Supporting Information

Comparative genomics guided discovery of two missing archaeal enzyme families involved in the biosynthesis of the pterin moiety of methanopterin and tetrahydrofolate

Valérie de Crécy-Lagard[1], Gabriela Phillip[s1], Laura L. Grochowski[2], Basma El Yacoubi1, Francis Jenney[3], Michael W.W. Adams[4], Alexey G. Murzin[5] and Robert H. White[2]

[1]Department of Microbiology and Department of Microbiology and Cell Science, University of Florida, P.O. Box 110700, Gainesville, FL 32611-0700.

[2]Department of Biochemistry (0308), Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

[3]Department of Basic Sciences, Georgia Campus, Philadelphia College of Osteopathic Medicine, Suwanee, GA 30024

[4]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602

[5]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK

Address correspondence to: Valérie de Crécy-Lagard, Department of Microbiology and Cell Science, University of Florida, P.O. Box 110700, Gainesville, FL 32611-0700; E-mail : vcrecy@ufl.edu, Tel: (352) 392 9416; Fax: (352) 392 5922

Supplemental data 1

Enzymatic assay of PF1278: absence of 7,8-dihydroneopterin aldolase (DHNA) activity.

The standard assay used for the measurement of DHNA enzymatic activity of the PF1278 derived enzyme was conducted in 25.5 µl reaction volume and included 39 µM protein, 40 mM TES/Cl^- buffer pH 7.4, 8 mM $MgCl_2$, 16 mM DTT, and 110 µM 7,8-dihydroneopterin supplemented with Fe(II), Zn(II), Co(II), Mn(II), Ni(II) or Cu(II) at 60 µM. Samples were sealed under argon and incubated for 10 min at 70°C. Following incubation, the reactions were quenched by the addition of 20 µl 1 M HCl. 7,8-Dihydroneopterin and 6-hydroxymethyl-7,8-dihydropterin in the incubation mixture were oxidized to the fluorescent neopterin and 6-hydroxymethylpterin by the addition of 8 µl of a saturated solution of iodine in methanol and incubated at room temperature for 30 min. Following oxidation, the samples were neutralized by the addition of 20 µl 1 M NaOH and excess iodine removed by reduction with 8 µl 1 M $NaHSO_3$. Following centrifugation, the samples were combined with water for a final volume of 1 mL and analyzed by HPLC as described in the methods section. No PF1278 dependent DHNA activity was observed under any condition.

Enzymatic assay of PF1278: PF1278 is a 7,8-dihydroneopterin monophosphate aldolase.

We then tested if PF1278 could harbor 7,8-dihydroneopterin monophosphate activity (Figure below). To 80 µl of a ~100 µM solution of 7,8-dihydroneopterin monophosphate in 50 mM *N*-[tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid (TES), pH 7.0, buffer containing 10 mM $MgCl_2$, and 20 mM DTT was added 10 µl of a solution of the PF1278 enzyme (~13 mg/ml) and the mixture was sealed under argon and incubated for 15 min at 70°C. 7,8-dihydroneopterin-monophosphate and 6-hydroxymethyl-7,8-dihydropterin in the incubation

mixture were oxidized to the fluorescent D-neopterin monophosphate and 6-hydroxymethylpterin by the addition of 8 µl of a saturated solution of iodine in methanol and incubated at room temperature for 30 min. Following oxidation, the samples were neutralized by the addition of 20 µl 1 M NaOH and excess iodine removed by reduction with 8 µl 1 M $NaHSO_3$. Following centrifugation, the samples were combined with water for a final volume of 1 mL and analyzed by HPLC as described herein. Addition of 200 µM Fe(II) or Zn(II) had no effect on the yield of 6-hydroxymethylpterin measured.

