Beilstein-Institut
Experimental Standard Conditions of Enzyme Characterization
September 12th – 16th, 2011, Rüdesheim/Rhein, Germany

Successes and Challenges in Functional Assignment in a Superfamily of Phosphatases

Karen N. Allen[1,*] **and Debra Dunaway-Mariano**[2,#]

[1]Department of Chemistry, Boston University, 590 Commonwealth Avenue, Boston, MA 02215 – 2521, U.S.A.

[2]Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, NM, 87131, U.S.A.

E-Mail: *drkallen@bu.edu and #dd39@unm.edu

Received: 17th September 2012/Published: 15th February 2013

Abstract

The explosion of protein sequence information from genome sequencing efforts requires that current experimental strategies for function assignment must evolve into computationally-based function prediction. This necessitates the development of new strategies based, in part, on the identification of sequence markers, including residues that support structure and specificity as well as a more informed definition of orthologues. We have undertaken the function assignment of unknown members of the haloalkanoate dehalogenase superfamily using an integrated bioinformatics/structure/mechanism approach. Notably, a number of members show "substrate blurring", with activity toward a number of substrates and significant substrate overlap between paralogues. Other family members have been honed to a specific substrate with high catalytic efficiency and proficiency. Our findings highlight the use of the cap domain structure and enzyme conformational dynamics in delineating specificity.

The Haloalkanoate Dehalogenase Superfamily (HADSF)

The "central dogma" of protein structure/function studies is that protein sequence dictates protein structure which, in turn defines protein function. Thus, evolution of any new function must be accompanied by the use of new sequence. In this way, as long as the "neutral drift" of sequences, allowing adoption of the same stable fold, can be distinguished from functionally transformative mutations, the prediction of function from sequence should be possible. In order to identify and assess sequence markers that support structure and specificity, we have undertaken the study of a large superfamily, comprised mostly of phosphoryltransferases, the haloalkanoate dehalogenase superfamily (HADSF) [1 – 3]. Because of the occurrence of the family in all domains of life and the number of homologues within each organism (27 in *E. coli* and 183 in *Homo sapiens* [4]) the members provide numerous examples of both orthologues to study determinants of specificity and paralogues to study function diversification.

The HADSF has successfully evolved several forms of chemical transformation and has experienced expansion through substrate space. Physiological substrates are varied in size and shape, ranging from phosphoproteins, nucleic acids, and phospholipids to phosphorylated disaccharides, sialic acids and terpenes to the smallest of the organophosphate metabolites, phosphoglycolate [5]. The diversity of substrates reflects the wide array of cellular roles of HADSF members including catalysis in biosynthetic pathways, reduction of antimetabolite levels, balancing nucleotide pools, elimination of toxins, and repair of damaged nucleic acids. The promiscuity and catalytic efficiency of the family matches the cellular function, with K_m values of $5 - 5000\,\mu M$ and with efficiency reflected by k_{cat}/K_m values ranging from $1 \times 10^3 - 1 \times 10^7\ M^{-1}s^{-1}$ (e.g. [6 – 9]).

Based on the work of Wolfenden [10], it is known that transfer of a phosphoryl group from a phosphate ester is demanding, such that the catalytic proficiency of the average HADSF member is 10^{17}. From this perspective, the catalytic machinery of the HADSF is well tuned to perform phosphoryl transfer. It is through the acquisition of structural appendages to the catalytic domain that the HADSF has succeeded in acting upon a vast array of substrates. The majority of the HADSF phosphatases consist of a conserved Rossmann domain (the catalytic domain) and a tethered "cap" domain that is variable in fold [3]. The catalytic scaffold is formed by four backbone segments located at the C-terminal end of the Rossmann-fold central sheet. The cap domain is inserted into either one of two loops of the catalytic scaffold (Figure 1).

