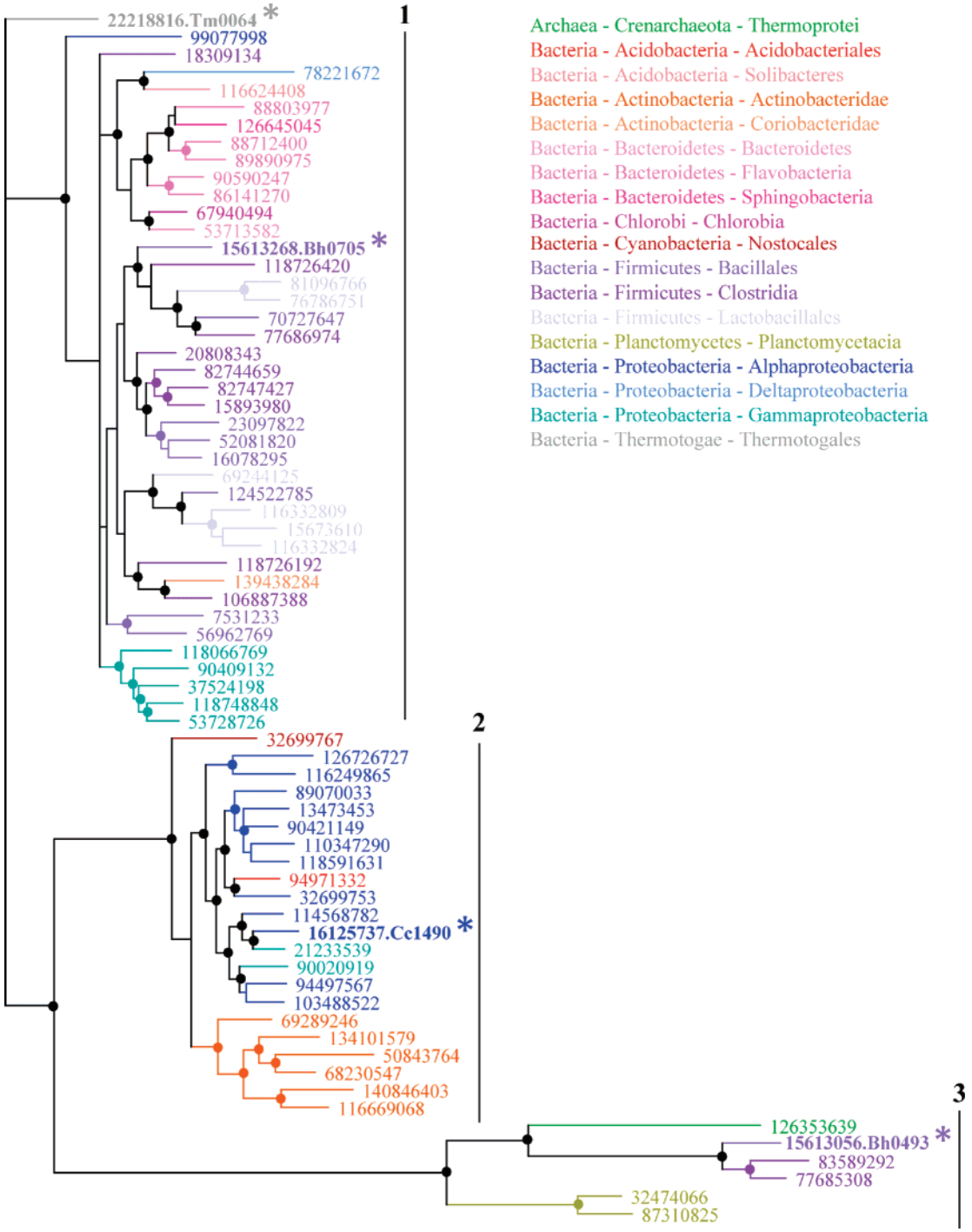


FIGURE 3: Bayesian phylogenetic tree of the proteins in the uronate isomerase-like sequence set. A representative set of sequences was selected by filtering the complete uronate isomerase sequence set to 70% identity. Sequences are listed according to their NCBI gi numbers. The structurally characterized uronate isomerase sequences from *T. maritima*, *B. halodurans*, and *C. crescentus* are indicated with species abbreviations and asterisks (*). The characterized URI from *E. coli* is not shown in the tree but belongs to group 1. Branch confidence values greater than 0.9 are indicated with circles. The three sequence clusters discussed in the text are indicated with numbers.

forms contain similar Bh0493 trimers. There is one protein trimer in the asymmetric unit of the tetragonal form, and there are four protein trimers in the asymmetric unit of the triclinic crystal form. Three molecules in every trimer are connected by a noncrystallographic 3-fold axis. Several ions were observed that sit on the local 3-fold axis of every trimer. The Na^+ ion on the local 3-fold axis in the tetragonal crystal form is coordinated by six water molecules, connected to protein side chains. This crystal form was crystallized from solutions containing NaCl. The triclinic Bh0493 crystal form had no NaCl in the crystallization solutions, and the analogous feature positioned on the local 3-fold axis was interpreted as Zn^{2+} as shown in Figure 5. The tetragonal crystal form additionally contains a Cl^- ion that sits on the local 3-fold axis of the trimer and is surrounded by the side chains of Lys 278 from three adjacent protein molecules. All of these ions additionally stabilize the internal structure of the trimers.

