DR. DIANA MARIE DOWNS (Orcid ID : 0000-0002-1564-6205)

Article type : Research Article

Functional characterization of the HMP-P synthase of *Legionella pneumophila* (Lpg1565)

Michael D. Paxhia[1], Michele S. Swanson[2] and Diana M. Downs[1*]

[1]Department of Microbiology, University of Georgia, Athens, GA 30606

[2]Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48019

[*] To whom correspondences may be addressed: Dr. Diana Downs, Department of Microbiology, University of Georgia, Athens, GA, USA, dmdowns@uga.edu

Running Title: HMP-P synthase in *Legionella pneumophila*

Keywords: Hydroxymethyl pyrimidine, HMP, Thiamine synthesis, *THI5*, *Legionella pneumophila, lpg1565*

ABSTRACT

The production of the pyrimidine moiety in thiamine synthesis, 2-methyl-4-amino-5-hydroxymethylpyrimidine phosphate (HMP-P), has been described to proceed through the Thi5-dependent pathway in *Saccharomyces cerevisiae* and other yeast. Previous work found that *Sc*Thi5 functioned poorly in a heterologous context. Here we report a bacterial ortholog to the yeast HMP-P synthase (Thi5) was necessary for HMP synthesis in *Legionella pneumophila*. Unlike *Sc*Thi5, *Lp*Thi5 functioned *in vivo* in *Salmonella enterica* under multiple growth conditions. The protein *Lp*Thi5 is a dimer that binds pyridoxal-5'-phosphate (PLP), apparently without a solvent-exposed Schiff base. A small percentage of *Lp*Thi5 protein co-purifies with a bound molecule that can be converted to HMP. Analysis of variant proteins both *in vivo* and *in vitro* confirmed that residues in sequence motifs conserved across bacterial and eukaryotic orthologs modulate the function of *Lp*Thi5.

IMPORTANCE

Thiamine is an essential vitamin for the vast majority of organisms. There are multiple strategies to synthesize and salvage this vitamin. The predominant pathway for synthesis of the pyrimidine moiety of thiamine involves the Fe-S cluster protein ThiC. An alternative pathway utilizes Thi5, a novel enzyme that uses PLP as a substrate. The Thi5-dependent pathway is poorly characterized in yeast and has not been characterized in Bacteria. Here we demonstrate that a Thi5-dependent pathway is necessary for thiamine biosynthesis in *Legionella pneumophila* and provide biochemical data to extend knowledge of the Thi5 enzyme, the corresponding biosynthetic pathway and the role of metabolic network architecture in optimizing its function.

INTRODUCTION

Thiamine pyrophosphate (TPP), the active form of Vitamin B_1, is a cofactor important for several enzymes in central metabolism including transketolase, pyruvate dehydrogenase, and α-ketoglutarate dehydrogenase. The cofactor is comprised of two moieties, 2-methyl-4-amino-5-hydroxymethylpyrimidine diphosphate (HMP-PP) and 4-methyl-5-β-hydroxyethylthiazole phosphate (THZ-P), that are independently synthesized and combined to form TPP (Jurgenson *et al.*, 2009). The pyrimidine precursor HMP-P is synthesized from an intermediate in the purine biosynthetic pathway (aminoimidazole ribotide) by the phosphomethylpyrimidine synthase ThiC (E.C. 4.1.99.17) an enzyme encoded in plants, archaea and most bacteria (Jurgenson *et al.*, 2009). However some organisms, notably

Saccharomyces cerevisiae, use Thi5 (HMP-P synthase) in place of ThiC to generate HMP-P. Labeling studies showed that the atoms of HMP originate from histidine and a B6 vitamer, presumably pyridoxal-5'-phosphate (PLP) (Tazuya *et al.*, 1989, Tazuya *et al.*, 1995, Ishida *et al.*, 2008). (Figure 1A) The hemiascomycetes clade of fungi contain multiple paralogs of *THI5*, and genetic studies in *S. cerevisiae* S288c showed any member of the *THI5/11/12/13* gene family was sufficient for the synthesis of HMP-P (Wightman & Meacock, 2003). Other fungi encode a single ortholog of *THI5*, and a *Schizosaccharomyces pombe* mutant lacking the *THI5* homolog *nmt1* is a thiamine auxotroph that specifically requires HMP for growth (Maundrell, 1990, Schweingruber *et al.*, 1991).

The Thi5 HMP-P synthase from lower eukaryotes (*e.g., Candida albicans* and *S. cerevisiae*) is a member of the periplasmic binding protein superfamily (COG0715), and three-dimensional crystal structures of these proteins have been solved (Lai *et al.*, 2012, Coquille *et al.*, 2012), and are available in protein databases (PDBs 4H65, 4H67, 4H6D, 4ESX). The *S. cerevisiae* Thi5 protein was purified with PLP in its active site and formed dimers in solution, an oligomeric state confirmed by its crystal structure (Coquille *et al.*, 2012). An *in vitro* study of the *C. albicans* Thi5 protein showed that the protein had HMP-P synthase activity, albeit unexpectedly low. When PLP and iron were provided in a reaction mixture, *Ca*Thi5 produced between 0.2-0.5 mol of HMP-P per mol of protein (Lai *et al.*, 2012). The small amount of product formed, in addition to the lack of stimulation of enzyme activity by histidine, suggested that a necessary molecule was supplied by the protein. *Ca*Thi5 variants H66G and H66N had no activity in the *in vitro* assay (Lai *et al.*, 2012) and other variants that altered residue H66 were inactive *in vivo* (Coquille *et al.*, 2012). In total, the data led authors of one study with *Ca*Thi5 to conclude that the enzyme used a suicide mechanism to donate a moiety of residue H66 to yield HMP-P (Lai *et al.*, 2012). The *in vitro* formation of HMP-P required a high concentration (600 μM) of *Ca*Thi5, the addition of excess ferrous iron (900 μM) and molecular oxygen (Lai *et al.*, 2012). The above conditions of the assay, and the low product yield achieved, make it formally possible that a component or condition relevant for Thi5 activity *in vivo* was not recapitulated in the *in vitro* work.

Despite the presumed *in vivo* availability of both histidine and pyridoxine substrates, *Sc*Thi5 fails to support growth of a *thiC* mutant of *S. enterica* on minimal glucose medium (Palmer *et al.*, 2015). The ease with which suppressors that restored growth were isolated suggested that the architecture of the metabolic network was impacting Thi5 activity. An additional study found that the Thi5-dependent thiamine synthesis in *S. cerevisiae* required an ammonium-dependent PLP synthase (E.C. 4.3.3.6; *SNZ2, SNZ3*) (Paxhia & Downs, 2019). Thus, it was formally possible that the lack of activity in a bacterial host

reflected the need for a PLP synthesis/delivery system specific for the Thi5 enzyme. Such enzymes are not present in *S. enterica*, which generates PLP via a DXP-dependent pathway (Figure1B). The study herein was initiated to address this possibility, and to better understand the cellular function and biochemical properties of a bacterial Thi5 protein. To gain insight to the impact of network structure on thiamine synthesis, we identified and focused on a *THI5* homolog encoded within the *Legionella pneumophila* thiamine biosynthesis operon (Rodionov *et al.*, 2002, Sahr *et al.*, 2012). This homolog is regulated by a thiamine-pyrophosphate riboswitch in the 5' untranslated region, and a study in *L. pneumophila* Paris demonstrated that transcription from this region is repressed by thiamine-pyrophosphate and activated by CsrA (Sahr *et al.*, 2017). The work herein verified the role of *lpg1565* (redesignated *thi5*) in thiamine synthesis in *L. pneumophila* and analyzed its ability to function in *S. enterica.* Further, data herein showed that *Lp*Thi5 co-purifies with a molecule that can be converted to HMP, and identified variants that have altered activity *in vivo.* In total, this work expands our understanding of the function of Thi5, and its integration in the metabolic network of the cell.

RESULTS AND DISCUSSION

***THI5* orthologs co-occur with DXP-dependent PLP synthesis in prokaryotes.** In order to test the simple hypothesis that lack of Thi5 function in *S. enterica* was due to an absence of an Snz/PdxS pathway for PLP biosynthesis, genomes in the IMG database containing *THI5* orthologs were found using iterative BLAST-P searches, and the phylogeny of those homologs as well as the presence of homologs of key PLP biosynthetic genes were assessed. A CCCXC motif distinguishes *Sc*Thi5p from the structural homolog *N*-formyl-4-amino-5-aminomethyl-2-methylpyrimidine-binding ThiY, and so this motif was used to identify putative *THI5* homologs that had been annotated as a substrate binding protein for an ABC transport system (Bale *et al.*, 2010). Bacterial homologs with the CCCXC motif were identified in the *Endozoicomonas*, *Fluorobacter* and *Legionella* genera. Orthologs identified without the N-terminal 60 amino acids were not considered, since this region of the protein contains residues important for coordinating PLP in the *Sc*Thi5p active site (Coquille *et al.*, 2012). Presence of homologs for PNP synthase (PdxJ; E.C. 2.6.99.2), or a subunit of the glutamine-hydrolyzing PLP synthase (PdxS, Snz; E.C. 4.3.3.6) were used to define the presence of the DXP-dependent or -independent pathway for PLP biosynthesis, respectively (Figure 2). Among Eukaryotes *THI5* homologs were only found in fungi. Each of the 180 fungal genomes with a putative *THI5* gene had a homolog of both *pdxS* and *pdxH*, indicating they used the DXP-independent pathway for PLP biosynthesis and could salvage B_6 vitamers, similar to *S. cerevisiae*. Two clades of γ-proteobacteria with *THI5* homologs were identified: four *Endozoicomonas*

species and 31 species within the Legionellaceae family. Each of these bacteria had homologs to *pdxJ* and *pdxH* indicating synthesis of PLP was by the DXP-dependent pathway and suggesting a potential for vitamer salvage. It was noted that many of the Legionellaceae species contained multiple paralogs of *pdxH* (29 of 31 species have 2-4 paralogs), a feature that was not investigated further. Thus, genomic analysis suggested Thi5 enzymes could function in the absence of a PdxS PLP synthase. While this finding appeared to negate the hypothesis that *Sc*Thi5 required this enzyme as a specific PLP delivery system to function *in vivo*, it is worth noting that non-orthologous replacement for PdxS was not ruled out by these analyses.

lpg1565* encodes a Thi5 ortholog and contributes to HMP synthesis in *L. pneumophila. A role for the *THI5* ortholog in a metabolic network using DXP-dependent PLP biosynthesis was investigated using *L. pneumophila. lpg1565* is adjacent to several open reading frames predicted to encode enzymes for thiamine biosynthesis. The operon includes homologs of *thiO, thiG* and *thiF* (*lpg1566, lpg1567, lpg1569,* respectively), and a short ORF with homology to *thiS*. Each of these proteins is involved in the synthesis of the thiazole moiety of thiamine in defined pathways. The operon also had an ORF homologous to a fusion protein between *thiD2* and *thiE* (*lpg1568*) (Figure 1C) (Sahr *et al.*, 2012). A mutant of *L. pneumophila* with an in-frame insertion-deletion of *lpg1565* was constructed and subjected to growth analyses. The Δ*lpg1565* strain with an empty vector (pJB98) required exogenous thiamine for growth on Modified Ristroph Medium (Figure 3A). A plasmid containing the *lpg1565* gene (pDM1631) was unable to complement the nutritional requirement of the Δ*lpg1565* strain, suggesting the insertion was polar on one or more downstream genes (data not shown). A plasmid carrying the entire locus (*lpg1565 – lpg1569*; pDM1632) restored growth of the Δ*lpg1565* strain in the absence of thiamine (Figure 3B). In contrast, a plasmid derived from pDM1632 that lacked *lpg1565* (pDM1633) failed to complement the nutritional defect of the Δ*lpg1565* strain. Full growth of the Δ*lpg1565* strain carrying pDM1633 was restored by the addition of 100 nM thiamine. Partial growth of this strain was restored with the addition of 2.5 mM HMP (Figure 3C). Together these results demonstrate that *L. pneumophila* requires the *lpg1565* gene product for thiamine biosynthesis, specifically to generate the pyrimidine moiety. Based on these data and the bioinformatic analyses, *lpg1565* was renamed *thi5* and the gene product designated *Lp*Thi5 throughout.

