

YbiV from *E. coli* K12 is a HAD Phosphatase

Anne Roberts^{1,2}, Seok-Yong Lee^{2,3}, Emma McCullagh⁴, Ruth E. Silversmith⁵,

David E. Wemmer^{1,2*}

¹ Department of Chemistry, University of California, Berkeley, Berkeley, California 94720, USA

² Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

³ Graduate Group in Biophysics, University of California, Berkeley, Berkeley, California 94720, USA

⁴ Department of Molecular and Cellular Biology, University of California, Berkeley, Berkeley, California 94720, USA

⁵ Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7290, USA

* Corresponding author:

David. E. Wemmer, Department of Chemistry, University of California, Berkeley, Berkeley, California 94720, Phone (510) 486-4318, Fax (510) 486-6059, DEwemmer@lbl.gov

Keywords: phosphatase, ybiV, phosphoaspartate intermediate, HAD superfamily

ABSTRACT

The protein YbiV from *Escherichia coli* K12 MG1655 is a hypothetical protein with sequence homology to the haloacid dehalogenase (HAD) superfamily of proteins. Although numerous members of this family have been identified, the functions of few are known. Using the crystal structure, sequence analysis, and biochemical assays, we have characterized ybiV as a HAD phosphatase. The crystal structure of YbiV reveals a two domain protein, one with the characteristic HAD hydrolase fold, the other an inserted α/α fold. In an effort to understand the mechanism we also solved and report the structures of YbiV in complex with beryllifluoride (BeF_3^-) and aluminum trifluoride (AlF_3) which have been shown to mimic the phosphorylated intermediate and transition state for hydrolysis, respectively, in analogy to other HAD phosphatases. Analysis of the structures reveals the substrate binding cavity, which is hydrophilic in nature. Both structure and sequence homology indicate ybiV may be a sugar phosphatase, which is supported by biochemical assays which measured the release of free phosphate on a number of sugar-like substrates. We also investigated available genomic and functional data in an effort to determine the physiological substrate.

Acknowledgements

We wish to thank Ho Cho who was involved in the initial identification of YbiV as a potential phosphatase and Sydney Kustu for helpful discussions regarding the potential function of YbiV, and for access to certain microarray data. Thanks also to Jeremy Glasner and the ASAP site at www.genome.wisc.edu for permission to cite microarray data available on their site. This work was supported by National Institutes of Health grant GM 62163.

Atomic Coordinates

The atomic coordinates and structure factors in this study have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>) (codes 1RLM, 1RLO, and 1RLT.)

Abbreviations

MAD, multiwavelength anomalous dispersion; PEG, polyethylene glycol; SeMet, selenomethionine; HAD, haloacid dehalogenase; AMP, adenosine monophosphate

INTRODUCTION

Members of the haloacid dehalogenase (HAD) superfamily of enzymes comprise a large and diverse set of hydrolases found in the genomes of bacteria, archaea, and eukarya. This family includes phosphatases, phosphonates, P-type ATPases, α -phosphoglucomutases, phosphomannomutases, and dehalogenases, which are involved in cellular processes ranging from amino acid biosynthesis to detoxification.¹⁻² Although sequence identity across the entire HAD family is less than 15%, members share three conserved sequence motifs and a common catalytic mechanism, involving the formation of a covalent intermediate.²⁻³

Several members of the phosphatase/phosphohydrolase class of HAD proteins have been identified. In these proteins, a conserved aspartic acid residue serves as the nucleophile in the phosphotransfer reaction, resulting in the formation of a phospho-aspartate intermediate.³ HAD phosphatases are identified by three unique motifs that are conserved in sequence and spatial proximity in the structures. In motif I, **DXDX(T/V)**, the first aspartate forms the phospho-protein intermediate. In Motif II, **S/TXX**, the conserved serine or threonine is involved in hydrogen bonding to the phosphoryl oxygen. Motif III, **K-(X)₁₈₋₃₀(G/S)(D/S)XXX(D/N)**, is involved in phosphoryl oxygen hydrogen bonding and coordination of the magnesium ion required for activity.³⁻⁴

Several HAD member structures, including a calcium pump ATPase from rabbit⁵, α -Phosphoglucomutase (PGM) from *Lactobacillus lactis*⁶, YrbI (a 3-Deoxy-D-manno-octulosonate 8-phosphatase) from *Haemophilus influenzae*⁷⁻⁸, phosphonoacetaldehyde hydrolase (PhnX) from *Bacillus cereus*⁹, L-2 Haloacid Dehalogenase (L2H) from *Pseudomonas sp.* YL¹⁰, TM0651 (a putative carbohydrate phosphatase) from *Thermotoga maritima*¹¹, human mitochondrial deoxyribonucleotidase (HMD)¹², and a recently identified phosphoglycolate

phosphatase (PGP) from *Thermoplasma acidophilum*¹³ have been solved by x-ray crystallography. Phosphoserine phosphatase (PSP) from is the most extensively studied of the HAD phosphatases, with its reaction pathway having been mapped out crystallographically (*M. jannaschii*).^{4,14} All the HAD structures reveal a hydrolase domain consisting of an α/α Rossman fold that brings the conserved motifs together in the active-site. The structures also reveal inserted or “cap” domains of divergent sequences and sizes that confer specificity on the enzymes and result in the diverse substrate range.^{1,4,9}

An analysis of the size and placement of these cap domains reveals three distinct classes.¹ The first includes proteins that contain an inserted domain after Motif I, the second includes a large inserted domain after Motif II, and the third contains little to no inserted domain. PSP, α -PGM, and HMD belong to the first class and all the structures reveal an inserted domain composed of a 4 or 5 helix bundle.^{6,12,14} Several proteins have been identified as belonging to the second class, including trehalose and sucrose phosphatases and two proteins, TM0651 and PGP, for which structures have recently been solved.^{11,13} The third class has the fewest members and includes histidinol phosphatase and magnesium dependent kinase I.¹ Despite the classification of hundreds of proteins as putative HAD phosphatases, the physiological functions of few have been pinpointed.

The genome of *Escherichia coli* K12-MG1655 (*E. coli* K12) provides the best available backdrop with which to investigate the global function of one of these uncharacterized HAD phosphatases. It is the best characterized organism in terms of genomic annotation, the mapping of its metabolic and biosynthetic pathways, and understanding of the elements that govern transcriptional regulation. Numerous centers are devoted to the systematic classification and functional annotation of every gene (PEC, <http://www.shigen.nig.ac.jp/ecoli/pec>,

<http://www.genome.wisc.edu/>, Ecogene, <http://bmb.med.miami.edu/Ecogene/EcoWeb>, Ecocyc <http://www.ecocyc.org/>, RegulonDB).¹⁵⁻¹⁸ HAD phosphatases also appear to be particularly abundant in bacteria. *E. coli* K12 contains at least 9 of the type II class, with only one, *otsB*, a trehalose phosphatase, having an experimentally verified function¹⁹ (<http://www.shigen.nig.ac.jp/ecoli/pec>.)

Here we report the characterization of the gene product from *E. coli ybiV* (b0822). YbiV belongs to the typeII class of HAD proteins, contains the conserved HAD motifs, and is a HAD phosphatase. We have investigated its function and mechanism using crystallographic, biochemical, and bioinformatics tools. In particular, we have solved the structure of YbiV in complex with beryllium fluoride (BeF_3^-) and aluminum fluoride (AlF_3), as stable mimics of the phosphorylated intermediate and transition state for hydrolysis, respectively. In both PSP and the bacterial receiver domains CheY and NtrC (related to HAD proteins by circular permutation), the BeF_3^- adduct has been shown to closely resemble the phospho-protein intermediate, binding occurring in the active-site in a geometry similar to the phosphoryl group.^{4,20}

MATERIALS and METHODS

Materials

Oligonucleotide primers for constructing the non-his tagged YbiV were purchased from Operon (Emeryville, CA). The Enzcheck Phosphate Assay (Molecular Probes) was used for kinetic studies. Reagents for cloning were purchased from New England Biolabs, Novagen, or Invitrogen. Other materials for expression, purification, and crystallization were purchased from Sigma, Amersham Biosciences, and Quiagen.

Cloning

The his-tagged construct of YbiV was created by insertion of the gene (obtained by PCR amplification of *E. coli* K12) into the pRSETA vector (Invitrogen) between the BamHI and HindIII sites. Crystallization of the his-tagged YbiV did not yield high-quality crystals. A non his-tagged construct was cloned into a pET21d vector (Novagen.) The NcoI site in pET21d resulted in a serine to alanine mutation at position 2. A single base mutation in the region covered by the reverse primer resulted in a serine to tyrosine mutation at residue 267. This mutation occurs after the last helix in YbiV and is 3 residues from the end of what could be visualized in the electron density map. Both structural comparisons and the retention of para-nitrophenyl phosphatase activity indicated this mutation did not significantly affect the structure or activity of the protein. All the reported activity assays were performed on the his-tagged construct of YbiV which contained the published sequence in addition to the histidine tag¹⁵.

Expression and Purification

The his-tagged construct was expressed in BL21(DE3) cells grown overnight at room temperature and purified by Ni-NTA agarose according to the Qiagen procedure. The pET21d construct was expressed in BL21(DE3) containing the plasmid pACYC. Expression was induced by addition of 1mM IPTG at an O.D. of 0.6-0.8. The cells were grown overnight at 37°C and harvested by centrifugation. The cells were resuspended in 50 mM Tris pH 8.0, 50mM NaCl, 1mM DTT, and 200 μ M PMSF, lysed by sonication and centrifuged. The resulting supernatant was applied to a DEAE column, and the flow through applied to a Q column (Amersham Biosciences). Both columns were pre-equilibrated with 50 mM Tris pH 8.0, 25 mM NaCl. The flow through from both columns was concentrated and applied to a Sephadex-75 sizing column in 50 mM Tris, pH 8.0, 500 mM NaCl, 1mM EDTA, and .02% NaN₃. The

purified protein was dialyzed against 50mM Tris pH 7.4, 100mM NaCl, and 1mM DTT, and concentrated to ~20 mgs/ml. Glycerol was added to 5%. Selenomethionine containing protein was grown according to published procedures and treated the same as native protein²¹.

