Insights into the Mechanism of Catalysis by the P−C Bond-Cleaving Enzyme Phosphonoacetaldehyde Hydrolase Derived from Gene Sequence Analysis and Mutagenesis†

Angela S. Baker,‡ Michael J. Ciocci,‡ William W. Metcalf,§,|| Jaebong Kim,‡ Patricia C. Babbitt,⊥ Barry L. Wanner,§ Brian M. Martin,# and Debra Dunaway-Mariano*,@

Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742, Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446, Clinical Neuroscience Branch, IRP, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892, and Department of Chemistry, University of New Mexico, Albuquerque, New Mexico 87131-1096

Received October 29, 1997; Revised Manuscript Received March 5, 1998

ABSTRACT: Phosphonoacetaldehyde hydrolase (phosphonatase) catalyzes the hydrolysis of phosphonoacetaldehyde to acetaldehyde and inorganic phosphate. In this study, the genes encoding phosphonatase in *Bacillus cereus* and in *Salmonella typhimurium* were cloned for high-level expression in *Escherichia coli*. The kinetic properties of the purified, recombinant phosphonatases were determined. The Schiff base mechanism known to operate in the *B. cereus* enzyme was verified for the *S. typhimurium* enzyme by phosphonoacetaldehyde−sodium borohydride-induced inactivation and by site-directed mutagenesis of the catalytic lysine 53. The protein sequence inferred from the *B. cereus* phosphonatase gene was determined, and this sequence was used along with that from the *S. typhimurium* phosphonatase gene sequence to search the primary sequence databases for possible structural homologues. We found that phosphonatase belongs to a novel family of hydrolases which appear to use a highly conserved active site aspartate residue in covalent catalysis. On the basis of this finding and the known stereochemical course of phosphonatase-catalyzed hydrolysis at phosphorus (retention), we propose a mechanism which involves Schiff base formation with lysine 53 followed by phosphoryl transfer to aspartate (at position 11 in the *S. typhimurium* enzyme and position 12 in the *B. cereus* phosphonatase) and last hydrolysis at the imine C(1) and acyl phosphate phosphorus.

Phosphonates constitute a class of naturally occurring organophosphorus compounds which differ in structure from the more predominate phosphate esters by the direct linkage of phosphorus to carbon (*1*−*3*). The P−C bond is resistant to acid- and base-catalyzed hydrolysis as well as to enzymes which catalyze the cleavage of P−O−C bonds in phosphate esters. Their similarity to phosphate ester structure, coupled with their stability, gives phosphonates a variety of biological properties which we have seen exploited in both natural products (e.g., antibiotics and phosphatase-resistant membrane components) and synthetics (e.g., insecticides, herbicides, and pharmaceuticals). Although phosphonates have been found in numerous organisms ranging from bacteria to mammals, active synthesis of phosphonates has thus far been demonstrated in only certain bacteria (*4*−*6*), a protozoan *Tetrahymena pyriformis* (*7*, *8*), and a mollusk *Mytilus edulis* (*9*). Phosphonate degradative pathways, on the other hand, have been demonstrated only in bacteria (*1*−*3*).

It has been speculated that phosphonate natural products are a vestige of an earlier era on earth when reduced forms of organophosphorus compounds predominated over organophosphates (*1*−*3*, *10*). If this is true, then the phosphonohydrolases are likely to have evolved long ago to allow the cycling of phosphorus between phosphonates and phosphates. Presently, three P−C bond-cleaving enzymes are known. The C−P lyase, which has a very broad substrate specificity, cleaves the C−P bond homolytically, yet ultimately produces inorganic phosphate and the corresponding hydrocarbon (*11*−*14*). The C−P lyase is the most widely distributed bacterial phosphonate-degrading enzyme, and it appears to serve as the main route for inorganic phosphate extraction from environmental phosphonates.

The second P−C bond-cleaving enzyme, phosphonoacetate hydrolase, was recently discovered in a strain of *Pseudomonas fluorescens* isolated from a waste/sludge treatment plant by screening for growth on phosphonoacetate (an antiviral/antibiotic agent) as the sole carbon and phosphorus source. This enzyme is specific for phosphonoacetate, producing, via a metal ion-assisted hydrolytic cleavage of the P−C bond, inorganic phosphate and acetate (*15*). The extent of its

† This work was supported by NIH Grants GM-36360 (D.D.-M.) and GM-35392 (B.L.W.) and Department of Energy Grant DE-FG03-96ER62269 (P.C.B.).

* Corresponding author. E-mail: dd39@unm.edu. Telephone: (505) 277-3383. Fax: (505) 277-6202.

‡ University of Maryland.

§ Purdue University.

|| Current address: Department of Microbiology, University of Illinois, Urbana, IL 61801-3704.

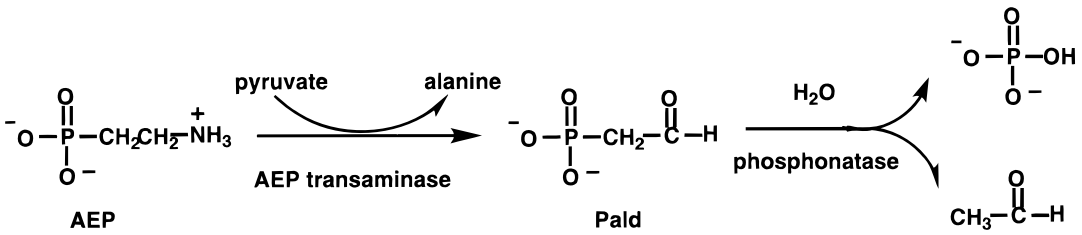
⊥ University of California.

National Institutes of Health.

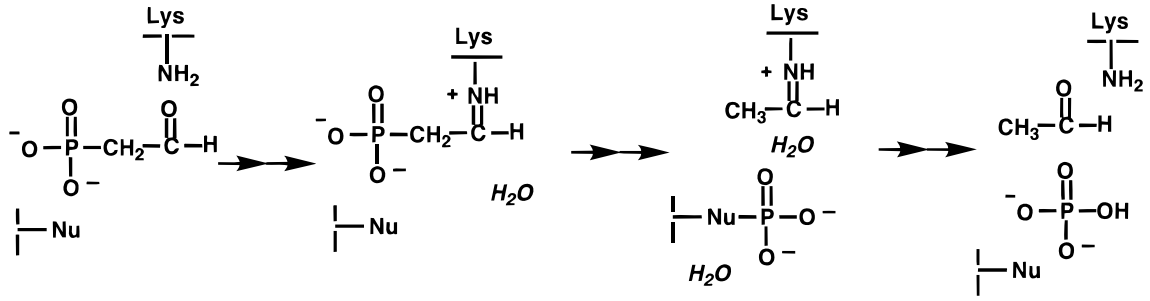
@ University of New Mexico.

S0006-2960(97)02677-9 CCC: $15.00
Published on Web 06/12/1998

Scheme 1: Chemical Steps of the 2-Aminoethyl Phosphonate Pathway (*18*, *19*)



Scheme 2: Proposed Mechanism of Phosphonatase Catalysis (*26*)



distribution throughout bacterial populations is not yet known.

The third P−C bond-cleaving enzyme, phosphonoacetaldehyde hydrolase (trivial name, phosphonatase) (*16*, *17*), is the topic of this paper. Phosphonatase serves to mediate the second step of the catabolism of 2-aminoethyl phosphonate (AEP)[1] (Scheme 1) (*18*, *19*), the most ubiquitous and abundant of the naturally occurring phosphonates (*1*). Phosphonatase is found in bacteria which specialize in AEP utilization, converting it into usable forms of nitrogen (alanine) and carbon (acetaldehyde) as well as phosphorus (inorganic phosphate).

Mechanistic studies carried out on *Bacillus cereus* phosphonatase have unveiled a remarkable catalytic strategy (represented in Scheme 2) wherein an unprotonated active site lysine residue ($pK_a = 5.9$) forms a Schiff base with phosphonoacetaldehyde (Pald), thus activating it for P−C bond cleavage (*20*−*23*).[2] The chemistry of this first partial reaction is analogous to that observed in acetoacetate decarboxylase catalysis wherein an uncharged active site lysine residue forms a Schiff base with acetoacetate, activating it for C−C bond cleavage (*24*, *25*).

The formation of phosphate from Pald occurs, in the second partial reaction, with overall retention of stereochemistry at phosphorus (as determined using chiral [^{17}O,^{18}O]-thiophosphonoacetaldehyde as the substrate), suggesting that the phosphoryl group is first transferred to an active nucleophile before it is intercepted by solvent water (*26*). The formation of a phosphoenzyme during the course of phosphoryl transfer is a common mechanistic theme among phosphotransferases, particularly among phosphatases.