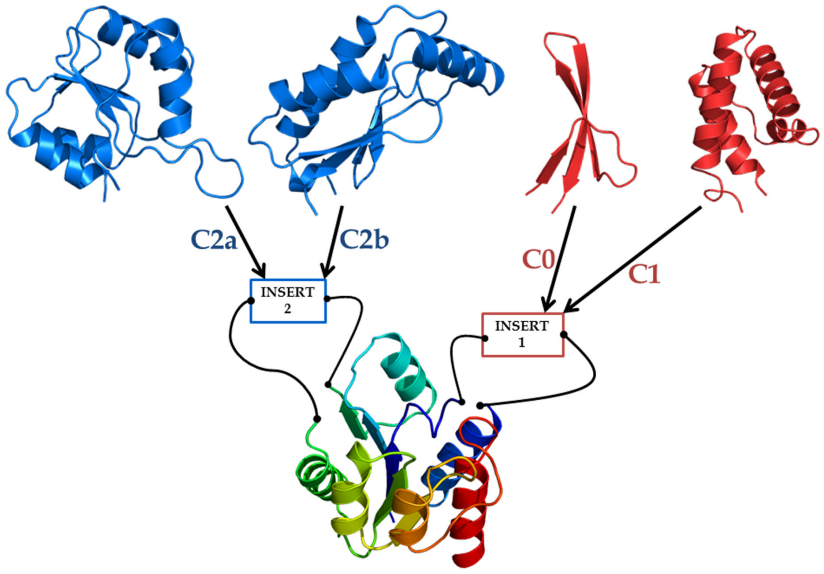

Figure 1. Cap assemblages of the HADSF can be broken down by insertion into the Rossman fold (shown as ribbon color ramped from blue to red) at site 1 of loops (type C 0) or alpha helical domains (type C 1), or at site 2 of mixed α/β domains (type C 2a and C 2b).

The catalytic scaffold binds the transferring phosphoryl group, whereas the cap domain 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. This trait is hypothesized to play a key factor in the evolution of the HADSF because the replacement of the substrate recognition residues that would switch the specificity from one physiological substrate to another would not perturb the environment of the catalytic residues, and hence their ability to stabilize the high energy transition states of the reaction pathway [4]. When the HADSF evolves new biochemical function by a switch of substrate there is no need for retooling the machinery responsible for catalysis. The beauty of this design is that one transition state fits all. Ultra-high resolution X-ray structures of a hexose phosphate phosphatase, BT4131 bound to transition-state analogues show that the catalytic domain residues in the HADSF form a "mold" in which the trigonal-bipyramidal transition states created during phosphoryl transfer are stabilized by electrostatic forces [11]. A conserved catalytic scaffold is used by which all members of the HADSF stabilize this transition-state geometry (Figure 2).