Crystallization and Data Collection

Crystals of YbiV were obtained using the hanging drop method. For the “native” protein, (referred to henceforth as protein crystallized without beryllium or aluminum fluoride) 2 μ l of concentrated protein solution were mixed with 2ul of reservoir solution, which contained 0.1M cacodylic acid (pH 5.3), 0.2 M sodium acetate, and 15 % (w/v) PEG 8k. Large crystals appeared within hours. For AlF₃ and BeF₃⁻ derivatized protein, MgCl₂ was added to the protein solution to 5mM Mg²⁺ and allowed to incubate, while BeCl₂ or AlCl₃ were combined with a 5x molar excess of NaF. These mixtures were added to the protein solution to a final concentration of 2mM. Crystals formed under the same conditions as the native protein but appeared in a couple of days. Selenomethionine derivatized protein crystallized under the same conditions but at higher PEG concentration (18%). Glycerol was used as a cryoprotectant (added to a final concentration of 20%) and the crystals were flash frozen in liquid nitrogen before data collection. All crystals belonged to the space group P2₁2₁2₁, had similar cell dimensions, and contained 4 protein molecules per asymmetric unit (see Table I.)

All data were collected at the Advanced Light Source on Beamlines 5.0.2 and 8.3.1. These beamlines are equipped with ADSC Quantum4 CCD Detectors. Data processing and scaling were carried out using HKL2000, DENZO, and SCALEPACK.²²

Structure Determination

The structure of YbiV was initially solved using three wavelenth MAD data from selenomethionine derivatized protein, from crystals that diffracted to 2.8 Å. The protein

sequence contains 8 methionines and 32 sites were expected based on four monomers per asymmetric unit. 25 of the 32 sites were found by the program SOLVE and refined using the program SHARP.²³⁻²⁴ The native and SeMet data were merged with the phases found from SOLVE and the program RESOLVE was used to improve the phase, for automated model building, and density modification.²⁵⁻²⁶ The model produced by RESOLVE was largely complete and was fully completed by manually adding residues according to the $2F_o - F_c$ density map produced by the program CNS, and visualized using the program O.²⁷ After model building the structure was refined against the native data using the program CNS.²⁸ Non-crystallographic symmetry averaging was employed until the final stages of refinement, when the NCS restraints were relaxed and each molecule was refined individually. For the AlF_3 derivatized protein, the native structure was used as a starting model and refined against the AlF_3 reflections using CNS. Any differences in the structures were adjusted manually to fit the AlF_3 $2F_o - F_c$ map, including the placement of aluminum fluoride in the active-site. For the BeF_3^- -derivatized protein, the molecular replacement protocol in CNS found three of the molecules and the fourth molecule was located manually after some refinement and examination of the $2F_o - F_c$ map. Refinement of the structures was carried out by manual adjustment followed by minimization using CNS. Ordered water molecules were picked using an automated script in CNS and then manually inspected.

Structure Analysis and Representation

Analysis of the structure quality was done using procheck from the CCP4 suite of programs.²⁹ Figures were made using Molscript, Raster3D, and Pymol.³⁰⁻³²

Phosphatase Assays

The ability of his-tagged YbiV to catalyze the dephosphorylation of various compounds was measured using the Enzchek Phosphate Assay (Molecular Probes) adapted to a 96-well plate format. YbiV reactions were carried out in plate wells at room temperature in a total volume of 50 μ L. Reactions were initiated by addition of substrate (15 mM) to a solution containing YbiV (0.024-3.0 μ M) in 50 mM MES, pH 5.4, 10 mM MgCl_2 . After 30 minutes, 200 μ L of quench/assay reagent [270 mM Tris, pH 7.5, 27 mM EDTA, and Enzchek assay components MESG (270 μ M) and purine nucleoside phosphorylase (1.5 U/mL) was added and mixed by repeated pipetting. For each substrate, a control reaction lacking YbiV was run in an identical fashion. Each plate also contained phosphate standards (0-300 μ M in 50 μ L reaction buffer). After fifteen minutes incubation with quench/assay reagent, absorbances at 360 nm were read using a Molecular Devices Spectramax 340 plate reader. All samples were run in triplicate and the values averaged. For each substrate, the rate of YbiV-dependent phosphate release was the difference between the rate for the YbiV-containing reaction and the control lacking YbiV. For several substrates, substrate concentration was varied (1-50 mM) and double reciprocal plots used to deduce K_m and k_{cat} values.

RESULTS and DISCUSSION

Quality of the Models

The structures of the native, BeF_3^- and AlF_3 bound YbiV were refined to 1.9, 2.0, and 2.2 Å, respectively, with pdb accession codes 1RLM, 1RLO, and 1RLT. Statistics for the data collection and refinement are given in Tables I and II. All the models were complete except for the first and or second residues, and the last residue. Regions of higher B-factors were located around loops (residues 151-161), and in turns (residues 178-181), and at the termini, depending

on crystallographic contacts present for a particular molecule in the asymmetric unit. The final models included magnesium ions, glycerol molecules, and water molecules. Beryllium or aluminum trifluoride were clearly visualized in the active-sites of those protein molecules whose crystallization buffers contained those additives.

Overall Structure and Structural Comparisons

The crystal structure of YbiV reveals a two domain protein, the characteristic hydrolase domain with an α/α Rossmann fold, and an α/α domain corresponding to the inserted domain (figures 1A and 1B). At the interface of the two domains is the active-site, which forms a negatively charged cavity (figure 1C). The hydrolase domain consists of a 6 stranded parallel β sheet surrounded by 6 alpha helices. The sheet is extended by a β hairpin the end of which leads to the inserted domain. The inserted domain occurs between Motifs II and III (as expected from sequence analysis) and extends from residues 80-190. The inserted domain contains a 6 stranded mixed β sheet surrounded by 3 β helices.

The topology of YbiV is very similar to that of TM0651, which has an unknown function but is speculated to have carbohydrate substrate(s).¹¹ When the full structures of TM0651 and native YbiV are overlayed, the C α rmsd is ~ 1.85 Å, with the hydrolase domains appearing slightly offset. If the domains are detached and overlayed independently, the rmsd reduces to 1.6 Å for the inserted and 1.3 Å for the hydrolase domains. This indicates that there may be some flexibility in the relative domain orientation depending on the state of the protein, a trait attributed to other HAD phosphatases.^{4,13} More generally, the structures of these two proteins are probably representative of many of the type II class of HAD hydrolases, with specificity achieved by altering the composition of the residues that line the inserted domain.

A search for structurally similar proteins on the DALI server yielded many other HAD hydrolases and proteins that contain the Rossman fold.³³ The proteins with the highest Z-scores were YrbI and PSP. Despite differences in the placement, size, and topology of the cap domains, the hydrolase domains of YbiV, YrbI, and PSP overlay with an r.m.s.d of 1.5 Å, reflecting the strong structural conservation of this domain among HAD proteins (figure 1D).^{7,14}

Inserted Domain and Substrate Binding

In all 4 monomers of the native YbiV crystals, at least one glycerol molecule is bound in the inserted domain in a hydrophilic cavity, forming several hydrogen bonds with residues that line the cavity. This is likely the binding pocket for the true substrate. Hydrogen bonds to the glycerol hydroxyls originate from inserted domain residues Tyr 130, Ser 150, Ser 178, and Asp 184, and hydrolase domain residues Asn 46 (Motif II) and Asn 68 (figure 2A). Lys 148 forms the back of the cavity, hydrogen bonding to the carbonyl of Gly 45 (Motif II) and forming a salt bridge with Glu 67. These residues line both sides of the cavity and suggest that a substrate with several polar groups is most likely to bind in it.

As in native crystals, there are 4 molecules per asymmetric unit in the AlF_3 and BeF_3^- bound structures. In each of these, two molecules very closely resemble the native structures, including the glycerol molecule bound in the inserted domain. In the other two molecules in each case, however, Phe 180 in the turn between strands 10 and 11 shifts over to stack against His 129, effectively closing off the inserted domain pocket. This shift tilts helix 5 out and slightly changes the orientation of strands 10 and 11 but otherwise leaves the structure unperturbed (figure 2B). This stacking interaction may be a means of preventing substrate from binding while the protein is phosphorylated. Analogously, a flexible hinge and disordered loop

in PSP clamp and order over the active-site in the phosphorylated, substrate, and product-bound states.⁴

Active-Site Architecture

Native Active-Site

A common feature of HAD phosphatases is the negatively charged environment of the active-site.^{11,13} In each of the native, beryllofluoride, and aluminum-fluoride bound structures of YbiV, this negative charge is tempered by the presence of a magnesium ion and several ordered water molecules that form an extensive hydrogen bonding network. For the native YbiV, no additional magnesium was added to the protein purification or crystallization buffers, indicating the ion co-purified with the protein.

In the native structure the Mg^{2+} ion is hexacoordinate with Asp 9 (the site of phosphorylation), the carbonyl of Asp 11 (Motif I), Asp 215, Ser 216 (Motif III) and watA and watB (figure 3A) as ligands. Lys 192 forms a salt bridge with Asp 219 (Motif III) (2.5 Å). Wat C hydrogen bonds to the carboxyl of Asp 11 while watD bridges Asp 9 and Ser 44 (Motif II.) As might be expected from sequence and structural conservation, the active-sites of HAD phosphatases are highly superimposable. Despite differences in sequence around Motif III, the active sites of magnesium-bound YbiV and cobalt bound YrbI superimpose well (figure 3B).⁷

BeF₃⁻ Bound Active-Site

Both response regulators and HAD phosphatases share the Rossman fold and use an active-site aspartate as the site of phosphorylation. In the response regulators NtrC and CheY, beryllofluoride (BeF₃⁻) has been shown to bind the active-site aspartate in a tetrahedral geometry similar to the phosphoryl group. Functional studies indicate that BeF₃⁻ binding induces the same conformational changes as phosphorylation and leads to downstream signalling.²⁰ This indicates

that BeF_3^- both structurally and functionally mimics the phosphorylated state. Among HAD phosphatases, BeF_3^- bound structures have been solved for both PSP and HMD in an effort to better understand their reaction pathways.^{4,12}

The $2F_o-F_c$ map of the beryllofluoride-bound active-site of YbiV reveals density corresponding to the 3 fluorines, which, together with O2 of Asp 9, form a tetrahedral structure around the beryllium (figure 4A). The magnesium ion is coordinated by F1, the carbonyl of Asp 11, Asp 215, watA, watB (which mediates coordination by both Ser 216 and Asn218) and Asp 9. F2 forms long hydrogen bonds to Lys 192 (2.80 Å), Asn 218 (3.00 Å) (Motif III), and the amide of Gly 45 (Motif II) (3.10 Å). F3 is hydrogen bonded to Ser 44 (Motif II) (2.60 Å). WatC, coordinated by the carbonyl of Ser 178 from the inserted domain and the carboxyl of Asp 11 (2.70 Å), is collinear with both the beryllium and O2 of Asp 9 at a distance of 3.50 Å (figure 4B). This active-site geometry is very similar to that seen in the BeF_3^- bound active-site of PSP. As in that structure, watC is in an ideal position to act as the hydrolytic water molecule, activated by the carboxylate of Asp 11 which acts as a base.⁴

