Beilstein-Institut

Experimental Standard Conditions of Enzyme Characterizations,
September 13th – 16th, 2009, Rüdesheim/Rhein, Germany

Evolution of New Specificities 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: 13th July 2010 / Published: 14th September 2010

Abstract

The evolution of new catalytic activities and specificities within an enzyme superfamily requires the exploration of sequence space for adaptation to a new substrate with retention of those elements required to stabilize key intermediates/transition states as well as the core enzyme fold. Phylogenetic analysis, mechanistic information, and structure determination are used to reveal novel ways in which the catalytic scaffold of a mechanistically diverse superfamily, the haloalkanoic acid dehalogenase enzyme superfamily, is tailored to new biochemical functions. Newly uncovered substrate specificities and activities in members of the superfamily are highlighted to explore the interplay of function and form. We provide evidence that core residues in this large enzyme family, form a "mold" in which the trigonal bipyramidal transition-states formed during phosphoryl transfer are stabilized by electrostatic forces.

Introduction

The Haloalkanoic Acid Dehalogenase Superfamily (HADSF) is a large superfamily of enzymes [1] with over 30,000 members to date [2]. Members are found in all three kingdoms of life, and are often observed several times within a given organism's genome (*e.g.* 29 in *E. coli* and 158 in humans) [3]. Although named for the first enzyme in the family to be

well characterized, the HADSF is primarily comprised of phosphohydrolases, with < 1 % of members catalyzing dehalogenation reactions [4]. These minority reactions at carbon centers include 2-L-haloalkanoic acid dehalogenase (C-Cl cleavage) [5] and azetidine hydrolase (C-N cleavage) [6] and the more prevalent phosphoryl transfer reactions include the phosphonoacetaldehyde hydrolase (C-P cleavage) [7 – 8], Ca^{2+}-ATPase (PO-P cleavage) [1], phosphoserine phosphatase and phosphomannomutase (CO-P) [9 – 10] (Figure 1). Within this panoply of reactions, HADSF members act on a wide variety of substrates varying in steric bulk, electrostatics, and polarity. Here, we address how the phosphoryl hydrolases in this enzyme family have adapted to work on varying substrates while retaining high catalytic efficiency.

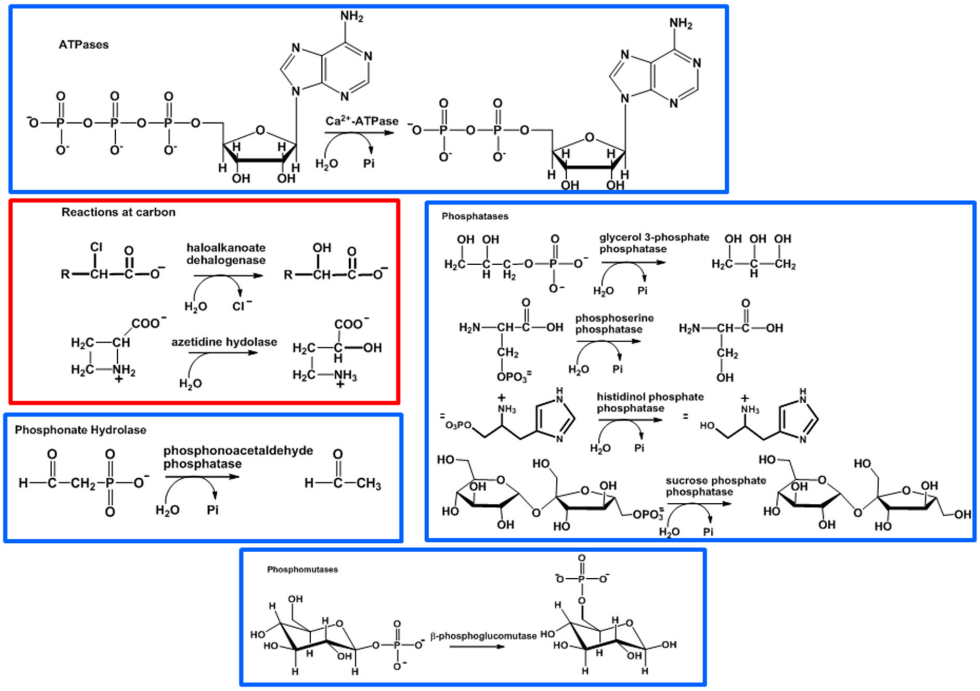

Figure 1. Representative reactions catalyzed by the HADSF

Conserved Chemical and Catalytic Mechanism

Examination of the sequence similarity in the superfamily uncovers high diversity in primary structure, with only 10 – 15 % sequence identity between homologues and closer to 30 – 40 % identity within the strictly conserved catalytic motifs used to identify HADSF members [11]. These four motifs are positioned on loops at the C-terminal end of the parallel beta strands of the core Rossmann fold common to all superfamily members (Figure 2). Within motif 1 resides both the catalytic Asp nucleophile [9] and the general acid/base Asp catalyst