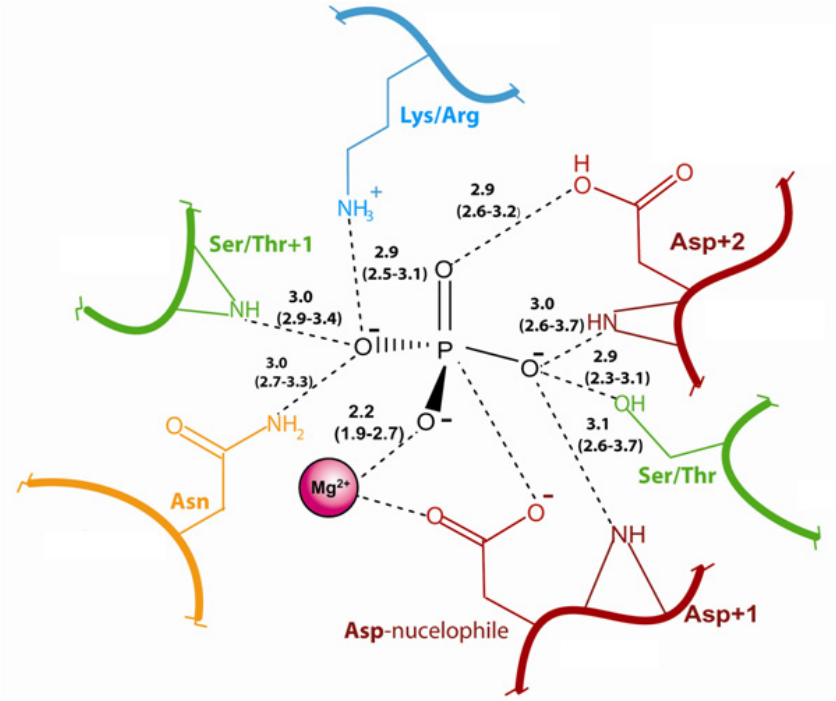

Figure 2. Composite scheme of the residues involved in forming the pentavalent trigonal–bipyramidal mold in the HADSF. Hydrogen bonds are depicted as dashed lines labeled with average bond lengths in Å (range of bond lengths in parenthesis). The notation +1 denotes the next amino acid in sequence. Loops/sequence motifs forming the active site from the Rossman fold are each colored differently (loop 1, red; loop 2, green; loop 3, blue; loop 4, red).

In concert with the residues contributed by the cap domain, dynamics between the catalytic domain and cap domain confer substrate specificity. Cap closure affords formation of the active site, which lays at the interface of the cap and the core domains. Ligand binding favours the cap closed form of the enzyme [12]. A classic example is that of β-phosphoglucomutase from *Lactoccocus lactis* (β-PGM), in which structural and kinetic analysis together have provided a model for the synchronization of cap closure with acid/base catalysis allowing the enzyme to discriminate between β-glucose 1-phosphate and water as the phosphoryl group acceptor [13]. Indeed mutation of one of the residues linking cap and core in β-PGM (Thr16) uncouples substrate binding from the positioning of the general acid/base, and changes the enzyme from a phosphomutase to a phosphohydrolase (the preference for hexose sugar over water as a phosphoryl acceptor changes from 6000:1 to 1:8 [14]).

Recognizing this intimate relationship between cap structure and enzyme function we have undertaken the function assignment of unknown members of the HADSF using an integrated approach through substrate profiling, structure analysis, computational substrate docking,

and bioinformatic analysis [15]. This work highlights the promise and challenges ahead in function assignment by uncovering cases where: 1) structure and kinetics go hand in hand to assign activity and ultimately function, 2) structure and kinetics gives few clues to function and 3) subtleties of isoforms and allosteric effectors make assignment challenging.

Structure and Function: Hand-in-Hand

The HADSF member from *Bacteroides thetaiotaomicron*, BT2127, could easily be mistaken for an orthologue of β-PGM (sequence identity 22%, similarity 42%) because of the similarity in overall fold but also from the conserved His-Lys diad and linker hinge residue Ser/Thr that is characteristic of β-PGM (Figure 3). Surprisingly, *in vitro* substrate activity screening shows that BT2127 is not a mutase nor an organophosphate hydrolase but rather it is an inorganic pyrophosphatase with a k_{cat}/K_m value for pyrophosphate of ~1×10^5 M^{-1} s^{-1} [16]. The *in vitro* kinetic data, together with the gene context, support the assignment of *in vivo* function as an inorganic pyrophosphatase. Orthologues of β-PGM occur in species with maltose phosphorylase or trehalose phosphorylase (which collaborate in the utilization of maltose or trehalose as a carbon and energy source) and BT2127 occurs in species with bifunctional L-fucose kinase/L-fucose-1-phosphate guanyltransferase which produces inorganic pyrophosphate and β-L-fucose-GDP (used in biosynthesis of surface capsular polysaccharides and glycoproteins). The X-ray structural analysis of wild-type BT2127 shows that BT2127 differs from β-PGM in that the cap does not take on the open conformation in the absence of substrate. Specifically, Glu47 from the cap domain makes a coordinate bond to the Mg^{2+} ion in the core domain, promoting the closed conformer [16]. Thus, substrate discrimination is based, in part, on active-site size restrictions imposed by the cap domain. In BT2127 the substrate range and structure/function relationship are clear, but this is not always the case in the HADSF.