AlF_3 Bound Active-Site

The AlF_3 -bound active site is very similar to that with the BeF_3^- bound (figure 4C). The hydrogen bond distance between Ser 44 and F2 has increased slightly from 2.6 Å to 3.0 Å. The largest changes occur around the aluminum fluoride. The distance from O2 of Asp 9 to the aluminum increases while the distance from watC to the aluminum has shortened significantly (from 3.4 to 2.3 Å.) This reflects the intermediate (planar) state of the phosphoryl group as it is hydrolyzed and agrees well with a recent structure which trapped a putative pentavalent phosphate intermediate of phospho-glucomutase (though an alternative explanation of the density as MgF_3^- has been proposed).³⁴⁻³⁶

Mechanism

The structure of YbiV in its native, BeF_3^- , and AlF_3^- bound forms suggests that YbiV shares a similar mechanism for phosphotransfer as the best characterized HAD phosphatase, PSP. The YbiV structures all contain the physiologically relevant Mg^{2+} ion that orients the HAD active-site motifs and facilitates subsequent binding of the phosphate group of the substrate. The binding of one or more glycerol molecules in the inserted domain reveals the substrate binding pocket, which is hydrophilic in nature. When the substrate is properly oriented in the active-site, Asp 9 nucleophilically attacks the phosphate group which results in protein phosphorylation. The phosphorylated intermediate is stabilized by coordination by Lys 192 (motif III), the amide of Gly 45 (motif II), and Mg^{2+} . At this point, the substrate likely acquires a proton either from a water molecule or a residue in the active-site and is released. In the absence of substrate Phe 180 shifts over to stack against His 129 on the opposite side of the inserted domain. This effectively closes off the substrate binding pocket while the protein is phosphorylated. A water molecule coordinated by the carbonyl of Ser 178 and the carboxylate of Asp 13 likely acts as the hydrolytic molecule after having a proton extracted by Asp 13. This is supported by the BeF_3^- and AlF_3^- bound structures which mimic the phosphorylated state of the protein, and the transition state of hydrolysis, respectively.

Sequence Comparison

A PSI-BLAST search of YbiV yields hundreds of hits from putative HAD hydrolases.³⁷ The highest scoring sequences (from 99 % to 51 % identity) are from closely related organisms (*Salmonella typhimurium*, *Shigella flexneri*, other *E. coli* strains) and include another protein from *E. coli* K12, YbjI. A sequence alignment of YbiV with homologs of varying degrees of sequence similarity is shown in figure 5.

Although many of the absolutely conserved residues in the alignment stem from the active-site-motifs, a careful inspection of the YbiV sequence alignment reveals that many of the residues lining the inserted domain are completely or strongly conserved among the closest homologs of YbiV (first 6 sequences in the alignment), indicating their potential functional relevance. These include residues N46, E67, N68, Y130, K148, S178, and D184, which are all involved in glycerol binding or structural integrity of the inserted domain (i.e. K148-E67 forming the back of the inserted domain cavity.) This suggests perhaps structurally similar substrates for these proteins. Although the two residues involved in the stacking interaction, Phe 180 and His 129, are not strictly conserved in the closest homologs, they are conserved as aromatic residues with YbjI from *E. coli* K12 as the lone exception.

The alignment also indicates factors that may alter substrate specificity. For instance S150 of YbiV, which lines the inserted domain and forms a hydrogen bond to glycerol, is replaced by a glycine in YbjI, resulting in less hydrogen-bonding potential.

After these highest scoring hits come hundreds of other type II hydrolases from bacteria, ranging in sequence identity from 20-40% with YbiV. Most of these proteins are classified as belonging to the COF-subfamily of typeII hydrolases, all predicted to have similar inserted domain topology to YbiV (family classifications: TIGR00099 www.tigr.org and IPR000150 www.ebi.ac.uk/interpro.) Proteins belonging to this subfamily are almost exclusively present in bacteria. *E. coli* K12 has 6, *Listeria monocytogenes* EGD-e has 12, *Lactococcus lactis* subsp. *lactis* has 8, etc. The COF-subfamily is a subset of subfamily IIB (TIGR01484, IPR006379). Other subfamilies of IIB include trehalose-phosphatases, sucrose-phosphate phosphatases, eukaryotic phosphomannomutases, and YedP hydrolase (a putative mannosyl-3-phosphoglycerate phosphatase).³⁸ Connections between these subfamilies are evident in the

albeit-lower scoring hits of the PHI-Blast search of YbiV, which include sucrose and trehalose phosphatases. The remaining four sequences in the alignment include sucrose phosphatase from maize ³⁹, trehalose phosphatase from *E. coli* K12, and Cof and YbhA, two of the remaining 4 predicted type IIB hydrolases from *E. coli* K12 (figure 5). It is clear that YbiV is more distantly related to these proteins. However, the overall homology in the placement of the motifs and sequence length is quite strong, indicating their structures are likely similar. Gaps in sequence, corresponding to secondary structure in the inserted domain, may indicate differences in the inserted domain topology, particularly for sucrose and trehalose phosphatases.

Catalytic Activity

Potential substrates for YbiV were tested in *in vitro* assays and ranked according to the rate at which inorganic phosphate (Pi) was released relative to the best substrate tested (Table III). YbiV had negligible activity towards phosphoserine, phosphothreonine, and phosphocreatine, some of the first substrates tested. The hydrophilic nature of the inserted domain (as revealed by the structure and bound glycerol) led us to try substrates with several polar groups. Glycerol-1-p and glycerol-2-p were hydrolyzed appreciably, with glycerol-1-p being hydrolyzed slightly faster. The best substrates in our test panel were ribose-5-p and glucose-6-p. Other sugar molecules, including 2-deoxy-glucose-6-p, mannose-6-p, and fructose-6-p are hydrolyzed appreciably, further indicating a terminal sugar phosphate as a likely candidate for the substrate. Sucrose-6-p and glucose-1-p hydrolyzed very poorly in comparison to the other substrates. Sucrose-6-p is a two-ringed sugar molecule that is too large to be accommodated in the active-site. Though glucose-1-p is also a terminal sugar phosphate, steric clashes made by placing C6 towards the top of the inserted domain may prevent it from being hydrolyzed. The other possible explanation is that this molecule cannot ring open to a linear

form, as opposed to glucose-6-p and ribose-5-p, which can. Particularly in their linear forms, these molecules fit nicely into the active-site, with approximately the correct length necessary to reach Asp 9 and form hydrogen bonds on either side of the cavity (figure 6). This suggested to us a linear sugar as a candidate for the substrate, though glucose-6-p and ribose-5-p are predominantly in their cyclized form in solution. Although 6-p-gluconate is linear, it hydrolyzed poorly, which may reflect the inability of the inserted domain to accommodate a molecule with a negative charge due to electrostatic repulsion with Asp 184. Sorbitol-6-p which is linear and differs from glucose-6-p only in the replacement of a carbonyl oxygen with a hydroxyl was also appreciably hydrolyzed, though about 5 times less rapidly than the best substrate, ribose-5-p. Thus, a cyclic sugar cannot be definitively ruled out as a possible substrate.

The kinetic constants (k_{cat} and K_m) were analyzed for ribose-5-p, glycerol-1-p, and glycerol-2-p (Table III). We observe values for K_m and k_{cat} ranging from approximately 4-6mM and 1-4 s⁻¹, respectively. This is very similar to the range of kinetic constants observed when YrbI from *H. influenzae* was first tested for phosphatase activity.⁷ However, a later study, which isolated 3-deoxy-d-manno-octulosonate 8-phosphatase activity from *E. coli* K12 (an activity previously unassociated with a specific gene) identified the source of activity as YrbI, with kinetic constants K_m and k_{cat} of 75 μ M and 175 s⁻¹, respectively. These constants better reflect the high specificity typically associated with enzymatic activity.⁸ Although the likeliest interpretation of the kinetic data is that the physiological substrate of YbiV has not been found, the possibility still exists that some enzymes may be designed to have poorer binding and/or catalytic efficiency as a means of regulating activity.

Genomic Location and Functional Genomics

Often genes that are functionally related are grouped within the genome or belong to the same transcriptional unit or operon. For instance, *otsB*, the *E. coli* K12 gene for trehalose phosphatase is in the same operon as *otsA*, the trehalose synthetase gene (<http://www.ecocyc.org/>). *YbiV* (b0822) is surrounded by genes that are seemingly diverse in nature, and is not predicted to form part of an operon (RegulonDB).¹⁸ Surrounding genes encode proteins that are annotated as putative ABC transporters (b0820), unidentified (b0821), pyruvate formate lyase (b0823), pyruvate formate lyase activator (b0824), and the start of the molybdenum uptake operon (b0826) (http://bayesweb.wadsworth.org/binding_sites).⁴⁰ Ten genes upstream is the *dps* gene (b0812), a global regulator under starvation conditions (www.ecocyc.org).¹⁷ While it has been shown that genes of diverse functions may belong to the same operon, the above information does not provide many clues to the function of *YbiV*.⁴¹ We also looked at genomic location of genes with gene products of > 20% sequence identity to *YbiV* (*YigL*, *YbjI*, *Cof*, *YidA*, *YbhA*). Many were in similarly diverse regions except for *YbhA*, the last gene in the molybdenum uptake operon. Interestingly, *YbjI* (b0844) (50% identity to *YbiV*) is located 22 genes downstream of *YbiV*, in a region with genes involved in the DEOR (deoxyribose) metabolic pathway (<http://www.shigen.nig.ac.jp/ecoli/pec>.)