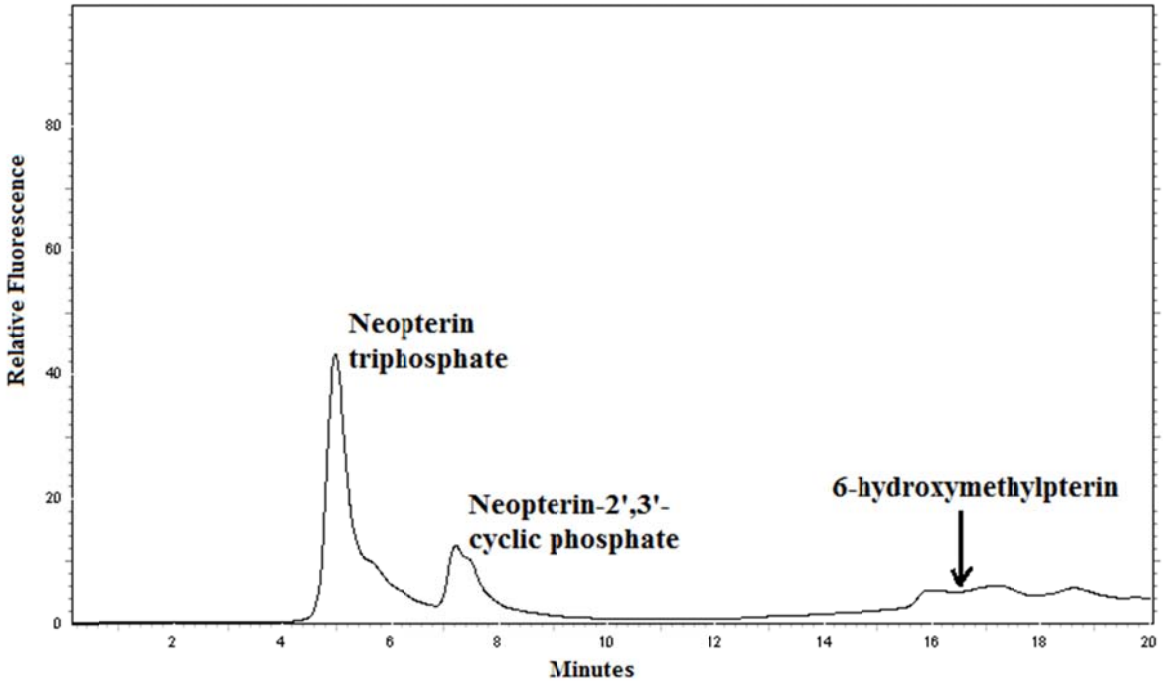
The other product of the reaction glycolaldehyde-phosphate, was synthesized by the periodate oxidation of DL-α-glycerol-P and converted into its O-(4-nitrobenzyl)hydroxylamine oximine derivative *(1)* for HPLC and mass spectral analyses. This material was identified by its ESI-MS with a MH^+ = 291.1 *m/z*; MH^+ = 313.2 *m/z*: and MNa_2^+ = 335.2 *m/z* and $M\text{-}H^-$ = 289.2 *m/z*. The product from the enzymatic reaction was converted into the same derivative and the known and unknown samples had the same retention times co-cromatographed upon HPLC analysis confirming that the other product of the reaction was glycolaldehyde-phosphate.

Enzymatic assay of PF1278: PF1278 is not a 7,8-dihydroneopterin triphosphate aldolase.

We tested if PF1278 could harbor 7,8-dihydroneopterin triphosphate activity. This involved the enzymatic production of 7,8-dihydroneopterin triphosphate from GTP using FolE2 from *Neisseria gonorrhoeae* (NgFolE2) provided by Dirk Iwata-Reuyl, Portland State U. Thus GTP, 5.7 mM, was incubated in 50 mM *N*-[tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid (TES), pH 7.0, buffer containing 20 mM DTT and 60 µg of NgFolE2 The sample was incubated under argon for 12 hr and the production of 7,8-dihydroneopterin triphosphate was followed by HPLC analysis as described above. To this reaction mixture was added 10 µl of a

solution of the PF1278 enzyme (~13 mg/ml) and the mixture was incubated for 45 min at 70° C. The sample was then assayed for 6-hydroxymethylpterin as described above. As shown on the Figure below, showed no 6-hydroxymethylpterin was produced.

The data demonstrated that PF1278 derived enzyme is exclusively a 7,8-dihydroneopterin monophosphate aldolase that produces glycolaldehyde-phosphate and 6-hydroxymethyl-7,8-dihydropterin from 7,8-dihydroneopterin monophosphate. The addition of 200 µM Fe(II) or Zn(II) to the incubation mixtures had no effect on the yield of 6-hydroxymethylpterin produced.