To gain further insight into the origin of the phosphonatase structure and mechanism, our goal was to clone and sequence the encoding gene and then to determine the X-ray crystal structure of the recombinant enzyme. Recently, one of our laboratories isolated the AEP pathway operon of *B. cereus* (*27*, *28*) and the other, the AEP pathway operon of *Salmonella typhimurium* (*29*−*31*). In the collaborative study reported in this paper, we cloned the phosphonatase genes from these two operons for high-level expression in *Escherichia coli*, determined the gene sequences, and compared the kinetic properties and catalytic pathways of the two purified enzymes. With the recent entry of the *Pseudomonas aeruginosa* phosphonatase sequence (GB access code U45309) (*32*) into GenBank, we are able to construct a three-sequence alignment for identification of conserved residues. These sequences were then used in our search of the protein sequence data banks for possible structural homologues that might provide us with clues regarding the evolution of phosphonatase catalysis. The results obtained and detailed in the text which follows suggest that phosphonatase is a member of a superfamily of enzymes, predominately hydrolases, whose basic backbone scaffold supports an active site Asp residue functioning in nucleophilic−covalent catalysis in some, if not all, family members.

MATERIALS AND METHODS

Determination of the N-Terminal Sequence of Phosphonatase Purified from B. cereus AI-2 and the S. typhimurium Phosphonatase Purified from E. coli Transformed with pWM67. The phosphonatase from *B. cereus* AI-2 was purified to homogeneity by using the procedure described in Olsen et al. (*21*). The *S. typhimurium* phosphonatase from an *E. coli* BW14329 transformant carrying pWM67 (*29*) was purified from cell free extract using the protoamine sulfate and ammonium sulfate precipitation steps described below followed by three column chromatographic steps (each carried out at pH 7.5 and 4 °C). These steps included DEAE-Sephadex A-50−120 column chromatography with a linear gradient of NaCl (0.15 to 0.4 M) in buffer A [50 mM HEPES buffer (pH 7.5) containing 10 mM $MgCl_2$ and 0.1 mM DTT], Sephadex G-150 column chromatography with buffer A, and hydroxyapatite column chromatography with a linear gradient

[1] Abbreviations: Pald, phosphonoacetaldehyde; AEP, 2-aminoethylphosphonate; ATPase, adenosine-5′-triphosphate phosphohydrolase; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonate; DTT, dithiothreitol; NADH, dihydronicotinamide adenine dinucleotide.

[2] Phosphonatase is known to require Mg^{2+} for catalytic activity. The Mg^{2+} does not participate directly in catalysis but rather functions in a structural role, stabilizing the active homodimer (*20*).

of K_2HPO_4 (0.00 to 0.004 M) in buffer A. The purified proteins were chromatographed on a SDS−PAGE gel and then blotted onto a polyvinyl difluoride transfer membrane (Millipore) using a Hoefer Semiphor transfer blotter at 0.8 mA/cm^2 for 45 min. The N-terminal sequence was determined by Edman degradation using an Applied Biosytem model 470 gas phase protein sequencer. The N-terminal sequence found for the *B. cereus* AI-2 phosphonatase is M-K-I-E-A-V-I-F-D-W-A-G-T-T-V-D-Y. The N-terminal sequence found for the *S. typhimurium* phosphonatase is M-N-R-I-H-A-V-I-L.

Cloning the B. cereus Phosphonatase Gene in E. coli for Sequence Analysis (Generation of pT7T3a18-HindIII/PstI-B. cereus and pSE 420-KpnI/NcoI-B. cereus Clones). DNA was isolated according to standard methods (*14*) from cells harvested from a 100 mL late log phase culture of *B. cereus* strain AI-2 (kindly provided by H. Rosenberg and grown as described in ref *16*). Restriction enzyme digests of the purified DNA were separated on agarose gels and analyzed by probe [^{32}P-labeled 5′-GGT GTT GCA ATT ACA GCA GAA GAA GCA CGT AAA CCA ATG GGT CTT CTT AAA ATT GAT CAT GTT CGT-3′ designed from the sequence of the active peptide reported by Olsen et al. (*21*)] hybridization to Southern blots (*33*). *Hin*dIII and *Pst*I (600 and 1200 units, respectively) were used to digest the purified DNA (1 mg in 1 mL of supplied buffer at 37 °C for 3 h), thus forming a 4 kb fragment containing the hybridization site. Following partial purification on an agarose gel, 180 ng of crude 4 kb *Hin*dIII/*Pst*I fragment and 100 ng of *Hin*dIII/*Pst*I-digested pT7T3a18 vector (Life Technologies) were ligated with 3 units of T4 DNA ligase (Promega) in 10 μL of supplied buffer (16 °C for 16 h). One microliter of the mixture was used to transform 20 μL of *E. coli* DH5a competent cells (Gibco). Positive colonies, grown on LB ampicillin plates treated with IPTG and X-gal, were identified with the ^{32}P-labeled oligonucleotide probe using Southern blot techniques and confirmed by *Hin*dIII/*Pst*I digest analysis of plasmid minipreps (Wizard Plus Minipreps, Promega). The *Hin*dIII/*Pst*I insert (2 μg) obtained from a positive recombinant (pT7T3a18-*Hin*dIII/*Pst*I-*B. cereus*) was treated along with 2 μg of the sequencing vector pSE 420 (Invitrogen) with 10 units each of *Kpn*I and *Nco*I (20 μL of Multicore buffer, Promega) at 37 °C for 2 h. Following purification on an agarose gel, the DNA fragments were ligated with 3 units of T4 DNA ligase in the supplied buffer. Following the transformation of competent *E. coli* DH5a cells, plating, and colony screening, a positive recombinant was identified and named pSE 420-*Kpn*I/*Nco*I-*B. cereus*. DNA sequencing was carried out using the dideoxy-chain termination method (*34*) in conjunction and either the Sequenase kit or (for regions where secondary structure proved problematic) the Taquence kit (U.S. Biochemicals).

Construction of the B. cereus Phosphonatase−pKK223-3 Subclone for Expression in E. coli JM105 Cells and the S. typhimurium Phosphonatase−pSE420 Subclone for Expression in E. coli TOP 10 Cells. B. cereus Phosphonatase Expression System. The phosphonatase gene was subcloned in the expression vector pKK223-3 by using PCR techniques. The pT7T3a18-*Hin*dIII/*Pst*I-*B. cereus* clone (described above) was used as the template, and the oligonucleotides 5′-CGAATTCATGGACAGAA-3′ and 5′-GTCTGCAGGT-CATGATCC-3′ were used as primers in the PCR to create an *Eco*RI restriction site at the 5′-end and a *Pst*I restriction site at the 3′-end of the phosphonatase gene. A 100 μL reaction solution, contained in a 0.5 mL thin-walled GeneAMP microfuge tube and consisting of 1 μg of plasmid DNA and (from the Promega PCR kit) dATP, dCTP, dGTP, and dTTP (each at 200 μM), 20 pmol of each primer, 2.5 units of Taq DNA polymerase, and 2.5 mM $MgCl_2$ in the supplied buffer, was overlaid with 50 mL of autoclaved silica oil. Temperature cycles were performed with a Thermolyne model Temptronic thermocycler. The PCR program consisted of 3 min at 94 °C followed by 25 cycles with denaturation at 94 °C, primer annealing at 45 °C for 1 min, extension at 72 °C for 1 min, and 5 min at 72 °C between cycles. The PCR product was purified by electrophoresis (low-melting point 1% agarose gel followed by treatment with the GENECLEAN II kit) and digested along with the pKK223-3 vector (5 μg each) with *Eco*RI and *Pst*I (20 units each in Promega Multicore buffer at 37 °C for 1 h). Following purification, the insert and vector were ligated with 6 units of T4 DNA ligase in 40 μL of supplied buffer and then used to transform competent *E. coli* JM105 cells which were spread on LB−agar plates containing 50 μg/mL ampicillin. Colonies resulting from overnight incubation at 37 °C were screened by *Eco*RI/*Pst*I restriction analysis of the isolated plasmid. The sequence of the cloned gene was verified by DNA sequencing.