The inability of HMP to restore robust growth to the *thi5* mutant suggested *L. pneumophila* lacked the salvage system characterized in *E. coli*, *S. enterica* and *S. cerevisiae*. In these organisms, exogenous HMP is incorporated into the biosynthetic pathway by the hydroxymethylpyrimidine kinase ThiD (E.C. 2.7.1.49). The presence of a *thiD2-thiE* fusion homolog (*lpg1568*) in *L. pneumophila*, rather than full

length *thiD* and *thiE* genes is likely to be the genomic difference that accounts for the weak HMP salvage observed. Other ThiD2-ThiE fusion proteins lack the HMP kinase activity associated with purified full-length ThiD while retaining HMP-P kinase activity. Thus a non-specific kinase may generate some HMP-P from exogenous HMP, allowing salvage via Lpg1568 in *L. pneumophila* (Thamm *et al.*, 2017). The ability to salvage thiamine reflects the presence of a homolog to a thiamine pyrophosphokinase (E.C. 2.7.6.2 ; ThiN) in *L. pneumophila* (Lpg2497), although a transporter for thiamine has not been identified.

***Lp*Thi5 functions in the *S. enterica* metabolic network.** The co-occurrence phylogenetic data above was not consistent with the simple hypothesis that Thi5 proteins require an Snz/PdxS-like protein to be functional. However, the phylogeny of Thi5 proteins clustering with distinct PLP biosynthesis pathways may suggest distinct structural features have evolved in Thi5 orthologs within metabolic networks of similar architecture. Therefore, we considered whether there were structural aspects of Thi5 that differ in concordance with the pathway for PLP biosynthesis in the organism, and whether structural differences could account for functional differences *in vivo*. As a first step, the functional complementation of a *thiC* mutant of *S. enterica* with a protein from a Eukaryote (*Sc*Thi5p) and the ortholog from a Prokaryote (*Lp*Thi5) were compared. *ScTHI5* was codon-optimized for expression in *S. enterica* and cloned into pTac85 under regulation of a Tac promoter (pDM1625). *Lpthi5* was similarly inserted into pTac85 to generate pDM1486. Both constructs were introduced into a *thiC* mutant of *S. enterica*, and growth was monitored on several carbon sources to test the response to different metabolisms (Figure 4). Neither *Lpthi5* nor *ScTHI5* restored growth on any carbon source without induced expression of the respective genes. However, when transcription was induced with IPTG (100 μM), *Lpthi5* (pDM1486), the *S. enterica* mutant grew on ribose, glucose, gluconate and galactose (Figure 4, Panel Ⅰ, A-D). In contrast, the codon-optimized *ScTHI5* conferred growth on ribose but not glucose and supported intermediate growth on gluconate and galactose (Figure 4, Panel Ⅱ, E-H).

When the thiazole moiety THZ (100 nM) was added to the growth medium, expression of *ScTHI5* allowed full growth of the *thiC* mutant on glucose, galactose and gluconate, and also increased the growth conferred by *Lpthi5*. Titration experiments on glucose medium showed the addition of THZ reduced the HMP concentration required for full growth of a *thiC* mutant from 25 nM to 10 nM (Supplemental Table 1). The growth stimulation by THZ is likely due to the kinetics of thiamine synthase (ThiE), which allows excess THZ-P to drive the reaction when HMP-PP levels are low (Backstrom *et al.*, 1995). Additionally, the *thiC* mutant required less HMP when using ribose as carbon source than when growing on glucose, also considered to be an indirect effect of metabolic flux differences (Supplemental Table 1). In total, the

data above support the conclusion that *Lp*Thi5 is an HMP-P synthase that is more active in the *S. enterica* metabolic network than is the *Sc*Thi5 enzyme.

***Lp*Thi5 is a pyridoxal-5'-phosphate binding protein.** *Lp*Thi5-His_6 was purified from *E. coli* and characterized for quaternary structure and cofactor occupancy, using size exclusion chromatography and UV-Vis spectroscopy, respectively. Based on protein standards of known molecular weight, *Lp*Thi5 was calculated to be 86 ± 5.8 kDa, reasonably close to the theoretical weight of a *Lp*Thi5 dimer (72.2 kDa) (Figure 5) and thus consistent with the designation of *Sc*Thi5p as a dimer based on size exclusion chromatography as well as its crystal structure (Coquille *et al.*, 2012).

After purification in buffer lacking PLP, *Lp*Thi5 had absorbance maxima at 325 and 425 nm, features that are characteristic of a tautomeric equilibrium of the Schiff-base of characterized PLP-binding proteins (Mozzarelli & Bettati, 2006) (Figure 6A). Dialysis of *Lp*Thi5 against HEPES buffer (50 mM, pH 7.5) containing PLP (200 μM) increased the absorbance maxima at 325 and 425 nm, indicating the as-purified protein was not fully occupied with PLP (Figure 6B). Denaturation of *Lp*Thi5 (100 μM in 200 μL) with NaOH (0.1 M), released PLP that was detected by the absorption maximum at 392 nm, characteristic of free PLP in HEPES (Geders *et al.*, 2012). The PLP occupancy of the as-purified protein was 29 ± 1 %, while dialysis against PLP-containing buffer raised the occupancy to 63 ± 2 %.

In a majority of PLP-binding proteins, treatment with $NaBH_4$ (1 mM) reduces the Schiff-base between the protein and PLP, resulting in a noticeable reduction in absorbance at 425 nm (Toney, 2011, Soniya & Chandra, 2018). Treatment of as-purified *Lp*Thi5 with up to 25 mM $NaBH_4$ did not affect the absorbance at 425 nm. Accordingly, if present, a Schiff-base between PLP and the protein was not solvent accessible (Figure 6A). Surprisingly, when a variant lacking the proposed catalytic residue H70 (*Lp*$Thi5_{H70A}$) was treated with $NaBH_4$ (25 mM), absorbance at 425 nm decreased significantly (Figure 6C). These data suggested the catalytic histidine residue might shield access to the Schiff base in the wild-type protein.

***Lp*Thi5 purifies with a bound metabolite.** Efforts to detect HMP-forming activity of *Lp*Thi5 with the addition of its presumed substrates, histidine and PLP (Tazuya *et al.*, 1995, Ishida *et al.*, 2008) under multiple conditions were not successful. Specifically, these conditions included incubations of 100 μM *Lp*Thi5 at 37 °C for 16 hours in 50 mM HEPES pH 7.5, aerobically or anaerobically, with the addition of 1-10X histidine, PLP or PLP + histidine. However, control experiments generated an unexpected result. When *Lp*Thi5 (20 nmol) in HEPES (50 mM, pH 7.5) was incubated overnight (16 hr) at 37 °C, a biologically active compound was released when the protein was denatured with heat (Figure 7A). The

biologically active molecule allowed growth of a *thiC* mutant, but not a *thiI* mutant strain of *S. enterica*; moreover, this activity was detected in samples with no added substrate. Based on these data the compound was not thiamine, which would have satisfied the requirement of both the *thiC* and the *thiI* mutants. Further, neither histidine, PLP, or histidine and PLP together detectably impacted the biological activity present after the incubation described (data not shown).

A series of experiments identified conditions required for the production of the biologically active compound. First, if a protein sample was pre-treated with acid (10 % TFA), base (0.1 M NaOH), or temperature (95 °C, 5 mins) prior to incubation at 37 °C, (i.e., T=0), no biological activity was detected. These data demonstrated time was required for formation, and that HMP was not present in the purification. Second, when the concentration of HEPES was 10, rather than 50 mM, no biological activity was detected. Third, formation of a biologically active molecule was dependent on buffer: Tris (40 mM, pH 7.5), Bis-Tris-Propane (40 mM, pH 7.5), or HEPES resulted in its formation, but incubation in MOPS (40 mM, pH 7.5) did not. Additionally, since the protein sample was exchanged into a buffer lacking PLP using a PD10 desalting column, the relevant molecule was likely bound to the enzyme. In total, these data suggested the molecule that purified with the protein was not the fully formed biologically active compound. It is formally possible that the protein is partially modified *in vivo*, and that this modified *Lp*Thi5 may serve as the substrate for the *in vitro* reaction as previously proposed for *Ca*Thi5 (Lai *et al.*, 2012).

HMP is released after incubation of Thi5. The bioassay results indicated that the active compound released from *Lp*Thi5 was either HMP, HMP-P or a molecule that the *S. enterica* could transport and convert to HMP-P. Indeed, MALDI-TOF MS identified the active molecule as HMP (Fig 7). The molecule released after an overnight incubation of 690 nmol of *Lp*Thi5 was concentrated with a C18 SepPak and separated by HPLC with a Luna C_{18} column. Fractions (0.5 mL) were collected across 30 minutes. From a sample incubated overnight, biological activity was found in fractions with a retention time of 20-21.5 minutes and the corresponding peak had UV-Vis spectral features similar to HMP. Moreover, a standard of authentic HMP eluted at a similar retention time. A control (T=0) sample generated no peak with a UV-Vis spectrum consistent with HMP(-P); nor did these collected fractions have biological activity (Figure 7B). The active fractions of the sample from an overnight incubation contained a mass of 140.0 Da when ionized by MALDI-TOF MS. This peak, present in the active but not an inactive fraction, was consistent with the monoisotopic mass of ionized HMP (Figure 7C). Together these data support the conclusion that *Lp*Thi5 co-purified with a molecule that generated HMP upon incubation in buffer. It was considered

possible the relevant molecule was HMP-P since it was not clear the phosphorylated form would be stable during the ionization protocol.

Iron decreases time required for the release of HMP(-P) from *Lp*Thi5. Initial experiments suggested that more than a 12-hour incubation was needed to generate detectable HMP from a purified Thi5 protein. Based on its putative metal binding motif (CCCXC) (Bale *et al.*, 2010), we tested the hypothesis that the reaction is catalyzed by metals. To do so, purified Thi5 was pre-incubated with a 10-fold excess of a series of metals, and reaction time was quantified. *Lp*Thi5 samples supplemented with metals were denatured after eight-hours at 37 °C and product formation was determined by bioassay. Of the metals tested (Mg^{2+}, Ni^{2+}, Mn^{2+}, Zn^{2+}, Co^{2+}, Ca^{2+}, Cd^{2+}, Cu^{2+}, and Fe), only Fe stimulated the release of HMP by *Lp*Thi5. Presence of either Fe (III) or Fe (II) salts increased the efficiency of HMP formation with *Lp*Thi5. Titration of 100μM protein with 50, 100, 200, 500 μM, or 1 mM iron showed that an Fe:protein ratio of 2:1 was optimal for product formation. When *Lp*Thi5 (100μM) and iron ammonium sulfate (200 μM) were present in buffer, HMP was detectable by bioassay after 2 hr at 37°C. In these conditions, HMP release increased over time plateauing at 4 hr at a level similar to that released with an overnight incubation in buffer alone (Figure 7D). Based on the ability of iron to increase the rate of product release, it was plausible that the generation of HMP in an overnight incubation was facilitated by iron that was in the *Lp*Thi5 protein preparation and was acting as a general electron sink. The requirement for buffer suggests that exchange of ions between the protein and the solvent, perhaps accelerated by iron, could be important for release of HMP. Bound iron in purified *Lp*Thi5 was quantified by Ferene assay (Kennedy *et al.*, 1984, Palmer & Downs, 2013). A sample of *Lp*Thi5 (70.8 μM) contained 545 ± 43 nM Fe, while a buffer-only control contained 140 nM Fe. These data indicated that 18 nmol *Lp*Thi5 contained only 0.10 ± 0.008 nmol of Fe, minimizing the possibility that iron was specifically bound to the purified protein. While not quantifiable precisely, analysis by bioassay suggested 5-8% of purified *Lp*Thi5 protein released HMP. The stoichiometry of HMP(-P) to *Lp*Thi5 cannot effectively distinguish between release of a previously bound molecule and the proposed suicide mechanism, as previously suggested (Lai *et al.*, 2012).