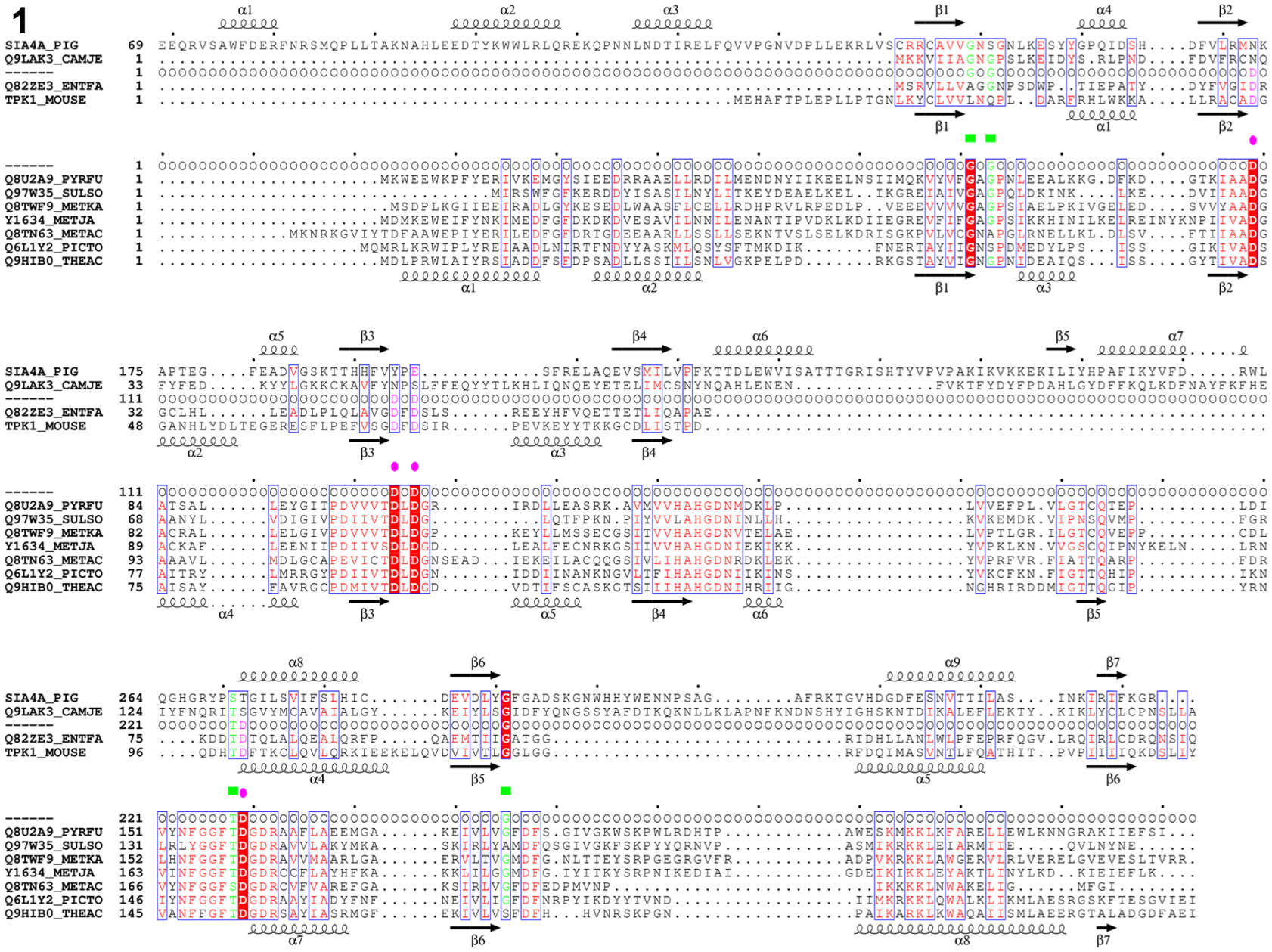


HPLC assay of the PF1278 gene product for 7,8-dihydroneopterin triphosphate aldolase activity. The trace was obtained after the assay was incubated for 45 min at 70° C. The neopterin-2',3'-cyclic phosphate is generated by the chemical cyclization of 7,8-dihydroneopterin triphosphate substrate.