Another genomics based tool for functional discovery is that of functional genomics. The two main approaches to functional genomics are monitoring expression levels of all genes in response to different external conditions, and making a knockout of each gene and determining the phenotype.⁴²⁻⁴⁴ The first type of experiment typically reports the ratio of mRNA transcripts present in response to growth conditions (different growth media, knockouts of transcription factors, regulators, stationary growth phase, etc.) compared to a control condition. Results under

different conditions can be clustered to group genes that might be involved in related pathways. Although universal standards for the interpretation of DNA microarray data are not yet available, such experiments have already proven useful in identifying the cellular functions or pathways of proteins of unknown function.⁴⁵ The second approach assesses the ability of a knockout to grow on many different types of media (sugars, amino acids, etc.). Both the ability to grow and the speed at which the cells propagate can be monitored colorimetrically and can be compared to wild-type cells⁴². This methodology has been validated by work on knockouts of “global regulators of carbon metabolism”. Systematic mutagenesis of every *E. coli* K12 gene is underway and each strain will be subjected to this type of analysis (<http://www.genome.wisc.edu/>). Series of multi-gene knockouts have already determined genes or regions required for *E. coli* survival, and these studies indicate that b0822 is not essential (<http://www.shigen.nig.ac.jp/ecoli/pec.>)

In the case of YbiV, published microarray data have not revealed a condition in which expression of YbiV is greatly induced or suppressed. (Among homologous proteins, the protein YidA is slightly upregulated during growth on glucose.⁴⁶) Although access to unpublished “raw” microarray data is becoming increasingly available, it is difficult to judge how meaningful changes in expression are or define a “threshold” for up or down regulation without a full statistical analysis. Even so, unpublished, calibrated microarray data from www.genome.wisc.edu/functional.htm shows an increase in *ybiV* mRNA production upon mutation of the cAMP Receptor Protein, CRP (a ubiquitous cyclic AMP receptor and global regulator of secondary carbon metabolism).^{42,47,48} Further investigation reveals a CRP binding site 56 bp upstream from the start of *ybiV*, in a noncoding region⁴⁹ (http://arep.med.harvard.edu/ecoli_matrices/sco/crp.dba). This combination of results suggests

CRP is involved in the regulation of expression of YbiV but it is difficult to glean more specific information from this finding.

The lack of significant enhancement or suppression of expression of YbiV has two possible explanations. The most obvious is that the proper conditions for the induction of expression have not yet been probed. The second is that the protein, for perhaps regulatory reasons, is constitutively expressed at low levels.

CONCLUSION

Mechanistic insights provided by the native, berylliofluoride, and aluminum-fluoride bound structures of YbiV, sequence homology, and biochemistry indicate that YbiV is a phosphatase with a sugar-like substrate. Although the physiological substrate could not be determined, the pronounced presence of homologous proteins in certain bacteria, and absence in other branches perhaps indicate the sugars are unique to these organisms, or that the pathway involved is unique to these organisms, and not yet discovered. Since small molecules are generally present and metabolically active in their phosphorylated forms, this leads to the question of the functional significance of these reactions, whether they are regulatory (and perhaps active at some intrinsically low rate), involved in recycling of molecules, or induced by certain types of stress, less obvious than those that induce the synthesis of trehalose for example. But, given the relatively close homology between the type II hydrolases present, it seems unlikely that these are non-specific phosphatases.

The combination of genomics, structural genomics, and functional genomics are complementary techniques. As in this case, structure and to some extent sequence analysis yielded a good indication of function, pinning down the nature of the reaction if not the specific substrate. This latter task may be much more difficult to achieve. Selective testing of

compounds is useful in narrowing the range of possible substrates, but high-throughput techniques such as expression profiling and phenotypic analysis of gene knockouts may be increasingly important in both determining the cellular substrate and its importance in the metabolic pathways of the cell.

References

1. Selengut JD. MDP-1 Is a New and Distinct member of the Haloacid Dehalogenase Family of Aspartate-Dependent Phosphohydrolases. *Biochemistry* 2001;40(42):12704-12711.
2. Koonin EV, Tatusov RL. Computer Analysis of Bacterial Haloacid Dehalogenases Defines a Large Superfamily of Hydrolases with Diverse Specificity. *Journal of Molecular Biology* 1994;244:125-132.
3. Collet J-F, Stroobant V, Pirard M, Delpierre G, Van Schaftingen E. A New Class of Phosphotransferases Phosphorylated on an Aspartate Residue in an Amino-terminal DXDX(T/V) Motif. *The Journal of Biological Chemistry* 1998;273(23):14107-14112.
4. Wang W, Cho HS, Kim R, Jancarik J, Yokota H, Nguyen H, Grigoriev IV, Wemmer DE, Kim S-H. Structural Characterization of the Reaction Pathway in Phosphoserine Phosphatase: Crystallographic "Snapshots" of Intermediate States. *Journal of Molecular Biology* 2002;319:421-431.
5. Toyoshima C, Nakasako M, Nomura H, Ogawa H. Crystal Structure of the Calcium Pump of Sarcoplasmic Reticulum at 2.6 Å Resolution. *Nature* 2000;405(6787):647-655.
6. Lahiri SD, Zhang G, Dunaway-Mariano D, Allen KN. Caught in the Act: The Structure of Phosphorylated *B*-Phosphoglucomutase from *Lactococcus lactis*. *Biochemistry* 2002;41(26):8351-8359.
7. Parsons JF, Lim K, Tempczyk A, Krajewski W, Eisenstein E, Herzberg O. From Structure to Function: YrbI from *Haemophilus Influenzae* (HI1679) is a Phosphatase. *Proteins: Structure, Function, and Genetics* 2002;46:393-404.
8. Wu J, Woodard RW. *Escherichia coli* YrbI is 3-Deoxy-D-manno-octulosonate 8-Phosphate Phosphatase. *Journal of Biological Chemistry* 2003;278(20):18117-18123.
9. Morais MC, Zhang W, Baker AS, Zhang G, Dunaway-Mariano D, Allen KN. The Crystal Structure of *Bacillus cereus* Phosphonoacetaldehyde Hydrolase: Insight into Catalysis of Phosphorus Bond Cleavage and Catalytic Diversification within the HAD Enzyme Superfamily. *Biochemistry* 2000;39(34):10385-10396.
10. Hisano T, Hata Y, Fugii T, Liu J-Q, Kurihara T, Esaki N, Soda K. Crystal Structure of L-2-Haloacid Dehalogenase from *Pseudomonas* sp. YL: an α/α Hydrolase Structure that is Different from the α/α Hydrolase Fold. *Journal of Biological Chemistry* 1996;271(34):20322-20330.
11. Shin DH, Roberts A, Jancarik J, Yokota H, Kim R, Wemmer DE, Kim S-H. Crystal Structure of a Phosphatase with a Unique Substrate Binding Domain from *Thermotoga maritima*. *Protein Science* 2003;12(7):1464-1472.
12. Rinaldo-Matthis A, Rampazzo C, Reichard P, Bianchi V, Nordlund P. Crystal Structure of a Human Mitochondrial Deoxyribonucleotidase. *Nature Structural Biology* 2002;9(10):779-787.

13. Kim Y, Yakunin AF, Kuznetsova E, Xu X, Pennycooke M, Gu J, Cheung F, Proudfoot M, Arrowsmith CH, Joachimiak A, Edwards A, Christendat D. Structure and Function-based Characterization of a New Phosphoglycolate Phosphatase from *Thermoplasma acidophilum*. Journal of Biological Chemistry 2003;in press, published online Oct. 10:pp. 42.
14. Wang W, Kim R, Jancarik J, Yokota H, Kim S-H. Structure of Phosphoserine Phosphatase from *Methanococcus jannaschii*, a Hyperthermophile, at 1.8 Angstrom Resolution. Structure 2001;9(1):65-71.
15. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. The Complete Genome Sequence of *Escherichia coli* K-12. Science 1997;277(5331):1453-1474.
16. Rudd KE. EcoGene: A Genome Sequence Database for *Escherichia coli* K-12. Nucleic Acid Research 2000;28(1):60-64.
17. Karp PD, Riley M, Saier M, Paulsen IT, Paley S, Pellegrini-Toole A. The EcoCyc Database. Nucleic Acid Research 2002;30(1):56-58.
18. Salgado H, Santos-Zavaleta A, Gama-Castro S, Millan-Zarate D, Diaz-Peredo E, Sanchez-Solano F, Perez-Reuda E, Bonavides-Martinez C, Collado-Vides J. RegulonDB (version 3.2):Transcriptional Regulation and Operon Organization in *Escherichia coli* K-12. Nucleic Acid Research 2001;29(1):72-74.
19. Kaasen I, Falkenberg P, Styrvoid OB, Strom Ar. Molecular Cloning and Physical Mapping of the otsBA Genes, Which Encode the Osmoregulatory Trehalose Pathway of *Escherichia coli*: Evidence that Transcription is Activated by katF (AppR). Journal of Bacteriology 1992;174(3):889-898.
20. Yan D, Cho H, Hastings CA, Igo MM, Lee S-Y, Pelton JG, Stewart V, Wemmer D, Kustu S. Beryll fluoride Mimics Phosphorylation of NtrC and other Bacterial Response Regulators. Proceedings of the National Academy of Sciences 1999;96(26):14789-14794.
21. Doublie S. Preparation of Selenomethionyl Proteins for Phase Determination. Methods in Enzymology 1997;276:523-529.
22. Otwinowski Z, Minor W, editors. Processing of X-ray Diffraction Data Collected in Oscillation Mode. Volume 276. New York: Academic Press; 1997. 307-326 p.
23. Terwilliger TC, Berendzen J. Automated MAD and MIR Structure Solution. Acta Crystallographica D 1999;55:849-861.
24. La Fortelle E, Bricogne G. Maximum-likelihood Heavy-atom Parameter Refinement: A Successful Combination of Tools for Macromolecular Structure Determination. In: Sweet RM, Carter Jr. CW, editors. Methods in Enzymology, Macromolecular Crystallography. Volume 276. New York: Academic Press; 1997. p 472-494.

25. Terwilliger TC. Maximum Likelihood Density Modification. *Acta Crystallographica D* 2000;56:965-972.
26. Terwilliger TC. Automated Main-Chain Model-Building by Template Matching and Iterative Fragment Extension. *Acta Crystallographica D* 2002;59:33-44.
27. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in These Models. *Acta Crystallographica* 1991;A47(110-119).
28. Brunger AT, Adams PD. Crystallography and NMR System: a New Software Suite for Macromolecular Structure Determination. *Acta Crystallographica D* 1998;54:905-921.
29. Laskowski RA, MacArthur MW. PROCHECK: a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography* 1993;26:283-291.
30. Kraulis EA. MOLSCRIPT: a Program to Produce both Detailed and Schematic Plots of Protein Structures. *Journal of Applied Crystallography* 1991;D50:869-873.
31. Merritt EA, Bacon DJ. Raster3D Photorealistic Molecular Graphics. *Methods in Enzymology* 1997;277:505-524.
32. DeLano WL. The PyMOL Molecular Graphics System. <http://www.pymol.org> 2002.
33. Holm L, Sander C. Protein Structure Comparison by Alignment of Distance Matrices. *Journal of Molecular Biology* 1993;233:123-138.
34. Lahiri SD, Zhang G, Dunaway-Mariano D, Allen KN. The Pentavalent Phosphorus Intermediate of a Phosphoryl Transfer Reaction. *Science* 2003;299(5615):2067-2071.
35. Blackburn GM, Williams NH, Gamblin SJ, Smerdon SJ. Comment on "The Pentacovalent Phosphorous Intermediate of a Phosphoryl Transfer Reaction". *Science* 2003;301:1184c.
36. Allen KN, Dunaway-Mariano D. Response to Comment on "The Pentacovalent Phosphorus Intermediate of a Phosphoryl Transfer Reaction". *Science* 2003;301:1184d.
37. Altschul SF, Madden TL, Schaffer AA, Zhang AA, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a New Generation of Protein Database Search Programs. *Nucleic Acids Research* 1997;25:3389-3402.
38. Empadinhas N, Marugg JD, Borges N, Santos H, da Costa MS. Pathway for the Synthesis of Mannosylglycerate in the Hyperthermophilic Archaeon *Pyrococcus horikoshii*. *Journal of Biological Chemistry* 2001;276(47):43580-43588.
39. Lunn JE, Ashton AR, Hatch MD, Heldt HW. Purification, Molecular Cloning, and Sequence Analysis of Sucrose-6F-Phosphate Phosphohydrolase from Plants. *Proceedings of the National Academy of Sciences* 2000;97(23):12914-12919.