Conserved residues are involved in the function of *Lp*Thi5 *in vivo* and *in vitro*. Thi5 variants that have individually altered conserved residues were surveyed for their functional properties. Previous work with *Sc*Thi5 showed alanine substitutions at the PLP-binding lysine residue binding (K62) or the adjacent histidine residue (H66) eliminated function *in vivo* (Coquille *et al.*, 2012). Likewise, neither *Lp*Thi5$_{K66A}$ or *Lp*Thi5$_{H70A}$ complemented a *thiC* strain of *S. enterica* (data not shown). Further, these purified variants did not to produce a biologically active compound when incubated in the presence of iron (data not

shown). Previous work with *Sc*Thi5 showed that alanine substitutions in the CCCXC motif compromised function *in vivo* (Coquille *et al.*, 2012). To corroborate the importance of this motif in *Lp*Thi5, each of the cysteines in the conserved CCCXC motif were changed to Ala, His or Ser by site directed mutagenesis in pDM1486. None of the twelve variants of *Lp*Thi5 complemented a *thiC* mutant when tested only with ribose as the sole carbon source (data not shown). Further, four of the variants (*Lp*Thi5$_{C191A}$, *Lp*Thi5$_{C192A}$, *Lp*Thi5$_{C193A}$, *Lp*Thi5$_{C195A}$) were purified, and none released a biologically active molecule after a 4 hr incubation in the presence of iron, or overnight in its absence, implying these conserved residues are essential for Thi5 activity *in vitro* and *in vivo* (data not shown).

The crystal structure of *Sc*Thi5p revealed a GEFG motif involved in hydrogen bonding with the phosphate group of PLP in the active site (Coquille *et al.*, 2012). This motif is strictly conserved between diverse homologs of Thi5 (Figure S2) and was modeled in *Lp*Thi5 (Figure 8B). Because of the conservation of this motif as well as residues lining the active site across Bacteria and Eukaryotes (K66, H70 in *Lp*Thi5) it is highly likely the proteins have the same mechanism. If the mechanism proposed for *Ca*Thi5 is correct, it would suggest H70 was a substrate for *Lp*Thi5. *Lp*Thi5 variants of the GEFG loop were tested for activity *in vivo*. A G121A variant failed to complement a *S. enterica thiC* mutant grown with glucose or ribose as carbon source. However, when expression of this protein was induced, the variant was not soluble, suggesting that the alanine substitution of the G121 residue impacted folding of *Lp*Thi5. Interestingly, alleles of *Lpthi5* encoding individual alanine substitutions in each position of the GEF motif supported growth of the *thiC* mutant on ribose medium (Figure 9 A-C) but eliminated complementation on glucose (Figure 9 D-F). Taken together with previous results demonstrating a reduced requirement for HMP on ribose compared to glucose (Table S1), the data suggest these variants had reduced activity *in vivo*. When overexpressed and purified, both *Lp*Thi5$_{G118A}$ and *Lp*Thi5$_{F120A}$ variants produced product that supported growth of an HMP auxotroph when incubated in buffer with iron (data not shown). These data indicated both *Lp*Thi5$_{G118A}$ and *Lp*Thi5$_{F120A}$ variants co-purified with the same molecule as wild-type *Lp*Thi5, consistent with detection of the enzyme's *in vivo* activity.

Conclusions. Prior to this work, the Thi5 pathway for the synthesis of HMP was characterized only in fungi. The data herein show *L. pneumophila* encodes a Thi5 pathway for HMP-P synthesis but lacks an efficient salvage pathway for HMP. *Lp*Thi5 appears to outperform the *S. cerevisiae* Thi5 ortholog in the heterologous metabolic network of *S. enterica*, even when the fungal locus was codon optimized for expression in this host. These differences may suggest structural variances within the PLP biosynthetic pathways of fungi vs. bacteria. When purified, *Lp*Thi5-His6 binds PLP, but the histidine residue in the

active site restricts solvent accessibility of the Schiff base. *Lp*Thi5 purifies with a molecule that when incubated with iron produces HMP. Neither the identity of this molecule, nor the formal possibility that it is a partially modified His70 residue were addressed here. We identified a conserved GEF motif that modulates function *in vivo* and corroborated the importance of other conserved amino acids predicted to be in the active site for function of this homolog both *in vitro* and *in vivo* (CCCXC, K66, H70). The results herein extend the understanding of the Thi5 pathway for HMP-P synthesis and confirm its presence beyond Eukaryotes into the Legionellaceae family of Bacteria.

EXPERIMENTAL PROCEDURES

Strains. Strains used in this study are derivatives of strain *Salmonella enterica* Serovar Typhimurium strain LT2, *Escherichia coli* K12*,* or *Legionella pneumophila* strain Lp02*.* Strains, plasmids and their source are listed in Table 1. An *lpg1565* deletion mutant in *L. pneumophila* strain Lp02, a thymidine auxotroph derived from Philadelphia-1 (Berger & Isberg, 1993), was constructed by recombineering as described previously (Bryan *et al.*, 2013). In brief, the *lpg1565* gene and its ~ 750 bp 5' and 3' flanking sequences were amplified using primers lpg1565F and lpg1565R (Table 1). The DNA product was cloned by standard methods into vector pGEM T-easy (Promega), creating pGEM-*lpg1565*. An *FRT*-flanked *cat* cassette encoding chloramphenicol resistance was amplified from pKD3 (Datsenko & Wanner, 2000) using primers lpg1565P0 and lpg1565P2 (Table 1), and the product purified. Replacement of the *lpg1565* allele with the *cat* cassette was accomplished by co-transforming *E. coli* strain DY330, which encodes λ-red recombinase, by electroporation with the purified *cat* fragment and pGEM-*lpg1565*. Candidate *E. coli* DY330 colonies harboring pGEM-*lpg1565::cat* were screened by PCR, and the corresponding recombinant plasmids then transformed into *E. coli* DH5α. Next, the recombinant allele *lpg1565::cat* was amplified by PCR using primers lpg1565F and lpg1565R (Table 1) and transferred to *L. pneumophila* strain Lp02 by natural transformation and chloramphenicol selection. Replacement of the Lp02 *lpg1565* locus with *cat* was confirmed by DNA sequencing.

Media and Chemicals. *E. coli* and *S. enterica* strains were routinely grown on Nutrient Broth (NB) containing 8 g/L Difco Nutrient broth and 5 g/L NaCl. For protein purification the cultures were grown in superbroth (SB; 32 g/L vegetable tryptone, 20 g/L yeast extract (Fisher Scientific), 5 g/L NaCl with 0.05 N NaOH). Solid media contained 1.5 % agar. Kanamycin (Kn) and Ampicillin (Ap) were added to rich media at 50, or 100 mg/L, respectively. Minimal media was No-carbon E salts (NCE) (Vogel & Bonner, 1956) with 1 mM $MgSO_4$, 0.1X trace minerals (Balch *et al.*, 1979) and 11 mM glucose, gluconate,

galactose or 13.2 mM ribose as a sole carbon source as indicated. 2-methyl-4-amino-5-hydroxymethylpyrimidine (HMP) was purchased from LabSeeker, Inc. (Wujiang City, China).

L. pneumophila strains were grown in ACES-buffered Feeley-Gorman (FG) broth (10 g/L ACES, 17.5 g/L Casein enzymatic hydrolysate, 3 g/L Beef extract, 0.4 g/L L-cysteine HCl•H_2O, 0.25 g/L ferric pyrophosphate, adjusted to pH 6.9 with KOH), and on ACES-buffered Yeast Extract plates containing 10 g/L ACES, 10 g/L yeast extract, 0.4 g/L L-cysteine HCl•H_2O, 0.25 g/L ferric pyrophosphate, adjusted to pH 6.9 with KOH, 1.7 % agar and 0.2 % activated charcoal (Feeley *et al.*, 1978, Pasculle *et al.*, 1980). Defined media for growth of *L. pneumophila* was Modified Ristroph Medium (MRM) (Ristroph *et al.*, 1981, Sauer *et al.*, 2005). 100 µg/mL thymidine was added to all media when growing thymine auxotrophs. When divalent cations were present in the incubation of *Lp*Thi5, they were added as the following salts; $MgSO_4$, $NiSO_4$, $MnCl_2$, $ZnSO_4$, $CoCl_2$, $CaCl_2$, $CdCl_2$, $CuSO_4$, $Fe(II)(NH_4)_2(SO_4)_2$, $Fe(III)_2(SO_4)_3$. Chemicals were purchased from Sigma-Aldrich, St. Louis, MO unless otherwise indicated.

Growth Analysis. Growth of bacterial strains was monitored at OD_{650} in a 96-well plate with a BioTek ELx808 plate reader. For *L. pneumophila*, strains were grown overnight in ACES-buffered FG (5 mL) to an OD_{650} = 1.5 (late-log) and pelleted (9400 x g for 5 min). The cell pellet was resuspended in an equal volume of ddH_2O and used to inoculate (5 %) MRM medium with indicated supplements. Plates were incubated at 37 °C with fast shaking, and data were plotted using Prism 7 (Graph Pad). *S. enterica* strains were grown overnight in NB Ap (2 mL) prior to pelleting and resuspension in an equal volume of 0.85 % NaCl. The cell suspension was used to inoculate (1 %) the indicated medium. Plates were incubated at 37 °C with medium shaking, and data were plotted using Prism 7 (Graph Pad).

Bioinformatics analyses. A BLAST-P search of the RefSeq Protein database (09/20/19) used the amino acid sequence encoded by *lpg1565* as a query. The default BLAST parameters were used, and the top 1000 hits were examined. Bacterial homologs with a conserved CCCXC motif were identified in the *Endozoicomonas*, *Fluorobacter* and *Legionella* genera. Based on this result, finished and permanent draft genomes on the IMG database of *Endozoicomonas*, *Fluorobacter* and *Legionella* genera of Bacteria and all available Eukaryotic genomes were queried to find *THI5* homologs (Chen *et al.*, 2019). Iterative BLAST-P searches using the default settings and the predicted amino acid sequences of *THI5* homologs from *L. pneumophila* (*lpg1565*), *Endozoicomonas elysicola* (B144DRAFT_03762) and *S. cerevisiae* (*THI5*; *YFL058W*) as the query identified 587 genomes that contained *THI5* homologs with a CCCXC motif. The amino acid sequences of the 35 non-redundant bacterial *THI5* homologs were downloaded from the IMG database in FASTA format, and the headers of the files were modified to place the organism

name first. Representative eukaryotic homologs from *S. cerevisiae* and *S. pombe* were also included in the phylogenetic analysis. Geneious Prime 2019.2.1 was used to align these 41 homologs. A MUSCLE Alignment with 100 iterations was used to generate a PHYLIP alignment file which was analyzed via Smart Model Selection PhyML to generate a phylogenetic tree using the Le-Gascuel (LG) substitution model (Edgar, Guindon *et al.*, 2010, Lefort *et al.*, 2017). The phylogenetic tree was annotated using the interactive Tree Of Life (iTOL) and Adobe Illustrator (Letunic & Bork, 2019). The presence or absence of homologs of genes encoding key enzymes involved in PLP biosynthesis and salvage (*pdxJ*, *pdxS*, and *pdxH*) were identified using the IMG Pathway Profile for the 215 non-redundant genomes with *THI5* homologs (Chen *et al.*, 2019). In cases where multiple genomes were found within the same species, only genomes from the original type strain were used to annotate putative PLP biosynthesis homologs.

Molecular Techniques. Plasmids were constructed and modified using standard molecular techniques. pTac85 (Marsh, 1986), pJB98 (Hammer & Swanson, 1999), and pET-28b(+) (Novagen) were isolated using the PureYield Plasmid MiniPrep System (Promega, Madison, WI). Q5 DNA polymerase (New England Biolabs, Ipswich, MA) was used to amplify DNA with primers synthesized by Integrated DNA Technologies, Coralville, IA or Eton Bioscience, Inc., Research Triangle Park, NC. PCR products were purified using the PCR purification kit (Qiagen, Venlo, Limburg, The Netherlands). Restriction endonucleases and ligase were purchased from New England Biolabs, Ipswich, MA.

