

*Proteins*. Author manuscript; available in PMC 2012 November 01.

Published in final edited form as:

Proteins. 2011 November; 79(11): 3099–3107. doi:10.1002/prot.23137.

# The X-ray Crystallographic Structure and Specificity profile of HAD Superfamily Phosphohydrolase BT1666: Comparison of Paralogous Functions in *B. thetaiotaomicron*<sup>†</sup>

Zhibing Lu<sup>1</sup>, Debra Dunaway-Mariano<sup>2</sup>, and Karen N. Allen<sup>1</sup>

<sup>1</sup>Department of Chemistry, Boston University, Boston, Massachusetts 02215

<sup>2</sup>Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, NM 87131-0001

# **Abstract**

Analysis of the haloalkanoate dehalogenase superfamily (HADSF) has uncovered homologues occurring within the same organism that are found to possess broad, overlapping substrate specificities and low catalytic efficiencies. Here we compare the HADSF phosphatase BT1666 from Bacteroides thetaiotaomicron VPI-5482 to a homologue with high sequence identity (40%) from the same organism BT4131, a known hexose-phosphate phosphatase. The goal is to find if these enzymes represent duplicated versus paralogous activities. The X-ray crystal structure of BT1666 was determined to 1.82 Å resolution. Superposition of the BT1666 and BT4131 structures revealed a conserved fold and identical active sites suggestive of a common physiological substrate. The steady-state kinetic constants for BT1666 were determined for a diverse panel of phosphorylated metabolites to define its substrate specificity profile and overall level of catalytic efficiency. Whereas BT1666 and BT4131 are both promiscuous, their substrate specificity profiles are distinct. The catalytic efficiency of BT1666 ( $k_{cat}/K_m = 4.4 \times 10^2~M^{-1}~s^{-1}$  for the best substrate fructose 1,6-(bis)phosphate) is an order of magnitude less than that of BT4131 ( $k_{cat}/K_m = 6.7 \times 10^{-5}$ 10<sup>3</sup> M<sup>-1</sup> s<sup>-1</sup> for 2-deoxyglucose 6-phosphate). The seemingly identical active-site structures point to sequence variation outside the active site causing differences in conformational dynamics or subtle catalytic positioning effects that drive the divergence in catalytic efficiency and selectivity. The overlapping substrate profiles may be understood in terms of differential regulation of expression of the two enzymes or a conferred advantage in metabolic housekeeping functions by having a larger range of possible metabolites as substrates.

### **Keywords**

Haloalkanoate dehalogenase superfamily; phosphohydrolase; phosphatase; *Bacteroides* thetaiotaomicron; substrate promiscuity; housekeeping; divergent evolution; BT1666; BT4131

<sup>&</sup>lt;sup>†</sup>This work was supported by N.I.H. grants GM61099 and U54 GM093342.

Address Correspondence to: Karen N. Allen Department of Chemistry Boston University 590 Commonwealth Avenue, Room 299 Boston, MA 02215-2521 Tel (617)-358-5544 drkallen@bu.edu Fax: (617) 353-6466 Debra Dunaway-Mariano Department of Chemistry and Chemical Biology University of New Mexico Albuquerque, NM 87111 Tel: (505) 277-3383 dd39@unm.edu Fax: (505) 277-6202.

# Introduction

The haloalkanoate dehalogenase superfamily (HADSF) is a large, ubiquitous and diverse superfamily of enzymes, the vast majority of which are phosphatases and ATPases. Multiple HADSF phosphatases exist within any given organism (for instance, 28 in *Escherichia coli*; 30 in *Mycobacterium tuberculosis*, 84 in *Caenorhabditis elegans*, 169 in *Arabidopsis thaliana*, and 183 in *Homo sapiens*) where they assume roles in biosynthetic and biodegradation pathways or perform a wide variety of housekeeping functions (1). Although phosphatases have evolved in other enzyme families, these families have not expanded to occupy the biological niches of the cell to nearly the extent seen with the HADSF phosphatases. A simple count of the phosphatases in *E. coli* shows that the HADSF phosphatases have been able to secure more biochemical jobs than the phosphatases from all six other phosphatase families together.

Earlier work on the HADSF phosphatases showed that the physiological substrates run the full gamut in size and shape from phosphoproteins, nucleic acids, and phospholipids to phosphorylated disaccharides, sialic acids, and terpenes to the smallest of the organophosphate metabolites, phosphoglycolate (2). It is through the acquisition of structural accessories to the catalytic domain that the HADSF has succeeded in covering this vast range of substrate structure(3). Moreover, the modular design of the HADSF phosphatase might play a key role in the high evolvability. The vast majority of the HADSF phosphatases consist of a conserved Rossmann domain (the catalytic domain) and a tethered "cap" domain (variable in fold) (Figure 1A). The catalytic scaffold, which is formed by four backbone segments located at the C-terminal end of the Rossmann-fold central sheet (Figure 1B), conserves the catalytic residues that mediate phosphate ester hydrolysis (Figure 1C). The mechanism conserved among HADSF phosphatases involves the nucleophilic attack of an aspartate on the phosphate of the phosphoryl group with general acid catalysis and subsequent hydrolysis of the aspartyl-phosphate intermediate. The cap domain is inserted into either one of two loops of the catalytic scaffold (Figure 1A). The catalytic scaffold alone binds the transferring phosphoryl group, whereas the cap domain alone binds the leaving group. Thus, HADSF phosphatase catalytic residues positioned by the catalytic scaffold are physically separated from the substrate recognition residues positioned on the cap domain. A priori, the substrate recognition site can evolve independent of the catalytic site. In addition, the cap domain provides a large surface, which can accommodate numerous substrate recognition residues. Consequently, the potential to recognize more than one substrate exists, consistent with the observation that the HADSF phosphatases tend to be promiscuous (4-5).