which are integral to the mechanism of phosphoryl transfer [12 – 13]. The chemical mechanism of phosphoryl transfer itself is retained throughout the superfamily, and there is evidence that the catalytic mechanism is also retained.

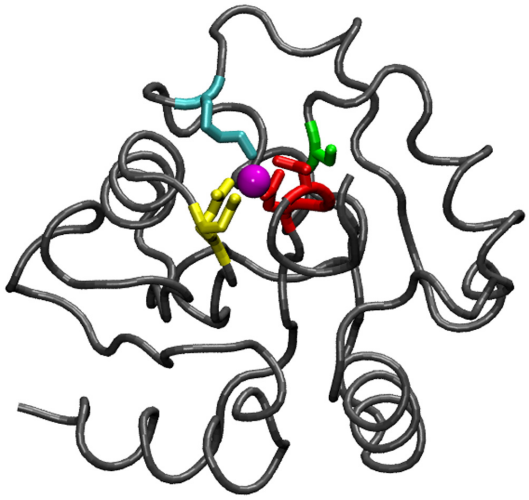

Figure 2. The conserved catalytic motifs of the HADSF phosphotransferases mapped onto the Rossmann fold catalytic scaffold (PDB 1LVH with cap domain removed for view into active site). Motif 1 (DXD) is depicted in red, motif 2 (S/T) in green, motif 3 (K/R) in cyan and motif 4 (DD, GDxxxD, or GDxxxxD) in yellow. The Mg^{2+} cofactor is depicted as a magenta sphere.

The commonality of the catalytic mechanism is supported by ultra-high resolution X-ray structures of transition-state analogues liganded to the enzyme hexose phosphate phosphatase (HPP) BT4131 from *Bacteroides thetaiotaomicron* VPI-5482 [13]. The complex of vanadate with HPP determined at 1.00 Å resolution assumes a trigonal bipyramidal coordination geometry with the nucleophilic Asp8 and one oxygen ligand at the apical position (Figure 3). Notably, the tungstate complex (1.03 Å resolution) assumes the same coordination geometry. The general acid/base residue Asp10 is critical to the stabilization of the trigonal bipyramidal species as evidenced by the collapse of the trigonal bipyramidal geometry in complexes of the Asp10Ala mutant complexed with vanadate (1.52 Å resolution) as well as tungstate (1.07 Å resolution). The ease of attaining the trigonal bipyramidal geometry for a given complex parallels the inhibition constants with K_i values for vanadate, tungstate and phosphate equal to 510 nM, 65 μM, and 5.2 mM, respectively. The core Rossmann fold stabilizes this trigonal bipyramidal transition state by engaging in favorable electrostatic interactions with the axial and equatorial atoms of the transferring phosphoryl group. Conserved backbone and side-chain interactions contributed by this scaffold were uncovered by a structural analysis of twelve liganded HADSF structures deposited in the protein data bank. Overall, these findings support the model that core domain residues in the

HADSF form a "mold" in which the trigonal bipyramidal transition states formed during phosphoryl transfer are stabilized by electrostatic forces, retaining a single catalytic mechanism for phosphoryl transfer.

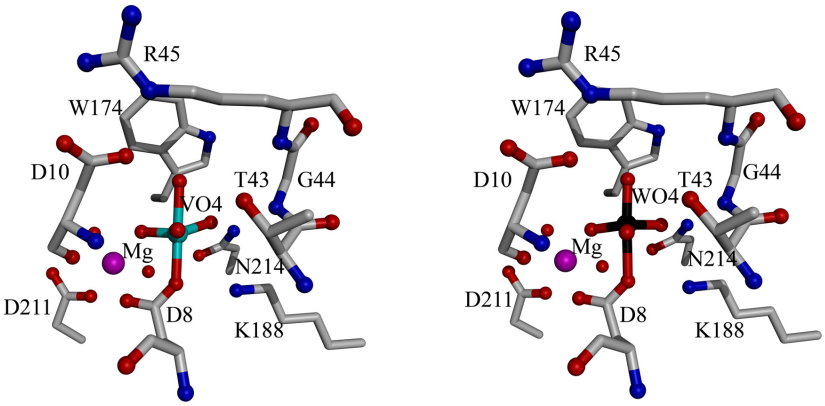

Figure 3. The trigonal bipyramidal complexes of wild type hexose phosphate phosphatase BT4131 from *Bacteroides thetaiotaomicron* VPI-5482 with vanadate (left) and tungstate (right) depicted as ball and stick and Mg^{2+} cofactor (magenta sphere).