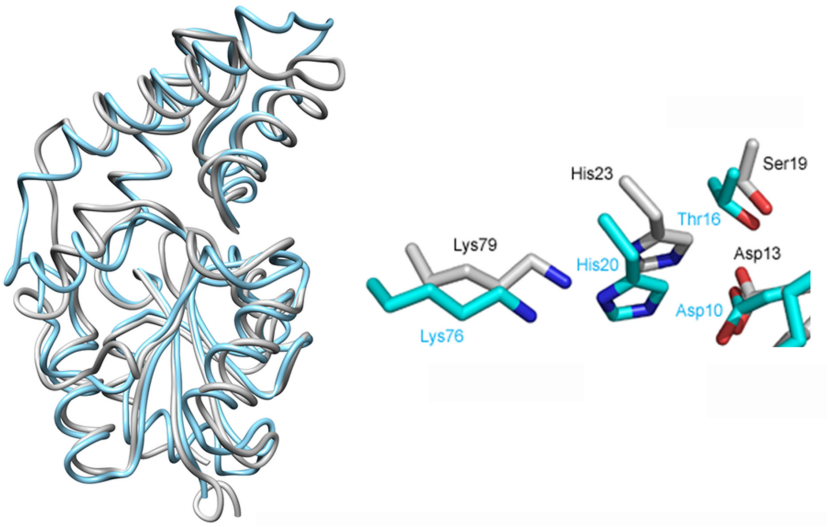

Figure 3. Superposition of BT2127 (PDB entry 3QX7) (gray) and *L. lactis* β-PGM (PDB entry 1O08) (cyan) showing the overall fold (left panel) and the conserved active-site residues (right panel).

Get a Clue: Structure and Kinetics may not Yield Clues to Function

A hexose phosphate phosphatase in the HADSF from *Bacteroides thetaiotaomicron,* BT4131, possesses broad substrate specificities and low catalytic efficiency [8]. Within *Bacteroides* a homologue, BT1666, shares 38.4% sequence identity with BT4131. We queried whether these enzymes represent duplicated versus paralogous activities. The X-ray crystal structure of BT1666 compared to that of BT4131 reveals a conserved fold and identical active sites, suggestive of a common physiological substrate [17]. However, this apparent structural similarity is not mirrored in the substrate specificity profiles of the two enzymes using a panel of common phosphorylated metabolites (Table 1). BT1666, like BT4131, is an enzyme which shows "substrate blurring" or similar activities for a number of structurally similar substrates; however, the substrate specificity profiles for the two enzymes are distinct. Moreover, BT1666 shows overall lower activity (by approximately one order of magnitude) toward most substrates. This finding is unexpected as the active sites are identical and patterns of sequence conservation extremely similar. We posit that sequence variation outside the active site causes differences in conformational dynamics or subtle positioning effects that alter catalysis and thus drive the divergence in efficiency and selectivity. *In vivo* the overlapping substrate profiles may be explained by differential regulation of expression of the two enzymes or may confer an advantage in housekeeping functions by having a larger range of possible metabolites as substrates. There exists, of

course, the possibility that the correct, specific substrate has not been uncovered for either enzyme. As seen in another example from the HADSF, subtleties of specificity can be introduced by effectors in the cellular milieu.

Table 1. Ratio of steady-state kinetic constants for BT4131 and BT1666 Catalyzed hydrolysis of phosphorylated small-molecule substrates in 50 mM HEPES containing 5 mM $MgCl_2$ (pH 7.0, 37 °C)

	BT1666/BT4131		
Substrate	k_{cat}/k_{cat}	K_m/K_m	$k_{cat}/K_m/k_{cat}/K_m$
D-glucose 6-phosphate	0.08	1.1	0.072
2-deoxy glucose 6-phosphate	0.02	3.3	0.006
glucosamine 6-phosphate	4.4	0.75	5.3
N-acetyl-glucosamine 6-phosphate	0.036	0.065	0.57
mannose 6-phosphate	0.23	1.3	0.18
fructose 6-phosphate	0.37	2	0.18
fructose 1,6-(bis)phosphate	78	0.62	125
arabinose 5-phosphate	0.027	0.43	0.063
ribose 5-phosphate	0.048	0.43	0.11
sorbitol 6-phosphate	0.078	0.96	0.081
DL-α-glycerol 3-phosphate	0.040	0.61	0.066
sucrose 6'-phosphate	8.5	0.45	19
trehalose 6-phosphate	5	–	–
ADP	28	6.3	4.5
pyridoxal 5'-phosphate	0.16	4.7	0.035
p-nitrophenyl-phosphate	0.66	0.66	1