Bh0493 adopts a $(\beta/\alpha)_8$ structural fold with a distorted barrel core. The Bh0493 monomer is composed of a $(\beta/\alpha)_7\beta$-

Table 1: Kinetic Parameters for Bh0493 and Bh0705 with D-Glucuronate and D-Galacturonate[a]

enzyme	substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
Bh0493	D-glucuronate	5.2 ± 0.3	0.05 ± 0.01	(1.1 ± 0.3) × 10^5
		3.6 ± 0.3	15 ± 5	(2.5 ± 0.8) × 10^2
	D-galacturonate	2.4 ± 0.1	0.05 ± 0.01	(4.7 ± 0.8) × 10^4
		0.9 ± 0.1	12 ± 5	72 ± 30
Bh0705	D-glucuronate	9.7 ± 0.1	0.33 ± 0.01	(2.9 ± 0.10) × 10^4
	D-galacturonate	1.2 ± 0.01	3.3 ± 0.1	(3.6 ± 0.12) × 10^2

[a] Kinetic constants were determined at 30 °C, pH 8.0 from a fit of the data to eq 1 or eq 2.

Table 2: Inhibition Constants with Mimics of the *cis*-Enediol Intermediate[a]

enzyme	inhibitor	K_i (μM)
Bh0493	D-arabinaric acid (**1**)	(5.5 ± 0.3) × 10^{-2}
	monohydroxamate derivative (**2**)	2.1 ± 0.1
Bh0705	D-arabinaric acid (**1**)	0.40 ± 0.02
	monohydroxamate derivative (**2**)	18 ± 2

[a] These constants were obtained at pH 8.0, 30 °C from a fit of the data to eq 3.

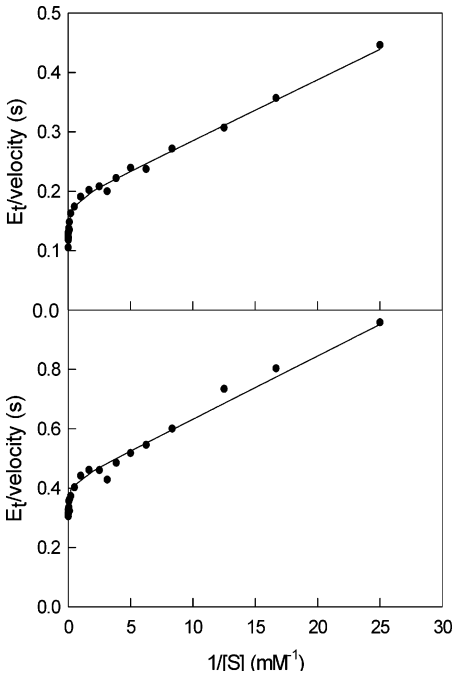


FIGURE 4: Double-reciprocal plot for the isomerization of D-glucuronate (top) and D-galacturonate (bottom) by Bh0493. The initial velocities were fit to eq 2.

barrel with N- and C-terminal extensions of the chain from both sides of the barrel. The following chain segments are included in the eight β-strands of the barrel: β1 (residues 21−28), β2 (137−143), β3 (165−170), β4 (219−224), β5 (253−260), β6 (295−301), β7 (318−323), and β8 (349−353). The N-terminal extension of the chain includes α-helix 6−20. The C-terminal extension includes two long distorted α-helices from residues 360−386 and 393−413. These three long α-helices form a mutual hydrophobic core that protrudes from the main body of the molecule. The long loop between strands β1 and β2 of the barrel contains eight helices. The chain segment 40−122 of this loop contains six helices, sticks out of the molecule, and can be considered as a separate domain B. The long interval between strands β3 and β4 of the barrel contains three helices. The remaining

Table 3: Data Collection and Refinement Statistics

	SeMet Bh0493	Native Bh0493
Data collection		
beamline	NSLS X4A	NSLS X4A
wavelength (Å)	0.97915	0.979
space group	$P4_122$	$P1$
no. of molecules in au	**3**	**12**
unit cell parameters		
a (Å)	133.372	121.274
b (Å)	133.372	120.967
c (Å)	196.821	126.241
α (deg)		78.82
β (deg)		78.67
γ (deg)		80.42
resolution (Å)[a]	2.4 (2.49−2.4)	2.0 (2.07−2.0)
unique reflns	64148	415887
completeness (%)	92.4 (89.7)	90.1 (78.4)
R_{merge}	0.067 (0.435)	0.088 (0.316)
av I/σ	31.5 (3.8)	24.9 (4.4)
Refinement		
resolution	25.0−2.4	25.0−2.0
R_{cryst}	0.231	0.226
R_{free}	0.246	0.239
rmsd for bonds (Å)	0.006	0.006
rmsd for angles (deg)	1.3	1.4
no. of protein atoms	10182	40812
no. of waters	196	963
no. of ions	3 Zn^{2+}, 1 Na^+, 1 Cl^-	16 Zn^{2+}
PDB ID	2Q6E	2QO8

[a] Numbers in parentheses indicate values for the highest resolution shell.