Reference

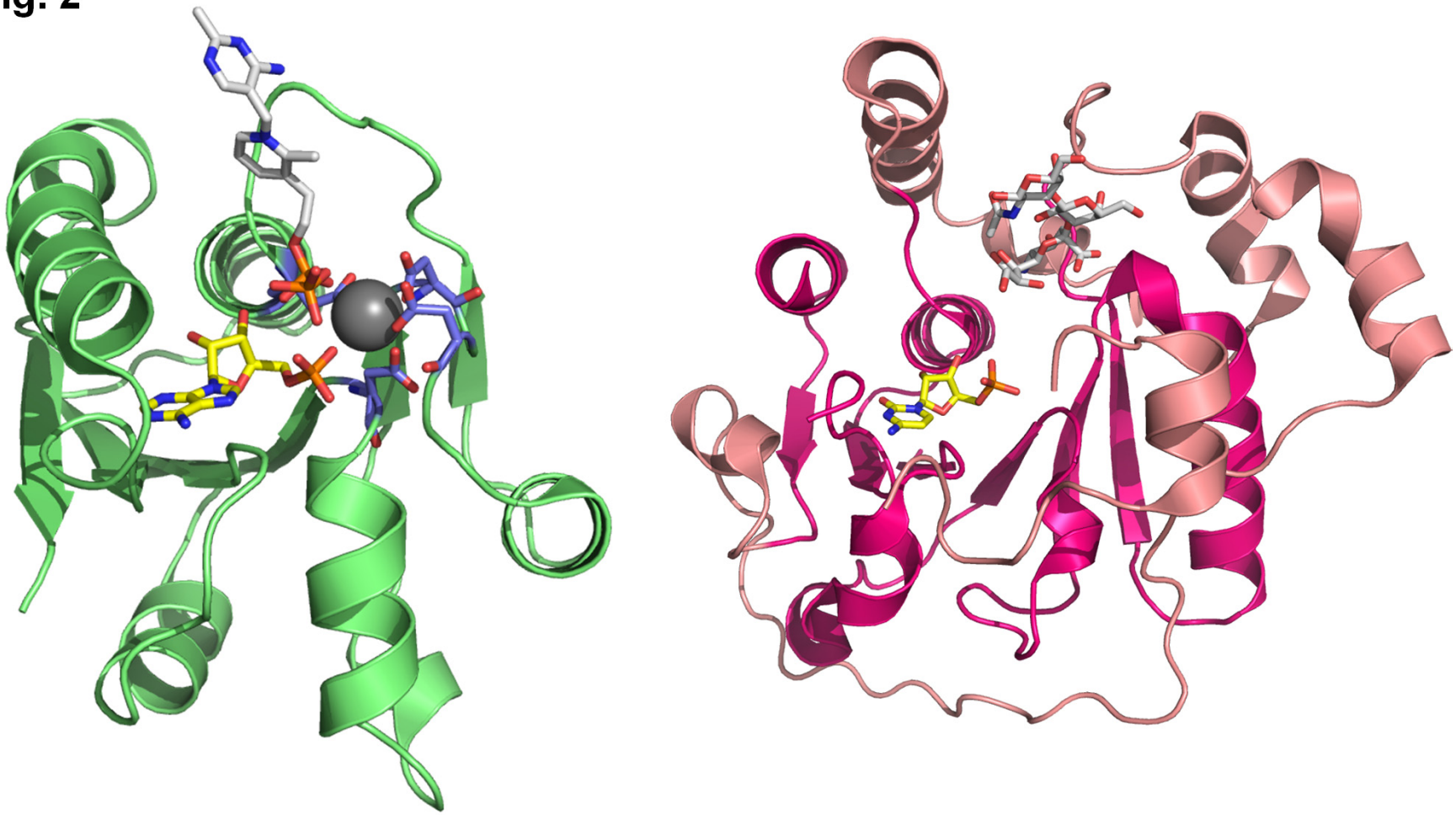
White, R. H. (2008) Biochemical origins of lactaldehyde and hydroxyacetone in *Methanocaldococcus jannaschii*, *Biochemistry 47*, 5037-5046.