Functional Changes Effected by Ligands

Human α-phosphomannomutase (α-PMM) catalyses the conversion of D-mannose 6-phosphate to α-D-mannose 1-phosphate which is required for GDP-mannose and dolichol-phosphate-mannose biosynthesis. These are essential constituents of N-glycosylation and glycosyl phosphatidylinositol membrane anchoring. Two isoforms PPM1 and PMM2 are found (65% sequence identity) with differential expression profiles (PMM2 is ubiquitous where PMM1 is expressed in brain and lungs) [18]. As might be supposed from their high sequence identity the two isozymes have a conserved fold and active-site structure (Figure 4) with root mean square deviations of 0.65 Å (for the cap) and 1.0 Å (for the core) when the two domains are overlaid separately [19].

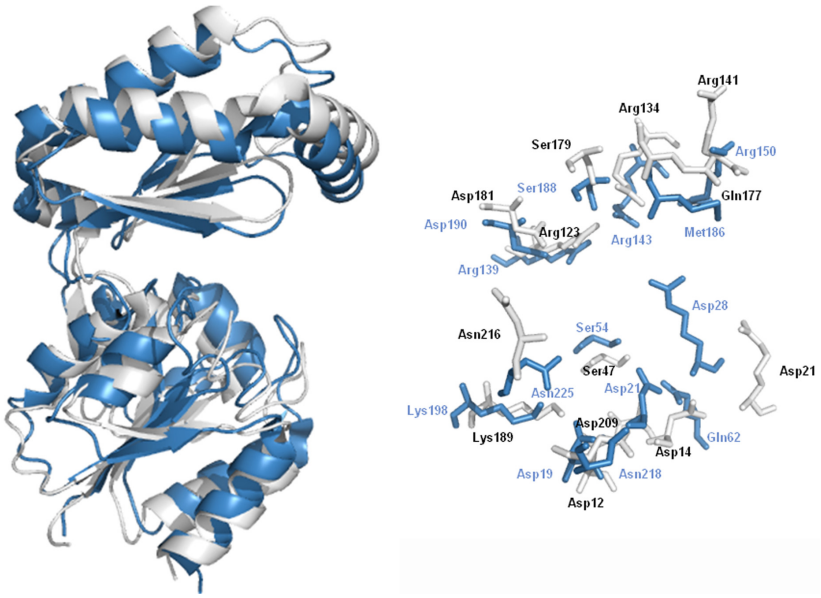

Figure 4. Superposition of PMM1 (PDB entry 2FUC) (gray) and PMM2 (PDB entry 2AMY; deposited in the Protein Data Bank by the Center for Eukaryotic Structural Genomics) (blue) showing the overall fold (left panel) and the active-site residues (right panel).

The isozymes are efficient catalysts of both substrates with high specificity as indicated by the magnitude of k_{cat}/K_m. Additionally, mutase activity requires that the bis-phosphorylated intermediates must dissociate from the active site (either within the enzyme or into bulk solvent) and then bind in the opposite orientation [20 – 22] so the bis-phosphorylated hexoses can also be considered substrates. Initially, it seems puzzling that such an essential enzyme would accept a second hexose, α-D-glucose 1-phosphate, as substrate. However, a classic observation brings to light a role for the glucose substrate. Two decades ago, a putative glucose-1,6-(bis)phosphatase brain enzyme was characterized [23 – 24] that is dependent for its activity on the presence of inosine mono-phosphate (IMP), the concentration of which increases in anoxia. Such an activity acts to restore glucose 6-phosphate levels and hence working ATP concentrations. More recently, the mystery phosphatase was identified as PMM1 [25]. Steady-state kinetics shows that IMP acts as an effector which shifts the activity of PMM1, but not PMM2, from that of a phosphomutases to a phosphohydrolase. Thus, in ischemic events, while PMM2 carries on the mutase activity to produce the α-mannose 1-phosphate essential to protein glycosylation, PMM1 switches to phosphohydrolase activity to restore glucose 6-phosphate levels for glycolysis. From a mechanistic perspective, we posit that IMP binds to the enzyme in place of the hexose phosphate, or to a secondary site to favour the conformation necessary for phosphoryl transfer, allowing water

to act as the phosphoryl acceptor. This elegant solution to a metabolic problem exemplifies the challenges of using sequence/structure and even *in vitro* kinetics to predict *in vivo* function.