40. McCue LA, Thompson W, Carmack CS, Lawrence CE. Factors Influencing the Identification of Transcription Factor Binding Sites by Cross-Species Comparison. *Genome Research* 2002;12:1523-1532.
41. Lyngstadaas A, Lobner-Olesen A, Grelland E, Boye E. The Gene for 2-Phosphoglycolate Phosphatase (gph) in *Escherichia coli* is Located in the Same Operon as Dam and at Least Five Other Diverse Genes. *Biochimica Biophysica Acta* 1999;1472:376-384.
42. Bochner BR, Gadzinski P, Panomitros E. Phenotype MicroArrays for High-Throughput Phenotypic Testing and Assay of Gene Function. *Genome Research* 2001;11:1246-1255.
43. Rhodius V, Van Dyk TK, Gross C, LaRossa RA. Impact of Genomic Technologies on Studies of Bacterial Gene Expression. *Annual Reviews of Microbiology* 2002;56(1):599-624.
44. Richmond CS, Glasner JD, Mau R, Jin H, Blattner FR. Genome-wide Expression Profiling in *Escherichia coli* K-12. *Nucleic Acid Research* 1999;27(19):3821-3835.
45. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemakes DD, Gachotte D, Chakraborty K, Simon J, Bard M, Friend SH. Functional Discovery via a Compendium of Expression Profiles. *Cell* 2000;102:109-126.
46. Gonzalez R, Tao H, Shanmugam KT, York SW, Ingram LO. Global Gene Expression Differences Associated with Changes in Glycolytic Flux and Growth Rate in *Escherichia coli* During the Fermentation of Glucose and Xylose. *Biotechnology Progress* 2002;18:6-20.
47. Allen TE, Herrgard MJ, Liu M, Qiu Y, Glasner JD, Blattner FR, Palsson BO. Genome-Scale Analysis of the Uses of the *Escherichia coli* Genome: Model-driven Analysis of Heterogeneous Data Sets. *Journal of Bacteriology* 2003;185(21):6392-6399.
48. Glasner JD, Liss P, Plunkett G, Darling A, Prasad T, Rusch M, Byrnes A, Gilson M, Biehl B, Blattner FR. ASAP, a Systematic Annotation Package for Community Analysis of Genomes. *Nucleic Acid Research* 2003;31(1):147-151.
49. Robison K, McGuire AM, Church GM. A Comprehensive Library of DNA-binding Site Matrices for 55 Proteins Applied to the Complete *Escherichia coli* K12 Genome. *Journal of Molecular Biology* 1998;284:241-254.
50. Beitz E. TeXshade: Shading and Labelling of Multiple Sequence Alignments Using Latex2_ε *Bioinformatics* 2000;16(2):135-139.

Figure legends

Figure 1A. Ribbon diagram of the structure of YbiV. The hydrolase domain is colored in grey, the inserted domain in blue. The nucleophilic aspartate is represented in green, the magnesium ion in magenta.

Figure 1B. Secondary structure topology of YbiV colored similarly to Fig. 1A.

Figure 1C. Overlay of the hydrolase domains of YbiV (grey), PSP (pink) and YrbI (blue.) The full structures of the PSP and YrbI monomers (top and bottom respectively) are shown to the right.

Figure 2A. Glycerol binding in the inserted domain. The glycerol is represented in yellow, the protein in green. The HAD motifs are labelled. Hydrogen bonds are represented by white dashed lines with the respective distances labelled above (in Angstroms). Residues with an * are conserved among the closest homologs to YbiV.

Figure 2B. Differences in conformation of the inserted domain in the native glycerol bound YbiV (left), and the non-glycerol bound BeF₃- YbiV (right.) In the absence of glycerol, Phe 180 shifts over to stack against His 129, closing off the binding pocket. The glycerol is shown in grey and red, the magnesium ion, magenta, and the berylliofluoride (right) in purple.

Figure 3A. Stereoview of the native active-site of YbiV. The magnesium ion is represented by a magenta sphere, water molecules as blue spheres. Hydrogen bonds are represented as dashed lines and numbers indicate distances in Angstroms. The active-site motifs are labelled.

Figure 3B. Overlay of the active-sites of YrbI from *H. influenzae* (green) and YbiV (blue.) The blue sphere represents magnesium, the green sphere, cobalt. The residue numbers are labeled, YbiV on the left, YrbI on the right.

Figure 4A. 2Fo-Fc electron density map of the BeF₃- bound active-site of YbiV. The magnesium is colored in magenta, water in red, and the beryllofluoride in yellow.

Figure 4B. Stick diagram of the active-site of beryllio-fluoride bound YbiV. The beryllium and fluorines are represented in yellow, the active-site motifs in green and Ser 178, from the inserted domain, in grey. Nitrogens and oxygens are blue and red, respectively. Dotted lines represent hydrogen bonds, with the corresponding distances labelled in Angstroms. Water molecules are represented by blue spheres, the magnesium ion by a magenta sphere.

Figure 4C. Stick diagram of the aluminum fluoride bound active-site of YbiV. The aluminum fluoride is represented in yellow. Colored similarly to figure 4B.

Figure 5. Sequence alignment of YbiV and its homologs, YbjI from *E. coli*, ECYBJI-[P75809](#), YbiV and YbjI from *Salmonella typhimurium* (STYBIV-[CAD05288](#) and STYBJI-[AAL19778](#)),

and proteins from *Lactobacillus gasseri* (LACG-[ZP_00046022](#)), *Listeria monocytogenes* (LIST-[NP_463949](#)), cof ([NP_414980](#)) and YbhA ([NP_415287](#)) from *E. coli* K12, sucrose phosphatase from maize [38] (ZMSPP) and trehalose phosphatase (otsB, [NP_416411](#)) from *E. coli*. 179 residues are omitted from ZMSPP for clarity. Fully conserved residues are white with black background, conserved are white with dark grey background, and alike are black with light gray background. The secondary structure for the hydrolase domain (black) and inserted domain (grey) is indicated over the alignment. Triangles note the residues involved in glycerol binding and/or line the inserted domain cavity. Arrows note the residues involved in the stacking interaction. Figure made with Texshade.⁵⁰

Figure 6. Ribose-5-phosphate (grey), the best performing substrate, modeled into the inserted domain of YbiV. The magnesium ion is represented by a magenta sphere.

Table I						
Data Collection Statistics						
Dataset	Native	BeF ₃ ⁻	AlF ₃	□1 (MAD)	□2	□3
Wavelength (Å)	1.000	1.000	1.000	.97872	.97890	.96373
Space Group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁		
unit cell (Å) a	71.82	72.28	71.76	72.05		
b	91.60	91.19	91.17	91.76		
c	186.56	176.41	183.85	187.02		
Resolution (Å)	(20-1.9)	(50-2.0)	(50-2.2)	(20- 2.8)	(20- 2.8)	(20- 2.8)
Completeness	98.6(98.3)	100(100)	99.7(100)	98(92.8)	98(94)	97(93)
I/□	23.5(4.5)	35(13)	31(5)	18(4.7)	19(6.3)	16(3.6)
Rsym ^a (%)	7.6(22.4)	11.2(10.8)	8 (25)	12.4(55.6)	10.6(38.5)	12.4(65.9)
Unique Refl.	96340	78528	61606	30807	30668	30741
^a Rsym = $\frac{\sum I_i - \langle I_i \rangle }{\sum I_i}$						

Table II				
Phasing and Refinement Statistics				
Phasing				
Resolution Range	Peak	Edge	Remote	All
20-2.8 Å				
<i>R</i> _{cullis} (iso) ^a		0.633	0.738	
<i>R</i> _{cullis} (ano) ^b	0.612	0.682	0.739	
FOM (acen/cen)				0.67/0.43
^a <i>R</i> _{cullis} (iso) = $\frac{\sum \Delta_{iso}(obs) - \Delta_{iso}(calc) }{\sum \Delta_{iso}(obs)}$ for centric reflections ^b <i>R</i> _{cullis} (ano) = $\frac{\sum \Delta_{ano}(obs) - \Delta_{ano}(calc) }{\sum \Delta_{ano}(obs)}$ for acentric reflections $\Delta_{ano}(obs)$ is the anomalous difference				
Refinement				
Dataset	Native	BeF₃⁻	AlF₃	
Resolution Range (Å)	(20-1.9)	(20-2.0)	(20-2.2)	
No. of Reflections used				
working	85387	69583	53778	
test	9467	7797	6059	
R-value (%) ^a	21.9	21.9	22.1	
Free R-value (%)	24.8	25.1	26.4	
Completeness (%) (highest shell)	97.1(94.9)	98.6(97.2)	96.5(90)	
Sigma cutoff	0.0	0.0	0.0	
RMS deviations from ideal geometry				
bond length	0.006	0.008	0.009	
angle (°)	1.6	1.5	1.8	
dihedral (°)	23.1	23.1	23.2	
improper (°)	0.99	1.06	1.22	
Temperature Factors (Å ²)				
Protein atoms	29.2	25.3	40.9	
Solvent	46.9	49.8	42.2	
No. of atoms				
protein	8536	8532*	8516	
water	577	454	396	
Mg ²⁺	4	4	4	
other	36	12	32	
Ramachandran plot stats (%)				
Most Favored	88.2	88.3	88.4	
Other Allowed	11.8	11.7	11.6	
^a R-value = $\frac{\sum F(obs) - F(calc) }{\sum F(obs)}$				
* (includes BeF ₃ ⁻ atoms as part of new residue)				

Table III		
Phosphatase Activity of YbiV		
substrate	V/[E](s ⁻¹)	% ribose-5-p
ribose-5-p	2.4	100.0
glucose-6-p	1.5	62.5
glycerol-1-p	0.59	24.6
2-deoxy-glucose-6-p	0.54	22.5
mannose-6-p	0.48	20.0
sorbitol-6-p	0.47	19.6
glycerol-2-p	0.47	19.6
fructose-6-p	0.37	15.4
p-tyr	0.033	1.4
6-p-gluconate	0.031	1.3
sucrose-6-p	0.0097	0.4
p-ethanolamine	0.0069	0.3
glucose-1-p	0.0057	0.2
ATP	0.0013	0.1
p-ser	0.0004	0.0
p-thr	0.0012	0.1
p-creatine	0	0.0
Kinetic Constants:		
	k _{cat} (s ⁻¹)	K _m (mM)
glycerol-1-p	0.9	6
glycerol-2-p	0.5	4.5
ribose-5-p	4	6

Table 3 Activity of ybiV towards a variety of potential substrates. The first column, V/[E], is velocity ($\mu\text{MP}_i/\text{s}$) divided by enzyme concentration. The second is the relative phosphatase rate vs. the best substrate, ribose-5-p. Kinetic constants for ribose-5-p and glycerol-1-p and glycerol-2-p are shown below the trial substrates.