Substrate promiscuity underlies high evolvability. When a gene encoding a promiscuous enzyme is duplicated the opportunity exists for a novel function to evolve within one of the two genes through mutations that enhance a low, intrinsic activity towards a novel substrate. In earlier work, we had identified two HADSF type C0 phosphatases in *Bacteroides thetaiotaomicron*, with 25% sequence identity which undoubtedly are related by a gene duplication event. The 2-keto-3-deoxy-8-phospho-D-manno-octulosonic acid (KDO-8-P) phosphatase of the lipid A pathway is posited to be the progenitor while the "new" enzyme is posited to be 2-keto-3-deoxy-D-glycero-D-galacto-9-phosphonononic acid (KDN-9-P) phosphatase. The KDO-8-P phosphatase displays a  $k_{cat}/K_m = 1.5 \times 10^4~M^{-1} s^{-1}$  towards its physiological substrate KDO-8-P and a low level of activity ( $k_{cat}/K_m = 6.7 \times 10^2~M^{-1} s^{-1}$ ) towards KDN-9-P. Conversely, KDN-9-P phosphatase displays a  $k_{cat}/K_m = 1.1 \times 10^4~M^{-1} s^{-1}$  towards its physiological substrate KDN-9-P and a low level of activity ( $k_{cat}/K_m = 2.0 \times 10^2~M^{-1} s^{-1}$ ) towards KDO-8-P (6-7).

The work described in this paper examines a second pair of B. thetaiotaomicron HADSF phosphatase homologues with high sequence similarity, BT1666 and BT4131 with the goal of finding whether they have paralogous functions. Previously, our laboratories had reported the structure and function of BT4131(5). BT4131 possesses a C2b type cap domain, which supports phosphatase activity towards hexose-phosphates generated in the course of chitin recycling. BT1666 shares 40% sequence identity and 75% sequence similarity with BT4131. Bioinformatic analysis was carried out to define the biological range and co-occurrence of BT1666 and BT4131 orthologues. The X-ray crystal structure and substrate specificity profile of recombinant BT1666 was determined for the purpose of relating change in substrate recognition with change in the structure of the substrate binding site. Whereas the two substrate binding sites proved to be essentially identical, the substrate specificity profiles are not and, most curious is the finding that the level of activity in BT1666 as defined by the magnitude of the k<sub>cat</sub>/K<sub>m</sub> value is substantially lower than that observed with BT4131. The implications of these findings in relation to structure based function assignment within the HADSF are examined. In addition, the distribution of one or the other paralogue versus both paralogues among Bacteroides species is explored within the context of the minimal level of catalytic efficiency required of a HADSF housekeeper, and the potential advantage bestowed by an additional housekeeper.

# **Materials and Methods**

# Gene cloning, construct design, protein expression, and purification

The cDNA encoding the BT1666 gene from B. thetaiotaomicron VPI-5482 was amplified by PCR using the genomic DNA of B. thetaiotaomicron VPI-5482 (a kind gift from Dr. J. Gordon (Washington University, St. Louis, MO)), *Pfu* Turbo DNA polymerase (Stratagene) and oligonucleotide primers (5'-CTTTGTAGCCCAAAAATATAGATCATATGAT, and 5'-GCTTTGTAGCCAAGGATCCTTAAGACATTT), containing the restriction endonuclease cleavage sites for Ndel and BamHI. The pET15B vector, cut by restriction enzymes NdeI and BamHI, was ligated to the PCR product that had been purified and then digested with the same restriction enzymes. The ligation product was used to transform E. coli JM109 competent cells (Stratagene), which were grown on an Ampicillin-containing agar plate. The selected colony was used to prepare plasmid DNA (Mini-Prep Kit from Qiagen). The gene sequence was confirmed by DNA sequencing carried out by the Tufts University Core Facility. The plasmid was transformed into E. coli Rosetta(DE3)plysS cells for protein expression. The transformed E. coli cells were used to inoculate 4 L of Luria-Bertani medium containing 100 μg/ml ampicillin at 37 °C. Following growth of the culture to an OD<sub>600</sub> value of 0.6-1.0, gene expression was induced at 15 °C with 0.4 mM isopropyl α-D-thiogalactopyranoside. Following a 4 h induction period, the *E. coli* cells (10 grams) were harvested and suspended in 100 mL of buffer A (300 mM NaCl, 50 mM phosphate (pH 7.5)). The cells were disrupted by sonication on ice and the lysates were clarified by centrifugation at 48,384 g at 4 °C. BT1666 was purified by affinity chromatography on a Ni-NTA column (Qiagen). Protein purity was assessed by SDS-PAGE analysis. The desired protein fractions were pooled and applied to a 120 mL Sephacryl-200 gel filtration column (GE Healthcare) equilibrated with buffer A. Purified BT1666 was concentrated at 4 °C using a 10 kDa cutoff Amicon Ultra Centrifugal filter (Millipore) in buffer A and stored at -80 °C. The final yield was 10 mg protein/1 g wet cells. The protein concentration was determined using the Bradford method.

### BT1666 molecular weight determination

The theoretical subunit molecular mass of recombinant BT1666 including the His<sub>6</sub>-tag and thrombin cleavage site was calculated as 30,111 Da by using the amino acid composition derived from the gene sequence and the EXPASY Molecular Biology Server program

Compute pI/MW. The subunit mass was determined to be 30,019 Da by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. The molecular weight of native BT1666 was estimated by FPLC gel filtration column chromatography against protein standards using a  $1.6~\rm cm \times 60~\rm cm$  Sephacryl S-200HR column (GE Healthcare) eluted at 4 °C with 50 mM HEPES, 100 mM NaCl, pH 7.5 buffer at a flow rate of 1 mL/min. A molecular weight of ~30 kDa for BT1666 was derived from the measured elution volume by extrapolation of the plot of the elution volume of the molecular weight standard vs. log molecular weight (range 13.7-220 kDa; GE Healthcare).

### Steady-state kinetic constant determination

The purified recombinant enzyme was concentrated with an Amicon Ultrafiltration apparatus (PM10) or Centricon-10 (Millipore), and dialyzed against buffer B (1 mM HEPES (pH 7.5)) before use in kinetic studies. The steady-state kinetic parameters ( $K_m$  and  $k_{cat}$ ) of phosphorylated substrates were determined from initial reaction velocities measured at varying substrate concentrations (from 0.5-  $5K_m$ ). The initial velocities were measured for reaction mixtures containing 5 mM MgCl<sub>2</sub> in 50 mM HEPES buffer (pH 7.0) at 37 °C. The assay methods used for the various substrates are described below. Absorbance measurements were performed with a Beckman DU800 UV-vis spectrophotometer. Data were fitted to with KinetAsystI to the Michaelis-Menten equation.