S. typhimurium Phosphonatase Expression System. *E. coli* BW14329 containing the *S. typhimurium phnR* to *phnX* gene cluster (*29*) was used as the source of the phosphonatase gene for cloning into the expression vector pSE420. Introduction of a unique *Nco*I restriction site at the initiation site of the phosphonatase gene was achieved by PCR using as outside primers 5′-CGCCATTCGTACCGCCC-3′ and 5′-CGCCATGGCCCAGGGACCGACC-3′ and as inside primers 5′-CGTGAATACGTGCCATGGTGGTTC-3′ and 5′-GGAACCACCATGGACGTATTCACG-3′. The 100 μL reaction solution, contained in a 0.5 mL thin-walled GeneAMP microfuge tube and consisting of 1 μg of plasmid DNA and dATP, dCTP, dGTP, and dTTP (each at 200 μM), 20 pmol of each primer, 2.5 units of Taq DNA polymerase, and 2.5 mM $MgCl_2$ in the supplied buffer, was overlaid with 50 μL of autoclaved silica oil and subjected to thermocycling. Five micrograms of the resulting 1 kb product of the secondary PCR and 5 μg of the pSE420 vector were digested with 20 units each of *Nco*I and *Kpn*I in Promega Multicore buffer (37 °C for 1 h). The purified insert and linearized vector were ligated with 6 units of T4 DNA ligase for 16 h at 16 °C and then used to transform competent *E. coli* TOP 10 cells. Positive colonies grown on the LB−ampicillin plates were identified by *Nco*I and *Kpn*I restriction digest analysis of plasmid minipreps. A clone having the correct gene orientation was identified on the basis of a 95-mer produced by *Cla*I digestion and named pSE420-*S. tryp.*-*Nco*I/*Kpn*I-phosphonatase. The gene sequence was verified by DNA sequencing.

Purification of the S. typhimurium Phosphonatase Expressed by the pASB2 Subclone in E. coli TOP 10 Cells and the B. cereus Phosphonatase Expressed by the pKK223-3 Subclone in E. coli JM105 Cells. A 150 mL culture of *E. coli* TOP 10 cells transformed with pASB2 (containing the *S. typhimurium* phosphonatase gene) or the *E. coli* JM105 cells transformed with pKK223-3 (containing the *B. cereus*

phosphonatase gene) were added to 6 L of LB−ampicillin (50 μg/mL) and incubated at 37 °C for 2.5 h until the absorbance of the culture measured at 600 nm reached 0.6−1.0. At this point, IPTG was added to a final concentration of 1 mM and the cells were incubated while they were shaken at 37 °C overnight and then harvested by centrifugation. Phosphonatase was purified from the supernatant of lysed cells [25 g of cell paste resuspended in 10 volumes of buffer A [50 mM HEPES buffer (pH 7.5) containing 10 mM $MgCl_2$ and 0.1 mM DTT] and passed through a French pressure cell press at 12 000 psi at 4 °C] using a modification of the method reported by LaNauze et al. (*20*). Accordingly, treatment of the supernatant with a 0.45% solution of protoamine sulfate (0.2 v/v) in buffer A at 0 °C was followed by centrifugation and then adjustment of the supernatant to 70% saturation with solid ammonium sulfate. The precipitated protein was collected by centrifugation, dialyzed overnight against buffer A, applied to a column of DEAE-Sephadex A-50−120 (5.0 cm × 60 cm) equilibrated with buffer A, and then eluted at 4 °C using a 1.8 L linear gradient of NaCl (0.15 to 0.40 M) in buffer A. Phosphonatase-containing fractions were pooled, concentrated at 4 °C with an Amicon protein concentrator fitted with an Amicon PM-10 membrane, and then chromatographed on a Sephadex G-75 column (2.0 cm × 52 cm) with buffer A at 4 °C. Phosphonatase-containing fractions were pooled, concentrated, and chromatographed on a Sephadex G-150 column (2.0 cm × 90 cm) with buffer A at 4 °C. The concentration of pooled phosphonatase fractions provided homogeneous enzyme as determined by SDS−PAGE.

Determination of the Steady State Kinetic Constants (k_{cat} and K_m) for Recombinant S. typhimurium and the B. cereus Phosphonatase. Acetaldehyde formation in 1 mL reaction solutions at 25 °C containing phosphonatase, 25−400 μM Pald, 0.13 mM NADH, 1 unit of yeast alcohol dehydrogenase, and 5 mM $MgCl_2$ in 100 mM HEPES (pH 7.0) was monitored at 340 nm ($\Delta\epsilon = 6.2$ mM^{-1} cm^{-1}). The K_m and V_{max} were obtained by analyzing the initial velocity data using eq 1 and the FORTRAN HYPERL program (*35*). The k_{cat} was determined from the V_{max} value by dividing it by the concentration of phosphonatase used.

$$v_o = V_{max}[S]/([S] + K_m) \quad (1)$$

where v_o is the initial velocity, K_m is the Michaelis constant for the substrate, V_{max} is the maximal velocity, and [S] is the concentration of the substrate.

Inactivation of Recombinant S. typhimurium Phosphonatase with Sodium Borohydride and Phosphonoacetaldehyde. Inactivation reactions were carried out using a procedure adapted from the protocol described by Olsen et al. (*21*). Phosphonatase (5 mM) was incubated for 30 s in 20 μL of 500 μM $NaBH_4$, 0 or 1 mM Pald, and 100 mM HEPES (pH 7.0 and 0 °C) and then transferred to a 1 mL assay solution containing 1 mM Pald, 0.13 mM NADH, 5 mM $MgCl_2$, and 1 unit of yeast alcohol dehydrogenase in 100 mM HEPES (pH 7.0 and 25 °C) to measure catalytic activity.

Site-Directed Mutagenesis of Lys53 in S. typhimurium Phosphonatase. The pASB2 subclone, prepared using the Wizard Miniprep Kit from Promega, served as template for the PCR. The primers used are as follows: K53L, 5′-GGCTGGGTCTATGGCAA-3′ and 5′-CCGACCCAGATACCGTTG-3′; K53A, 5′-GCTGGGTGCCTGGCAACAC-3′ and 5′-GTGTTGCCAGGCACCCAGC-3′; and K53R, 5′-GCTGGGTCGATGGCAAC-3′ and 5′-GTTGCCATCGACCCAGC-3′. The PCRs, the purification of the PCR products, and the ligation, transformation (*E. coli* TOP 10 cells), and screening steps were carried out as described previously for the phosphonatase gene subcloning. The sequence of each mutant was confirmed by DNA sequencing and the level of phosphonatase expressed examined by SDS−PAGE analysis of the lysed cells.

Sequence Analysis. The three phosphonatase sequences were used to search a nonredundant database comprised of GenBank CDS translations and the PDB, SwissProt, and PIR databases. All database searches were performed at the NCBI using the BLAST network server as permitted through the Sequence Analysis and Consulting Service at the University of California at San Francisco. Default parameters were used as described in the BLAST version 1.4 documentation files except that expectation values ranged from 10 to 50 and up to 499 sequence hits were obtained from each Blastp search (*36)*. From the database searches, relevant sequences with known and/or possibly related functions were selected for use as queries in additional BLAST searches.

From these searches, sequences that were found by more than two query sequences were assigned to the set of candidate phosphonatase homologues; from these, a divergent set of sequences was generated such that no sequence was more than 50% identical to any other sequence in the set. These sequences were multiply aligned using PILEUP (Wisconsin Genetics Computer Group (GCG)). Only the relevant domains or modules of multi-modular proteins were included in the multiple alignment. Conserved motifs apparent in the multiple alignment were used to evaluate superfamily membership for these and other sequences that were subsequently identified from additional database searches.

The set of likely homologues defined by this procedure was subjected to an automated congruence analysis step using the program Shotgun (*37*). The Shotgun algorithm performs identical Blastp searches on all candidate sequences, saving, in this case, 400 sequence hits for each query. The sequence hits found by all of the Blastp searches were then sorted by the number of query sequences that found each hit. The output of the algorithm includes accession numbers and brief header information, accompanied by the names of the query sequences that found that hit, the score and significance values of the highest-scoring HSP, and the number of HSPs found.

A master alignment was constructed using the sequences corresponding to the following accession numbers: a64499, U32784, Y10275, u62737, z11591N, l05781N, x65083N, x97024N, p38773, p41277, u69493, u45309, phosBc, p24069, x66249, s74078, u67781Y, q04541, p54607, d90815, z70730, q08623, u66356, p35924, u33322, d90907, p42509, p32662, p40852, u67781Z, z70724, p31467, a49101, u35731, and p10368. One sequence representing each known enzyme activity and the three phosphonatase sequences were selected for the final alignment from which the motifs represented in Figure 2 were taken.