Figure 1A

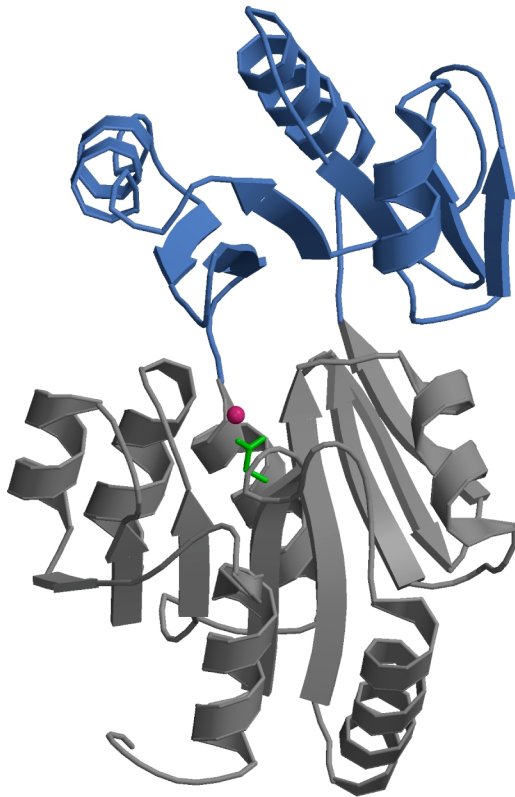


Figure 1B

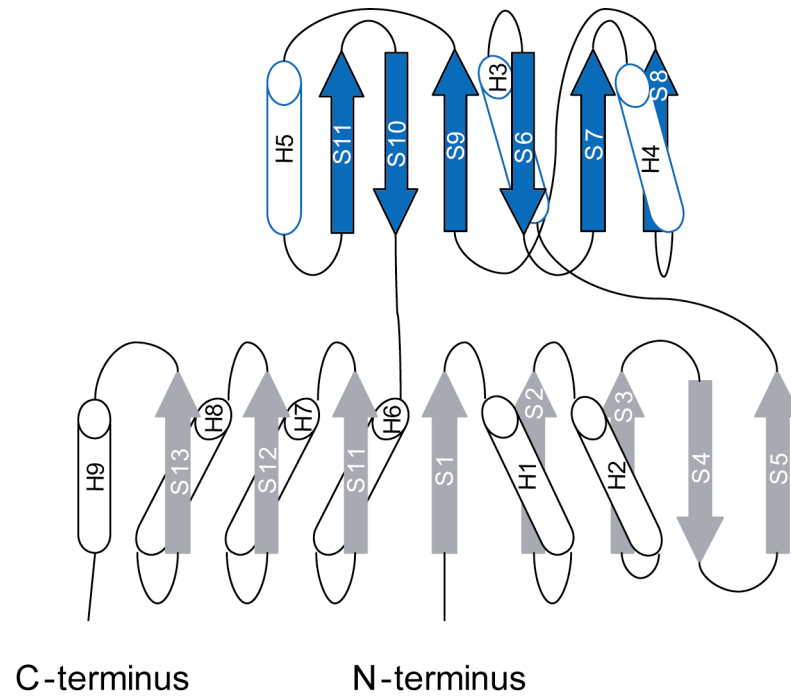


Figure 1C

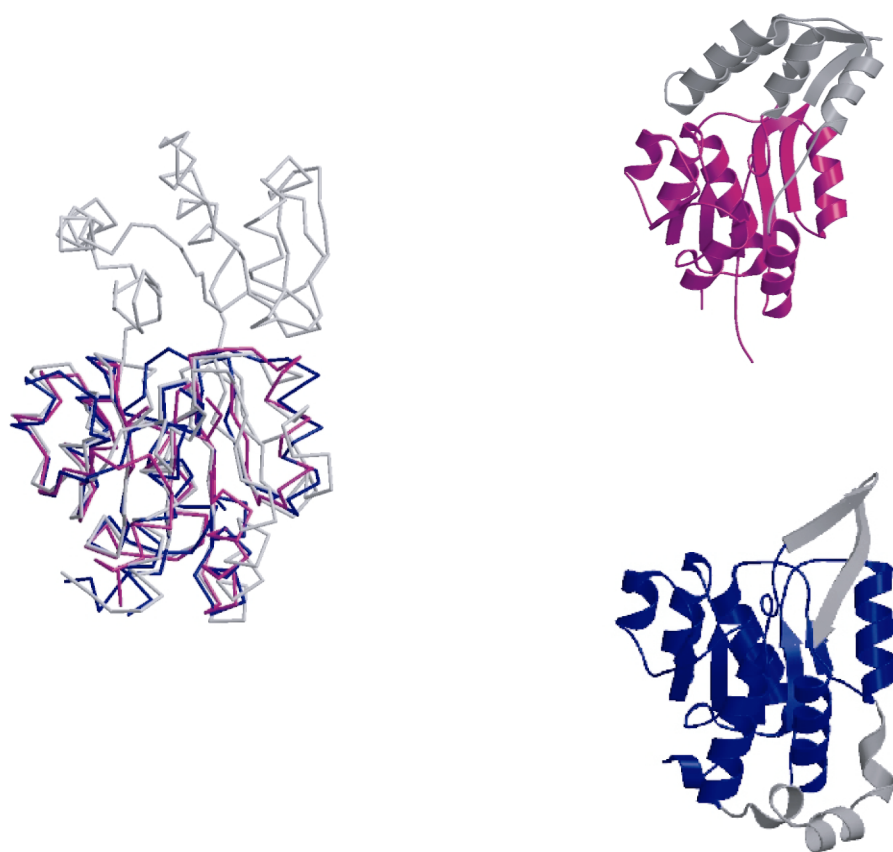


Figure 1D

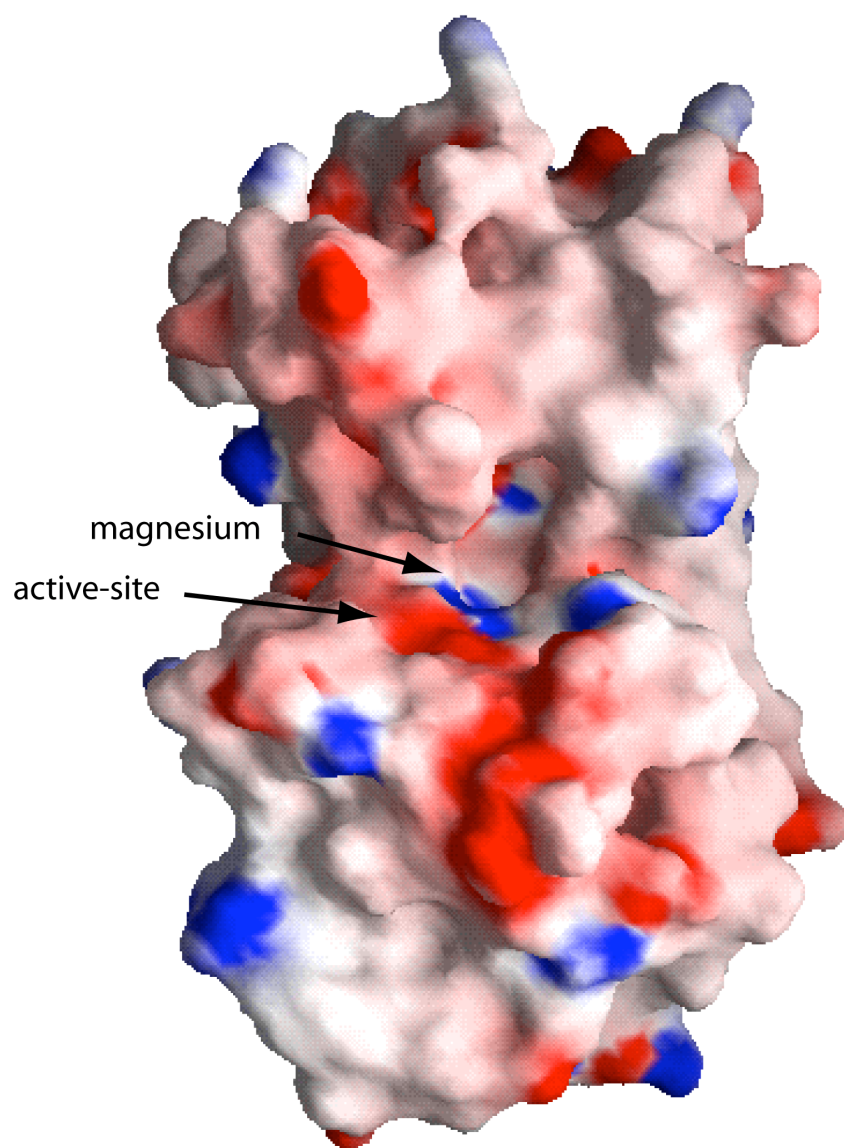


Figure 2A

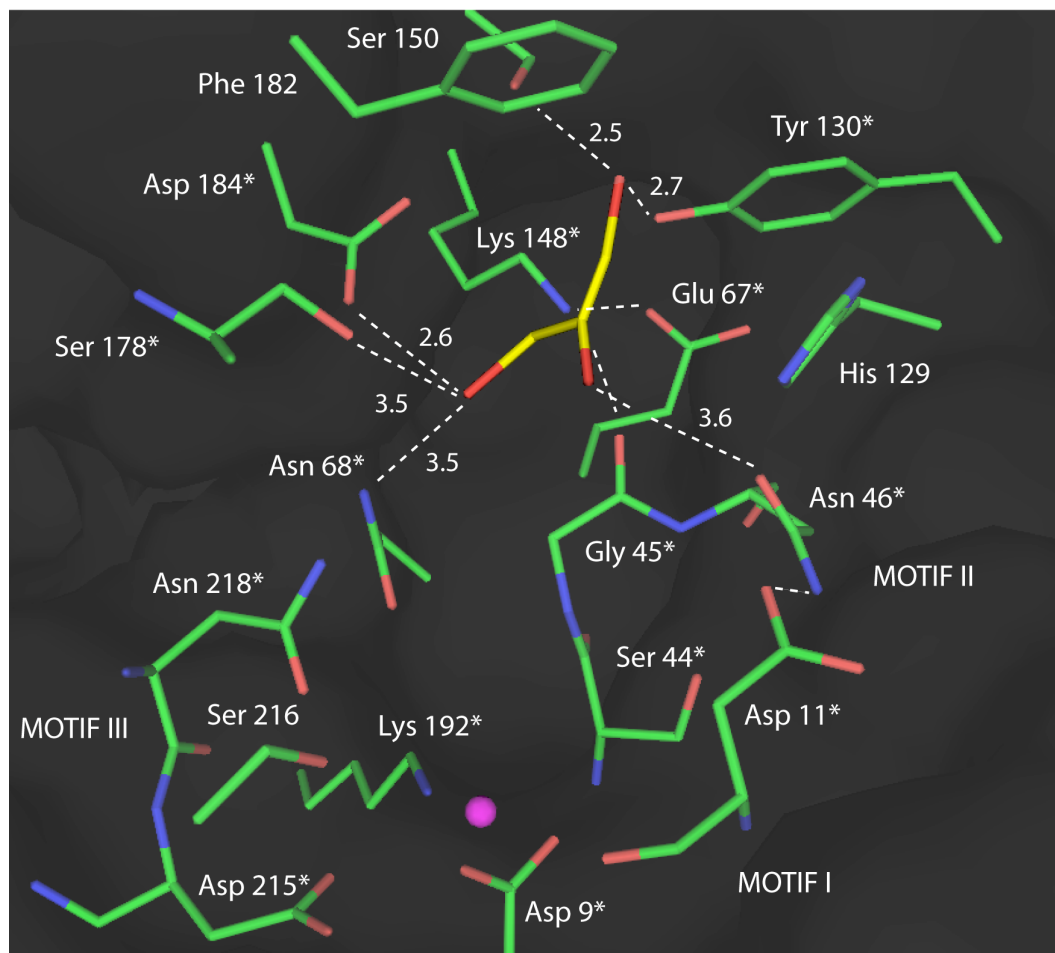


Figure 2B

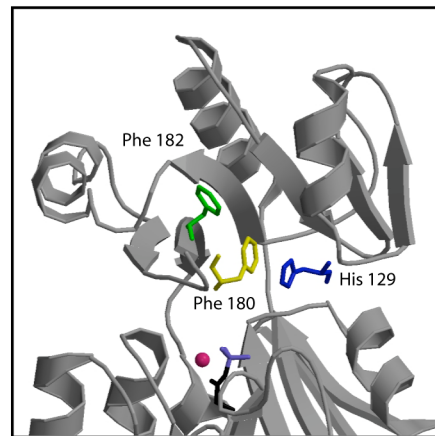
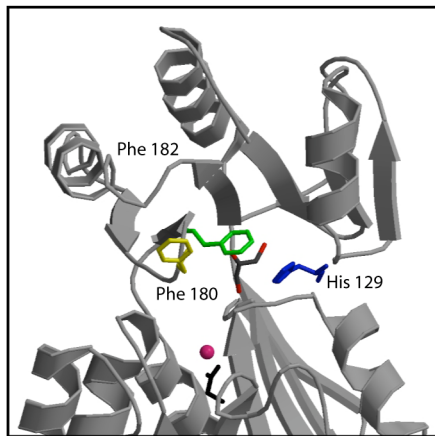