# **Activity assays**

Phosphate ester hydrolysis for all substrates was monitored using Biomol green (BIOMOL International) to measure the concentration of inorganic phosphate formed in the reaction solutions. The 1 mL assay mixture, containing 50 mM HEPES (pH 7.0), 5 mM MgCl<sub>2</sub>, various concentrations of substrate, and 0.023  $\mu$ M BT1666 was incubated at 37 °C and aliquots removed at various times for phosphate determination. Times corresponding to initial rates (<10% of substrate consumes) were used to insure initial rates. In parallel, the background level of phosphate release was measured using a control reaction mixture, which excluded BT1666. For analysis, 100  $\mu$ L of the mixture were added to 1 mL of Biomol green reagent. After 60 min incubation at room temperature, the absorbance of the solution at 620 nm was measured.

The rate of *p*-nitrophenyl phosphate (pNPP) hydrolysis was also determined by monitoring the increase in absorbance at 410 nm ( $\Delta \epsilon = 18.4 \text{ mM}^{-1} \text{ cm}^{-1}$ ) at 37 °C. The 0.5 mL assay mixtures contained 50 mM HEPES (pH 7.0), 5 mM MgCl<sub>2</sub>, and various concentrations of pNPP. The kinetic constants determined from this assay agreed with those determined using the fixed time phosphate assay described above.

### pH-rate profile determination

All reactions were carried out at 37 °C using the following buffers and pH values/ranges: 50 mM sodium acetate, pH 4.5, 50 mM Bis-tris, pH 5.0-6.5; 50 mM Hepes, pH 7.0-7.5; 50 mM Tris, pH 8.0-8.5; and 50 mM bicine, pH 9.0. The substrate mannose 6-phosphate was used at 3 mM ( $\sim$ K<sub>m</sub>) to achieve subsaturating substrate (V<sub>max</sub>/K<sub>m</sub>) conditions. All reaction solutions contained 5 mM MgCl<sub>2</sub>. The initial reaction rates were measured using the discontinuous assay as described above.

# Crystallization and data collection

Purified BT1666 was concentrated to 68 mg/mL in 1 mM HEPES (pH 7.5) for crystallization. Crystals of BT1666 were grown by the hanging-drop vapor-diffusion method where drops were formed by mixing 1  $\mu$ L of protein solution with 1  $\mu$ L of well solution. The initial BT1666 crystals were obtained from the Index Screen (Hampton Research). The

refined conditions consisted of 0.1 M Bis-Tris pH 6.5, 0.2 M ammonium sulfate and 25% w/v PEG 3350, at 25 °C. Crystals appeared in approximately one week. The crystals were protected for low-temperature data collection by transfer to 100% Paratone-N (Hampton Research) for 5-10 min before data collection at 100 K. Data were collected on a Rigaku-RU-300 X-ray generator and R-Axis IV++ image plate located at Boston University School of Medicine. Data from a single BT1666 crystal,  $0.3 \times 0.2 \times 0.05$  in dimension, were collected and processed to 1.82 Å resolution using HKL2000 (8). BT1666 crystallized in space group  $P2_12_12_1$  with unit cell dimensions a=52.8 Å, b=70.7 Å, c=72.4 Å. The data collection statistics are reported in Table 1.

### Molecular replacement and crystallographic refinement

The phase problem was solved via the molecular replacement method using BT4131 (5) as the search model (40% sequence identity and 75% sequence similarity to BT1666). The program MOLREP (9) in the CCP4 program suite was used to solve the rotation and translation functions, yielding a solution with a correlation coefficient of 37 % and an R-factor of 59 % at 2.1 Å resolution. The asymmetric unit contains one BT1666 molecule. Alternating rounds of manual rebuilding and refinement were performed using the molecular graphics program COOT (10) followed by minimization and simulated annealing in PHENIX (11). During refinement, the water molecules and ligands were added once the  $R_{\rm free}$  was below 30%. The final protein model was analyzed using composite-omit electrondensity maps calculated with CNS (12). The stereochemistry was monitored with PROCHECK (13). Analysis of the Ramachandran plot showed that 97.2% of the residues fall in the most favored regions with 2.8% in the additionally allowed regions and with no residues falling in the generously allowed or disallowed regions. The refinement statistics are reported in Table 1.

The final model is well ordered, and comprises all 258 amino acid residues. The construct used for crystallization contained an N-terminal His<sub>6</sub>-tag and thrombin cleavage site (MGSSHHHHHHSSGLVPRGSH). The final protein model also included the last 10 residues of the His<sub>6</sub>-tag/thrombin cleavage site (SSGLVPRGSH). Refinement and final model statistics are given in Table 1. The graphic figures of the structure were produced using PyMOL, Molscript (14) and rendered by POVRAY (15).

### **Bioinformatic Analysis**

BT1666 and BT4131 orthologues were identified by carrying out BLAST searches of the NCBI genome databank (http://www.ncbi.nlm.nih.gov/sutils/genom\_table.cgi). The cutoff used for assigning orthologues was >51% sequence identity as determined by a pair-wise amino acid sequence alignment. Multiple sequence alignments were carried out in ClustalW (http://www.ch.embnet.org/software/ClustalW.html) and rendered in ESPript (http://espript.ibcp.fr/ESPript/).