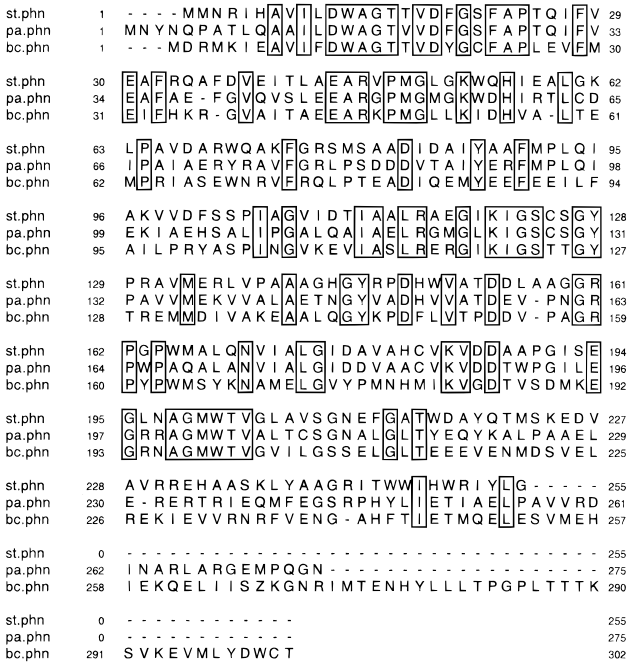


FIGURE 1: Alignment of the sequences of phosphonatase from *B. cereus* (bc.phn), *S. typhimurium* (st.phn; access code U69493), and *P. aeruginosa* (pa.phn; access code 45309) was made using the GCG program PileUp and boxed using SeqVu (J. Gardner, 1995, SeqVu, The Garvan Institue of Medical Research).

RESULTS AND DISCUSSION

B. cereus Phosphonatase Gene Isolation and Sequence Determination. The *B. cereus* phosphonatase gene was first cloned on a 4 kb fragment generated from a *Hin*dIII/*Pst*I digest of *B. cereus* DNA using the plasmid vector pT7T3a18 and *E. coli* DH5a competent cells. This clone, named pT7T3a18-*Hin*dIII/*Pst*I-*B. cereus* was used to generate the sequencing clone, pSE 420-*Kpn*I/*Nco*I-*B. cereus*, containing the phosphonatase gene on a ca. 1 kb *Kpn*I/*Nco*I fragment of the 4 kb *Hin*dII/*Pst*I fragment in the plasmid vector pSE 420 (*E. coli* DH5a). The sequence (GenBank access code 2623262) of the DNA fragment consisted of 15 bases preceding the start codon ATG of an open reading frame encoding a protein of 266 amino acids (Figure 1) and a calculated molecular weight of 30 361 (and calculated p*I* of 4.8). The region corresponding to the N-terminal sequence (MKIEAVIFDWAGTTVDY) obtained by automated amino acid sequencing of wild-type phosphonatase isolated from *B. cereus* indicates that the first three amino acids are removed from the N terminus of the nascent peptide by post-translational modification. The stretch of sequence determined for the tryptic peptide (GVAITAEEARKP-MGLLKIDHV) isolated from a digest of the *B. cereus* phosphonatase which had been modified at the catalytic lysine residue with radiolabeled acetonyl phosphonate (to form the Schiff base adduct) and $NaBH_4$ (to reduce the Schiff base adduct to lysine *N*′-2-propylphosphonate) (*22*) matches residues 37−57 and identifies lysine 53 as the lysine residue responsible for Schiff base formation with the Pald during catalytic turnover.

B. cereus Phosphonatase Gene Expression and Protein Purification and Characterization. The phosphonatase gene was subcloned from pT7T3a18-*Hin*dIII/*Pst*I-*B. cereus* pT-PHP into the plasmid pKK223-3 vector for high-level expression in *E. coli* JM105. The protocol (illustrated in Table 1) used to purify the recombinant enzyme from this clone (pKK223-3-*B. cereus*-phosphonatase/*E. coli* JM105) was based on that used for the purification of the wild-type enzyme by LaNauze et al. (*20*). SDS−PAGE analysis of the purified enzyme verified its homogeneity and 31 kDa subunit size. The yield of pure enzyme from the clone is ca. 1 mg per gram of cells which represents a 143-fold increase from the 0.007 mg of enzyme obtained from 1 g of *B. cereus* A-I2 cells. The K_m of 33 ± 2 μM and k_{cat} of 920 min^{-1} were determined at pH 7.0 and 25 °C using initial velocity techniques. These values are close to those reported by LaNauze et al. (*20*) for the wild-type enzyme: $K_m = 40$ μM and $k_{cat} = 1000$ min^{-1}.

S. typhimurium Phosphonatase Gene Subcloning and Expression and Protein Purification. The *S. typhimurium phn R* to *phnX* gene cluster had been cloned (pWM67) and sequenced earlier (*31*) (GenBank access code U69493). The N terminus of the phosphonatase gene product was determined by automated amino acid sequencing to be MNRI-HAVIL. This sequence allowed us to locate the phosphonatase gene on the sequenced insert of the pWM67 clone so that it could be subcloned into an efficient expression vector. Accordingly, a 764 bp open reading frame which encodes a 254-amino acid protein of 27 328 Da was identified. The *S. typhimurium* phosphonatase sequence is represented in the sequence alignment shown in Figure 1. For high-level expression in *E. coli* TOP 10 cells, the gene was subcloned using the plasmid vector pSE 420, requiring that the second amino acid from the N terminus be changed from Asn to Ala.[3] Purification of the phosphonatase from the pSE 420-*S. typhimurium* phosphonatase/*E. coli* TOP 10 cell expression system provided homogeneous enzyme of ca. 27 kDa in a yield of 2 mg per gram of cells (Table 2). The K_m of 40 ± 2 μM and k_{cat} of 1400 min^{-1} were determined at pH 7.0 and 25 °C using initial velocity techniques. These steady state kinetic constants are quite close in value to those reported for the *B. cereus* enzyme.

S. typhimurium Phosphonatase Active Site Lysine−Schiff Base Formation and Mutation. The utilization of a Schiff base mechanism of catalysis in the *S. typhimurium* phosphonatase was tested by treating the enzyme with the reducing agent $NaBH_4$ in the presence of Pald. Previous studies with the *B. cereus* enzyme had indicated that an imine forms between the catalytic Lys and Pald during the course of the enzymatic reaction (*20*−*22*). The Schiff base adduct undergoes a loss of the phosphoryl group to form the lysine *N*-ethylene amine which, in turn, is hydrolyzed to the free enzyme and the acetaldehyde product. In the presence of the $NaBH_4$, however, the lysine *N*-ethylene amine is reduced to the stable lysine *N*-ethyl amine, thus preventing further catalytic turnover by the enzyme. In this study, we found that treatment of the *S. typhimurium* phosphonatase with Pald and $NaBH_4$ at 0 °C for 30 s resulted in complete loss of activity. In contrast, the control reaction, in which Pald had been excluded, retained 92% of the original enzyme activity.

[3] The introduction of a unique *Nco*I restriction site at the initiation site of the phosphonatase gene was not possible unless the Asn codon was altered. The Ala codon was chosen since it satisfied the sequence requirements and allowed us to make a conservative substitution for the Asn.

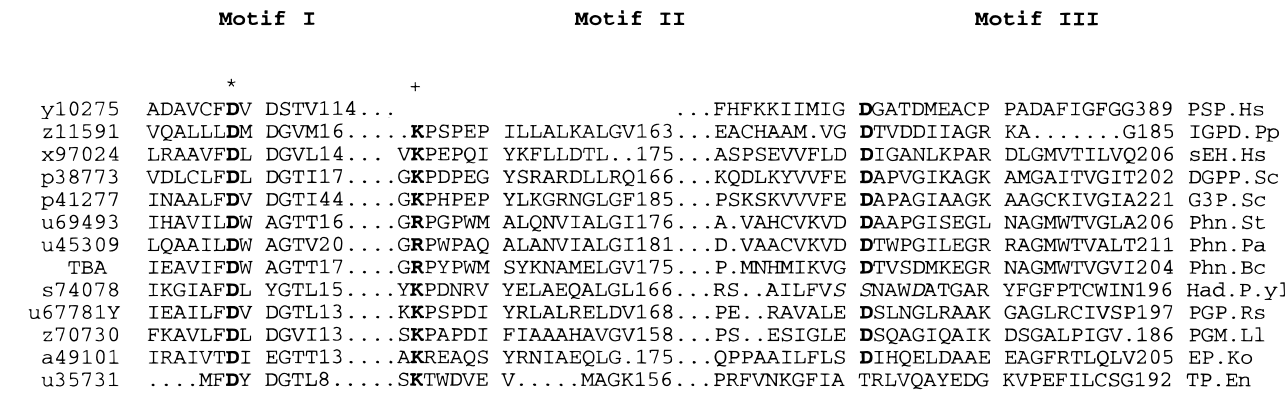


FIGURE 2: Three conserved motifs derived from the alignment of the phosphonatase/phosphatase superfamily sequences (see Materials and Methods for details). The positions of the motifs within the sequences are indicated by residue numbers. The sequences shown are as follows: PSP.Hs, phosphoserine phosphatase from *Homo sapiens* (access code 10275) (*50*); IGPD.Pp, the N-terminal domain of the putative bifunctional imidazole glycerolphosphate dehydratase:histidinol phosphatase from *Phytophthora parasitica* (access code Z11591; direct submission to the GenBank by P. Karlovsky in 1992); sEH.Hs, soluble epoxide hydrolase from *H. sapiens* (access code X97024) (*59*); DGPP.Sc, 2-deoxyglucose-6-phosphate phosphatase from *Sa. cerevisiae* (access code P38773) (*60*); G3P.Sc, glycerol-3-phosphate phosphatase from *Sa. cerevisiae* (access code p41277) (*61*); Phn.St, phosphonatase from *S. typhimurium* (access code U69493); Phn.Pa, phosphonatase from *P. aeruginosa* (access code 45309); Phn.Bc, phosphonatase from *B. cereus*; Had.Pyl, 2-haloalkanoic dehalogenase from *Pseudomonas* sp. strain YL (access code S74078) (*62*); PGP.Rs, 2-phosphoglycolate phosphatase (access code u67781Y) (*42*); PGM.Ll, β-phosphoglucomutase from *Lactococus lactis* (access code Z70730) (*49*); EP.Ko, enolase-phosphatase from *Klebsiella oxytoca* (access code a49101) (*63*); and Tp.En, trehalose phosphatase from *Emericella nidulans* (access code u35731) (*57*).