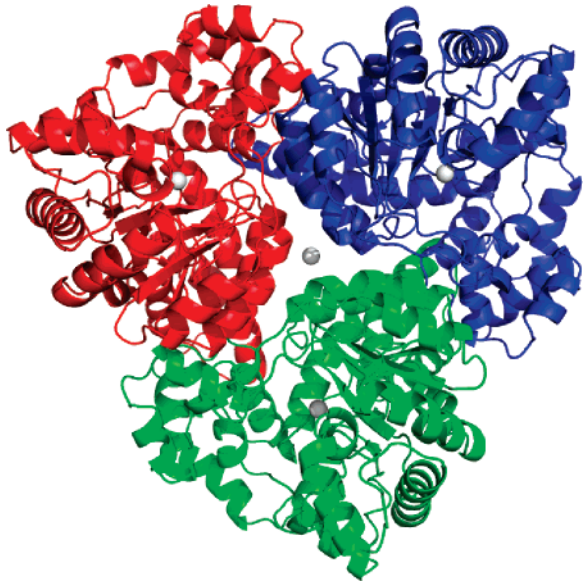


FIGURE 5: Ribbon representation of the trimeric form of Bh0493 where each color represents an individual subunit. The zinc ions are shown as gray spheres.

intervals between the β-strands of the barrel each contain a single helix. The structure of β-strand 8 is distorted. In this β-strand, Ile-350 forms two main chain hydrogen bonds with Ile-321 and Gly-323 from β-strand 7, but there are no hydrogen bonds to β-strand 1, and thus, the barrel is not completely closed.

The active site is located at the C-terminal end of the barrel in large domain A of the Bh0493 monomer and is open to external solvent. The three active sites are positioned at approximately the vertices of the trimer, distant from the 3-fold axis and from each other. The Zn^{2+} ion is bound in the active site of every monomer and is coordinated by His-

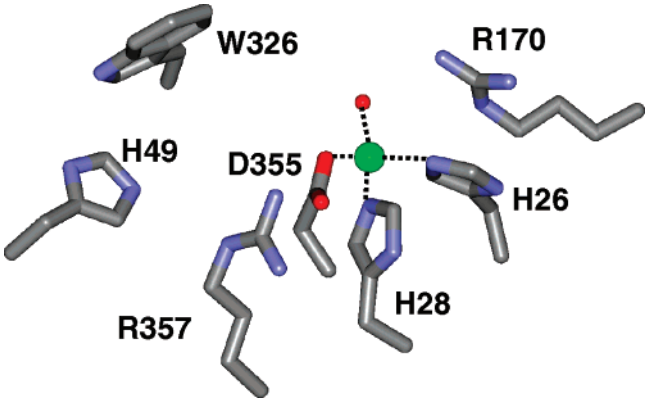


FIGURE 6: Close-up view of the active site region of Bh0493 showing the relative orientation of the zinc (green sphere), the four metal ligands, and other conserved residues that are likely candidates for the binding of substrate and catalytic transformations.

26 and His-28 from the strand β1 and by Asp-355 from the loop at the end of strand β8 as illustrated in Figure 6. The fourth ligand for Zn is a water molecule. The interatomic distances between the Zn in the active site to these four ligands are 2.3, 2.5, 2.3, and 2.0 Å, respectively. The other residues whose side chains may be important in substrate binding are His-49 and Tyr-50 from domain B, Arg-170 from the strand β3, and Arg-357 from the loop after β8.

The Bh0493 trimer is mainly stabilized by interactions between the C-terminal helices 360−386 and 393−413 of one monomer and the domain B (40−122) of the adjacent monomer. These interactions include main chain−main chain hydrogen bonds and side chain−main chain hydrogen bonds. Additional trimer stabilizing effects are interactions between the loops β5−β6, β6−β7, and β7−β8 of every monomer with the corresponding loops of neighboring monomer within the Bh0493 trimer.

DISCUSSION

Since Holm and Sander first identified a diverse set of amidohydrolases related to adenosine deaminase, phosphotriesterase, and urease in 1997 (*12*), increasingly sophisticated sequence search methods as well as the sequencing of additional genomes have revealed that the AHS contains far more diversity than previously thought, both in terms of function and in terms of sequence (Figure 1A). For example, α-amino-β-carboxymuconate-ϵ-semialdehyde decarboxylase (ACMSD) and related enzymes, recently identified as part of the AHS (*31*), catalyze nonoxidative decarboxylase reactions. Although these reactions likely share some mechanistic similarity with that of canonical AHS members, they are substantially different than the amide and ester hydrolysis reactions catalyzed by core superfamily members, and not surprisingly, these sequences are quite different from those of the core superfamily members. As shown by the network representation of the AHS in Figure 1A, at an *E*-value cutoff where most of the amidohydrolase families identified by Holm and Sander are tightly clustered, ACMSD and related sequences form a distinct, although connected, cluster. The uronate isomerase-like sequences are even more divergent than those of ACMSD and the related decarboxylases. Even at the permissive *E*-value cutoff used in Figure 1A, they do not connect to any other AHS sequences. Uronate isomerase is also an outlier of the superfamily in terms of function, being the only characterized superfamily member that catalyzes the interconversion of aldose and ketose functional groups (*13*).