### **Results and Discussion**

# **Substrate Specificity of BT1666**

The gene encoding BT1666 was expressed from a pET15B vector bearing a N-terminal  ${\rm His_6}$ -tag (BT1666-pET15B plasmid) in Rosetta (DE3)plysS competent cells with induction at low temperature (15 °C) overnight. The (>95%) pure  ${\rm His_6}$ -tagged BT1666 was stable and soluble in 300 mM NaCl and 50 mM phosphate buffer, pH 7.5. Gel-filtration chromatography demonstrated that  ${\rm His_6}$ -tagged BT1666 is a monomer in solution (see the Materials and Methods section for details) as is its paralogue BT4131 (5). Purified BT1666 phosphohydrolase activity was screened against a number of phosphorylated small-molecule substrates (Table 2). Of the compounds tested, BT1666 shows similar activity toward D-

fructose 1,6-(bis)phosphate, trehalose 6-phosphate, N-acetyl-glucosamine 6-phosphate, D-glucosamine 6-phosphate, mannose 6-phosphate, ribose 5-phosphate, D-arabinose 5-phosphate with  $k_{cat}/K_m$  values ranging between  $1.1\times 10^2~M^{-1}~s^{-1}$  to  $4.4\times 10^2~M^{-1}~s^{-1}$ . The activity for the best substrate is low compared to that expected for enzymes involved in primary metabolism  $(k_{cat}/K_m > 1\times 10^5~M^{-1}~s^{-1})$  but also lower than for the best substrates of BT4131  $(k_{cat}/K_m = 1.8\times 10^3~M^{-1}~s^{-1}$  -  $6.7\times 10^3~M^{-1}~s^{-1})$  and other HADSF enzymes involved in secondary metabolism (eg.  $\sim 1\times 10^4~M^{-1}~s^{-1}$  for E.~coli~NagD~(4))). Thus, BT1666 does not appear to be a very efficient catalyst. Because low catalytic efficiency might be an artifact of the tethered His\_6-tag associated with the recombinant protein, BT1666 was prepared without the tag and assayed. The activity of the native BT1666 was no greater. We also measured the pH rate profile (using V/K conditions with substrate mannose 6-phosphate, Figure S1) to determine if the optimal pH for BT1666 catalysis might significantly deviate from the solution pH used in the substrate activity screen (pH 7.5, similar to that used previously for BT4131 (pH 7.0)). Although the activity of BT1666 is higher at slightly acidic pH, the activity at pH 7.0 and pH 7.5 is only three-fold lower. Thus, the catalytic efficiency of BT1666 is less than that of BT4131, even for the best substrates.

BT1666 shows similar specificity constants toward seven of the sugar substrates assessed. However, BT1666 is not non-specific, as the  $k_{cat}/K_m$  for substrates such as  $\beta$ -D-glucose 6phosphate is ~100 fold lower than that of the best substrates. Comparison of the specificity profile previously reported for BT4131 (Table 2) (5) to that of BT1666 reveals some divergence in substrate recognition. For instance, the substrates D-fructose 1,6-(bis)phosphate and sucrose 6'-phosphate show 125 and 19 times higher k<sub>cat</sub>/K<sub>m</sub> for BT1666 compared to BT4131, respectively, but 2-deoxy-D-glucose 6-phosphate and D-arabinose 5phosphate show ~150 and ~20 fold higher specificity, respectively, for BT4131 compared to BT1666 (Table 2). Notably, the greatest contribution to the difference in specificity is in k<sub>cat</sub> not K<sub>m</sub>. Whereas the BT4131 gene context (juxtaposed to a chitobiase gene) together with the preference for hexose phosphates implicates a possible role in chitin recycling, the broad substrate range also raised the possibility of a housekeeping role. Gene context gives no clue as to the possible physiological function of BT1666; it is most frequently co-located with a putative phosphate acetyltransferase and a putative asparaginyl-tRNA synthetase. Overall, BT1666 is more active towards larger substrates than is BT4131, and the two together cover a larger substrate range than either alone can provide, which may confer a selective advantage in a general housekeeping function. The housekeepers would act by removing inhibitory phosphorylated metabolites which if they build to high concentrations, could prove to be cytotoxic.

### Structure determination

In order to understand the structural basis for their overlapping, but distinct, substrate profiles, the structure of BT1666 was determined. The final model of BT1666, including all 258 residues and the last 10 residues from the N-terminal His $_6$ -tag (SSGLVPRGSH), was refined to 1.82 Å resolution with a  $R_{\rm work}$  of 16.4% and  $R_{\rm free}$  of 19.8%. All residues are well-defined (see Figure S2 for sample electron density) with an average B-factor for the protein of 23.4 Å $^2$ . The final BT1666 model includes one  $Mg^{2+}$  ion, which occupies the cofactor binding site found in all HAD phosphotransferases (16). Because the crystallization buffer contained 0.2 M sulfate, a sulfate ion was modeled in electron density observed at the BT1666 active site where a phosphate ion or the phosphoryl group of the substrate binds in the structures of other HADSF members(17). Two sulfate ions were also modeled at the enzyme surface.

BT1666, which is composed of two distinct domains connected by two linker regions (Figure 2A), belongs to the C2b class of HAD members (Figure 1A) (5). The larger domain is the "conserved core" domain that houses the phosphoryl-transfer site. The slightly smaller

domain is the cap domain, which in other C2b members docks on the core domain to cover the active site and provide determinants of substrate specificity (18-21). The core domain is a modified Rossmann fold similar to that found in BT4131, including six parallel  $\beta$ -sheets surrounding by six  $\alpha$ -helices. The cap is a mixed alpha/beta fold with  $\alpha\beta\beta(\alpha\beta\alpha\beta)\alpha\beta\beta$  topology. As compared to other HAD members, BT1666 shows a unique feature in that there is a peptide from a neighboring molecule penetrating the active site (Figure 3). Analysis of the peptide shows that it comprises the final ten residues of the N-terminal His6tag of the adjacent symmetry related BT1666 monomer. The peptide is well ordered with B-factors of 25.2 Ų compared to those of the remaining protein structure (23.3 Ų; calculated excluding the ten residues of the peptide). This finding may give some insight into the mode of substrate binding of capped members of the HADSF that show phosphoprotein phosphohydrolase activity, including the cofilin-activating phosphatase chronophin (22), and Eyes absent (Eya) phosphatase responsible for dephosphorylating the C-terminal tyrosyl residue of histone H2A.X (23).

Based on similarity in sequence and structure BT1666 and BT4131 clearly share common ancestry and might be the products of gene duplication (Figure 2B; Figures S3-S5 for sequence alignments). Comparison of BT4131 and BT1666 shows that the protein structures have 1.93 Å rmsd and 40 % sequence identity overall, 1.95 Å rmsd and 25% sequence identity for the cap domain and 0.84 Å rmsd and 48% sequence identity for the core domain (notably the rmsd is biased by the fact that BT4131 was the model used for molecular replacement). The finding that the sequence conservation between the cap domains is considerably lower than that observed between the core domains indicates that BT4131 and BT1666 are diverging in function, as it is the cap that determines the physiological substrate(s). It is thus the cap domain that should be examined to understand the differences in substrate specificity between the two paralogues (*vide supra*).