The Cap Confers Substrate Specificity

If the phosphoryl transfer mechanism is common among family members, how then is specificity conferred? The positioning of the four catalytic motifs provides an open active site that can be accessed by macromolecules (in enzymes such as magnesium dependent phosphatase 1 [14], T4 polynucleotide kinase/phosphatase [15], and RNA polymerase II C-terminal domain phosphatases [16], defined as C0 members) or adapted to small molecules using a cap domain. Thus the HADSF has a "modular" design, with the phosphoryl binding site conferred by the core domain being spatially distinct from the leaving-group binding site conferred by the cap domain. By far the most common cap assemblage (> 65%, Chetanya Pandya, unpublished results) in the HADSF is the C1 type cap wherein an alpha helical domain, which can vary in size from two helices to more than eight helices, is inserted between motifs 1 and 2. Analysis of these C1 members shows that the residues interacting with the leaving group originate from a single substrate specificity loop (Figure 4) [17]. The C2 cap is the second most populated type of cap assemblage, which is further divided into C2a and C2b subfamilies as defined by the topology of the α/β fold prototypical of these members. Depending on the size of the cap domain, there are up to two substrate specificity loops in C2 HADSF members [18 – 19].

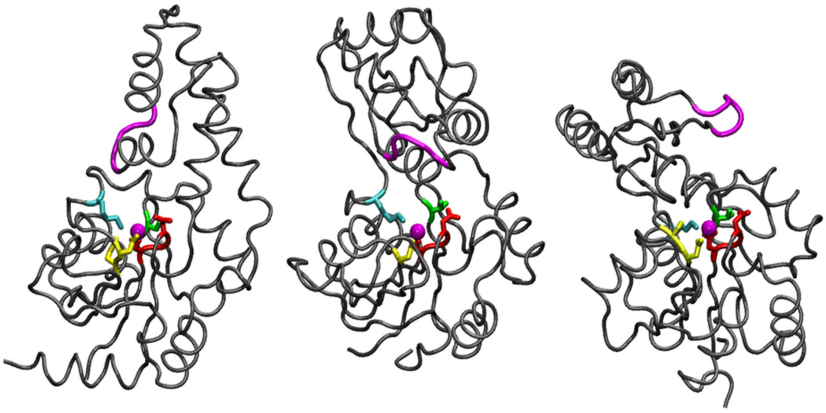

Figure 4. The substrate specificity loops (magenta) of type C1 (PDB 1LVH, left), C2a (PDB 2C4N, center) and C2b (PDB 1TJ5, right) HADSF cap assemblages. The catalytic motifs are colored as in Figure 2.

In addition to affording specificity residues in a single module, the dynamic movement of the cap domain relative to the core domain allows for ligand-induced cap closure [20] and, while giving access to the active site in the cap-open form, provides a solvent exclusive environment in the cap-closed form. In the cap closed form, the enzyme encapsulates the substrate such that the cavity remaining approximates the size and shape of the substrate. This feature has in fact been utilized to predict substrates in HADSF enzymes of unknown function [18]. The nature of cap closure differs between C1 and C2 caps, though in both cases the cap and core move as rigid bodies on two flexible linkers. The C1 caps move on a mechanical hinge (similar to the opening and closing of a clam) while the C2 caps close with a screw-like motion of the cap over the core. Thus, the cap domain is seemingly integral to specificity and chemical environment of the phosphoryl transfer reactions of HADSF members acting on small substrates.