Given currently available tools, enzymes can often be reliably assigned to a superfamily, especially when X-ray structures are available to augment sequence information. However, classifying superfamily sequences into appropriate families each associated with a specific overall reaction can be much more challenging, especially within superfamilies in which many divergent families catalyzing different overall reactions conserve active site residues common to all of the superfamily members. In such cases, even closely related enzymes may catalyze different overall reactions (*32*), and conversely, quite distantly related enzymes may catalyze the same overall reaction (*33*). The problem of functional inference in such superfamilies has been described in some detail for the muconate lactonizing enzyme subgroup of the enolase superfamily (*33*). This is also an issue in the AHS, where, for example, in silico docking approaches were required to identify the function of protein Tm0936 from *T. maritima* (PDB ID 1p1m and 1j6p) as an adenosine and thiomethyl adenosine deaminase (*34*). By sequence similarity, this protein most resembles the large chlorohydrolase and cytosine deaminase subgroup within the AHS but shows no sequence similarity with adenosine deaminases.

Even within the canonical members of the URI family, substantial divergence is apparent, with the pairwise sequence identity among the top three sequences shown in the alignment given in Figure 2 ranging from 35 to 52% and the sequence identities between these sequences and that from *C. crescentus* falling to between 22 and 27%. The new outlier uronate isomerase from *B. halodurans*, Bh0493, is even more distant, with a pairwise sequence identity to the other sequences in Figure 2 ranging from 17 to 22%. As shown by the network representation in Figure 1B, it is only at an *E*-value threshold of 10^{-10} that the outlier uronate isomerases, including Bh0493, connect to the tightly clustered canonical uronate isomerases. At a slightly more stringent *E*-value cutoff of 10^{-12}, the outlier uronate isomerases form a distinct group, losing their connection to the canonical uronate isomerases (not shown). However, because Bh0493 is so distant from even the canonical uronate isomerases, it is useful for identifying minimum sequence requirements for URI activity. The few active site residues conserved across Bh0493 and the canonical uronate isomerases (Figure 2) are likely candidates for residues required for substrate binding and catalysis. These residues include the conserved HxH motif from β-strand 1, the aspartate from β-strand 8, and a WWF motif from β-strand 7. We have demonstrated in unpublished experiments with the uronate isomerase from *E. coli* that residues corresponding to His-26, His-28, His-49, Arg-170, Asp-355, and Arg-357 in Bh0493 are critical for the functioning of this enzyme. These residues are conserved in the sequence alignment of Bh0493 with other authentic uronate isomerases (Figure 2).

The identification of Bh0493 as a uronate isomerase is further supported by the gene context within the genome of *B. halodurans*. Presented in Figure 7 is a schematic representation of the open reading frames in the vicinity of the gene that encodes for Bh0493. The protein Bh0492 is annotated as a tagaturonate oxidoreductase (UxaB), and Bh0490 is an altronate dehydratase (UxaA). The gene for

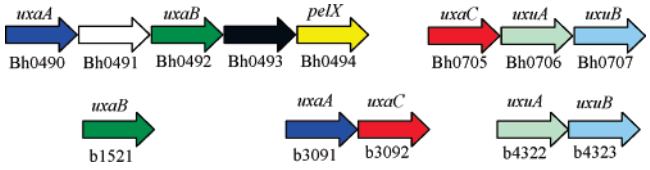


FIGURE 7: Chromosomal arrangement of the genes that code for enzymes involved in the metabolism of D-glucuronate and D-galacturonate. (top) *B. halodurans* and (bottom) *E. coli* K12. The relative lengths of each gene are not drawn to scale.

Bh0492 is homologous to the *uxa*B gene in *E. coli*, and the two proteins have an amino acid sequence identity of 38%. Likewise, Bh0490 has a sequence identity of 54% to altronate dehydratase from *E. coli* (UxaA). The co-location of these three genes suggests an operon for the metabolism of D-galacturonate to pyruvate and D-glyceraldehyde-3-phosphate (Scheme 1). This conclusion is further supported by the annotation of Bh0494 as an exopolygalacturonate lyase (PelX).