The relative disposition of the respective cap and core domains in BT1666 and BT4131 differ. Specifically, BT1666 is in a cap-open conformation whereas BT4131 is in a cap-closed conformation. The cap-open conformation allows substrate to bind and product to dissociate. The cap-closed conformation is required for catalytic turnover (24-25). The superposition of the BT1666 and BT4131 domains separately (to remove the effect of open and closed conformations) revealed that cap and core domain residues that comprise the active site are identical (Figure 4). In view of the differences in the specificity profiles and in the catalytic efficiencies of the two enzymes, this was an unexpected finding. The differences in "second-shell" residues in the cap could change the size and shape of the active site, either by interactions with active-site residues or by affecting the fit between cap and core. In BT4131, because of the closed cap conformation, there is an enclosed area which resembles the possible substrates in size and shape(5). However, because of the capopen conformation, the size and shape of the active-site cavity cannot be assessed for BT1666. Thus, it is not possible to ascertain whether the differences in substrate specificity are due to differences in the active site volume.

In addition to "static" effects of sequence variation on the structure of the active sites of the two paralogues, it is possible that the dynamics of the two differ. Protein dynamics at the level of large substrate-induced conformational changes as well as more subtle rearrangements as the enzyme approaches the transition-state have been implicated in both the catalytic efficiency of enzymes and in the selectivity between two substrates(26-28). To examine patterns of conservation that might be linked to protein scaffold dynamics, the sequence similarity of each protein was mapped onto each structure (29). The variation in conservation follows the expected pattern with highest conservation of catalytic residues in the core domain and those few cap domain residues that help form the active site (Figure 2C and 2D). Although the lower catalytic efficiency and differing specificity profiles of BT4131

and BT1666 may well be due to conformational dynamics, the pattern of residue conservation does not highlight underlying structural changes to which these differences can be ascribed.

### Conclusion

Comparison of the protein structures and overlapping substrate profiles highlights the similarities more than the differences between BT1666 and BT4131. The question remains, why would the bacterium retain two such similar catalysts? First, the substrates profiles between these enzymes, though overlapping, do differ. Although the  $k_{cat}/K_m$  values for these substrates are low, and hence might have little physiological relevance, the catalytic proficiency  $((k_{cat}/K_m) \, / \, k_{uncat})$  and rate acceleration  $(k_{cat}/k_{uncat})$  are high and thus can provide some advantage under selection processes. In addition, the retention of two similar enzymes under the control of different promoters for expression could allow the recruitment of these activities under differing conditions(4).

It has been hypothesized that ancient enzymes possessed broad specificities such that a small number of catalysts could fill multiple metabolic roles in the cell(30). Divergence of these generalist enzymes, through gene duplication, mutation, and natural selection (31), has led to the formation of enzyme families, with individual members that are highly specific and proficient (although the order of events is unclear(32)). Based on their high overall sequence identity (40%) one can speculate that BT4131 and BT1666 represent a gene duplication event followed by limited divergence and honing toward specific substrates. The sequence conservation among orthologues of each enzyme (33 % for BT1666, 27 % for BT4131) does not allow speculation as to which was the "original" and which was the "copy". Neither does the distribution of the two enzymes (Table S1) which shows that out of the 28 Bacteroides species possessing either enzyme, 15 bear BT4131 alone, 6 bear BT1666 alone and 7 bear both enzymes. Indeed, given the exceedingly high frequency of lateral gene transfer among bacteria, both within the genus Bacteroides and among Bacteroides species and other gram-positive bacteria in the human colon(33), these paralogues may have been acquired together, or in separate events. Such lateral gene transfer events may introduce enzyme activities isofunctional to those already encoded by the recipient genome(34). Notably, it has been demonstrated that high and low efficiency forms of phosphoglycerate mutase stemming from two different enzyme families (isologous enzymes) present in E. coli have overlapping and complementary roles in the cell. Overexpression of the low efficiency form of phosphoglycerate mutase can complement deletion of the high efficiency form. This functional redundancy is thought to be best tolerated in larger bacterial genomes (>3.7 Mb) and B. theataiotaomicron falls in this category at 4.6 Mb(35). In the case where genes are essential, functional redundancy is a prerequisite for any subsequent selective loss of one gene. In the case of phosphoglycerate mutase this necessity leads to the detection of nonhomologous genes in the ancestral genome that encode enzymes capable of carrying out the same reaction. In the case of BT4131 and BT1666, it appears that divergence of a common ancestor or gene duplication in some combination with lateral gene transfer has led to the phenomenon of closely functioning, if not isofunctional paralogues in Bacteroides thetaiotaomicron where gene regulation can allow either selective usage or enhanced production of one paralogue to fill a role in secondary metabolism.

# **Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

# Abbreviations used are

**Bis-Tris** bis-(2-hydroxy-ethyl)-amino-tris(hydroxymethyl)aminomethane

**Eya** Eyes absent

**HADSF** haloalkanoate dehalogenase superfamily

**HEPES** 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

(KDN-9-P) 2-keto-3-deoxy-D-glycero-D-galacto-9-phosphonononic acid

**KDO-8-P** 2-keto-3-deoxy-8-phospho-D-manno-octulosonic acid

pNPP p-nitrophenyl phosphatermsd root-mean-square deviation

## References

1. Allen KN, Dunaway-Mariano D. Markers of fitness in a successful enzyme superfamily. Curr Opin Struct Biol. 2009; 19:658–665. [PubMed: 19889535]