Sup Fig. 1



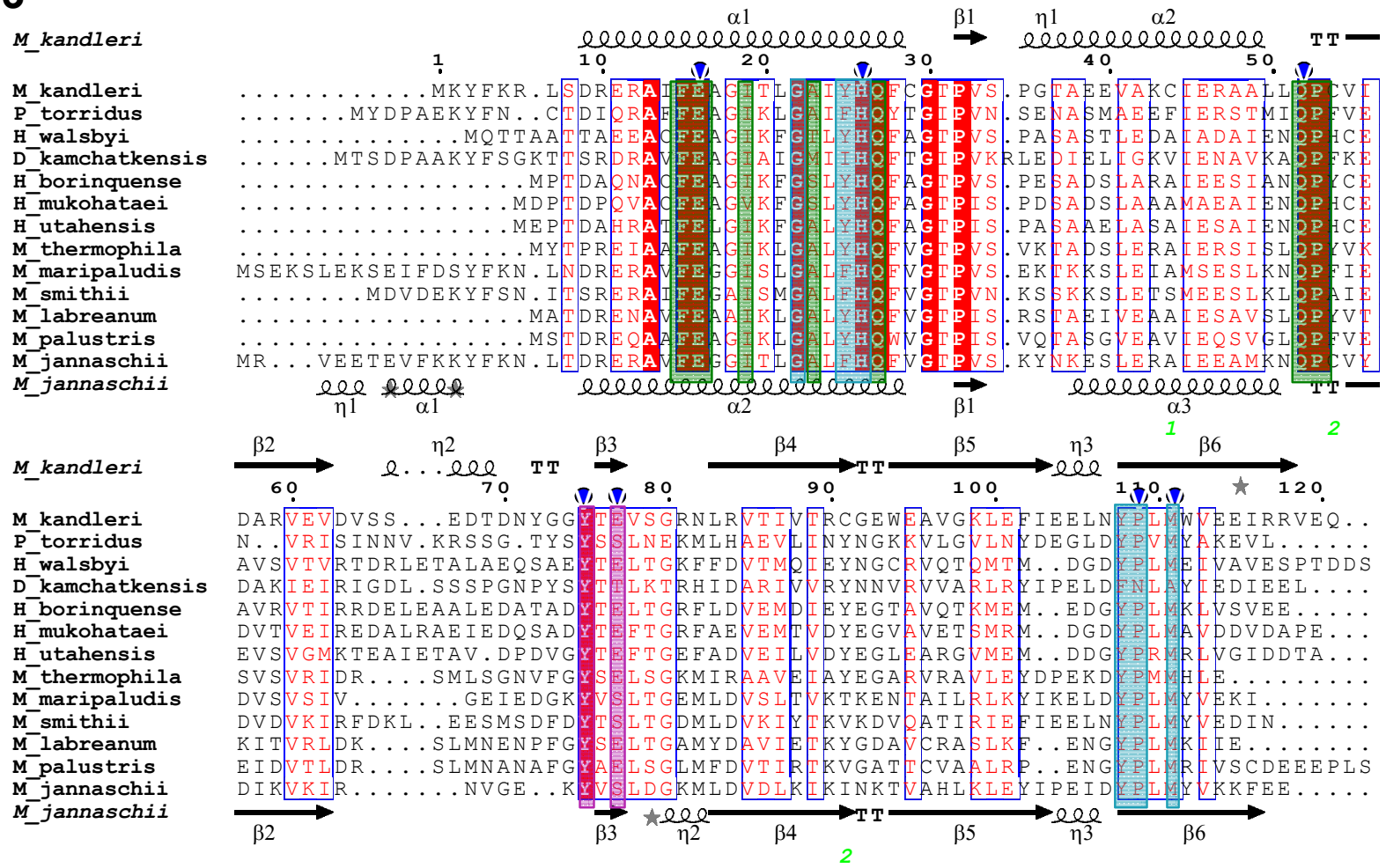
Structural based alignment of CO1634 (MptE), thiamin pyrophosphokinase (TPK) and GST-II sialyltransferase. The structure-based alignment of porcine ST3Gal-I (2WNB), *Campylobacter jejuni* CST-II (2X61), TPK family protein from *Enterococcus faecalis* (3MEL) and mouse TPK1 (2F17) is co-aligned with the multiple sequence alignment of the MptE family. The secondary structures of the mammalian sialyltransferase and TPK are shown, respectively, above and below their sequences. The predicted secondary structure of MptE is shown below the family sequences. Green squares mark the structurally essential positions in the NTP-binding site; magenta dots indicate the metal ion-binding aspartate residues in TPK and MptE.