Table 1: Purification of *B. cereus* Phosphonatase Expressed with the pASB5 Subclone in *E. coli* JM105 Cells[a]

step	total protein (mg)	total activity (U)[b]	specific activity (U/mg)	recovery (%)	purification (*x*-fold)
extract from 11 g of cells (6 L of culture)	8600	1300	0.16	100	1
protomine sulfate	2230	1100	0.50	85	3
ammonium sulfate (0−70%)	800	1104	1.40	85	6
DEAE-Sephadex A-50	46	733	16.0	55	103
Sephadex g-75	12	500	17.4	16	112

[a] The protein concentration was monitored at 280 nm and estimated according to $OD_{280} = 1$ mg/mL. [b] One unit (U) of enzyme activity was defined as the amount of enzyme required to produce 1 mmol of inorganic phosphate per minute in 50 mM HEPES, 10 mM $MgCl_2$, and 0.1 mM DTT at pH 7.0 and 25 °C.

Table 2: Purification of *S. typhimurium* Phosphonatase Expressed with the pASB2 Subclone in *E. coli* TOP 10 Cells[a]

step	total protein (mg)	total activity (U)[b]	specific activity (U/mg)	recovery (%)	purification (*x*-fold)
extract from 11 g of cells (6 L of culture)	3900	1400	0.4	100	1
protamine sulfate	890	1300	1.5	93	4
ammonium sulfate (0−70%)	500	1100	2.2	76	6
DEAE-Sephadex A-50	61	570	9.4	40	25
Sephadex G-75	44	500	11.0	34	31
Sephadex G-150	25	300	13.0	23	33

[a] The protein concentration was monitored at 280 nm and estimated according to $OD_{280} = 1$ mg/mL. [b] One unit (U) of enzyme activity was defined as the amount of enzyme required to produce 1 mmol of inorganic phosphate per minute in 50 mM HEPES, 10 mM $MgCl_2$, and 0.1 mM DTT at pH 7.0 and 25 °C.

These results confirm the expected, that catalysis in the *S. typhimurium* phosphonatase proceeds, as it does in the *B. cereus* phosphonatase, by a Schiff base mechanism.

From the alignment of the sequences of the *S. typhimurium* and *B. cereus* phosphonatases depicted in Figure 1, it can be surmised that Lys53 is the catalytic lysine residue of the *S. typhimurium* enzyme. This was confirmed by mutating this residue to Arg and observing the complete loss of catalytic activity (zero activity detected at a 100-fold wild-type enzyme concentration) in the K53R mutant (yield, 0.7 mg of mutant protein per gram of cells). While no tertiary structural characterization of the K35R mutant was carried out, the observation that its chromatographic behavior and stability toward proteolysis and activity loss match that of the wild-type enzyme suggested that it has retained the native fold.

Comparison of the B. cereus, S. typhimurium, and P. aeruginosa Phosphonatase Sequences. The alignment of the sequences of the phosphonatases from *B. cereus*, *S. typhimurium*, and *P. aeruginosa* shown in Figure 1 reveal that the *S. typhimurium* and *P. aeruginosa* phosphonatase sequences are 52% identical and that the *S. typhimurium* and *B. cereus* phosphonatase sequences are 38% identical as are the *P. aeruginosa* and *B. cereus* phosphonatase sequences. Twenty-six percent of the residues are identical across all three proteins. In addition to Lys53, possible candidates for catalytic residues include nine Asp or Glu, five Arg or Lys, one His, six Thr, Ser, or Tyr, two Trp, and one Asn.

The fact the sequences of the phosphonatases from *S. typhimurium* and *P. aeruginosa* are more closely related to each other than either of them is to the *B. cereus* phosphonatase sequence suggests that the phosphonatase gene is likely to be of ancient origin. Studies of evolutionary relatedness, based on 16S rRNA sequence comparisons, reveal that Gram-negative and Gram-positive bacteria diverged from a common ancestor very early in evolution (ca. 2.5 billion years ago) (*38*). Thus, the Gram-positive *B. cereus* is evolutionarily more distant from the Gram-negative *S. typhimurium* and *P. aeruginosa* than the latter two are from one another. Therefore, if the phosphonatase gene arose before the divergence of Gram-negative and Gram-positive bacteria, we would expect that the *B. cereus* phosphonatase sequence would show considerably less identity when paired with either the *S. typhimurium* or *P.*

aeruginosa phosphonatase sequence. In addition, the distant sequence relationships found to exist between the phosphonatases and the other members of their enzyme superfamily (see discussion below) also suggest the divergence of phosphonatase from a very early ancestral structure.

Insight into the Evolution and Mechanism of Phosphonatase Catalysis Derived from Known Catalytic Functions in Structural Homologues. Through the rapid elucidation of enzyme superfamilies, we are seeing a growing volume of evidence that new enzyme catalysts evolve from protein scaffolds that support related chemistry. Well-documented examples of enzyme superfamilies whose members maintain a common fold that mediates similar chemistry (e.g., through stabilization of a common tetrahedral oxyanion, aci−acid, or thiol−enolate reaction intermediate) across a varied range of reactions and substrates include the α/β fold hydrolase superfamily, the enolase superfamily, and the 2-enoyl-CoA superfamily (for a recent review of these families, see ref *39*). The studies show that, by interpreting the structure−function relationships of an enzyme within the context of a well-characterized enzyme superfamily, one can gain insight into the origin and mechanism of catalysis associated with that enzyme. It was with this goal in mind that the phosphonatase sequences were used to search available sequence databases for structural homologues.

Identification of the Enzyme Superfamily to Which Phosphonatase Belongs. Although the phosphonatase sequences showed statistically significant matches with each other, no clearly significant matches were found between the phosphonatase sequences and any other protein in the database. However, when used as queries, each of the phosphonatase sequences found several of the same sequences at scores that ranged from just below statistical significance to not statistically significant ($P = 0.00012-1.000000$). These included phosphoglycolate phosphatases, 1-phospho-2,3-diketohexane-2-hydrate enolase-1-phospho-2-hydroxy-3-keto-1-hexene phosphatase (trivial name, enolase phosphatase) of the methionine salvage pathway, glycerol-3-phosphate phosphatase, the N-terminal ("phosphatase") domain of some bifunctional imidazole glycerol-phosphate dehydratase:histidinol phosphatase enzymes of the histidine biosynthetic pathway, certain haloalkanoic dehalogenases, phosphoserine phosphatases, and many reading frames of unknown function. When database searches were performed using members of these families, in turn, as queries, additional potential candidates for superfamily membership were identified, including the N-terminal domain of mammalian cytosolic epoxide hydrolases (having a yet undetermined catalytic activity as discussed below), β-phosphoglucomutase, 2-deoxyglucose-6-phosphate phosphatases, and trehalose-6-phosphate phosphatases. Automated congruence analysis using the program Shotgun (*37*) showed that all of the enzyme families listed above were found by at least 4 of 16 putative superfamily members (at any statistical significance) with very few false positives.[4] From the set of candidate phosphonatase homologues, a divergent set of sequences was generated such that no sequence was more than 50% identical to any other sequence in the set. These sequences were multiply aligned using PILEUP [Wisconsin Genetics Computer Group (GCG)] (for the accession numbers and methodology used, see Materials and Methods). Because the sequences aligned are so distantly related (sequence identities are as low as 15−20% for many paired sequences), the overall alignment cannot be considered accurate. Nonetheless, the global relatedness of these sequences is obvious. Three regions of conserved sequence, similarly positioned in each sequence (corresponding to residues 6−17 for motif I and stretching from residue 158 to residue 204 for motifs II and III in the *B. cereus* phosphonatase sequence), are an indication of a conserved scaffold. From the original alignment of 35 sequences (in which the Asp10 residue discussed below is stringently conserved), the sequences encoding proteins of unknown function were removed to leave a subset of 13 sequences (three phosphonatase sequences and one sequence representing each of the 10 catalytic functions listed above). The three regions of conserved motifs are shown for these sequences in Figure 2.