- 2. Allen KN, Dunaway-Mariano D. Phosphoryl group transfer: evolution of a catalytic scaffold. Trends Biochem Sci. 2004; 29:495–503. [PubMed: 15337123]
- Burroughs AM, Allen KN, Dunaway-Mariano D, Aravind L. Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. J Mol Biol. 2006; 361:1003–1034. [PubMed: 16889794]
- 4. Tremblay LW, Dunaway-Mariano D, Allen KN. Structure and activity analyses of Escherichia coli K-12 NagD provide insight into the evolution of biochemical function in the haloalkanoic acid dehalogenase superfamily. Biochemistry. 2006; 45:1183–1193. [PubMed: 16430214]
- Lu Z, Dunaway-Mariano D, Allen KN. HAD superfamily phosphotransferase substrate diversification: structure and function analysis of HAD subclass IIB sugar phosphatase BT4131. Biochemistry. 2005; 44:8684–8696. [PubMed: 15952775]
- 6. Lu Z, Wang L, Dunaway-Mariano D, Allen KN. Structure-function analysis of 2-keto-3-deoxy-D-glycero-D-galactonononate-9-phosphate phosphatase defines specificity elements in type C0 haloalkanoate dehalogenase family members. J Biol Chem. 2009; 284:1224–1233. [PubMed: 18986982]
- 7. Wang L, Lu Z, Allen KN, Mariano PS, Dunaway-Mariano D. Human symbiont Bacteroides thetaiotaomicron synthesizes 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN). Chem Biol. 2008; 15:893–897. [PubMed: 18804026]
- 8. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology. 1997; 276:307–326.
- 9. Vagin A, Teplyakov A. MOLREP: an Automated Program for Molecular Replacement. Journal of Applied Crystallography. 1997; 30(Part 6):1022–1025.
- Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
- Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr. 2002; 58:1948– 1954. [PubMed: 12393927]
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr. 1998; 54:905–921. [PubMed: 9757107]
- 13. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 1993; 26(part 2):283–291.
- 14. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst. 1991; 24:946–950.

15. Fenn TD, Ringe D, Petsko GA. a program for model and data visualization using persistence of vision ray-tracing. Journal of Applied Crystallography. 2003; 36:944–947.

- Zhang G, Morais MC, Dai J, Zhang W, Dunaway-Mariano D, Allen KN. Investigation of metal ion binding in phosphonoacetaldehyde hydrolase identifies sequence markers for metal-activated enzymes of the HAD enzyme superfamily. Biochemistry. 2004; 43:4990–4997. [PubMed: 15109258]
- Lu Z, Dunaway-Mariano D, Allen KN. The catalytic scaffold of the haloalkanoic acid dehalogenase enzyme superfamily acts as a mold for the trigonal bipyramidal transition state. Proc Natl Acad Sci U S A. 2008; 105:5687–5692. [PubMed: 18398008]
- Kim Y, Yakunin AF, Kuznetsova E, Xu X, Pennycooke M, Gu J, Cheung F, Proudfoot M, Arrowsmith CH, Joachimiak A, Edwards AM, Christendat D. Structure- and function-based characterization of a new phosphoglycolate phosphatase from Thermoplasma acidophilum. J Biol Chem. 2004; 279:517–526. [PubMed: 14555659]
- 19. Rao KN, Kumaran D, Seetharaman J, Bonanno JB, Burley SK, Swaminathan S. Crystal structure of trehalose-6-phosphate phosphatase-related protein: biochemical and biological implications. Protein Sci. 2006; 15:1735–1744. [PubMed: 16815921]
- Shin DH, Roberts A, Jancarik J, Yokota H, Kim R, Wemmer DE, Kim SH. Crystal structure of a phosphatase with a unique substrate binding domain from Thermotoga maritima. Protein Sci. 2003; 12:1464–1472. [PubMed: 12824492]
- 21. Silvaggi NR, Zhang C, Lu Z, Dai J, Dunaway-Mariano D, Allen KN. The X-ray crystal structures of human alpha-phosphomannomutase 1 reveal the structural basis of congenital disorder of glycosylation type 1a. J Biol Chem. 2006; 281:14918–14926. [PubMed: 16540464]
- Gohla A, Birkenfeld J, Bokoch GM. Chronophin, a novel HAD-type serine protein phosphatase, regulates cofilin-dependent actin dynamics. Nat Cell Biol. 2005; 7:21–29. [PubMed: 15580268]
- 23. Krishnan N, Jeong DG, Jung SK, Ryu SE, Xiao A, Allis CD, Kim SJ, Tonks NK.

  Dephosphorylation of the C-terminal tyrosyl residue of the DNA damage-related histone H2A.X is mediated by the protein phosphatase eyes absent. J Biol Chem. 2009; 284:16066–16070.

  [PubMed: 19351884]
- 24. Dai J, Finci L, Zhang C, Lahiri S, Zhang G, Peisach E, Allen KN, Dunaway-Mariano D. Analysis of the structural determinants underlying discrimination between substrate and solvent in beta-phosphoglucomutase catalysis. Biochemistry. 2009; 48:1984–1995. [PubMed: 19154134]
- Dai J, Wang L, Allen KN, Radstrom P, Dunaway-Mariano D. Conformational cycling in betaphosphoglucomutase catalysis: reorientation of the beta-D-glucose 1,6-(Bis)phosphate intermediate. Biochemistry. 2006; 45:7818–7824. [PubMed: 16784233]
- Kellinger MW, Johnson KA. Nucleotide-dependent conformational change governs specificity and analog discrimination by HIV reverse transcriptase. Proc Natl Acad Sci U S A. 2010; 107:7734– 7739. [PubMed: 20385846]
- 27. Hammes-Schiffer S, Benkovic SJ. Relating protein motion to catalysis. Annual Review of Biochemistry. 2006; 75:519–541.
- 28. Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T. Hidden alternative structures of proline isomerase essential for catalysis. Nature. 2009; 462:669–673. [PubMed: 19956261]
- 29. Garcia-Boronat M, Diez-Rivero CM, Reinherz EL, Reche PA. PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. Nucleic Acids Res. 2008; 36:W35–41. [PubMed: 18442995]
- Jensen RA. Enzyme recruitment in evolution of new function. Annu Rev Microbiol. 1976; 30:409–425. [PubMed: 791073]
- 31. Ohno, S., editor. Evolution by gene duplication. Vol. xv. Allen & Unwin/Springer-Verlag; London-New York: 1970.
- 32. Khersonsky O, Tawfik DS. Enzyme promiscuity: a mechanistic and evolutionary perspective. Annu Rev Biochem. 2010; 79:471–505. [PubMed: 20235827]
- 33. Shoemaker NB, Vlamakis H, Hayes K, Salyers AA. Evidence for extensive resistance gene transfer among Bacteroides spp. and among Bacteroides and other genera in the human colon. Appl Environ Microbiol. 2001; 67:561–568. [PubMed: 11157217]