In nearly all microorganisms the isomerization of the two uronic acids, D-glucuronate and D-galacturonate, are catalyzed by the same enzyme. This is certainly true for *E. coli* and *T. maritima*. However, in the genome of *B. halodurans*, there is another gene that is annotated as *uxa*C, and apparently, this enzyme also catalyzes the isomerization of uronic acids. The gene for this protein (Bh0705) is found adjacent to two other genes currently annotated as *uxu*B (Bh0707) and *uxu*A (Bh0706). The enzymes encoded by these two genes are expected to be D-mannonate oxidoreductase (UxuB) and D-mannonate dehydratase (UxuA). The amino acid sequence of Bh0705 is 53% identical with the prototypical uronate isomerase from *E. coli* (UxaC, b3092), while Bh0706 is 38% identical with the sequence for the authentic mannonate dehydratase from *E. coli* (UxuA, b4322). However, the protein Bh0707 has no apparent homology to the *uxu*B gene product from *E. coli* (b4323). Nevertheless, this enzyme does have a high sequence identity to the gene product of y*jm*F from *Bacillus subtilis* and *uxu*B from *Bacillus stearothermophilus*. Both of these proteins have been annotated as D-mannonate oxidoreductases (*35*, *36*). Mannonate dehydrogenases generally belong to the long-chain alcohol dehydrogenase superfamily, but the protein sequence of *yjmF* is identified as a member of the short-chain alcohol dehydrogenase superfamily. The protein from *B. subtilis* was first recognized as a mannonate oxidoreductase because it was found in a cluster of genes that are part of the hexuronate catabolic pathway. The identification of Bh0707 as a mannonate dehydrogenase and Bh0706 as a mannonate dehydratase indicates that Bh0705 is in an operon for the metabolism for D-glucuronate. Therefore, it is apparent that in *B. halodurans*, there are separate pathways for the metabolism of D-glucuronate and D-galacturonate. Bh0705 is in the operon for D-glucuronate, whereas Bh0493 is in the operon for D-galacturonate.

The two putative uronate isomerase enzymes from *B. halodurans*, Bh0493 and Bh0705, were expressed, purified, and characterized. It was found that the enzyme Bh0705 will catalyze the isomerization of both D-glucuronate and D-galacturonate. However, the enzyme is significantly more active with D-glucuronate than with D-galacturonate. The value of k_{cat}/K_m for the isomerization of D-glucuronate (2.9×10^4 M^{-1} s^{-1}) is approximately 2 orders of magnitude higher than it is for the isomerization of D-galacturonate (3.6×10^2 M^{-1} s^{-1}). This indicates that Bh0705, unlike the homologous enzyme from *E. coli*, is relatively specific for the isomerization of D-glucuronate. The corresponding enzyme from *E. coli* (b3092) has been shown to isomerize D-glucuronate with a k_{cat}/K_m value of 4×10^5 M^{-1} s^{-1} and D-galacturonate with a k_{cat}/K_m value of 2×10^5 M^{-1} s^{-1} (*13*). Therefore, the lone uronate isomerase from *E. coli* cannot discriminate between these two epimeric sugars.

The enzyme Bh0493 can catalyze the isomerization of D-glucuronate and D-galacturonate with a similar catalytic activity for either compound. The observed double-reciprocal plots for Bh0493 are nonlinear for either substrate with an apparent activation at high substrate concentrations. The dependence of the observed catalytic activity as a function of substrate concentration was fit to eq 2. At low substrate concentrations of D-glucuronate, the value of k_{cat}/K_m is 1.1×10^5 M^{-1} s^{-1}, whereas at low concentrations of D-galacturonate, the value of k_{cat}/K_m is 4.7×10^4 M^{-1} s^{-1}, and thus, there is very little discrimination between these two acid sugars. At saturating concentrations of D-glucuronate, the effective turnover number is ~8.8 s^{-1} ($5.2 + 3.6$ s^{-1}) and somewhat less with D-galacturonate (3.3 s^{-1}). The reason for the nonlinearity in the double-reciprocal plots is unknown, but it could represent a homotropic cooperativity between subunits. These results validate the hypothesis that Bh0705 catalyzes the isomerization reaction in the D-glucuronate pathway and that Bh0493 is responsible for the isomerization reaction that initiates the utilization of D-galacturonate, although this enzyme catalyzes the isomerization of both compounds.

Two compounds that mimic the proposed *cis*-enediol intermediate in the isomerization of D-glucuronate and D-galacturonate were tested as inhibitors of the reactions catalyzed by Bh0493 and Bh0705. D-Arabinaric acid can be considered as a mimic for either compound depending on how the dicarboxylic acid is oriented within the active site. This compound was found to be an excellent competitive inhibitor for the isomerization of D-glucuronate with either enzyme. With Bh0493, the K_i value was found to be 55 nM, whereas with Bh0705, the competitive inhibition constant was 400 nM. The competitive inhibition constants for the monohydroxamate derivative of D-arabinaric acid are considerably weaker for both enzymes relative to the effects of D-arabinaric acid. For Bh0493 and Bh0705, the observed K_i values are 2 and 18 μM, respectively. The monohydroxamate derivative made for this investigation is an analogue of the *cis*-enediol of D-galacturonate rather than of D-glucuronate. This partially explains the weaker binding of the hydroxamate with Bh0705, but it does not explain the weaker binding, relative to D-arabinaric acid, with Bh0493. In any event, the inhibition of these compounds is consistent with the formation of a *cis*-enediol-like reaction intermediate during the isomerization of the aldose substrate to the ketose product.