Sup Fig. 2



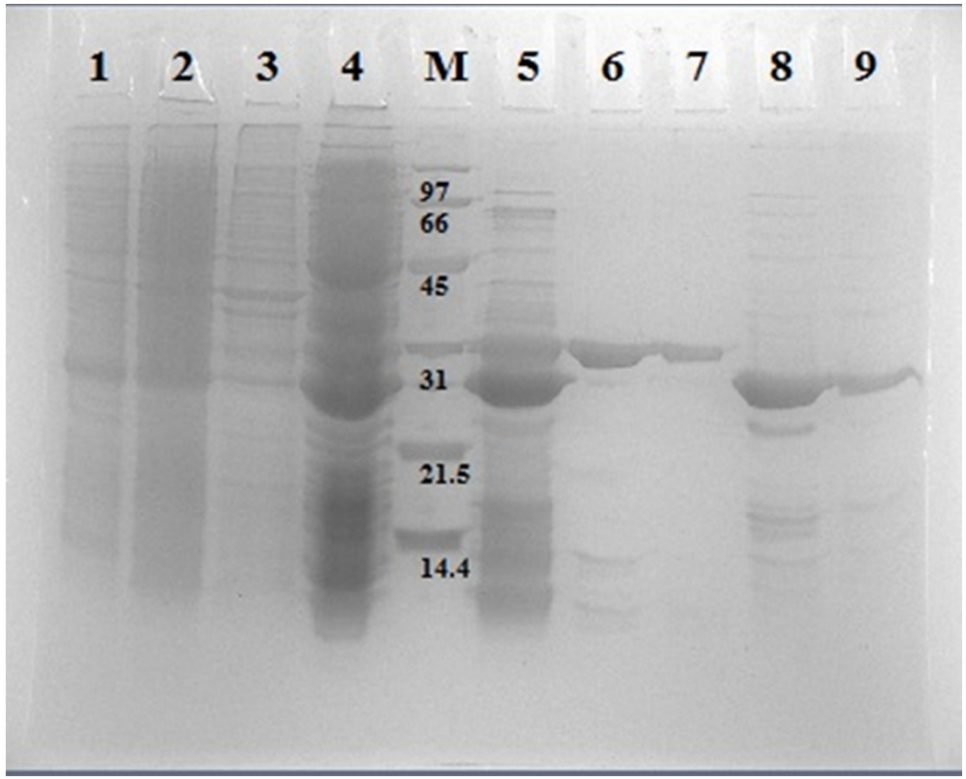
Structural comparison of the thiamin pyrophosphokinase (TPK) catalytic domain and GST-II sialyltransferase . Structural comparison of the thiamin pyrophosphokinase (TPK) catalytic domain (PDB entry 2F17, *left*) and GST-II sialyltransferase (2X61, *right*). The TPK-like core of GST-II is highlighted with dark pink. Similarly bound products, AMP and CMP, respectively, are shown with yellow carbons; other bound ligands with white carbons. Sphere depicts magnesium ion, coordinated by four conserved aspartate (slate carbons) and phosphate groups.

Sup Fig. 3



Structure based multisequence alignment of COG2098 family members. Secondary structure elements from *M. kandleri* COG2098 crystal structure (PDB ID 2IEC) are shown on top, and secondary structure elements from *M. jannaschii* COG2098 crystal structure (PDB ID 2OGF) are shown on bottom. Blue marks residues used involved in interacting with 8-oxoG. Active sites residues were colored using the same color scheme that in Fig. 3C. These are (2IEC numbering):Chain A (green): Phe 19, Glu 20, Ile 23, Ala 27, Gln 31, Gln 56, Pro 57; Chain B (cyan): Gly 26, Tyr29, His 30, Tyr 112, Pro 113, Met 115;Chain C (magenta): Tyr 79, Glu 81.

Sup Fig. 4



SDS-PAGE analysis of the purification of the MJ1634 gene product. Lane M represents the protein markers used to calibrate the gel, the masses listed under the bands are listed in kDa. Lane 1 is the un-induced cells and Lane 2 is the induced cells containing the MJ1634 gene product. Lane 3 and 4 is the extracted MJ1634 gene product present in the cell pellet and soluble fraction after sonication and centrifugation. Lane 5 is the 70 ^{o}C heated soluble MJ1634 gene product prior to purification. Lanes 6 through 9 are four of the MonoQ purified fractions believed to contain the MJ1634 gene product. The large band seen in lanes 5, 8, and 9 represents the chloramphenicol acetyltransferase (CAT) protein from *E. coli.* The MJ1634 gene product was identified from bands in lanes 6 and 7 by MALDI TOF/TOF analysis to be in the major band seen in lanes 6 and 7. The gel is representative of the purification of all the proteins described in this work.