Conclusions

The explosion of sequences made available by genome sequencing efforts requires that current experimental strategies for function assignment must evolve into computationally-based function prediction. However, such a development must first find its basis in the expansion of the experimental database of sequence markers, notably, residues that support catalysis and specificity. Examination of the HADSF has allowed the analysis of a wide variety of orthologues and paralogues to develop the structure/function language. Remarkably, the analysis has revealed that while many enzymes are honed to perfection in terms of substrate selectivity, others show broad ''substrate blurring'' and even overlap of substrates between paralogues within a cell. This concept extends to isozymes where differential metabolic demands seem to play a part in retaining two very similar enzymes. In order to move from function in the test tube to role in the cell, auxiliary information such as regulation of transcription, cell environment, and presence and identity of regulating metabolites may be necessary.

Acknowledgement

This work was supported by NIH U54 GM093342 to K.N.A. and D.D-M.

References

[1] Collet, J.F., van Schaftingen, E. Stroobant, V. (1998) A new family of phosphotransferases related to P-type ATPases. *Trends Biochem. Sci.* **23**(8):284. doi: http://dx.doi.org/10.1016/S0968-0004(98)01252-3.

[2] Koonin, E.V. and Tatusov, R.L. (1994) Computer Analysis of Bacterial Haloacid Dehalogenases Defines a Large Superfamily of Hydrolases with Diverse Specificity: Application of an Iterative Approach to Database Search. *J. Mol. Biol.* **244**(1):125 – 132. doi: http://dx.doi.org/10.1006/jmbi.1994.1711.

[3] Burroughs, A.M., *et al.* (2006) Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *J. Mol. Biol.* **361**(5): 1003 – 34. doi: http://dx.doi.org/10.1016/j.jmb.2006.06.049.

[4] Allen, K.N. and Dunaway-Mariano, D. (2009) Markers of fitness in a successful enzyme superfamily. *Curr. Opin. Struct. Biol.* **19**(6):658 – 65.
doi: http://dx.doi.org/10.1016/j.sbi.2009.09.008.

[5] Allen, K.N. and Dunaway-Mariano, D. (2004) Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends Biochem. Sci.* **29**(9):495 – 503.
doi: http://dx.doi.org/10.1016/j.tibs.2004.07.008.

[6] Tremblay, L.W., Dunaway-Mariano, D. Allen, K.N. (2006) 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* **45**(4):1183 – 93.
doi: http://dx.doi.org/10.1021/bi051842j.

[7] Wang, L., *et al.* (2010) Divergence of biochemical function in the HAD superfamily: D-glycero-D-manno-heptose-1,7-bisphosphate phosphatase (GmhB). *Biochemistry* **49**(6):1072 – 81.
doi: http://dx.doi.org/10.1021/bi902018y.

[8] Lu, Z., Dunaway-Mariano, D. and Allen, K.N. (2005) HAD superfamily phosphotransferase substrate diversification: structure and function analysis of HAD subclass IIB sugar phosphatase BT4131. *Biochemistry* **44**(24):8684 – 96.
doi: http://dx.doi.org/10.1021/bi050009j.

[9] Wang, L., *et al.* (2008) Human symbiont Bacteroides thetaiotaomicron synthesizes 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN). *Chem. Biol.* **15**(9):893 – 7.
doi: http://dx.doi.org/10.1016/j.chembiol.2008.08.005.

[10] Wolfenden, R. and Snider, M.J. (2001) The depth of chemical time and the power of enzymes as catalysts.*Acc. Chem. Res.* **34**(12):938 – 45.
doi: http://dx.doi.org/10.1021/ar000058i.

[11] Lu, Z., Dunaway-Mariano, D. Allen, K.N. (2008) 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.* **105**(15):5687 – 92.
doi: http://dx.doi.org/10.1073/pnas.0710800105.