34. Foster JM, Davis PJ, Raverdy S, Sibley MH, Raleigh EA, Kumar S, Carlow CK. Evolution of bacterial phosphoglycerate mutases: non-homologous isofunctional enzymes undergoing gene losses, gains and lateral transfers. PLoS One. 2010; 5:e13576. [PubMed: 21187861]

35. Shaheduzzaman SM, Akimoto S, Kuwahara T, Kinouchi T, Ohnishi Y. Genome analysis of Bacteroides by pulsed-field gel electrophoresis: chromosome sizes and restriction patterns. DNA Res. 1997; 4:19–25. [PubMed: 9179492]

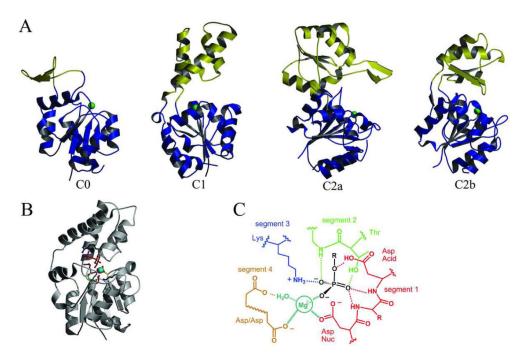
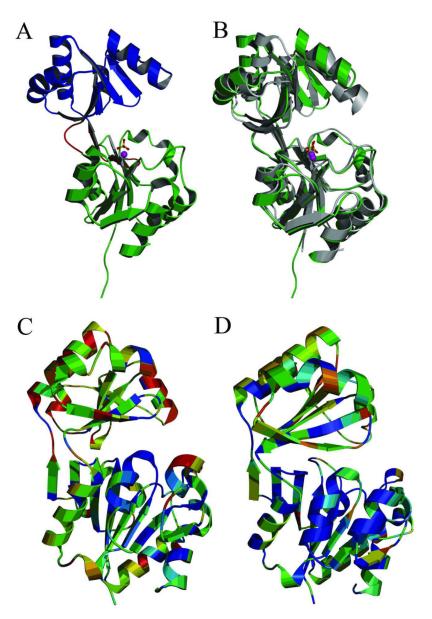
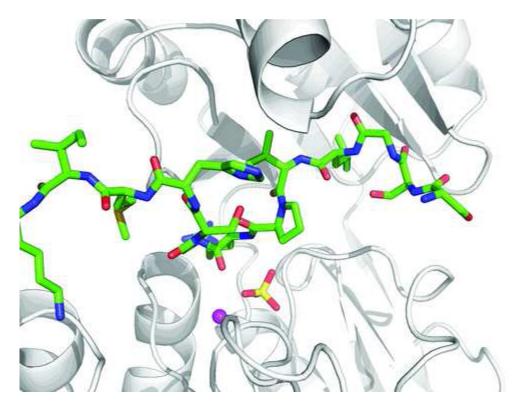


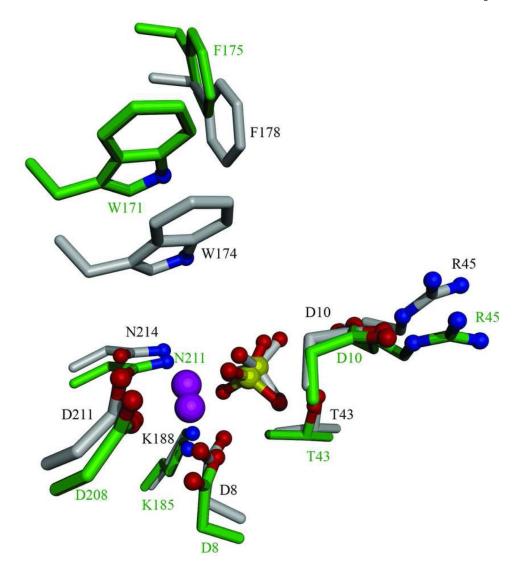
Figure 1. (A) The HADSF subfamilies-C1 and C2 are distinguished by the insertion points of the cap (gold) in the Rossmann fold (blue) (C1 after  $\beta$ -strand 1, C2 after  $\beta$ -strand 2, C0 has a minimal insert in either site). (B) Liganded  $\beta$ -PGM with catalytic segments colored as in (C), substrate in purple and one of the substrate-binding cap residues in black with Mg²+ in cyan.. (C) Chemdraw representation of the HADSF conserved catalytic scaffold.



**Figure 2.**(A) Ribbon diagram of BT1666 liganded to the cofactor Mg<sup>2+</sup> (magenta sphere) and sulfate ion (ball and stick). The core domain is colored green and cap domain is colored blue with the two interdomain linkers colored red. (B.) An overlay of the structures of BT1666 (green) and BT4131 (grey). Ribbon diagram of (C) BT1666 and (D) BT4131 with the sequence variability calculated from a multiple sequence alignment of each enzyme mapped onto the structure from blue (conserved) to red (variable).



**Figure 3.** The active site of BT1666 depicted as a ribbon diagram with the N-terminal 10 residues of the adjacent symmetry related monomer depicted as ball and stick.



**Figure 4.** Overlay of the active site of BT1666 with that of BT4131 colored as in Fig 2B. Residues W171/174, F175/F178 are contributed by the cap domain.

Data Collection Statistics	
Resolution (highest resolution shell) (Å)	50.00-1.82(1.89-1.82)
X-ray Source	CuKa
Wavelength (Å)	1.5418
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimension (Å)	a=52.8 b=70.7 c= 72.4
Reflections Observed (unique)	24886 (2449)
Completeness (%)	99.9 (100)
R <sub>merge</sub> <sup>a</sup> (%)	10.8 (55.7)
Ι/σ (Ι)	15.5 (3.0)
Redundancy	7.2 (7.0)
Refinement Statistics	
No. of protein residues/water atoms per asu	2063/332
No. Mg <sup>2+</sup> /sulfate ions per asu	1/3
Number of reflections (work/free)	24832/1264
R <sub>work</sub> /R <sub>free</sub> (%)	16.4/19.8
Resolution (Å)	18.1-1.82
Average B-factor (Å <sup>2</sup> )	25.1
Protein	23.4
$\mathrm{Mg}^{2+}$	8.5
Sulfate	37.8
Water	35.0
Root mean square deviation	
Bond length (Å)	0.008
Bond angle (°)	1.127

 $<sup>{}^{</sup>a}R_{merge} = \sum\nolimits_{hkl} \sum\nolimits_{i} |I_{hkl,i} - \langle I_{hkl} \rangle I| / \sum\nolimits_{hkl} \sum\nolimits_{i} |I_{hkl,i}|, \text{ where } \langle I_{hkl} \rangle \text{ is the mean intensity of the multiple } I_{hkl,\ i} \text{ observations for symmetry-related reflections.}$ 

Table 2

Steady-state kinetic constants for HPP and BT1666 catalyzed hydrolysis of phosphorylated small-molecule substrates in 50 mM HEPES containing 5 mM MgCl<sub>2</sub> (pH 7.0, 37 °C).