A literature search revealed that Koonin and Tatusov (*41*) had previously grouped 2-haloalkanoic dehalogenase, 1,2-haloalkanoic dehalogenase, 2-phosphoglycolate phosphatase, cytosolic epoxide hydrolase, enolase-phosphatase, imidazole glycerol-phosphate dehydratase:histidinol phosphatase, *p*-nitrophenyl phosphatase, phosphoserine phosphatase, and trehalose-6-phosphate phosphatase into the "haloacid superfamily" and had identified essentially the same three regions of highly conserved residues represented in Figure 2. In addition, possible structural relationships existing between "subgroups" of these proteins have also been identified by other workers (*42*−*45*). The findings from these independent studies are further evidence that the proteins we have grouped together in the phosphonatase family are indeed related.

A Possible Link between the Chemistry Catalyzed by the Members of the Superfamily and the Conserved Motifs. At present, only one member of this superfamily of enzymes has a known three-dimensional structure. This enzyme is the L-2-haloalkanoic dehalogenase from *Pseudomonas* sp. strain YL. As shown by the X-ray crystal structure of the apoenzyme represented in Figure 3, this dehalogenase folds into a small, four-α-helix bundle domain and a larger core domain containing a six-strand β-sheet flanked by five α-helices (*46*). The active site is located at the domain−domain interface, which in the absence of bound substrate forms a large crevice. For the moment, we will presume that this fold is representative of the folds of the other superfamily members and, hence, that the positions of the three conserved motifs (illustrated in Figure 2) found in this structure will also be representative. In Figure 3, the region (residues 9−13) surrounding the stringently conserved residue (Asp10) of motif I is represented in red, the region (residues 149−153) surrounding the stringently conserved positively charged residue (Lys151) of motif II in blue, and the region (residues 174−181) surrounding the highly conserved Asp residue (which is a Ser 176 in this protein) of motif III in green. Each of these motifs contributes to core domain loops, located in the active site.

[4] A test of congruence analysis was conducted using the recently described enolase superfamily (*40*). Shotgun analysis using 17 divergent members of the superfamily as the query set found all known members of the superfamily. In addition, all hits found by three or more of the query sequences were verified as superfamily members using the criteria described in ref *40*.

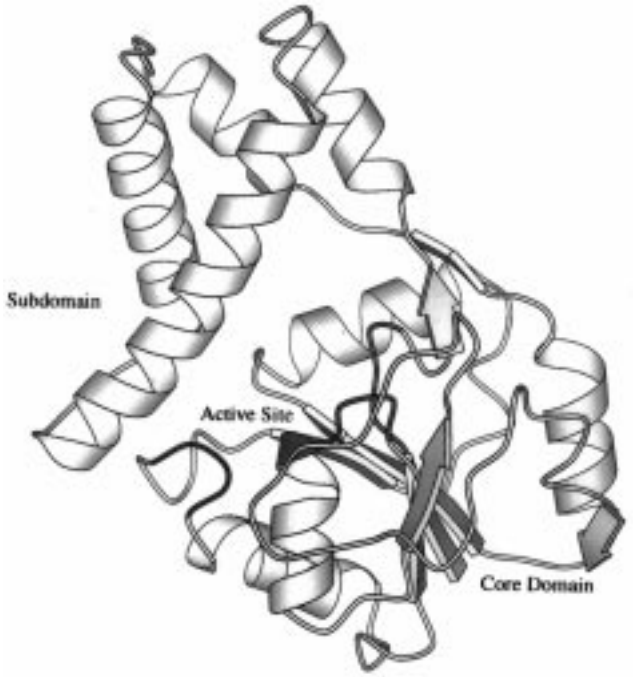


FIGURE 3: Molscript (*64*)-generated representation of the three-dimensional structure of *Pseudomonas* sp. strain YL 2-haloacid dehalogenase using the coordinates reported in ref *46*.

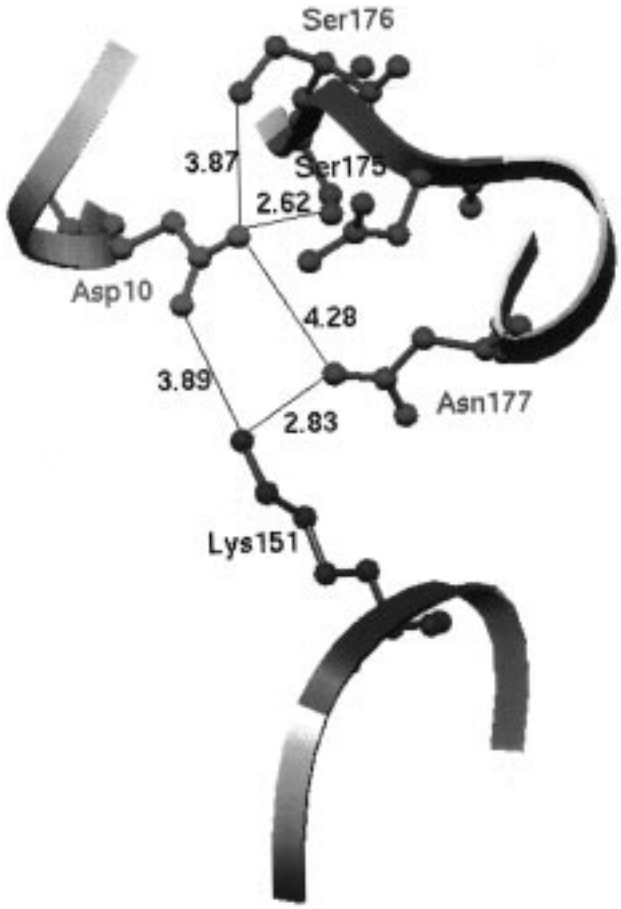


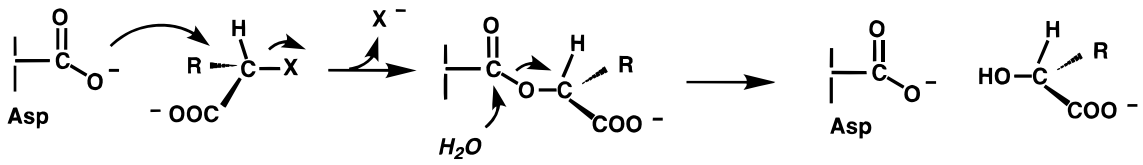
FIGURE 4: Active site region of *Pseudomonas* sp. strain YL 2-haloacid dehalogenase using the coordinates reported in ref *46* and InsightII.

Mechanistic studies (*47*) of the haloalkanoic dehalogenase have shown that the stringently conserved Asp10 of motif I functions in covalent catalysis, displacing the halide from C(2) of the 2-haloalkanoic substrate and forming an alkyl enzyme intermediate (see Scheme 3). The location of Asp10 in the active site is shown in Figure 4 (the two water molecules observed in the active site, surrounding the carboxylate group, are not shown). In the second partial reaction catalyzed by the dehalogenase, an activated water molecule attacks the acyl carbon, forming a tetrahedral oxyanion intermediate which then eliminates the 2-hydroxyalkanoic acid product and regenerates the Asp10 carboxylate group. The results obtained from mutagenesis (*48*) and substrate docking experiments (*46*) have led investigators to suggest Lys151 of motif II and Ser175, Asn177, and Asp180 of motif III (see Figure 4) as additional catalytic groups. Ser176, which is also shown in the active site picture of Figure 4, is not required for catalytic activity in the dehalogenase (the Ser176Ala mutant retains 92% of the wild-type activity), but we show it here for the purpose of further discussion provided below.

The reactions catalyzed by the members of the superfamily having known functions include phosphate monoester hydrolysis (the seven different phosphatases), phosphonate hydrolysis (the phosphonatases), and phosphoryl group transfer (the β-phosphoglucomutase). The evidence that each of these reactions involves covalent catalysis by the stringently conserved Asp10 rests on the mechanistic work that has been conducted to date on this enzyme superfamily. Catalysis by four of the enzymes phosphonatase (*26*), β-phosphoglucomutase (*49*), serine phosphate phosphatase (*50*), and phosphoglycolate phosphatase (*51*) is known to proceed via phosphoenzyme intermediates (the mechanisms are shown in Schemes 2 and 4). Stability studies have been carried out on the phosphoenzymes produced during catalytic turnover by the phosphoserine and 2-phosphoglycolate phosphatases to show that a phosphoacyl linkage, resulting from phosphoryl transfer to an active site carboxylate (Asp10, no doubt), is formed. The chemical pathways for the other phosphatases have not yet been examined, but we suggest that they also proceed via phosphoenzyme intermediates.