The first reported crystal structure in the uronate isomerase subfamily of the AHS was from *T. maritima*. The enzyme Bh0493, with an overall sequence identity of 19% to Tm0064, is structurally similar to this enzyme with an rmsd of 2.3 Å. Like Tm0064, the quaternary structure of Bh0493 is a homotrimer organized as a pinwheel (Figure 5). Members of the AHS are generally found in oligomerization states with

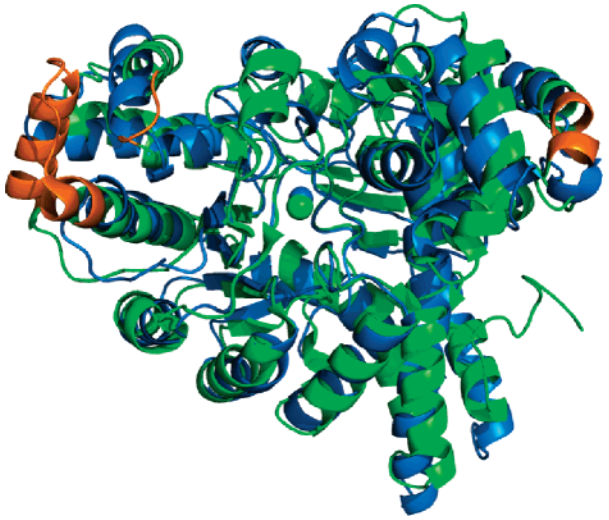


FIGURE 8: Overlay of the ribbon representation of the two structures, Tm0064 (green) and Bh0493 (blue). In orange are the nonoverlapping regions between the two structures. The green sphere represents the putative metal ion in Tm0064, and the blue sphere represents the zinc ion in Bh0493.

an even number of subunits. For example, phosphotriesterase and dihydroorotase exist as homodimers (*37*, *38*). NagA from *E. coli* is tetrameric (*39*), whereas cytosine deaminase and isoapartyl dipeptidase adopt hexameric and octameric oligomerization states (*40*, *41*). Uronate isomerase is thus the first example for a trimeric quaternary arrangement of subunits within this superfamily of enzymes.

Structural comparisons of Bh0493 with Tm0064 indicate that the tertiary fold of Bh0493 is also organized into two distinct domains (Figure 8). Domain A assumes the amidohydrolase-like structural fold (TIM-barrel) and is the site for the catalytic transformations. Domain B is composed mainly of α-helices and loops and possesses a unique structural fold. The structural alignment of domain B for these two enzymes indicates a nearly identical fold with the exception of one helix being slightly longer in the *T. maritima* enzyme. Comparison of domain A also reveals a highly conserved structural fold in these two enzymes. However, one major difference between the two structures is an extra helix in Tm0064 of over 20 residues that occurs at the end of β-strand 4. The structural alignment also demonstrates that the histidine contributed from the end of β-strand 5 (highly conserved in nearly all of the enzymes in the AHS) is not conserved in Bh0493, suggesting that this histidine is not important for catalytic activity. Aside from the HxH motif at the end of β-strand 1 and the aspartate at β-strand 8, several other conserved residues are present in this alignment, including His-49, Arg-170, Arg-357, and the WWF motif at the end of β-strand 7. Although the details of the chemical mechanism for the enzymatic isomerization of D-glucuronate and D-galacturonate are uncertain, it is quite likely that these specific residues will be found to play critical roles in the catalytic activity of the uronate isomerases. A sequence alignment including all uronate isomerase-like sequences shows that the residues identified previously are highly (in most cases, completely) conserved, supporting their importance for catalysis and/or substrate recognition (Figure 2). Additional positions within the active site appear to be conserved within one of the three uronate isomerase groups identified based on the phylogenetic tree but not across the entire uronate isomerase sequence set (Figure 2), possibly indicating different strategies for substrate recognition within each group.

B. halodurans may not be the only organism with separate enzymes for the isomerization of D-glucuronate and D-galacturonate. As noted previously, there are five organisms that have two genes for a URI family protein and one organism that has three such genes. *Alkaliphilus metalliredigenes* QYMF, also an alkaliphilic bacterium like *B. halodurans*, has two members in the uronate isomerase sequence set, one of which (gi: 77686974) has 48% sequence identity to the prototypical uronate isomerase of *E. coli* and the other of which (gi: 77685308) has 57% identity to the outlier uronate isomerase Bh0493. The gene encoding protein gi: 77686974 is located near homologues of *E. coli* mannonate dehydrogenase (UxuB, gi: 77686972) and mannonate dehydratase (UxuA, gi: 77686973), enzymes specific for the breakdown of D-glucuronate. Although the *A. metalliredigenes* genome has not been fully assembled, two fragments homologous to *E. coli* UxaB are found near 77685308, the second putative UxaC (gi: 77685309 and 77685310). Full-length homologues of *E. coli* UxaB and UxaA are also found in the genome (gi: 77686989 and 77686988, respectively). The homology of the two putative *A. metalliredigens* UxaCs to the *B. halodurans* UxaCs, along with genomic context information, suggest that *A. metalliredigenes* has distinct isomerases for the metabolism of D-glucuronate and D-galacturonate.