The *THI5* homolog *lpg1565* was amplified by PCR with *L. pneumophila* gDNA as a template using primers LpTHI5 *Nco*I F and LpTHI5 *Sal*I R, listed in Table 1. The amplified product was purified, digested with *Nco*I and *Sal*I and ligated into pTac85 (Marsh, 1986), resulting in pDM1486, which was confirmed by sequencing. The *lpg1565* coding sequence was cloned with additional nucleotides encoding a C-terminal His_6 tag into the *Nco*I/*Sal*I sites of pTac85. The relevant insert was then cloned into pET-28b(+) at the *Nco*I/*Sal*I sites to generate plasmid pDM1630, which expressed *Lp*Thi5-His_6 fusion protein for purification. *ScTHI5,* with flanking *Nco*I and *Sal*I sites, was codon-optimized for translation in *E. coli*, purchased from Genscript and ligated into pTac85 to generate pDM1625. Primers for site-directed mutagenesis were designed using the Agilent QuikChange Primer Design webtool and are listed in Table 1. Variants were created following instructions from the QuikChange II Site Directed Mutagenesis kit (Agilent Technologies, Inc., Santa Clara, CA), and confirmed by sequencing (Eton Bioscience, Inc, Research Triangle Park, NC).

To construct pDM1631, the *lpg1565* locus was amplified by PCR from gDNA using primers lpg1565 comp F and lpg1565 comp R (Table 1). The fragment was purified, digested with *Bam*HI and

*Sac*I and ligated into pJB98. The *lpg1565-lpg1569* genes that constitute an operon (Sahr *et al.*, 2012) were PCR amplified with gDNA as a template, using primers ThiOperon_Fwd and ThiOperon_Rev (Table 1). The amplified product was purified, digested with *Kpn*I and *Xba*I and ligated into pJB98 to generate pDM1632. The cloned operon was modified to remove *lpg1565* using the Q5 Site-directed mutagenesis kit (New England Biolabs, Ipswitch, MA) with primers NEBQC-thiOF and NEBQC-thi5R (Eton Bioscience, Inc, Research Triangle Park, NC). This manipulation placed the starting codon for *lpg1566* where the annotated starting codon for *lpg1565* is found and generated pDM1633, which was confirmed by sequencing.

Protein Purification. A culture (100 mL) of *E. coli* BL21-AI carrying pDM1630 was grown overnight at 30 °C in NB Kn, and 4 flasks of 1.5 L of SB Kn + pyridoxine (1 mM) were inoculated (1 %). Each of four cultures were grown at 37 °C with shaking (200 rpm). When the OD_{650} reached 0.6, temperature was lowered to 22 °C and arabinose and IPTG were added to a final concentration of 0.2 % and 1 mM, respectively. Incubation continued for 19-20 hours prior to harvesting by centrifugation. Typical cell yield was 8 g/L under these conditions. 100 g of cells were resuspended to a total volume of 200 mL in Buffer A [50 mM HEPES (Fisher Scientific), 300 mM NaCl, 20 mM Imidazole (Fisher Scientific), 1 mM TCEP (Gold Biotechnology), pH 7.5 at 4 °C] with DNAse (0.025 mg/mL), lysozyme (1 mg/mL) and phenylmethylsulfonyl fluoride (0.1 mg/mL) and incubated on ice for one hour. The cell suspension was lysed at 20 kpsi using a Constant Systems Limited One Shot (United Kingdom), and cell lysate was cleared at 48,000 x g (50 min, 4 °C). The cell-free extract was passed through a 0.45 µM PVDF filter (Millipore) and injected onto two pre-equilibrated 5 mL HisTrap HP Ni-Sepharose columns connected in sequence. The protein was washed with 5 column volumes of Buffer A, 5 column volumes of 4 % Buffer B (Buffer A + 480 mM Imidazole (Fisher Scientific), pH 7.5 at 4 °C) and finally eluted from the column with a gradient of Buffer B from 4 % to 100 % over 10 column volumes. Fractions containing *Lp*Thi5 as determined by SDS-PAGE were combined and concentrated by centrifugation using a 10 kDa filter (Millipore), exchanged into 50 mM HEPES buffer with 10 % glycerol, 1 mM TCEP, pH 7.5, using a PD10 column following the manufacturer's instructions (GE Healthcare), flash-frozen in liquid nitrogen and stored at -80 °C until use. Protein concentration was determined by extinction coefficient using the theoretical molecular weight and A_{280} extinction coefficient of *Lp*Thi5-His_6 as determined by the ExPASy Protparam database ($\varepsilon_{280} = 30870\ M^{-1}\ cm^{-1}$) (Gasteiger *et al.*, 2003). A typical purification yielded *Lp*Thi5-His_6 that was > 85 % pure as determined by densitometry (Figure S1).

Determination of PLP and iron content. The PLP was released from *Lp*Thi5 by denaturing the protein. Fifteen nmol of *Lp*Thi5 at a purity of 95 % was treated in 50 mM HEPES, pH 7.5 with 0.1 M NaOH in a total volume of 150 µL, the protein removed using a Nanosep 10 kDa spin-filter (PALL), and the absorbance of the supernatant measured at 392 nm. PLP was quantified using the empirically determined extinction coefficient of PLP (ε_{392}) of 1.965 x 10^6 M^{-1} cm^{-1} in 50 mM HEPES pH 7.5. Occupancy was calculated based on the percent purity from the starting protein preparation.

Iron associated with *Lp*Thi5 was determined by Ferene assay (Kennedy *et al.*, 1984, Palmer & Downs, 2013). Briefly, purified *Lp*Thi5 was concentrated using an Amicon Ultra 0.5 mL 30 kDa desalting filter at 14,000 x g for five minutes and resuspending the concentrated protein with 400 µL of 50 mM HEPES, pH 7.5 treated with 2 % Chelex. After three cycles of concentration and dilution with the Chelex-treated HEPES, *Lp*Thi5 concentration was determined by extinction coefficient, ε_{280} = 30870 M^{-1} cm^{-1}. HCl (0.06 N, 50 µL) was added to 19 nmol *Lp*Thi5 in 25 mM HEPES (25 µL) and incubated at 80 °C for 10 minutes. The following were added sequentially with mixing by vortex between additions: 0.96 M ammonium acetate (125 µL), 0.2 M ascorbic acid (25 µL), 87 mM SDS (25 µL), and 30 mM Ferene (25 µL). Precipitated protein was removed by centrifugation at 9000 x g for 5 minutes and absorbance of the supernatant was measured at 593 nm. Iron content was determined based on the extinction coefficient of Ferene complexed with Fe^{2+} (ε_{593} = 35.5 x 10^3 M^{-1} cm^{-1}).

Size exclusion chromatography. The quaternary structure of purified *Lp*Thi5 (which released bioactive molecule) was assessed by size exclusion chromatography (Hong *et al.*, 2012). A BioRad NGC Chromatography System with an SEC 650 column (BioRad) was equilibrated with 2 column volumes of HEPES (50 mM), TCEP (1 mM) pH 7.5 at 4 °C. Absorbance at 280 nm was monitored to detect when the protein was eluting. Injection of blue dextran (50 µL, 3 mg/mL) in duplicate with a flow rate of 1 mL/min determined the void volume (V_0) was 8.9 mL. The elution volumes (V_e) of duplicate samples of BioRad Gel Filtration Standards (50 µL) were used to establish a standard curve of K_{av} ($\frac{V_e - V_0}{V_{column} - V_0}$) vs $\log_{10}$ (molecular weight) (Figure 5). Three samples, each containing 17 nmol *Lp*Thi5, were injected at a flow rate of 1 mL/min, and the elution volume was used to determine the K_{av}. The molecular size of *Lp*Thi5 was determined by interpolation from the standard curve ($K_{AV} = -0.231\log(MW) + 0.732$; $R^2 = 0.992$), and standard deviation was determined from the deviation in retention time.

Detection of HMP. *Bioassay.* Purified *Lp*Thi5 (20 nmol) in 200 µL HEPES (50 mM, pH 7.5) was incubated at 37 °C. At T=0, 4, 16 hr or indicated time, the sample was shifted to 95 °C for 5 min, and

denatured protein was pelleted by centrifugation (17,000 x g, 5 minutes). When indicated, ferrous ammonium sulfate hexahydrate was added to 200 μM in the protein sample. The supernatant was evaluated for its ability to stimulate growth of a *thiC* or *thiI* mutant strain of *S. enterica* on minimal medium. Samples (5 μL) were spotted on soft agar embedded with the relevant mutant strain overlaid on minimal medium. Growth was assessed after ~16 hr incubation at 37 °C. A *thiI* mutant responds to thiamine, or TPP, while a *thiC* mutant responds to thiamine, TPP, HMP, HMP-P.

High Performance Liquid Chromatography (HPLC) and MS. Supernatants judged active and inactive by bioassay were subjected to HPLC, monitoring absorbance at 270 nm, which is a lambda max for HMP. Initial control experiments showed that a peak (270 nm) correlated with a bioactive molecule released from an overnight incubation of 3 mg *Lp*Thi5. After a 16 hr incubation of 3 mg *Lp*Thi5 at 37 °C, supernatant (2 mL) resulting from heat treatment was applied to a C18 Sep-Pak Plus cartridge (Waters) for concentration prior to HPLC analysis. Six cartridges were conditioned using 5 mL acetonitrile (ACN) followed by 10 mL 50 mM HEPES, pH 7.5. Five 1 mL fractions were eluted with acetonitrile and collected, evaporated, and tested for biological activity by bioassay. Biologically active fractions (Fractions 1 and 2 from each column) were combined, evaporated, suspended to a total volume of 0.8 mL in water, and 100 μL fractions were sequentially injected onto the HPLC. A 3 mg sample of *Lp*Thi5 (T=0) was treated in parallel. Fractions collected from the C18 Sep-Pak Plus cartridges were dehydrated, resuspended in 100 μL water and injected onto the HPLC for analysis.

Concentrated samples were separated by reversed-phase HPLC using a Shimadzu LC20-AT delivery system equipped with a 250 x 4.6 mm Luna C_{18} column (Phenomenex). The UV-Vis spectrum was monitored over time using a photodiode array detector (Shimadzu). Samples were eluted with a flow rate of 1 mL/min with a gradient of water and ACN; 100 % water for 10 minutes, a gradient from 0 – 15 % ACN over 10 minutes, and 15 % ACN for 10 minutes. Fractions (0.5 mL) collected over 30 min were dehydrated and resuspended in 50 μL water. Fractions that contained biological activity were determined by bioassay (spotting 1 μL onto a soft-agar overlay containing a *thiC* strain as an indicator and observing growth). The relevant fractions, eluting from 19.5 – 22 min, were pooled and analyzed by MALDI-TOF MS. Fractions over the same retention time from control samples, which had no activity in the bioassay, were independently pooled. Concentrated fractions and an HMP-P standard were analyzed by MALDI-TOF MS using a Bruker Autoflex (TOF) mass spectrometer at the Proteomics and Mass Spectrometry facility at UGA. Fractions containing biological activity from 25 mg of *Lp*Thi5 that had been incubated for 16 hours were combined from several HPLC runs to facilitate identification by MALDI-TOF MS.

ACKNOWLEDGEMENTS

We thank Thomas Knight for the construction of the *lpg1565* mutant and Vincent Starai for the gift of *Legionella pneumophila* genomic DNA. This work was supported by an award from the competitive grants program at the NIH (GM095837) to DMD and a Graduate Research Fellowship Grant (DGE-1443117) from the NSF to MDP. The authors state that they have no conflicts of interest in presenting this work.