Figure 3A

Figure 3B

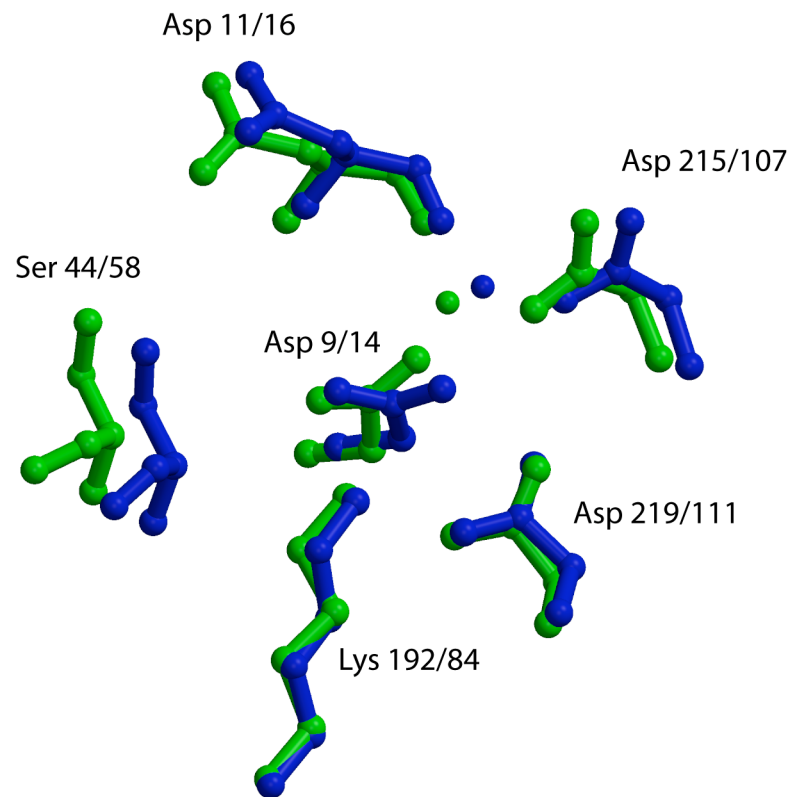


Figure 4A

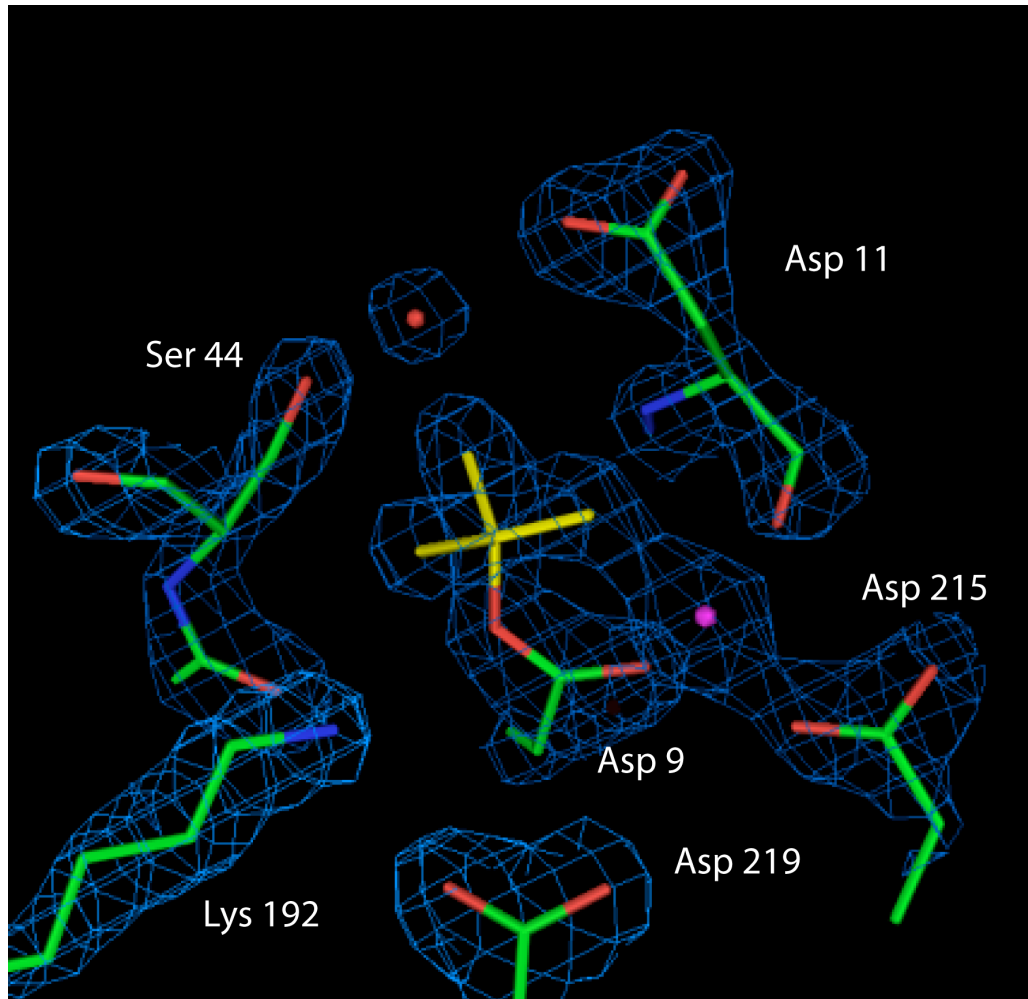


Figure 4B

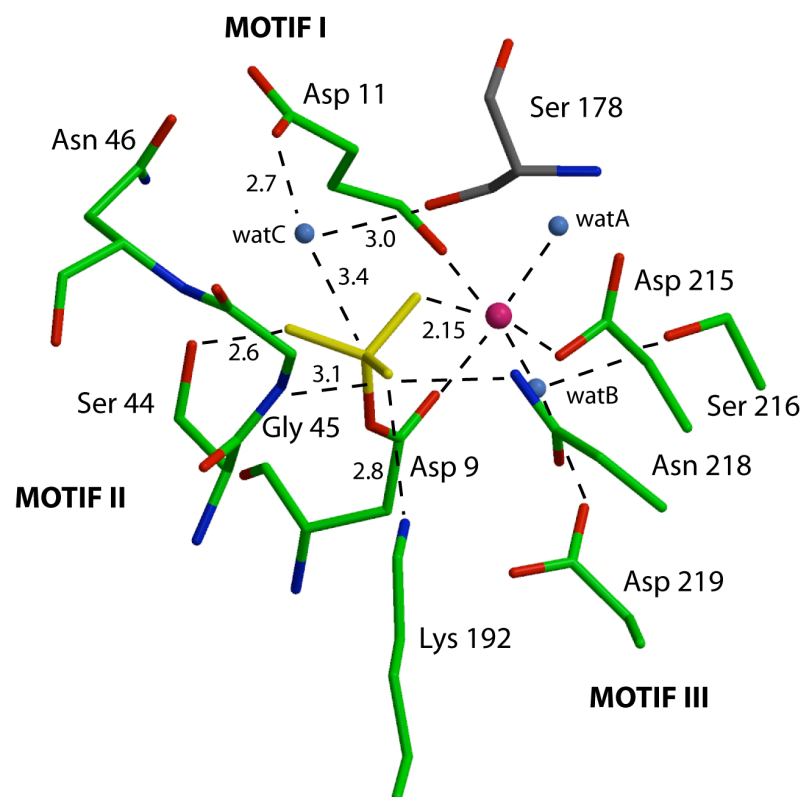


Figure 4C

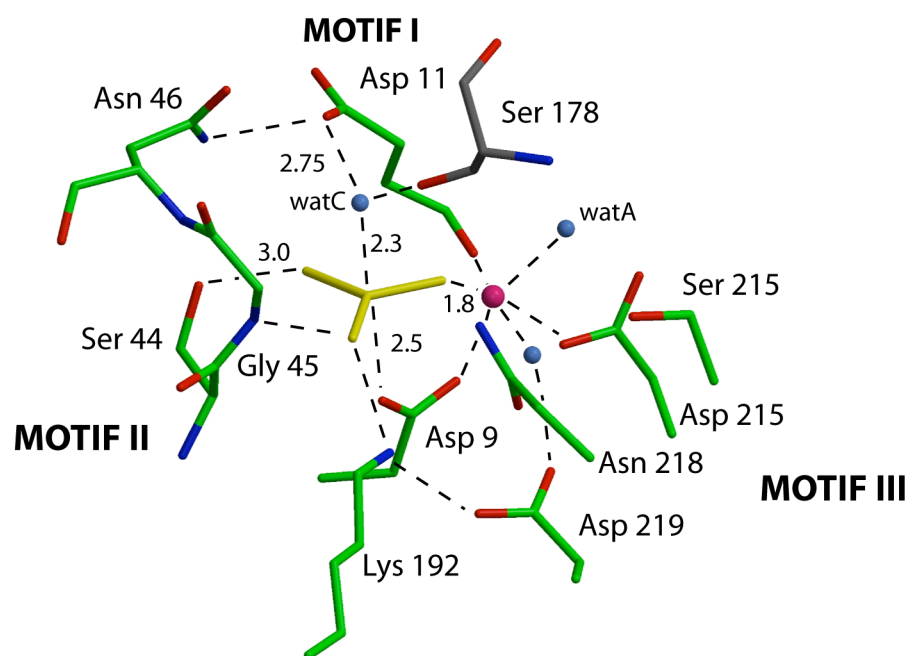


Figure 5

		Motif I		Motif II	
ECYBIV	MSVKVITVDMMDGTFLE..DAKTYNQPRFMA.QYQELKKRGTK.....FVVASGN	***		*Δ	46
STYBIV	MTVNVVVTDMDGTFELD..DAKQYDRVRFMA.QYQELKKRMIE.....FVVASGN				46
ECYBJI	MSIKLIAVDMMDGTFELS..DQKTYNRDRFMA.QYQMKKRGIR.....FVVASGN				46
STYBJI	MSIKLIAVDMMDGTFELS..DQKTYNRDRFMA.QYQMKKRGIR.....FVVASGN				46
LACG	MTKLIADVMDGTFELR..DDKSYDEEKGAN.IYQELEKRMIM.....FTVASGN				45
LIST	MIRASVDMMDGTFELD..GNGEYDRARFEK.IYAEIMKRGTK.....FIVASGN				45
COF	MARLAFDMMDGTFELM..PDHHLGEKTLS..TLARIREDDIT.....LTFATGR				44
YBHA	MTTRVIALDMDGTLIT..PKTLLPSSTIEA..LARAREAGYQ.....LIIVTGR				45
ZMSPP	MDKLSGSARLMIVSDLDHIMVDHHEENLSLLR.....FGALWESVYCEDSLVVFSTGR				54
QTSB	MTEPLTETPELSAKYAWFFDLDTLAEIKPHPDQVVVPDNLGLQLLATASDG...ALALISGR				62
ECYBIV	QYYQIISFFPEL.KDEISFVAENGALVYEHCKQLFHGETT..RHESRIVIGELIK.DKQLNMFVAC	ΔΔ			107
STYBIV	QYYQIISFFPEL.KDEISFVAENGALVYEHCKQLFHGETT..RHESRIVIGELIK.DKQLNMFVAC				107
ECYBJI	QYYQIISFFPEI.ANEIAFVAENGWVVEGKQVFNGETS..KDAFATVVEHLIT.RPEVEITAC				107
STYBJI	QYYQIISFFPEI.ANEIAFVAENGWVVEGKQVFNGETT..KADFQAVVEHLIT.RADVEITAC				107
LACG	QYYQITTFEKNF.P.TVIYVAENGALVRTQKILALHAFS..EESVKKIEDFLIK.QAELQFLVS				105
LIST	QYYQIKSFFFGK.DEELFYVAENGAVIFHQELRSVNRFD..ERLVQKILWTLIQDYHDLQVILQ				107
COF	HALEMQHILGALSLEYLITGNGTRVHSLCELLHRDDTP..ADVAEILYQQWDTRASMHFND				107
YBHA	HHVAIHPEYQALALDTPATCCNGTYLYDYHAKTVLEADPMP..VIKALQLIEMEN.EHHINGLMY				107
ZMSPP	SPTLYKELR.....KQKPMILTPDITIMSVGTEITYGEAMVP.DDGWEEYLNKWDNR...NIVVA				109
QTSB	SMVEIDALAK..PYRFPLAGVHGAERRDINGKTHIVH....LPDAIARDISVQHT.....				112
ECYBIV	GLQSAIVSENAPEAFVLMAKHYHR.....LKPVKDYQEI.DDVLFKFSNLPLDE.QIPLVIDK	ΔΔ			164
STYBIV	GLKSAIVSKMAPETFVLMAKHYHR.....LQPVNDYHDI.DDILFKFSNLPLDE.QIPLVIDK				164
ECYBJI	GKNSAYTLKRYDDAMKTVAEMYHR.....LEFYDNFDMNLE.DIFFKFGNLSDE.LIPQVQKA				164