Clostridium beijerincki NCIMB 8052 and *Saccharophagus degradans* 2−40 also contain multiple uronate isomerase-like sequences, with a genome context that suggests that one may be specific for the isomerization of D-glucuronate and the other for D-galacturonate. However, the putative D-galacturonate isomerases from these organisms do not cluster with those from *B. halodurans* and *A. metalliredigenes* or with each other. If they are, indeed, specific for D-galacturonate, this specificity appears to have evolved multiple times within the uronate isomerase-like group.

Other proteins homologous to the outlier uronate isomerase Bh0493 include protein sequences from *Moorella thermoacetica* ATCC 39073 (gi: 83589292), *Blastopirellula marina DSM3645* (gi: 87310825), *Caldivirga maquilingensis* IC-167 (gi: 126353639), and *Rhodopirellula baltica SH 1* (gi: 32474066). These organisms are unusual because they do not appear to have a protein homologous to the prototypical uronate isomerase in *E. coli*. The *M. thermoacetica* Bh0493 homologue is located near genes homologous to *E. coli* UxuB and UxuA, indicating that it may be a genuine uronate isomerase. The remaining sequences, however, cannot be validated based on genomic context. The function of the *C. maquilingensis* sequence is particularly open to question, as this is the only nonbacterial organism with a uronate isomerase-like sequence.

From kinetic and structural studies, we have determined that the protein Bh0493 from *B. halodurans* is a uronate isomerase and a highly divergent member of the AHS. In *B. halodurans*, the gene encoding uronate isomerase had been identified previously as Bh0705 based on sequence identity. Initial velocity studies of Bh0705 and Bh0493 with the substrates D-glucuronate and D-galacturonate indicate that Bh0705 is relatively specific for D-glucuronate, while Bh0493 can utilize either substrate with a similar efficiency. Bh0493

is located adjacent to genes that are annotated as tagaturonate oxidoreductase, altronate dehydratase, and exopolygalacturonate lyase (an operon for D-galacturonate metabolism), while the gene for Bh0705 is located near mannonate dehydratase and mannonate oxidoreductase (an operon for d-glucuronate metabolism). Sufficient information is not available to hypothesize as to why *B. halodurans* and some other organisms have more than one URI gene or whether this leads to a selective advantage for these organisms. Nor is the evolutionary path that connects this highly divergent family to the AHS clear. However, identification of this outlier Bh0493 sequence as a uronate isomerase extends the boundaries of the superfamily further into the "twilight" of functional and structural divergence than has been previously recognized.