AUTHOR CONTRIBUTIONS

Authors contributed to; (i) the conception or design of the study (MDP, DMD), (ii) the acquisition, analysis, or interpretation of the data (MDP, MSS, DMD) and (iii) writing of the manuscript (MDP. DMD).

DATA AVAILABILITY

The data that supports the findings of this study are available in the supplementary material of this article, and additional information is available upon request from the authors.

REFERENCES

Backstrom, A.D., McMordie, R.A.S., and Begley, T.P. (1995) Biosynthesis of Thiamin I: The Function of the thiE Gene Product. *J American Chem Soc* **117**: 2351-2352.

Balch, W.E., Fox, G.E., Magrum, L.J., Woese, C.R., and Wolfe, R.S. (1979) Methanogens: reevaluation of a unique biological group. *Microbiol Rev* **43**: 260-296.

Bale, S., Rajashankar, K.R., Perry, K., Begley, T.P., and Ealick, S.E. (2010) HMP binding protein ThiY and HMP-P synthase THI5 are structural homologues. *Biochemistry* **49**: 8929-8936.

Berger, K.H., and Isberg, R.R. (1993) Two distinct defects in intracellular growth complemented by a single genetic locus in *Legionella pneumophila*. *Mol Microbiol* **7**: 7-19.

Bryan, A., Abbott, Z.D., and Swanson, M.S. (2013) Constructing unmarked gene deletions in *Legionella pneumophila*. *Methods Mol Biol* **954**: 197-212.

Chen, I.A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J.R., Seshadri, R., Smirnova, T., Kirton, E., Jungbluth, S.P., Woyke, T., Eloe-Fadrosh, E.A., Ivanova, N.N., and Kyrpides, N.C. (2019) IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res* **47**: D666-D677.

Coquille, S., Roux, C., Fitzpatrick, T.B., and Thore, S. (2012) The last piece in the vitamin B1 biosynthesis puzzle: structural and functional insight into yeast 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate (HMP-P) synthase. *J Biol Chem* **287**: 42333-42343.

Datsenko, K.A., and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**: 6640-6645.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.

Feeley, J.C., Gorman, G.W., Weaver, R.E., Mackel, D.C., and Smith, H.W. (1978) Primary isolation media for Legionnaires disease bacterium. *J Clin Microbiol* **8**: 320-325.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**: 3784-3788.

Geders, T.W., Gustafson, K., and Finzel, B.C. (2012) Use of differential scanning fluorimetry to optimize the purification and crystallization of PLP-dependent enzymes. *Acta crystallographica. Section F, Structural biology and crystallization communications* **68**: 596-600.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307-321.

Hammer, B.K., and Swanson, M.S. (1999) Co-ordination of *Legionella pneumophila* virulence with entry into stationary phase by ppGpp. *Mol Microbiol* **33**: 721-731.

Hong, P., Koza, S., and Bouvier, E.S. (2012) Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *J Liq Chromatogr Relat Technol* **35**: 2923-2950.

Ishida, S., Tazuya-Murayama, K., Kijima, Y., and Yamada, K. (2008) The direct precursor of the pyrimidine moiety of thiamin is not urocanic acid but histidine in *Saccharomyces cerevisiae*. *J Nutr Sci Vitaminol (Tokyo)* **54**: 7-10.

Jurgenson, C.T., Begley, T.P., and Ealick, S.E. (2009) The structural and biochemical foundations of thiamin biosynthesis. *Annu Rev Biochem* **78**: 569-603.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**: 845-858.

Kennedy, M.C., Kent, T.A., Emptage, M., Merkle, H., Beinert, H., and Munck, E. (1984) Evidence for the formation of a linear [3Fe-4S] cluster in partially unfolded aconitase. *J Biol Chem* **259**: 14463-14471.

Lai, R.Y., Huang, S., Fenwick, M.K., Hazra, A., Zhang, Y., Rajashankar, K., Philmus, B., Kinsland, C., Sanders, J.M., Ealick, S.E., and Begley, T.P. (2012) Thiamin pyrimidine biosynthesis in Candida albicans : a remarkable reaction between histidine and pyridoxal phosphate. *J Am Chem Soc* **134**: 9157-9159.

Lefort, V., Longueville, J.E., and Gascuel, O. (2017) SMS: Smart Model Selection in PhyML. *Mol Biol Evol* **34**: 2422-2424.

Letunic, I., and Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**: W256-W259.

Marsh, P. (1986) Ptac-85, an *E. coli* vector for expression of non-fusion proteins. *Nucleic Acids Res* **14**: 3603.

Maundrell, K. (1990) nmt1 of fission yeast. A highly transcribed gene completely repressed by thiamine. *J Biol Chem* **265**: 10857-10864.

Mozzarelli, A., and Bettati, S. (2006) Exploring the pyridoxal 5′-phosphate-dependent enzymes. *The Chemical Record* **6**: 275-287.

Palmer, L.D., and Downs, D.M. (2013) The thiamine biosynthetic enzyme ThiC catalyzes multiple turnovers and is inhibited by S-adenosylmethionine (AdoMet) metabolites. *J Biol Chem* **288**: 30693-30699.

Palmer, L.D., Paxhia, M.D., and Downs, D.M. (2015) Induction of the Sugar-Phosphate Stress Response Allows Saccharomyces cerevisiae 2-Methyl-4-Amino-5-Hydroxymethylpyrimidine Phosphate Synthase To Function in *Salmonella enterica*. *J Bacteriol* **197**: 3554-3562.

Pasculle, A.W., Feeley, J.C., Gibson, R.J., Cordes, L.G., Myerowitz, R.L., Patton, C.M., Gorman, G.W., Carmack, C.L., Ezzell, J.W., and Dowling, J.N. (1980) Pittsburgh pneumonia agent: direct isolation from human lung tissue. *J Infect Dis* **141**: 727-732.

581 Paxhia, M.D., and Downs, D.M. (2019) SNZ3 Encodes a PLP Synthase Involved in Thiamine Synthesis
582 in *Saccharomyces cerevisiae*. *G3 (Bethesda)* **9**: 335-344.
583 Ristroph, J.D., Hedlund, K.W., and Gowda, S. (1981) Chemically defined medium for *Legionella*
584 *pneumophila* growth. *J Clin Microbiol* **13**: 115-119.
585 Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., and Gelfand, M.S. (2002) Comparative genomics of
586 thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. *J Biol Chem* **277**:
587 48949-48959.
588 Sahr, T., Rusniok, C., Dervins-Ravault, D., Sismeiro, O., Coppee, J.Y., and Buchrieser, C. (2012) Deep
589 sequencing defines the transcriptional map of *L. pneumophila* and identifies growth phase-
590 dependent regulated ncRNAs implicated in virulence. *RNA Biol* **9**: 503-519.
591 Sahr, T., Rusniok, C., Impens, F., Oliva, G., Sismeiro, O., Coppee, J.Y., and Buchrieser, C. (2017) The
592 *Legionella pneumophila* genome evolved to accommodate multiple regulatory mechanisms
593 controlled by the CsrA-system. *PLoS Genet* **13**: e1006629.
594 Sauer, J.D., Bachman, M.A., and Swanson, M.S. (2005) The phagosomal transporter A couples
595 threonine acquisition to differentiation and replication of *Legionella pneumophila* in
596 macrophages. *Proc Natl Acad Sci U S A* **102**: 9924-9929.
597 Schweingruber, A.M., Dlugonski, J., Edenharter, E., and Schweingruber, M.E. (1991) Thiamine in
598 *Schizosaccharomyces pombe*: dephosphorylation, intracellular pool, biosynthesis and transport.
599 *Curr Genet* **19**: 249-254.
600 Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H.,
601 Remmert, M., Soding, J., Thompson, J.D., and Higgins, D.G. (2011) Fast, scalable generation of
602 high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.
603 Soniya, K., and Chandra, A. (2018) Free energy landscapes of prototropic tautomerism in pyridoxal 5'-
604 phosphate schiff bases at the active site of an enzyme in aqueous medium. *J Comput Chem* **39**:
605 1629-1638.
606 Tazuya, K., Azumi, C., Yamada, K., and Kumaoka, H. (1995) Pyrimidine moiety of thiamin is
607 biosynthesized from pyridoxine and histidine in *Saccharomyces cerevisiae*. *Biochem Mol Biol*
608 *Int* **36**: 883-888.
609 Tazuya, K., Yamada, K., and Kumaoka, H. (1989) Incorporation of histidine into the pyrimidine moiety
610 of thiamin in *Saccharomyces cerevisiae*. *Biochim Biophys Acta* **990**: 73-79.

611 Thamm, A.M., Li, G., Taja-Moreno, M., Gerdes, S.Y., de Crécy-Lagard, V., Bruner, S.D., and Hanson,
612 A.D. (2017) A strictly monofunctional bacterial hydroxymethylpyrimidine phosphate kinase
613 precludes damaging errors in thiamin biosynthesis. *Biochem J* **474**: 2887-2895.
614 Toney, M.D. (2011) Controlling reaction specificity in pyridoxal phosphate enzymes. *Biochim Biophys*
615 *Acta* **1814**: 1407-1418.
616 Vogel, H.J., and Bonner, D.M. (1956) Acetylornithase of *Escherichia coli*: partial purification and some
617 properties. *J. of Biol Chem* **218**: 97-106.
618 Wightman, R., and Meacock, P.A. (2003) The THI5 gene family of *Saccharomyces cerevisiae*:
619 distribution of homologues among the hemiascomycetes and functional redundancy in the
620 aerobic biosynthesis of thiamin from pyridoxine. *Microbiology* **149**: 1447-1460.

621

622 **Table 1 – Plasmids, Bacterial strains and primers used in this study**

Plasmid Name	**Description**
pDM1486	pTac85-*lpg1565*
pDM1625	pTac85-*ScTHI5* (codon-optimized)
pDM1630	pET28b-lpg1565-His_6
pJB98	Amp^R Thy^+ (Hammer & Swanson, 1999)
pDM1631	pJB98-*lpg1565*
pDM1632	pJB98-*lpg1565-69*
pDM1633	pJB98-*lpg1566-69*
Strain	**Genotype**
S. enterica	
DM15269	*ΔthiC1225 ΔaraCBAD* / pDM1486
DM16449	*ΔthiC1225 ΔaraCBAD* / pDM1625
L. pneumophila	
DMLp6	*thyA rpsL hsdR lpg1565::Cm* / pJB98
DMLp7	*thyA rpsL hsdR lpg1565::Cm* / pDM1632
DMLp8	*thyA rpsL hsdR lpg1565::Cm* / pDM1633