Predicting Specificity from Sequence

If this is the case, one might envision that examination of sequence alone might be used to predict if a cap exists and thus whether a given member acts on a small molecule substrate (C1 or C2 member) or on a phosphorylated protein (a capless, C0 member). Unfortunately, such a model is an oversimplification. The lack of a cap domain does not predict an open active site because oligomerization of the core domain can provide both specificity residues from an adjacent core domain as well as encapsulation of the active site environment (Figure 5). A prototypical example of such a C0 cap member is 2-keto-3-deoxy-D-glycero-D-galacto-9-phosphononononic acid phosphate (KDN9P) phosphatase from *Bacteriodes thetaiotaomicron*, which acts in the biosynthetic pathway of the 9-carbon alpha-keto acid 2-keto-3-deoxy-D-glycero-D-galactononononic acid [21]. Such polyhydroxylated α-keto acids

are incorporated into cell-surface glycoproteins and glycolipids in both prokaryotic and eukaryotic organisms. The structure of KDN9P phosphatase complexed with the sialic acid N-acetyl neuraminate and vanadate to 1.63 Å resolution reveals a structure in which a small insert in the same position as the C1 cap domain allows tetramerization with active sites positioned at the subunit-subunit interface [21]. Indeed, seven out of the ten interactions made directly or through water between enzyme and the N-acetyl-neuraminic acid leaving group are made by the adjacent protomer core acting as a cap. The efficiency of this type of active site construction is demonstrated by the activity against the physiological substrate KDN9P ($K_m = 0.10$ mM, $k_{cat} = 1.2\ s^{-1}$, $k_{cat}/K_m = 1 \times 10^4\ M^{-1}s^{-1}$). Moreover, the substantively lower second-order rate constants against other related sugars such as the 8-carbon acid, 2-keto-3-deoxy-D-manno-8-phospho-octulosonic acid ($k_{cat}/K_m = 2 \times 10^2\ M^{-1}s^{-1}$) shows the ability of the adjacent core residues to confer substrate discrimination. Notably, the tetrameric assembly adopted by KDN9P phosphatase does not appear to undergo significant substrate-induced rearrangement, unlike the typical cap-core movement observed in HADSF C1 and C2 phosphatases. Such motion is therefore not a requirement of substrate specificity or catalytic efficiency in the HADSF.

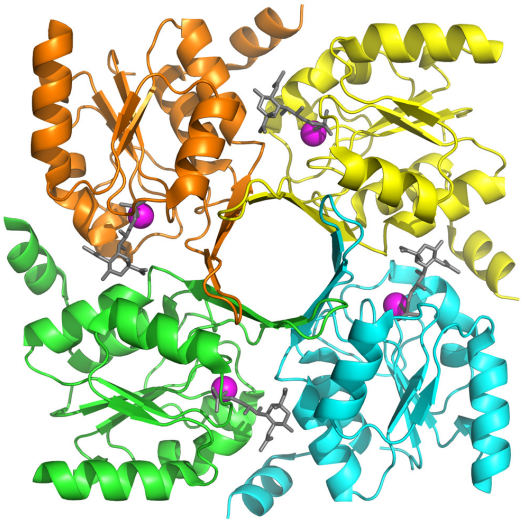

Figure 5. The structure of KDN9P phosphatase depicted as ribbons (colored by subunit) complexed with Mg^{2+} (magenta sphere), N-acetyl neuraminate and vanadate (grey sticks).

The adoption of a large cap domain to acquire substrate specificity elements versus a small insert to allow oligomerization (and use of the core for this same purpose) are not the sole mechanisms used by the HADSF to bind phosphate monoesters. The structure of D-*Glycero*-D-*manno*-heptose-1,7-(bis)phosphate phosphatase (GmhB) from *E. coli* bound to the substrate D-*Glycero*-D-*manno*-heptose-1,7-(bis)phosphate to 2.2 Å resolution reveals that, in place of a cap domain, the GmhB catalytic site is elaborated by three peptide inserts or

loops in the core domain of each GmhB monomer that pack to form a continuous binding surface around the substrate leaving group [22] (Figure 6). Examination of the solvent-accessible surface of the substrate-liganded structure shows that the substrate leaving group forms a ‘‘plug’’ that occludes the passage between bulk solvent and the phosphoryl-transfer site. Together, the three inserts and bulky substrate leaving group shield the catalytic site from solvent. The question then arises, are the inserted segments sufficient to afford substrate specificity and catalytic efficiency?