SUPPORTING INFORMATION AVAILABLE

Full alignment of amino acid sequences for uronate isomerases. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Gerlt, J. A., and Babbitt, P. C. (2000) Can sequence determine function? *Genome Biol. 5*, 1−10.
2. Eisen, J. A. (1998) Phylogenomics: Improving functional predictions for uncharacterized genes by evolutionary analysis, *Genome Res. 8*, 163−167.
3. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D., and Yeates, T. O. (1999) Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles, *Proc. Natl. Acad. Sci. U.S.A. 96*, 4285−4288.
4. Marcotte, E. M., Pelligrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999) Detecting protein function and protein−protein interactions from genome sequences, *Science (Washington, DC, U.S.) 285*, 751−753.
5. Chia, J., and Kolatkar, P. R. (2004) Implications for domain fusion protein−protein interactions based on structural information, *BMC Bioinf. 5*, 161.
6. Enright, A. J., Iliopoulos, I., Kyrpides, N. C., and Ouzounis, C. A. (1999) Protein interaction maps for complete genomes based on gene fusion events, *Nature (London, U.K.) 402*, 86−90.
7. Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D., and Maltsev, N. (1999) The use of gene clusters to infer functional coupling, *Proc. Natl. Acad. Sci. U.S.A. 96*, 2896−2901.
8. Marti-Arbona, R., Xu, C., Steele, S., Weeks, A., Kuty, G. F., Siebert, C. M., and Raushel, F. M. (2006) Annotating enzymes of unknown function: *N*-Formimino-L-glutamate deiminase is a member of the amidohydrolase superfamily, *Biochemistry 45*, 1997−2005.
9. Friedberg, I., Jambon, M., and Godzik, A. (2006) New avenues in protein function prediction, *Protein Sci. 15*, 1527−1529.
10. Roodveldt, C., and Tawfik, D. S. (2005) Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily, *Biochemistry 44*, 12728−12736.
11. Siebert, C. M., and Raushel, F. M. (2005) Structural and catalytic diversity within the amidohydrolase superfamily, *Biochemistry 44*, 6383−6391.
12. Holm, L., and Sander, C. (1997) An evolutionary treasure: Unification of a broad set of amidohydrolases related to urease, *Proteins 28*, 72−82.
13. Williams, L., Nguyen, T., Li, Y., Porter, T. N., and Raushel, F. M. (2006) Uronate isomerase: A nonhydrolytic member of the amidohydrolase superfamily with an ambivalent requirement for a divalent metal ion, *Biochemistry 45*, 7453−7462.
14. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389−3402.
15. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks, *Genome Res. 13*, 2498−2504.
16. Pegg, S. C., Brown, S. D., Ojha, S., Seffernick, J., Meng, E. C., Morris, J. H., Chang, P. J., Huang, C. C., Ferrin, T. E., and Babbitt, P. C. (2006) *Biochemistry 45*, 2545−2555.
17. Edgar, R. C. (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity, *BMC Bioinf. 5*, 113.
18. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Gerrin, T. E. (2004) UCSF Chimera: A visualization system for exploratory research and analysis, *J. Comput. Chem. 25*, 1605−1612.
19. Jewett, A. I., Huang, C. C., and Ferrin, T. E. (2003) MINRMS: An efficient algorithm for determining protein structure similarity using root-mean-squared distance, *Bioinformatics 19*, 625−634.
20. Li, W., Jaroszewski, L., and Godzik, A. (2002) Tolerating some redundancy significantly speeds up clustering of large protein databases, *Bioinformatics 18*, 77−82.
21. Altekar, G., Dwarkadas, S., Huelsenbeck, J. P., and Ronquist, F. (2004) Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference, *Bioinformatics 20*, 407−415.
22. Ronquist, F., and Huelsenbeck, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models, *Bioinformatics 19*, 1572−1574.
23. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, in *Methods in Enzymology* (Carter, C. W. J., Sweet, R. M., Abelson, J. N., and Simon, M. I., Eds.) pp 307−326, Academic Press, New York.
24. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 55*, 849−861.
25. Terwilliger, T. C. (2000) Maximum-likelihood density modification, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 56*, 965−972.
26. Perrakis, A., Morris, R, and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol. 6*, 458−463.
27. Jones, A. T. (1985) Interactive computer graphics: FRODO, *Methods Enzymol. 115*, 157−171.
28. Brunger, A. T., Adams, P. D., Clore, G. M., DeLAno, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography and NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54*, 905−921.
29. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) Likelihood-enhanced fast translation functions, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 61*, 458−464.
30. Cook, P. I., and Cleland, W. W. (2007) Initial velocity studies in the absence of added inhibitors, in *Enzyme Kinetics and Mechanism*, p 111, Garland Science Publishing, New York.
31. Liu, A., and Zhang, H. (2006) Transition metal-catalyzed nonoxidative decarboxylation reactions, *Biochemistry 45*, 10407−10411.
32. Seffernick, J. L., de Souza, M. L., Sadowsky, M. J., and Wackett, L. P. (2001) Melamine deaminase and atrazine chlorohydrolase: 98% identical but functionally different, *J. Bacteriol. 183*, 2405−2410.
33. Glasner, M. E., Fayazmanesh, N., Chiang, R. A., Sakai, A., Jacobson, M. P., Gerlt, J. A., and Babbitt, P. C. (2006) Evolution of structure and function in the *o*-succinylbenzoate synthase/*N*-acylamino acid racemase family of the enolase superfamily, *J. Mol. Biol. 360*, 228−250.
34. Hermann, J., Marti-Arbona, R., Fedorov, E., Fedorov, A., Almo, S., Shoichet, B. K., and Raushel, F. M. (2007) Structure-based activity prediction for an enzyme of unknown function, *Nature (London, U.K.) 448*, 775−779.
35. Rivolta, C., Soldo, B., Lazarevic, V., Joris, B., Mauel, C., and Karamata, D. (1998) A 35.7 kb DNA fragment from the *Bacillus subtilis* chromosome containing a putative 12.3 kb operon involved in hexuronate catabolism and a perfectly symmetrical hypothetical catabolite-responsive element, *Microbiology (Reading, U.K.) 144*, 877−884.
36. Shulami, S., Gat, O., Sonenshein, A. L., and Shoham, Y. (1999) The glucuronic acid utilization gene cluster from *Bacillus stearothermophilus* T-6, *J. Bacteriol. 181*, 3695−3704.
37. Benning, M. M., Kuo, J. M., Raushel, F. M., and Holden, H. M. (1994) Three-dimensional structure of phosphotriesterase: An enzyme capable of detoxifying organophosphate nerve agents, *Biochemistry 33*, 15001−15007.

38. Thoden, J. B., Phillips, G. N., Neal, T. M., Raushel, F. M., and Holden, H. M. (2001) Molecular structure of dihydroorotase: A paradigm for catalysis through the use of a binuclear metal center, *Biochemistry 40*, 6989−6997.
39. Hall, R. S., Brown, S., Fedorov, A. A., Fedorov, E. V., Xu, C., Babbitt, P. C., Almo, S. C., and Raushel, F. M. (2007) Structural diversity within the mononuclear and binuclear active sites of *N*-acetyl-D-glucosamine-6-phosphate deacetylase, *Biochemistry 46*, 7953−7962.
40. Ireton, G. C., McDermott, G., Black, M. E., and Stoddard, B. L. (2002) The structure of *Escherichia coli* cytosine deaminase, *J. Mol. Biol. 315*, 687−697.
41. Thoden, J. B., Marti-Arbona, R., Raushel, F. M., and Holden, H. M. (2003) High-resolution X-ray structure of isoaspartyl dipeptidase from *Escherichia coli*, *Biochemistry 42*, 4874−4882.

BI7017738