Primer Name	**Sequence**
LpTHI5 NcoI F	TAGGCCATGGCGATGTCATCACTAAAATCC
LpTHI5 SalI R	TAGGGTCGACTTAATTTTCAAGACAACAGGCG
LpTHI5-HisTag SalI R	TAGGGTCGACTTAGTGATGGTGATGGTGATGATTTTCAAGACAACAGGCG
LpTHI5 K66A F	tgcaacagtatgaatcattgccgcaacaccgaaatctactgttcccagg
LpTHI5 H70A	ttttgctttggctgcaacagtagcaatcattgctttaacaccgaaatc
LpTHI5 C191A	cgatcaattagctggtttaggcgcttgttgtttctgctcaatacaa
LpTHI5 C192A	cgatcaattagctggtttaggctgtgcttgtttctgctcaatacaatttatt
LpTHI5 C193A	tcaattagctggtttaggctgttgtgctttctgctcaatacaatttattgtt
LpTHI5 C195A	ctggtttaggctgttgttgtttcgcctcaatacaatttattgttcctga
LpTHI5 G118A	ggaaacgagtaggttatatcgccgaattcggcaaaaaaattat
LpTHI5 E119A	gaaacgagtaggttatatcggcgcattcggcaaaaaaattattgatg
LpTHI5 F120A	aataatttttttgccggcttcgccgatataacctactcgtttcccaaca
LpTHI5 G121A	gtaggttatatcggcgaattcgccaaaaaaattattgatgatttgg
LpTHI5 C191S	cgatcaattagctggtttaggcagctgttgtttctgctcaatacaa
LpTHI5 C192S	cgatcaattagctggtttaggctgtagctgtttctgctcaatacaatttatt
LpTHI5 C193S	tcaattagctggtttaggctgttgtagcttctgctcaatacaatttattgtt
LpTHI5 C195S	ctggtttaggctgttgttgtttcagctcaatacaatttattgttcctga
LpTHI5 C191H	cgatcaattagctggtttaggccattgttgtttctgctcaatacaa
LpTHI5 C192H	cgatcaattagctggtttaggctgtcattgtttctgctcaatacaatttatt
LpTHI5 C193H	tcaattagctggtttaggctgttgtcatttctgctcaatacaatttattgtt
LpTHI5 C195H	ctggtttaggctgttgttgtttccattcaatacaatttattgttcctga
lpg1565F	AAATTAAGCGGGAATCGAAGTGTAGC
lpg1565R	AAATATGAGGTAAAAATTTCCAGGTCT
lpg1565P0	TTTAAAATAATAACATAAGGAGTTATGGCGATGTCATGTGTAGGCTGGAGCTGCTTC
lpg1565P2	AATGCCTGCTGCCATAATTAATTTTCAAGACAACAGCATATGAATATCCTCCTTAGTTCC

lpg1565 comp F (BamHI)	AAAGGATCCCCTCCATATATCCAATCTCGCAAG
lpg1565 comp R (SacI)	AAAGAGCTCCCGCATTATGTAATACAAAAGCCA
thiOperon_fwd	attcttcgagctcggtacccCCTCCATATATCCAATCTC
thiOperon_rev	gtcgactctagaggatccccTATCTGCTAAATAATATTGCCG
NEBQC-thiOF	ATGCGAGCAGGCATTGTA
NEBQC-thi5R	AACTCCTTATGTTATTATTTTAAATATTATGGAG

* Underlining designates nucleotides encoding six additional histidine residues at the c-terminus

FIGURE LEGENDS

Figure 1 – Pathways for HMP-P, PLP biosynthesis. A) Two biosynthetic routes for the synthesis of thiamine pyrophosphate (TPP) are schematically represented with a focus on the source of HMP-P. Gene products are indiecatead by the reactions they catalyze. Those depicted in blue are present in *S. enterica* and not *L. pnuemophila.* Those in red, are found in *L. pnuemophila* and other organisms as described in the text. In *S. enterica,* ThiC converts 5'-aminoimidizole ribotide (AIR) into HMP-P, while Thi5 synthesizes HMP-P from histidine and a B_6 vitamer. Incorporation of atoms derived from histidine and the B_6 vitamer into HMP-P are shown (Ishida *et al.*, 2008, Tazuya *et al.*, 1995, Tazuya *et al.*, 1989). B) Two pathways for PLP biosynthesis are schematically represented with emphasis on the final steps. Enzymes belonging to the DXP-dependent pathway are highlighted in red, and exemplified by *S. enterica*. PdxH is present in many organisms as part of a salvage pathway and its presence does not indicate the DXP-dependent pathway is present. The enzymes of the DXP-independent pathway are highlighted in green and exemplified by *S. cerevisiae* (Snz/Sno) and *B. subtilis* (PdxS/PdxT). C) Organization of the thiamine biosynthetic locus in *L. pneumophila*. The putative annotations and locus tags for coding regions homologous with other thiamine biosynthetic enzymes are shown. The TPP riboswitch in the 5'UTR of *lpg1565* is indicated (Sahr *et al.*, 2017). Abbreviations: 3-amino-1-hydroxyacetone phosphate, AHAP; deoxyxylulose-5-phosphate, DXP; glutamine, Gln; ribose-5'-phosphate, R5P; glyceraldehyde-3-phosphate, G3P; pyridoxine-5'-phosphate, PNP; pyridoxal-5'-phosphate, PLP; 5'-aminoimidizole ribotide, AIR; 2-methyl-4-amino-5-hydroxymethylpyrimidine, HMP; 4-amino-5-

hydroxymethylpyrimidine phosphate, HMP-P; 4-amino-5-hydroxymethylpyrimidine diphosphate, HMP-PP; 4-methyl-5-β-hydroxyethylthiazole, THZ; 4-methyl-5-β-hydroxyethylthiazole phosphate, THZ-P; thiamine, Thi; thiamine diphosphate, TPP; thiamine phosphate, TMP.

Figure 2 –Phylogeny of Thi5 homologs and associated PLP biosynthetic pathways. The phylogenetic relationship of Thi5p homologs among Bacteria (blue lines) and select Eukaryotes (red lines) in the IMG database containing the CCCXC motif. Annotated KEGG functions for *pdxJ* (K03474; E.C. 2.6.99.2) and *pdxS* (K06215; E.C. 4.3.3.6) in each genome are highlighted. Organisms in which mutants lacking the homolog are demonstrable HMP auxotrophs are highlighted in bold. Organisms which contain multiple Thi5 homologs are annotated with an asterisk. The scale represents percent genetic distance between homologs.

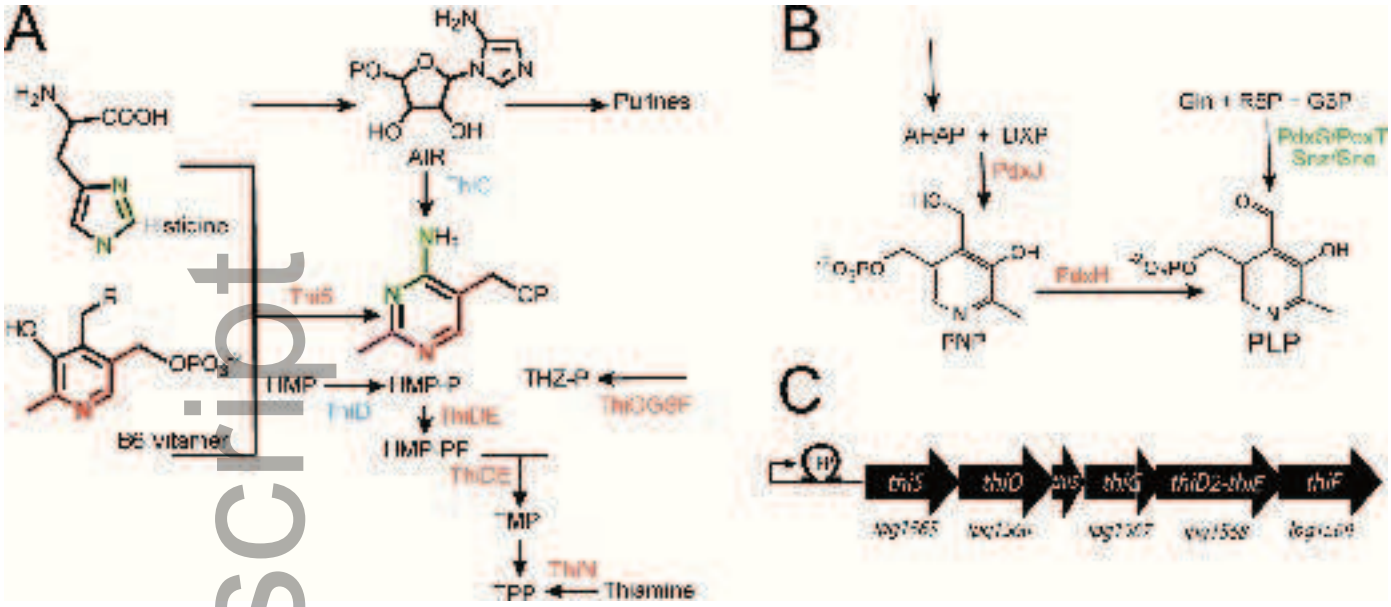
Figure 3 – *lpg1565* contributes to HMP synthesis in *L. pneumophila.* A mutant of *L. pneumophila* with an insertion-deletion of *lpg1565* containing (A) pJB98 (vector only), (B) pDM1632 (pJB98-*lpg1565-9*), or (C) pDM1633 (pJB98-*lpg1566-9*) was grown on MRM supplemented with i) no additions, i) HMP (2.5 mM) or iii) thiamine (100 nM) as indicated. Error bars indicate the standard deviation of three independent biological replicates.

Figure 4 – *Lp*Thi5 can generate HMP in *S. enterica.* Strains of *S. enterica* lacking *thiC* and containing either *Lpthi5* (Panel I) or a codon-optimized version of *ScTHI5* (Panel II) were grown on a minimal medium containing either 11 mM glucose (A, E), 13.2 mM ribose (B, F), 11 mM galactose (C, G) or 11mM gluconate (D, H). Medium was otherwise supplemented with nothing (open circles), 100 μM IPTG (solid squares), 100 μM IPTG and 100 nM THZ (solid inverted triangles) or 100 nM thiamine (solid circles). Growth was measured by optical density at 650 nm over time. Error bars indicate the standard deviation of three independent biological replicates.

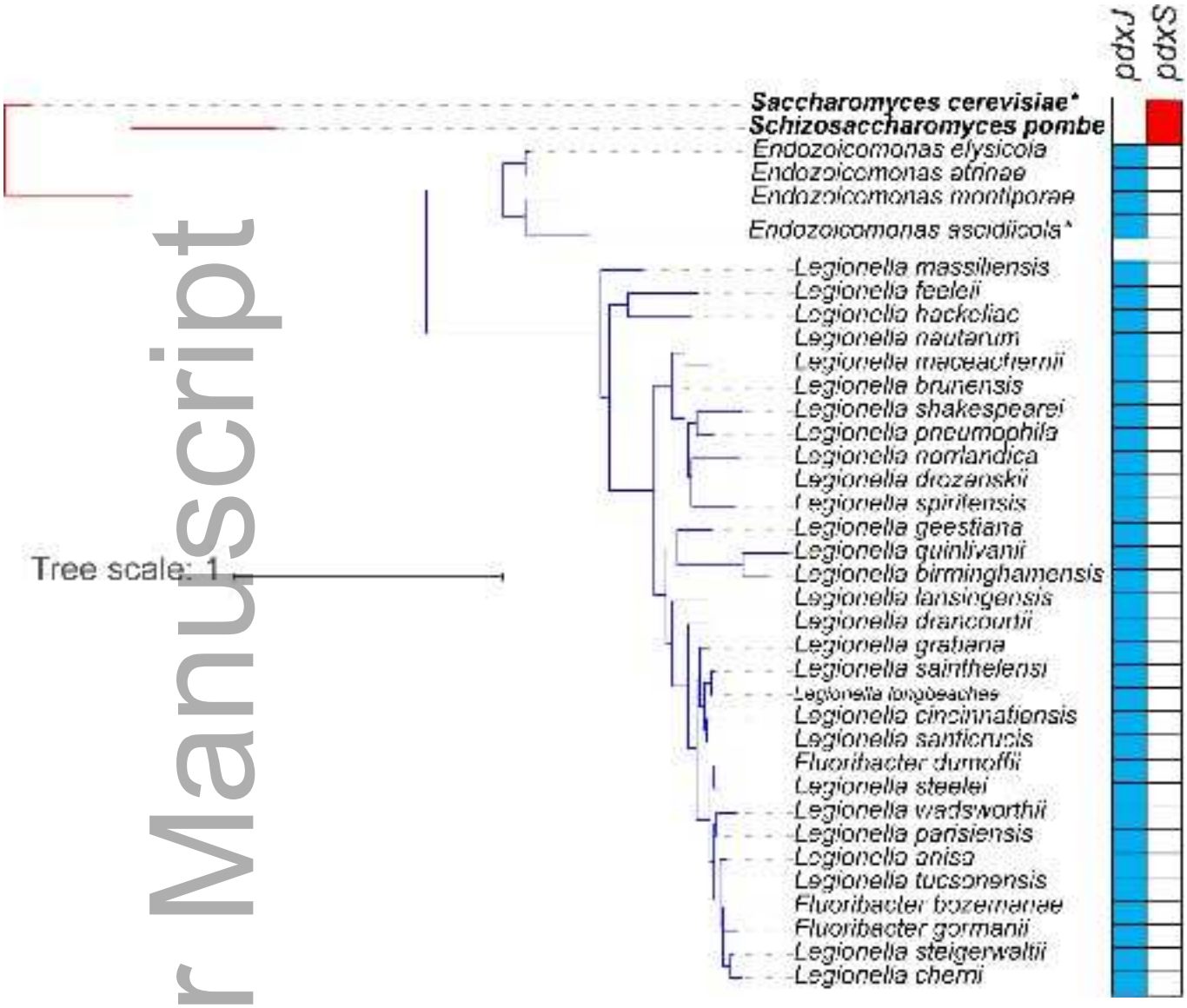
Figure 5 – *Lp*Thi5 is dimeric in solution The oligomeric state of *Lp*Thi5 in solution was determined by size exclusion chromatography. An SEC 650 column was equilibrated with 50 mM HEPES + 1 mM TCEP, pH 7.5. Molecular weight standards or *Lp*Thi5 (50 μL) were injected and eluted at a flow rate of 1mL/min. A) Aliquots (50 μL) of thyroglobuline (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), and myoglobin (17 kDa) were used as size standards to generate the standard curve and interpolate the size of *Lp*Thi5. The position of *Lp*Thi5 is represented by a square. B) Absorbance (280 nm) was followed over 20 mL of elution buffer after *Lp*Thi5 was injected on the column. The lone peak eluted at 13 mL.