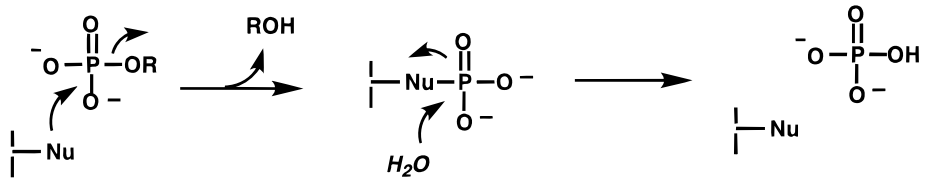
As illustrated in Figure 4, the arrangement of the three active site loops in the haloalkanoic dehalogenase structure allows the placement of several polar side chains (from residues 151, 175−177, and 180) within reach of the central active site Asp nucleophile. This would appear to be an ideal arrangement for catalysis of hydrolytic nucleophilic substitution in a variety of substrates because different arrangements of acid and base groups could be used to accommodate

Scheme 3: Chemical Steps of the 2-Haloalkanoic Dehalogenase Catalysis (*46*−*48*)

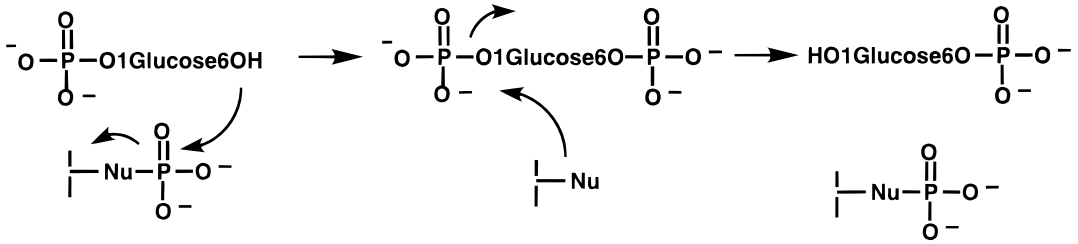


Scheme 4: Proposed Chemical Pathways for β-Phosphoglucomutase Catalysis (*49*) and for Catalysis by 2-Phosphoglycolate Phosphatase (*51*) and Phosphoserine Phosphatase (*50*)

A. *2-Phosphoglycolate phosphatase and phosphoserine phosphatase*



B. β- *Phosphoglucomutase*



different substrate structures. For instance, either position 180 or 176 might be used for placement of a carboxylate side chain to function in the activation of the water nucleophile (see Figure 4). Similarly, an electropositive group which is needed to interact with the charged leaving group might be positioned at position 151, 180, or 177 depending on the structure of the substrate and whether a metal ion cofactor is used in catalysis.

While the conserved motifs shown in Figure 2 provide support for the assignment of these proteins to a single structural superfamily, the full-length sequences are generally so divergent that it is difficult to obtain a reliable alignment of the overall sequences. Such divergence is, however, to be expected given the disparate enzyme families represented in the superfamily. Whereas all of the superfamily members have correlated elements of common structure and function, each has independently evolved the appropriate structural variations to mediate their unique and highly specific functions. As three-dimensional structural information becomes available for additional superfamily members, the manner in which the differences and similarities existing in their specific structures correlate with the differences and similarities in the chemistries catalyzed will be revealed.

Possible Link between the Conserved Aspartate of Superfamily Motif I and Nucleophilic Catalysis in Phosphonatase. If the invariant Asp residue (Figure 2) is conserved by all enzyme family members to function in nucleophilic catalysis, then we might expect Asp12 of the (*B. cereus*) phosphonatase sequence (Figure 1) to function as the initial phosphoryl acceptor, represented by Nu in Scheme 2. In view of the known stereochemistry of the phosphonatase reaction (*26*), retention of configuration at phosphorus and hydrolysis of the acyl−phosphoenzyme formed must necessarily occur at phosphorus. Indeed, the sarcoplasmic ATPase, which is known to catalyze the hydrolysis of ATP via formation of an acyl−phosphoenzyme intermediate (Asp351) (*52*), does so with overall retention of stereochemistry at the γ-phosphoryl phosphorus (*53*). Thus, for the ATPase, hydrolysis of the mixed anhydride occurs at phosphorus and not at carbon, consistent with our expectations of the regiochemistry of hydrolysis of the hypothetical acyl−phosphoenzyme formed during phosphonatase turnover.

Of the substrates undergoing catalyzed hydrolysis by the members of the enzyme superfamily, the phosphonatase substrate (Pald) presents the greatest mechanistic challenge because of the P−C bond cleavage required for product formation. The C−Cl bond cleavage catalyzed by the dehalogenase and P−O bond cleavage catalyzed by the phosphatases are comparatively simpler. The halide leaving group requires little in the way of stabilization by the dehalogenase, and the alkoxide anions displaced from the phosphate esters are easily neutralized by phosphatase-catalyzed proton transfer synchronized with P−O bond cleavage. The phosphonatase-catalyzed displacement of the enolate from Pald is avoided by initial Schiff base formation with the Lys53 residue. This initial partial reaction requires acid and base catalysis as well as a followup hydrolysis step for liberating the acetaldehyde from the enamine formed (see Scheme 2). Assuming for the moment that the phosphonatase fold is similar to the fold represented in Figure 3, we initially predict that the catalytic machinery involved in Schiff base formation and enamine hydrolysis is located on the subdomain, positioned at the domain−domain interface. Determination of the X-ray crystal structure of phosphonatase should provide further insight into how the superfamily ancestral structural scaffold evolved to accommodate the divergent and complex chemistry catalyzed by the phosphonatases.

A Note on the Use of the Superfamily Fold as a Functional Domain. The concept that proteins are constructed in a modular fashion, with separate domains performing different functions, is a by now a well-established one (*54*). In this context, we are intrigued to find cases in this superfamily where a member protein appears as a functional domain in

a larger multifunctional protein. For instance, in *E. coli* and *S. typhimurium*, the histidinol phosphatase (*55*) of the histidine biosynthetic pathway appears as an N-terminal appendage of the imidazole glycerolphosphate dehydratase to form a bifunctional enzyme, whereas in yeast (*56*), these two enzymes are monofunctional. The *Phytophthora nicotianae* enzyme (GB access code z11591; direct submission to the GenBank by P. Karlovsky in 1992) also appears to be bifunctional with the C-terminal domain functioning as the imidazole glycerolphosphate dehydratase. The N-terminal domain of the *Ph. nicotianae* sequence could possibly function as a histidinol phosphatase, although it appears from multiple alignment evaluation to be more closely related to the phosphonatase/phosphatase superfamily than are the other histidinol phosphatases.

In the mammalian cytosolic epoxide hydrolase (*45*), the C-terminal domain belongs to the α/β fold hydrolase family and is responsible for catalysis of epoxide hydrolysis (coincidentally, using an active site Asp residue in covalent catalysis). The N-terminal domain, shown in the sequence alignment presented in Figure 2, belongs to the phosphonatase superfamily; however, the function of this domain has not yet been determined (*45*).

Finally, the trehalose-6-phosphate phosphatase of *Saccharomyces cerevisiae* is a two-domain protein in which the C-terminal domain functions as the phosphatase and is homologous in size and sequence to the trehalose phosphatases of *E. coli* and *Aspergillus nidulans* (*57*). The N-terminal domain, on the other hand, is homologous to the trehalose synthase which catalyzes the first step of the two-step trehalose biosynthetic pathway. It has been suggested that during evolution a gene duplication and fusion event may have occurred, yielding a phosphatase with a synthase-like domain (*58*).

Like the trehalose-6-phosphate phosphatase of *Sa. cerevisiae*, the bifunctional imidazole glycerolphosphate dehydratase:histidinol phosphatase enzyme and the cytosolic epoxide hydrolase described above are also likely products of gene fusion. From these examples, we learn that the phosphonatase superfamily scaffold is stable both as an independent unit and as a domain within a larger protein structure. Evidently, it has been used in a variety of combinations as a functional and structural module in the evolution of new protein functions.

ACKNOWLEDGMENT

We thank Mr. Wen-hai Zhang and Ms. Hong Xiang for their assistance in the preparation of the figures used in this paper.