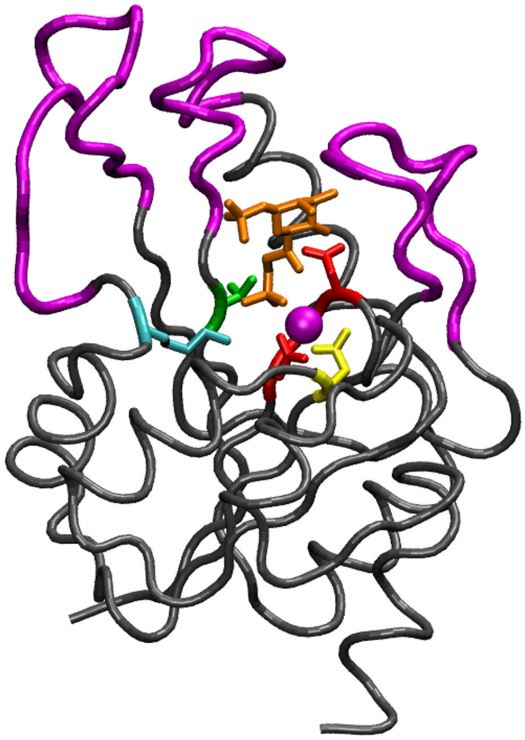

Figure 6. The structure of GmhB depicted as ribbons complexed with the substrate β-D-glycero-D-manno-heptose-1,7-(bis)phosphate (gold sticks). The substrate specificity loops inserted into the Rossmann core domain are shown in magenta. Motifs colored as in Figure 2.

This query is best approached by inspecting the steady-state kinetic properties of the enzyme. First, *E. coli* GmhB is highly efficient against its natural substrate β-D-glycero-D-manno-heptose 1,7-bisphosphate with $k_{cat}/K_m = 7 \times 10^6$ $M^{-1}s^{-1}$. Notably, orthologues of GmhB are utilized in the pathways for production of D-glycero-D-mannoheptose 1α-GDP [23] and L-glycero-D-manno-heptose 1β-ADP [24]. The pathways provide activated glycero-manno-heptose units for incorporation into membrane surface glycoproteins and glycolipids [25 – 29]. GmhB orthologues have been honed for hydrolysis of the anomer corresponding to that of the respective pathway kinase that generates the bis-phosphorylated substrate.

Consequently, *E. coli* GmhB shows a 100:1 preference for the β-anomer over the α-anomer while the *Bacteriodes thetaiotaomicron* GmhB has a 5-fold preference for the α-anomer over the β-anomer [30]. Site-directed mutagenesis demonstrates that conserved residues in the three inserted segments are integral to substrate binding and that the contribution to binding and activity differs between anomers [22]. Overall, the inserted segments are necessary and sufficient to confer anomeric selectivity, substrate specificity, and catalytic efficiency.

Conclusion

The HADSF is distinctive among enzyme superfamilies in that the elements responsible for the conserved chemistry are structurally and spatially distinct from those responsible for substrate specificity. Interactions between the core domain amino-acid side chains, main chain, and the substrate phosphoryl group make a ''mold'' for the trigonal bipyramidal transition state common to the phosphoryl transferases in the superfamily. Despite this conserved catalytic mechanism, a wide variety of substrates is allowed by utilizing determinants to bind the substrate leaving group, provided by inserted domains, loop extensions or segments that allow oligomerization and utilization of adjacent protomers. This modular approach to binding of new substrates is proposed to underlie successful evolution in the HADSF.

Acknowledgement

This work was supported by NIH GM061099 (to DD-M and KNA).

References

[1] Collet, J.F., van Schaftingen, E., and 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] Bateman, A. *et al*. (2004) The Pfam protein families database. *Nucleic Acids Res*. **32**(Database issue):D 138 – 41. doi: http://dx.doi.org/10.1093/nar/gkh121.

[3] 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.

[4] 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.

[5] Liu, J.Q. *et al.* (1995) Reaction mechanism of L-2-haloacid dehalogenase of Pseudomonas sp. YL. Identification of Asp10 as the active site nucleophile by 18O incorporation experiments. *J. Biol. Chem.* **270**(31):18309 – 12.

[6] Gross, C., Felsheim, R., and Wackett, L.P. (2008) Genes and enzymes of azetidine-2-carboxylate metabolism: detoxification and assimilation of an antibiotic. *J. Bacteriol.* **190**(14):4859 – 64. doi: http://dx.doi.org/10.1128/JB.02022-07.

[7] Baker, A.S. *et al.* (1998) Insights into the mechanism of catalysis by the P-C bond-cleaving enzyme phosphonoacetaldehyde hydrolase derived from gene sequence analysis and mutagenesis. *Biochemistry* **37**(26):9305 – 15. doi: http://dx.doi.org/10.1021/bi972677d.

[8] Morais, M.C. *et al.* (2000) The crystal structure of bacillus cereus phosphonoacetaldehyde hydrolase: insight into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD enzyme superfamily. *Biochemistry* **39**(34):10385 – 96. doi: http://dx.doi.org/10.1021/bi001171j.