Figure 6 – *Lp*Thi5 purifies with pyridoxal-5'-phosphate. Recombinant *Lp*Thi5-His_6 was purified, a portion of the protein preparation was dialyzed against HEPES (50 mM, pH 7.5) containing PLP (200

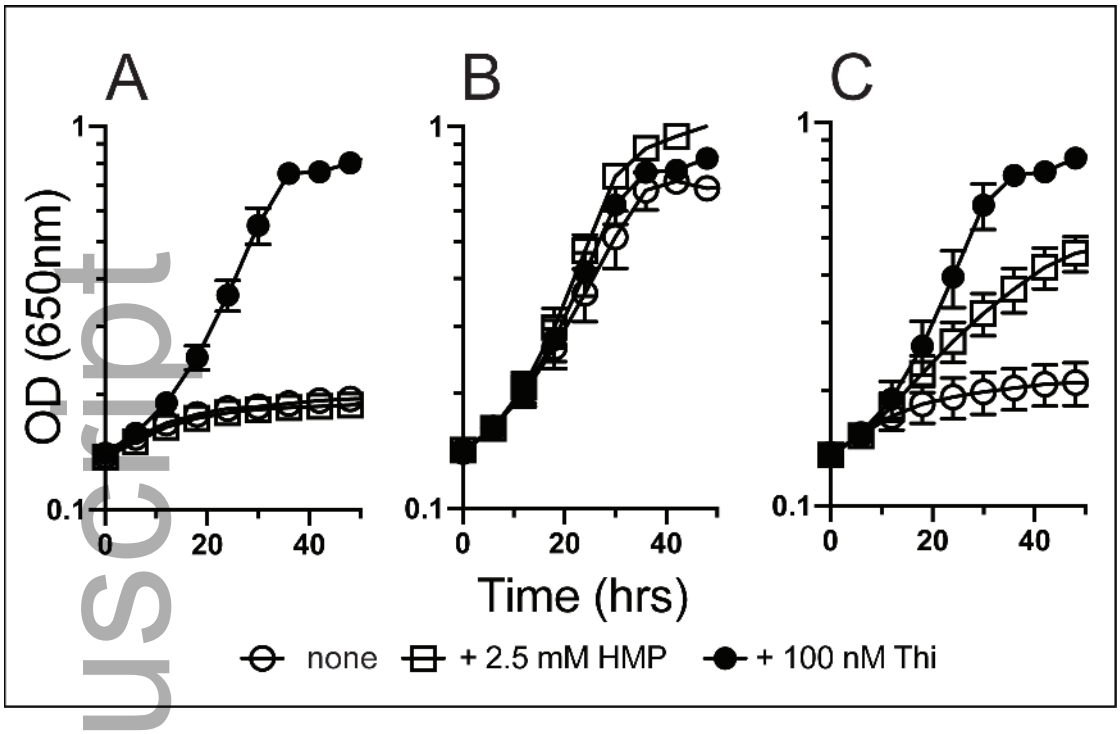
676 μM) and TCEP (1 mM) at 4 °C overnight. Unbound PLP was removed using a PD-10 desalting column,
677 and both preparations were characterized by UV-visible spectroscopy. Three spectra were taken of 100
678 μM *Lp*Thi5 (A), *Lp*Thi5 + PLP (B) and *Lp*Thi5$_{H70A}$ + PLP (C), untreated (solid line), treated with 25 mM
679 $NaBH_4$ (dotted line), treated with 0.1 M NaOH (dashed line).
680 **Figure 7 – Purified *Lp*Thi5 releases HMP** (**A**) *Lp*Thi5 protein in buffer was denatured with heat before
681 incubation (T=0), and after overnight incubation (O/N). In each case the supernatent was spotted on a
682 lawn of a *thiC* mutant of *S. enterica* embedded in soft agar is shown. HMP was spotted as a control.
683 Growth was detected as turbidity after 16 hr incubation at 37 °C. (**B**) *Lp*Thi5 (0.75 mg) was incubated
684 overnight in buffer (black line) or not (hashed line) before denaturing protein at 95 °C for 5 minutes. The
685 supernatents were injected onto an HPLC and absorbance was monitored over time at 270 nm. The star
686 indicates the peak that contained activity in the overnight sample, and the inset shows the UV-Vis
687 spectrum of this peak. (**C**) MALDI-TOF of authentic HMP-P, the biologically active and the inactive
688 fractions from *Lp*Thi5 samples described are shown. (**D**) Impact of iron in the buffer bioassay for the
689 production of HMP(-P) after a four hour incubation with iron.
690 **Figure 8 – *Lp*Thi5 and *Sc*Thi5p share structural motifs in the active site.** (**A**) Amino acid sequences
691 of *Lp*Thi5 and *Sc*Thi5p were aligned by CLUSTAL Omega (Sievers *et al.*, 2011), similar residues are
692 highlighted in grey and identical residues are highlighted in black. Conserved residues interrogated in this
693 study are boxed. (**B**) A structural homology model was generated with Phyre2 (Kelley *et al.*, 2015), and
694 residues in the CCCXC motif or those predicted to coordinate PLP in the active site are modeled as sticks
695 (K66, H70, G118, E119, F120) and visualized with PyMOL.
696 **Figure 9 – *Lpthi5* variants have reduced ability to complement *thiC* mutant on glucose.** Strains of *S.*
697 *enterica* lacking *thiC* and containing a plasmid expressing *Lpthi5* alleles encoding *Lp*Thi5$_{G118A}$ (A,D),
698 *Lp*Thi5$_{E119A}$ (B,E), or *Lp*Thi5$_{F120A}$ (C,F) under the regulation of P_{tac} were grown on a minimal medium.
699 Panels A-C contained ribose (13.2 mM) as a carbon source, while panels D-F contained glucose (11 mM)
700 as a carbon source. In each case the medium included no additions (open circles), 100 μM IPTG (solid
701 squares) or 100 nM thiamine (solid circles). Growth was determined by following optical density at 650
702 nm over time. Error bars indicate the standard deviation of three independent biological replicates