[12] Zhang, G., *et al.* (2002) Kinetic Evidence for a Substrate-Induced Fit in Phosphonoacetaldehyde Hydrolase Catalysis. *Biochemistry* **41**(45):13370 – 13377.
doi: http://dx.doi.org/10.1021/bi026388n.

[13] Zhang, G., *et al.* (2005) Catalytic cycling in beta-phosphoglucomutase: a kinetic and structural analysis. *Biochemistry* **44**(27):9404 – 16.
doi: http://dx.doi.org/10.1021/bi050558p.

[14] Dai, J., *et al.* (2009) Analysis of the structural determinants underlying discrimination between substrate and solvent in beta-phosphoglucomutase catalysis. *Biochemistry* **48**(9):1984 – 95.
doi: http://dx.doi.org/10.1021/bi801653r.

[15] Gerlt, J.A., *et al.* (2001) The Enzyme Function Initiative. *Biochemistry* **50**(46):9950 – 62.
doi: http://dx.doi.org/10.1021/bi201312u.

[16] Huang, H., *et al.* (2011) Divergence of structure and function in the haloacid dehalogenase enzyme superfamily: *Bacteroides thetaiotaomicron* BT2127 is an inorganic pyrophosphatase. *Biochemistry* **50**(41):8937 – 49.
doi: http://dx.doi.org/10.1021/bi201181q.

[17] Lu, Z., Dunaway-Mariano, D. and Allen, K.N. (2011) The X-ray crystallographic structure and specificity profile of HAD superfamily phosphohydrolase BT1666: comparison of paralogous functions in *B. thetaiotaomicron. Proteins* **79**(11):3099 – 107.
doi: http://dx.doi.org/10.1002/prot.23137.

[18] Pirard, M., *et al.* (1999) Kinetic properties and tissular distribution of mammalian phosphomannomutase isozymes. *Biochem. J.* **339**(1):201 – 207.
doi: http://dx.doi.org/10.1042/0264-6021:3390201.

[19] Silvaggi, N.R., *et al.* (2006) The X-ray crystal structures of human alpha-phosphomannomutase 1 reveal the structural basis of congenital disorder of glycosylation type 1a. *J. Biol. Chem.* **281**(21):14918 – 26.
doi: http://dx.doi.org/10.1074/jbc.M601505200.

[20] Dai, J., *et al.* (2006) Conformational cycling in beta-phosphoglucomutase catalysis: reorientation of the beta-D-glucose 1,6-(Bis)phosphate intermediate. *Biochemistry* **45**(25):7818 – 24.
doi: http://dx.doi.org/10.1021/bi060136v.

[21] Rhyu, G.I., Ray, W.J., Jr. and Markley, J.L. (1984) Enzyme-bound intermediates in the conversion of glucose 1-phosphate to glucose 6-phosphate by phosphoglucomutase. Phosphorus NMR studies. *Biochemistry* **23**(2):252 – 60.
doi: http://dx.doi.org/10.1021/bi00297a013.

[22] Naught, L.E. and Tipton, P.A. (2005) Formation and reorientation of glucose 1,6-bisphosphate in the PMM/PGM reaction: transient-state kinetic studies. *Biochemistry* **44**(18):6831 – 6.
doi: http://dx.doi.org/10.1021/bi0501380.

[23] Guha, S.K. and Rose, Z. B. (1982) Brain glucose bisphosphatase requires inosine monophosphate. *J. Biol. Chem.* **257**(12):6634 – 7.

[24] Guha, S.K. and Rose, Z. B. (1983) Role of inosine 5'-phosphate in activating glucose-bisphosphatase. *Biochemistry* **22**(6):1356 – 61. doi: http://dx.doi.org/10.1021/bi00275a006.

[25] Veiga-da-Cunha, M., *et al.* (2008) Mammalian phosphomannomutase PMM1 is the brain IMP-sensitive glucose-1,6-bisphosphatase. *J. Biol. Chem.* **283**(49):33988 – 93. doi: http://dx.doi.org/10.1074/jbc.M805224200.