REFERENCES

1. Mastelerz, P. (1984) in *Natural Products Chemistry* (Zulewski, R. I., and Skolik, J. J., Eds.) p 171, Elsevier, Amsterdam.
2. Hilderbrand, R. L. (1983) *The Roles of Phosphonates in Living Systems*, CRC, Boca Raton, FL.
3. Horiguchi, M. (1984) in *Biochemistry of Natural C-P Compounds* (Hori, T., Horiguci, M., and Hayashi, H., Eds.) Maruzen, Ltd., Kyoto, Japan.
4. Hidaka, T., Mori, M., Imai, S., Hara, O., Nagaoka, K., and Seto, H. (1989) *J. Antibiot. 42*, 491−494.
5. Hidaka, T., Goda, M., Kuzuyama, T., Takei, N., Makoto, H., and Seto, H. (1995) *Mol. Gen. Genet. 249*, 274−280.
6. Nakashita, H., Shimazu, A., Hidaka, T., and Seto, H. (1992) *J. Bacteriol. 174*, 6857−6861.
7. Bowman, E., McQueney, M., Barry, R. J., and Dunaway-Mariano, D. (1988) *J. Am. Chem. Soc. 110*, 5575−5576.
8. Seidel, H. M., Freeman, S., Seto, H., and Knowles, J. R. (1988) *Nature 335*, 457−458.
9. Kim, A., Kim, J., Martin, B. M., and Dunaway-Mariano, D. (1998) *J. Biol. Chem. 273*, 4443−4448.
10. De Graaf, R. M., Visscher, J., and Schwartz, A. W. (1997) *J. Mol. Evol. 44*, 237−241.
11. Wackett, L. P., Shames, S. L., Venditti, C. P., and Walsh, C. T. (1987) *J. Bacteriol. 169*, 710−717.
12. Chen, C.-M., Ye, Q.-Z., Zhu, Z., Wanner, B. L., and Walsh, C. T. (1990) *J. Biol. Chem. 265*, 4461−4471.
13. Lee, K.-S., Metcalf, W. W., and Wanner, B. L. (1992) *J. Bacteriol. 174*, 2501−2510.
14. Wanner, B. L. (1994) *Biodegradation 5*, 175−184.
15. McGrath, J. W., Wisdom, G. B., McMullen, G., Larkin, M. J., and Quinn, J. P. (1995) *Eur. J. Biochem. 234*, 225−230.
16. La Nauze, J. M., Rosenberg, H., and Shaw, D. C. (1970) *Biochim. Biophys. Acta 212*, 332−350.
17. Dumora, C., Lacoste, A. M., and Cassigne, A. (1983) *Eur. J. Biochem. 133*, 119−125.
18. La Nauze, J. M., and Rosenberg, H. (1968) *Biochim. Biophys. Acta 165*, 438−447.
19. Lacoste, A. M., and Neuzil, E. (1969) *C. R. Seances Acad. Sci., Ser. D 269*, 254−257.
20. La Nauze, J. M., Coggins, J. R., and Dixon, H. B. F. (1977) *Biochem. J. 165*, 409−411.
21. Olsen, D. B., Hepburn, T. W., Moos, M., Mariano, P. S., and Dunaway-Mariano, D. (1988) *Biochemistry 27*, 2229−2234.
22. Olsen, D. B., Hepburn, T. W., Lee, S.-L., Martin, B. M., Mariano, P. S., and Dunaway-Mariano, D. (1992) *Arch. Biochem. Biophys. 296*, 144−151.
23. Schmidt, D. E., and Westheimer, F. H. (1971) *Biochemistry 10*, 1249−1253.
24. Laursen, R. A., and Westheimer, F. H. (1966) *J. Am. Chem. Soc. 88*, 3426−3430.
25. Kokesh, F. C., and Westheimer, F. H. (1971) *J. Am. Chem. Soc. 93*, 7270−7274.
26. Lee, S.-L., Hepburn, T. W., Schwartz, W. H., Ammon, H. L., Mariano, P. S., and Dunaway-Mariano, D. (1992) *J. Am. Chem. Soc. 114*, 7346−7354.
27. Baker, A. S. (1996) Ph.D. Thesis, University of Maryland, College Park, MD.
28. Ciocci, M. J. (1996) Ph.D. Thesis, University of Maryland, College Park, MD.
29. Jiang, W., Metcalf, W. W., and Wanner, B. L. (1995) *J. Bacteriol. 174*, 2501−2510.
30. Wanner, B. L. (1996) in *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biololgy* (Neidhart, F. C., Curtiss, R. I., Schaecter, M., and Umbarger, H. E., Eds.) pp 1357−1381, American Society for Microbiology, Washington, DC.
31. Metcalf, W. W., Jiang, W., and Wanner, B. L. (1998) (manuscript in preparation).
32. Dumora, C., Marche, M., Doignon, F. M., Aigle, M., Cassaigne, A., and Crouzet, M. (1997) *Gene 197*, 405−412.
33. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Habor, NY.
34. Sanger, F., Milken, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463.
35. Cleland, W. W. (1979) *Methods Enzymol. 63*, 103.
36. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol. 215*, 403−410.
37. Pegg, S., and Babbitt, P. C. (1998) (manuscript in preparation).
38. Woese, C. R. (1987) in *Bacterial Systematics* (Logan, N. A., Ed.) p 79, Blackwell Scientific Publications, Boston.
39. Babbitt, P. C., and Gerlt, J. A. (1997) *J. Biol. Chem. 272*, 30591−30594.
40. Babbitt, P. C., Hasson, M., et al. (1996) *Biochemistry 35*, 16489−16501.

41. Koonin, E. V., and Tatusov, R. L. (1994) *J. Mol. Biol. 244*, 125−132.
42. Gibson, J. L., and Tabita, F. R. (1997) *J. Bacteriol. 179*, 663−669.
43. Schäferjohann, J., Yoo, J.-G., et al. (1993) *J. Bacteriol. 175*, 7329−7340.
44. Lyngstadaas, A., and Løbner-Olesen, A. (1995) *Mol. Gen. Genet. 247*, 546−554.
45. Beetham, J. K., Grant, D., Arand, M., Garbarino, J., Kiyosue, T., Pinot, F., Oesch, F., Belknap, W. R., Shinozaki, K., and Hammock, B. D. (1995) *DNA Cell Biol. 14*, 61−71.
46. Hisano, T., Hata, Y., Fujii, T., Liu, J.-Q., Kurihara, T. H., Esaki, N., and Soda, K. (1996) *J. Biol. Chem. 271*, 20322−20330.
47. Liu, J.-Q., Kurihara, T., Miyagi, M., Esaki, N., and Soda, K. (1995) *J. Biol. Chem. 270*, 18309−18312.
48. Kurihara, T., Liu, J.-Q., Nardi-Dei, V., Koshikawa, H., Esaki, N., and Soda, K. (1995) *J. Biochem. 117*, 1317−1322.
49. Qian, N. Y., Stanely, G. A., Hahn-Hagerdal, B., and Radstrom, P. (1994) *J. Biol. Chem. 176*, 5304−5311.
50. Collet, J.-F., Gerin, I., Rider, M. H., Veiga-da-Cunha, M., and Van Schaftingen, E. (1997) *FEBS Lett. 408*, 281−284.
51. Seal, N. S., and Rose, Z. B. (1987) *J. Biol. Chem. 262*, 13496−13500.
52. Maruyama, K., and MacLennan, D. H. (1988) *Proc. Natl. Acad. Sci., U.S.A. 85*, 3314−3318.
53. Webb, M. R., and Trentham, D. R. (1981) *J. Biol. Chem. 256*, 4884−4887.
54. Brandon, C., and Tooze, J. (1991) in *Introduction to Protein Structure*, p 144, Garaland Publishing, New York.
55. Chiariotti, L., Nappo, A. G., Carlomagono, M. S., and Bruni, C. B. (1986) *Mol. Gen. Genet. 202*, 42−47.
56. Parker, A. R., Moore, T. D. E., Edman, J. C., Schwab, J. M., and Davison, V. J. (1994) *Gene 145*, 135−138.
57. Borgia, P. T., Miao, Y., and Dodge, C. L. (1996) *Mol. Microbiol. 20*, 1287−1296.
58. Kaasen, I., McDougall, J., and Strom, A. R. (1994) *Gene 145*, 9−15.
59. Beetham, J. K., Tian, T., and Hammock, B. D. (1993) *Arch. Biochem. Biophys. 305*, 197−201.
60. Randezgil, F., Blasco, A., Preito, J. A., and Sanz, P. (1995) *Yeast 11*, 1233−1240.
61. Norbeck, J., Pahlam, A.-K., Akhtar, N., Blomberg, A., and Adler, L. (1996) *J. Biol. Chem. 271*, 13875−13881.
62. Nardi-Dei, V., Kurihara, T., Okamura, T., Liu, J.-Q., Koshikawa, H., Ozaki, H., Terashima, Y., Esaki, N., and Soda, K. (1994) *Appl. Environ. Microbiol. 60*, 3375−3380.
63. Balakrishnan, R., Frohlich, M., Rahaim, P. T., Backman, K., and Yocum, R. Y. (1993) *J. Biol. Chem. 268*, 24792−24795.
64. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−956.

BI972677D