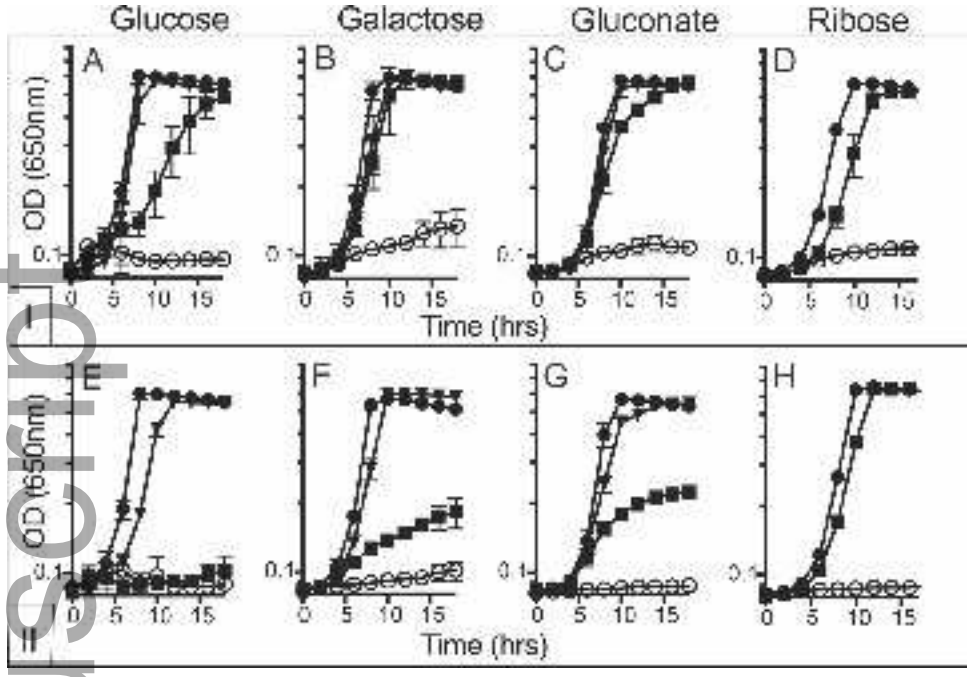
mmi_14622_f1.tif



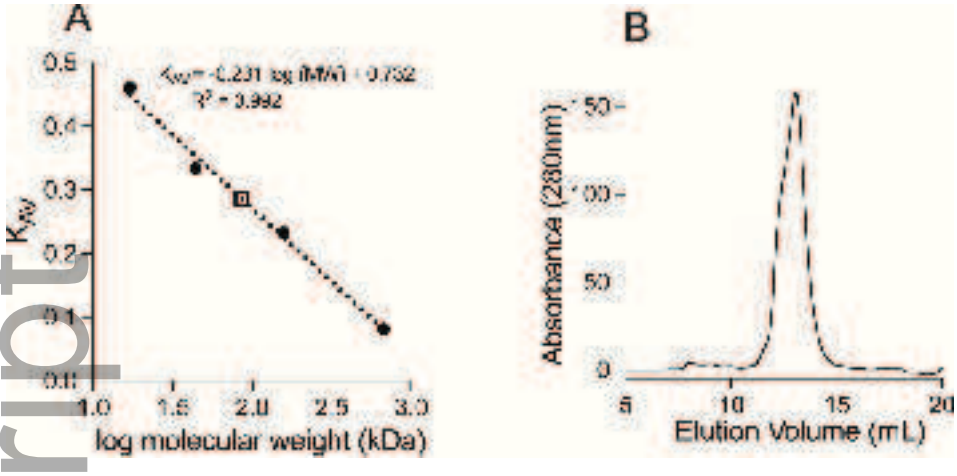
mmi_14622_f2.tif



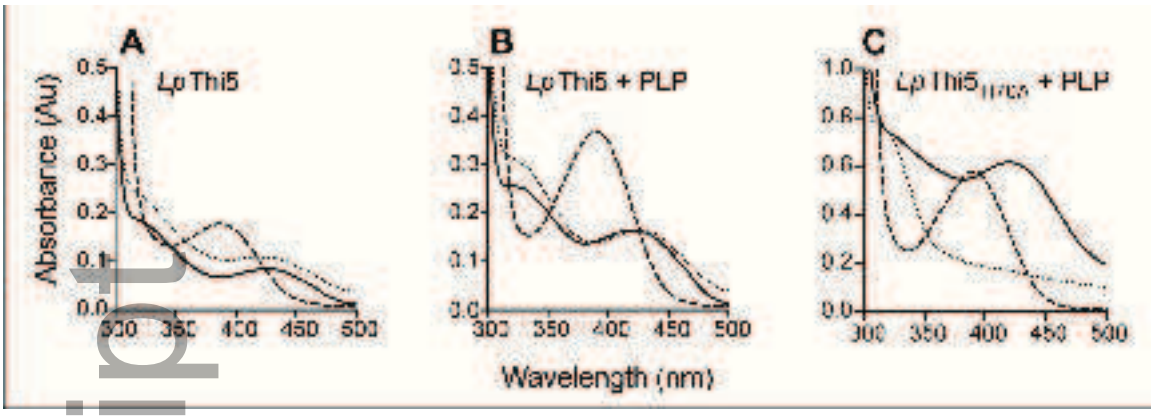
mmi_14622_f3.tif



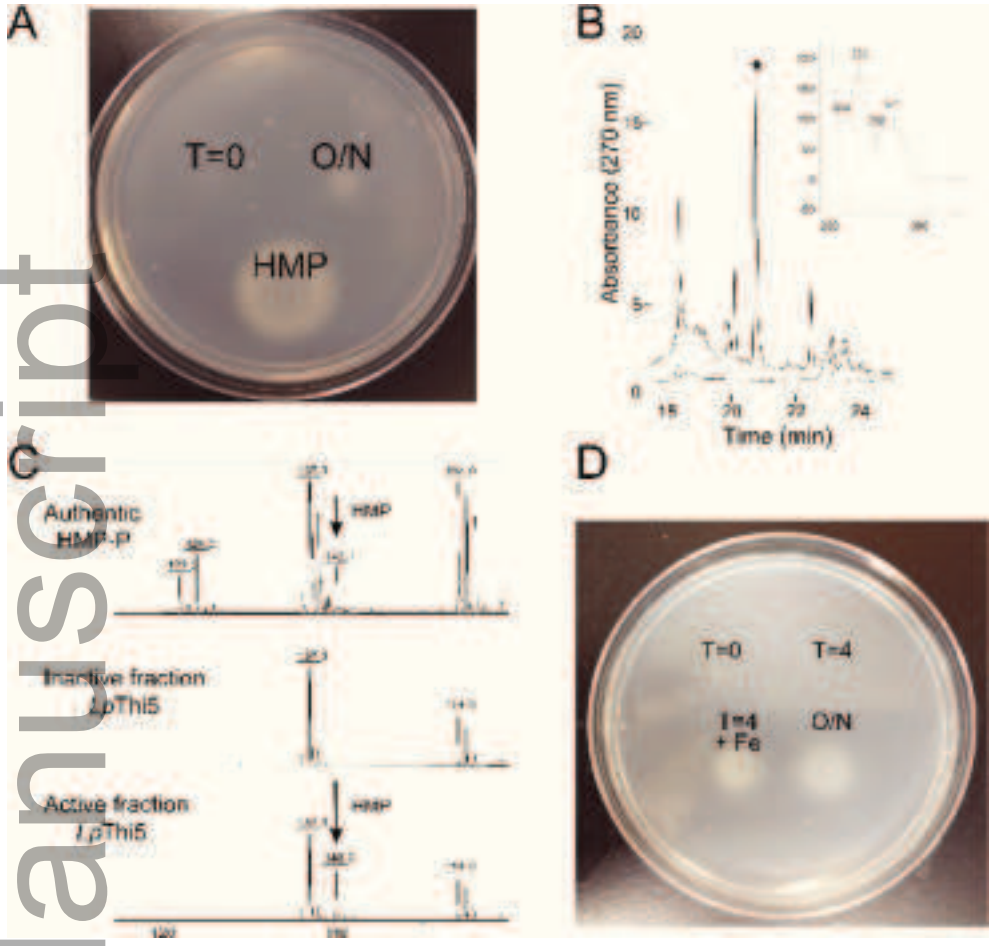
mmi_14622_f4.tif



mmi_14622_f5.tif



mmi_14622_f6.tif

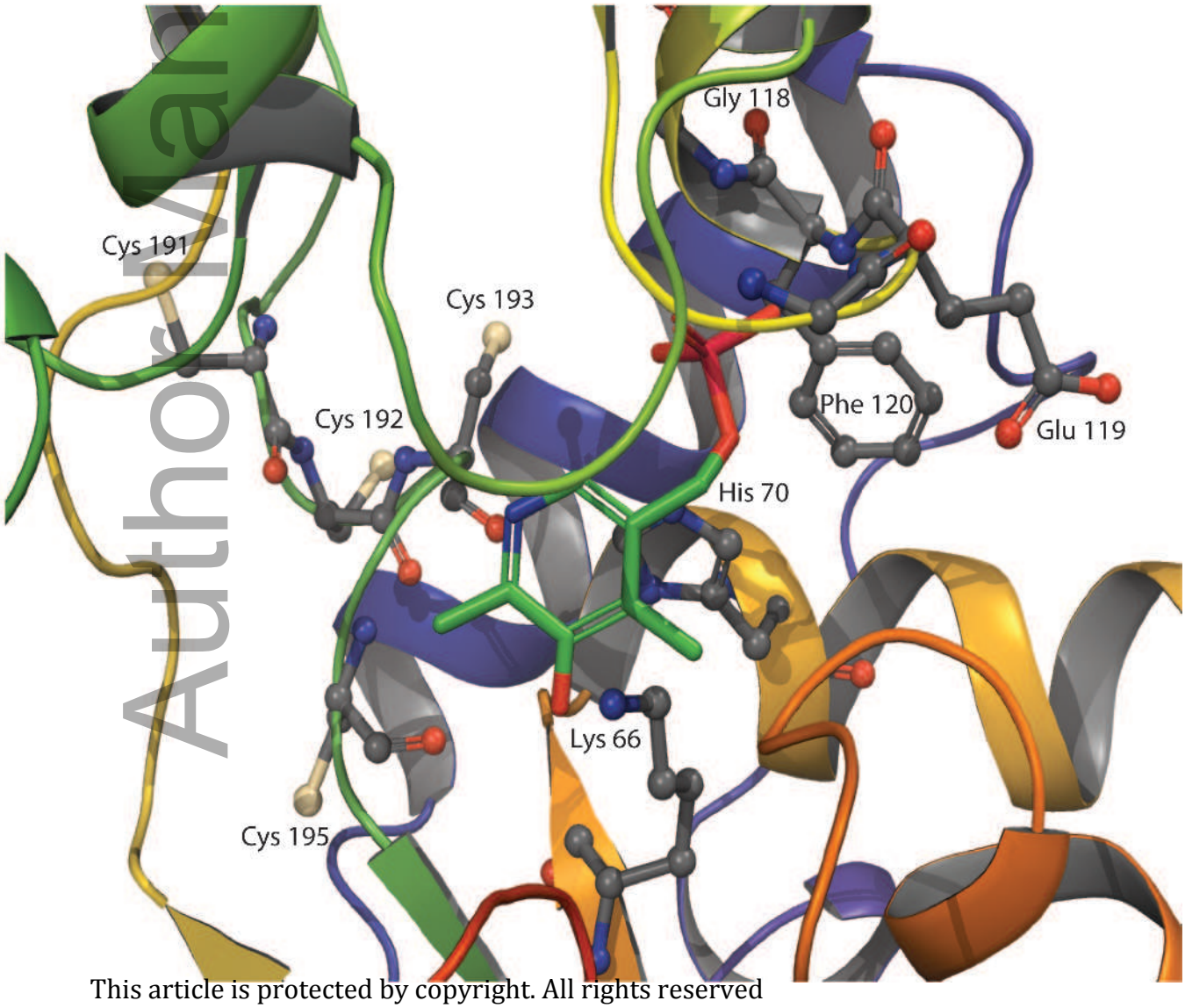


mmi_14622_f7.tif

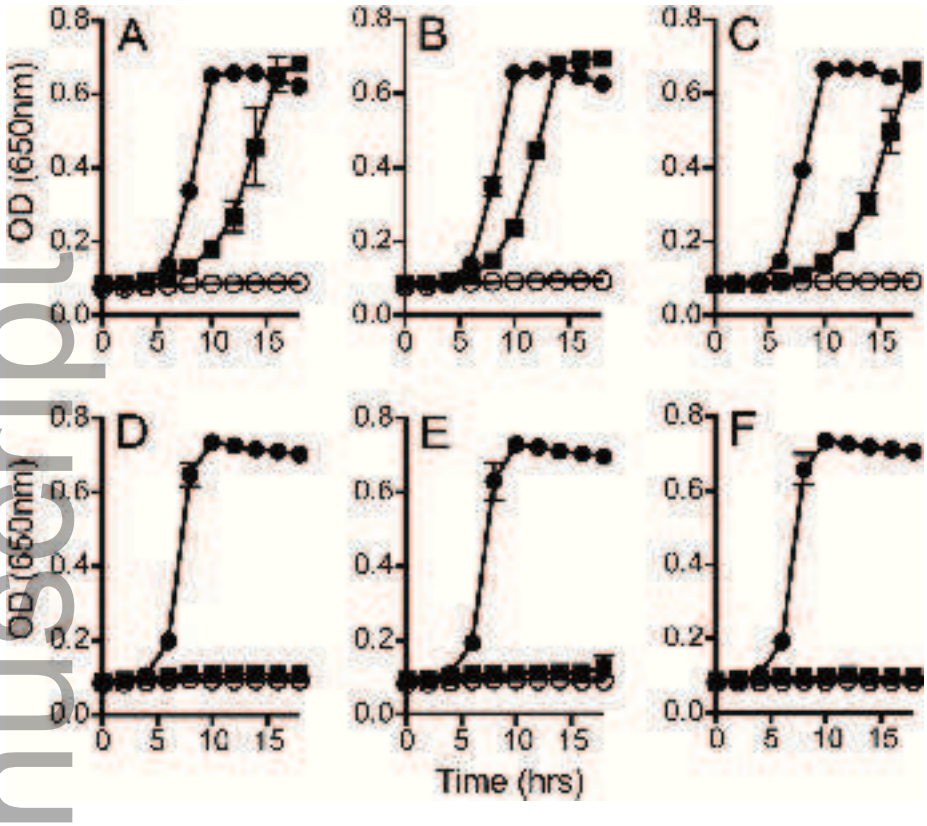
A

```
ScThi5p     1 ----MSTDKITFLLNWQPTPYHIPIFLAQTKGYFKEQGLDMAILEPTNPSDVTELIGSGK
LpThi5      1 MAMSSLKSRVTLLLNWYTNPYHTPILVAQQLGFYAEEGIKLAILEPADPSDVTEIVGLGT

ScThi5p    57 VDMGLKAMIHTLAAKARGFPVTSVASLLDEPFTGVLYLKGSGITEDFQSLKGKKIGYVGE
LpThi5     61 VDFGVKAMIHTVAAKAKGYPVTSIGTLLDEPPTGLIALKSSGIN-SFQDIVGKRVGYIGE

ScThi5p   117 FGKIQIDELTKHYGMKPEDYTAVRCGMNVAKYIIEGKIDAGIGIECMQQVELEEYLAKQG
LpThi5    120 FGKKIIDDLASLAGIDPTSYKTVRIGMNVTDAIYRDVIDTGIGFINFQKVELEHLC----

ScThi5p   177 RPASDAKMLRIDKLACLGCCCFCTVLYICNDEFLKKNPEKVRKFLKAIKKATDYVLADPV
LpThi5    176 ---GETVFLRIDQLAGLGCCCFCSIQFIVPETTL-KQPELVKGFLRATQRGAAYTTEKPE

ScThi5p   237 KAWKEYIDFKPQLNNDLSYKQYQRCYAYFSSSLYNVHRDWKKVTGYGKRLAILPPDYVS-
LpThi5    232 EAYELLCQAKPQLRTPLYQKIFTRTLPFFSRTLINVDRDWDKVGRYTKHLNIIDEHFDIS

ScThi5p   296 -NYTNEYLSWPEPEEVSDPLEAQRLMAIHQEKCRQEGTFKRLALPA
LpThi5    292 QCYTNRFLPDTPYSDL-KP-----------IACCLEN---------
```

B



mmi_14622_f8.tif



mmi_14622_f9.tif