[9] Collet, J.-F. *et al.* (1997) Human -3-phosphoserine phosphatase: sequence, expression and evidence for a phosphoenzyme intermediate. *FEBS Letters* **408**(3):281 – 284. doi: http://dx.doi.org/10.1016/S0014-5793(97)00438-9.

[10] 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.

[11] 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. *Journal of Molecular Biology* **244**(1):125 – 132. doi: http://dx.doi.org/10.1006/jmbi.1994.1711.

[12] 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.

[13] Lu, Z., Dunaway-Mariano, D., and 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.

[14] Peisach, E. *et al.* (2004) X-ray crystal structure of the hypothetical phosphotyrosine phosphatase MDP-1 of the haloacid dehalogenase superfamily. *Biochemistry* **43**(40):12770 – 9. doi: http://dx.doi.org/10.1021/bi0490688.

[15] Galburt, E.A. *et al.* (2002) Structure of a tRNA repair enzyme and molecular biology workhorse: T4 polynucleotide kinase. *Structure* **10**(9):1249 – 60. doi: http://dx.doi.org/10.1016/S0969-2126(02)00835-3.

[16] Kamenski, T. *et al.* (2004) Structure and mechanism of RNA polymerase II CTD phosphatases. *Mol. Cell* **15**(3):399 – 407. doi: http://dx.doi.org/10.1016/j.molcel.2004.06.035.

[17] Lahiri, S.D. *et al.* (2004) Analysis of the substrate specificity loop of the HAD superfamily cap domain. *Biochemistry* **43**(10):2812 – 20. doi: http://dx.doi.org/10.1021/bi0356810.

[18] 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.

[19] Tremblay, L.W., Dunaway-Mariano, D. and 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.

[20] 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.

[21] Lu, Z. *et al.* (2009) Structure-function analysis of 2-keto-3-deoxy-D-glycero-D-galactononononate-9-phosphate phosphatase defines specificity elements in type C0 haloalkanoate dehalogenase family members. *J. Biol. Chem.* **284**(2):1224 – 33. doi: http://dx.doi.org/10.1074/jbc.M807056200.

[22] Nguyen, H.H. *et al.* (2010) Structural determinants of substrate recognition in the HAD superfamily member D-glycero-D-manno-heptose-1,7-bisphosphate phosphatase (GmhB). *Biochemistry* **49**(6):1082 – 92. doi: http://dx.doi.org/10.1021/bi902019q.

[23] Kneidinger, B. *et al.* (2002) Biosynthesis pathway of ADP-L-glycero-beta-D-manno-heptose in *Escherichia coli. J. Bacteriol.* **184**(2):363 – 9.
doi: http://dx.doi.org/10.1128/JB.184.2.363-369.2002.

[24] Kneidinger, B. *et al.* (2001) Biosynthesis of nucleotide-activated D-glycero-D-manno-heptose. *J. Biol. Chem.* **276**(24):20935 – 44.
doi: http://dx.doi.org/10.1074/jbc.M100378200.

[25] Kosma, P. *et al.* (1995) Glycan structure of a heptose-containing S-layer glycoprotein of *Bacillus thermoaerophilus. Glycobiology* **5**(8):791 – 6.
doi: http://dx.doi.org/10.1093/glycob/5.8.791.

[26] Messner, P. (1996) Chemical composition and biosynthesis of S-layers, in *Crystalline Bacterial Cell Surface Proteins*, Sleytr, U.B., *et al.*, Editors. R. G. Landes/Academic Press: Austin, TX. p. 35 – 76.
doi: http://dx.doi.org/10.1016/B978-012648470-0/50007-4.

[27] Schaffer, C. and Messner, P. (2004) Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology. *Glycobiology* **14**(8):31R-42R.
doi: http://dx.doi.org/10.1093/glycob/cwh064.

[28] Schaffer, C. *et al.* (1999) Complete glycan structure of the S-layer glycoprotein of *Aneurinibacillus thermoaerophilus* GS4 – 97. *Glycobiology* **9**(4):407 – 14.
doi: http://dx.doi.org/10.1093/glycob/9.4.407.

[29] Valvano, M.A., Messner, P. and Kosma, P. (2002) Novel pathways for biosynthesis of nucleotide-activated glycero-manno-heptose precursors of bacterial glycoproteins and cell surface polysaccharides. *Microbiology* **148**(Pt 7):1979 – 89.

[30] 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.