STYBJI	GKNSAYTLKRYDDALKVVAAMYHR.....LEFYDNFDMNI.MDVFFKFGNMITDE.RIPEVQAA				164
LACG	GVESAYFPVQFTDDYYEISNKKYYR.....LKKIHDFSEI.DDKILKFSISCPDE.KTAYYVDF				162
LIST	GKNSAYLLKAADPDFKPAKKYFE.....LQEVESFDVLPDITFIKFAIDVEVA.KTGQIVED				165
COF	..DGWFTGKEIPALLQFVYSGR.....YQIIDVKKMPLGSVTKICFCGDHD.DLTRLQIQ				161
YBHA	VDDAMVYEHPTGHVIRTSNWAQTLPEQRPTFTQVSLAETAQQVMVWKFALTHD.DLPQLQHF				171
ZMSPP	ETV.....FSELKLQPET.EQRPHKVSFFVDDK.NAQEVTKS				145
QTSB	VIAQYPGAELAKGMAFADHYRQAPQHEDALMT				145
ECYBIV	LHVAIDGIMKPVTSFGF...FIDLIIPGLHKGANGISRLKRLDLSPQ...NVVAIGDSGNDAEML	ΔΔ			223
STYBIV	LHVSIDGIMKPVTSFGF...FIDLIIPGLHKGANGISRLKRLNRSPQ...NVVAIGDSGNDAEML				223
ECYBJI	LHEAIGDIMVSVHTGNG...SIDLIIPGVHKGANGRLQLKLIGIDS...EVVVFQDGGNDIEML				223
STYBJI	LHDAIGDIMVPVHTGYG...SIDLIIPGVHKGANGRLQLQRGIDS...EVVVFQDGGNDIEML				223
LACG	LKKSIGDYCNVTSSGHG...DIDLIIPGTHKAGHLELGKVLDIPLS...EMTAFQDGGNDLEML				221
LIST	LNQTFVGEIRAVSSGHG...SIDLIIPGVTKGNAIQQLNEMQVAPE...DLLAPGDAMNDIEML				224
COF	LYEATIGER...AHLCFSATDCLEVLPGVCMGAALTVMTHQLGLSLR...DCMAFGDAMNDREML				220
YBHA	GKHVEHELGLLECEWSWHDQVPIARGCNSKGRITKWVEAQGWSME...NVVAFQDGFNDISML				231
ZMSPP	VAERIDKCGLDKTIYSGGQDLILPQGAGGQALAYLEKLSSCGKPPNNTLVCGDSGNDALF				210
QTSB	LAQRITQIWPQMALQQ.GKCVVEIKPRGTSKGEAIAAFMQEAPFIGR...TPVFLGDDLTDESGF				206
ECYBIV	KMA...RYSFAMGNA.AENIKQIARYATDDN.....NHEGALNVTQAVLDNTSPFMS				271
STYBIV	KMA...HYSFAMGNA.ADNIKALSRYHTDDN.....NHQGALNVTQAVLDGTDPE				269
ECYBJI	RQAG...FSFAMENA.GSAVVAANKYRAGSM.....NREGVLDVLDKVLKHEAPFDQ				271
STYBJI	RQAG...FSFAMSHA.SEAVAANKYRAGSM.....NREGVLDIIDSVLNNEPPENV				271
LACG	KEFG...DGVAMSNA.NPVLFKVADHTTTSN.....NEQGVLTYTEENMIL				262
LIST	QLTS...NSYAMLES.SPEVLAAKHAVPSM.....KEAGVLDVLEEYMKKSQRQ				270
COF	VSV...GSGTIMGN.AMPQLRRELPHLPVIG...HCRNQAVSHYLTHWLDYPHLPYSPE				272
YBHA	EAA...GTGVAMGN.ADDAVKRAMIVIGD...N...TTDSIAQFYSHI				272
ZMSPP	SIPG...VHGVMVSN...AQEELLQWYTEM.....AKDNPK..IHS				244
QTSB	AVVNRLLGGMVKIGTG...ATQNSWRLAGVPDVWSWLEMITTAALQKRENNRSDDYESFSRSI				266

Figure 6