		BT1666			BT4131 <sup>a</sup>			BT1666/BT4131	T4131
substrate	$k_{\mathrm{cat}}\left(\mathrm{s}^{\mathrm{1}}\right)$	$\mathbf{K}_{\mathbf{m}}$ (mM)	$k_{cat}/\mathrm{K_m}$ ( $\mathrm{M}^{\text{-1}}\mathrm{s}^{\text{-1}}$ )	$k_{\rm cat}  ({ m s}^{-1})$	K <sub>m</sub> (mM)	$k_{cat}/\mathrm{K_m}$ ( $\mathrm{M}^{ ext{-}1}\mathrm{s}^{ ext{-}1}$ )	$k_{ m cat}$ / $k_{ m cat}$	$\rm K_m/K_m$	$k_{cat}/{ m K_m}/k_{cat}/{ m K_m}$
D-glucose 6-P	$2.9 (\pm 0.3) \times 10^{-1}$	$1.2 (\pm 0.3) \times 10^{1}$	$2.4 \times 10^1$	$3.6 \pm 0.2$	1.1 $(\pm 0.2) \times 10^1$	$3.3 \times 10^2$	80.0	1.1	0.072
2-deoxy glucose-6-P	$5.4 (\pm 0.3) \times 10^{-1}$	1.3 (± 0.2) × 10 <sup>1</sup>	$4.2\times10^{1}$	$2.6 (\pm 0.1) \times 10^{1}$	$3.9 \pm 0.4$	$6.7\times10^3$	0.02	3.3	900'0
glucosamine 6-P	$1.2\pm0.3$	5 ± 3	$2.4\times10^2$	$2.7~(\pm~0.2) \times 10^{-1}$	$6.7 \pm 0.9$	$4.0\times 10^{1}$	4.4	0.75	5.3
NAc-glucosamine 6-P	1.3 (± 0.1) × 10 <sup>-1</sup>	$1.1\pm0.2$	$1.2\times10^2$	$3.6\pm0.5$	$1.7 (\pm 5) \times 10^{1}$	$2.1\times10^2$	9:00	0.065	<i>LS</i> :0
mannose 6-P	$3.0 (\pm 0.3) \times 10^{-1}$	$2.5 \pm 0.5$	$1.2 \times 10^2$	$1.3 \pm 0.1$	$1.9 \pm 0.3$	$7.0 \times 10^2$	0.23	1.3	0.18
fructose 6-P	3.3 (± 0.3) × $10^{-1}$	8 ± 2	$4.1\times10^{1}$	$9 \ (\pm 1) \times 10^{-1}$	4 ± 1	$2.7 \times 10^2$	0.37	2	0.18
fructose 1,6-(bis)P	$1.1\pm0.1$	$2.5\pm0.5$	$4.4 \times 10^2$	1.4 (± 0.2) × 10 <sup>-2</sup>	4 ± 1	3.5	82	0.62	125
arabinose 5-P	$2.6 (\pm 0.1) \times 10^{-1}$	$1.5\pm0.1$	$1.7\times10^2$	$9.6\pm0.7$	$3.5 \pm 0.5$	$2.7\times10^3$	0.027	0.43	690.0
ribose 5-P	$4.2 (\pm 0.6) \times 10^{-1}$	$2.1\pm0.5$	$2.0\times10^2$	$8.7\pm0.4$	$4.9 \pm 0.6$	$1.8\times10^3$	0.048	0.43	0.11
sorbitol 6-P	$4.2 (\pm 0.6) \times 10^{-1}$	$6.5 \pm 3$	$6.5\times10^{1}$	$5.4 (\pm 0.2)$	$6.8 \pm 0.5$	$8.0\times10^2$	820.0	96.0	0.081
DL-α-glycerol 3-P	$4.2 (\pm 0.3) \times 10^{-1}$	$5.3 \pm 0.7$	$7.9 \times 10^1$	$1.04~(\pm~0.02)\times10^{1}$	$8.7 \pm 0.4$	$1.2\times10^3$	0.040	0.61	990'0
sucrose 6'-P	1.1 (± 0.1) × 10 <sup>-1</sup>	$1.4\pm0.3$	$7.9\times10^{1}$	1.3 (± 0.1) × 10-2	$3.1 \pm 0.1$	4.2	8.5	0.45	19
trehalose 6-P	$3.0 (\pm 0.3) \times 10^{-1}$	$2.2\pm0.4$	$1.4\times10^2$	$\sim 6 \times 10^{-2}$			2		
ADP	$4.5 (\pm 0.6) \times 10^{-1}$	$1.0 (\pm 2) \times 10^{1}$	$4.5\times10^{1}$	$1.6 (\pm 0.1) \times 10^{-2}$	$1.6 \pm 0.4$	$1.0\times10^{1}$	28	6.3	4.5
pyridoxal 5'-P	$2.9 (\pm 0.1) \times 10^{-1}$	$4.1\pm0.5$	$7.0\times 10^{1}$	$1.8\pm0.1$	$0.88 \pm 0.08$	$2\times 10^3$	0.16	4.7	0.035
pNPP	$5.5 (\pm 0.7) \times 10^{-2}$	$0.51\pm0.1$	$1.1 \times 10^2$	8.3 (± 0.2) × $10^{-2}$	$0.77 \pm 0.05$	$1.1\times10^2$	99:0	0.66	1

a from (5)