Supplementary information

Structural and mechanistic characterization of L-histidinol phosphate phosphatase from the PHP family of proteins

Swapnil V. Ghodge, Alexander A. Fedorov, Elena V. Fedorov, Brandan Hillerich, Ronald Seidel, Steven C. Almo and Frank M. Raushel

Synthesis of *N*-formyl-L-histidinol phosphate. The synthesis was carried out in two steps: N-formylation of L-histidinol, followed by phosphorylation to the desired product. N-formylation of L-histidinol was conducted using the method of Hosseini-Sarvari and coworker (*S1*). In brief, 80 mg (1 mmol) of zinc oxide and 0.22 mL (6 mmol) of formic acid were mixed in a 10 mL round-bottom flask. 428 mg (2 mmol) of L-histidinol hydrochloride was added and the reaction mixture was stirred for 3 hours at 70 °C in an oil bath. The reaction mixture was then cooled to room temperature, diluted with water and the formic acid was neutralized with NH_4OH. The solvent was evaporated *in vacuo*. The residual solid was dissolved in 10 mM NH_4HCO_3, filtered and the pH was adjusted to ~8.2 using NH_4OH. This solution was passed through DOWEX 50W x8 cation exchange column, pre-equilibrated with 10 mM NH_4HCO_3, pH 8.0, and washed with the same buffer. The flow-through was collected and the solvent was evaporated *in vacuo*. This procedure yielded 190 mg white solid (yield ~ 56%). ^{1}H-NMR δ 8.29 (s, 1H), 7.87 (s, 1H), 7.06 (s, 1H), 4.09 (m, 1H), 3.53 (d, 1H), 3.43 (d, 1H), 2.85 (d, 1H), 2.67 (d, 1H); mass, calculated for $C_7H_{11}N_3O_2$ 169.0851 Da, found 170.0732 Da(M+H).

The phosphorylation of *N*-formyl-L-histidinol was carried out in the presence of polyphosphoric acid using a method described to synthesize L-histidinol phosphate (*S2*). 180 mg of *N*-formyl-L-histidinol was mixed with 1.5 g of polyphosphoric acid and the mixture was heated at 100 °C in an oil bath for one hour. The reaction was quenched with ice-cold water, neutralized using barium hydroxide, and the supernatant solution was purified using Dowex 1x8 anion exchange resin. The fractions from the anion exchange chromatography were analyzed using UV absorbance at 220 nm. The peak fractions were pooled together and the solvent was evaporated *in vacuo*. The procedure yielded a mixture of *N*-formyl-L-histidinol phosphate and

L-histidinol phosphate, along with inorganic phosphate. This was confirmed using ^{1}H and ^{31}P-NMR, and ESI-MS. Mass, calculated for $C_7H_{12}N_3O_5P$ 249.0515 Da, found 248.0576 Da (M-H). The mixture was used without further purification, and the relative concentrations were determined by ^{31}P-NMR using methyl phosphonate as an internal standard.

References

S1. Hosseini-Sarvari, M., and Sharghi, H. (2006) ZnO as a new catalyst for N-formylation of amines under solvent-free conditions. *J. Org. Chem. 71*, 6652-6654.

S2. Wang, L., Huang, H., Nguyen, H.H., Allen, K.N., Mariano, P.S., and Dunaway-Mariano, D. (2010) Divergence of biochemical function in the HAD superfamily: D-glycero-D-manno-heptose-1,7-bisphosphate phosphatase (GmhB). *Biochemistry 49*, 1072-1081.

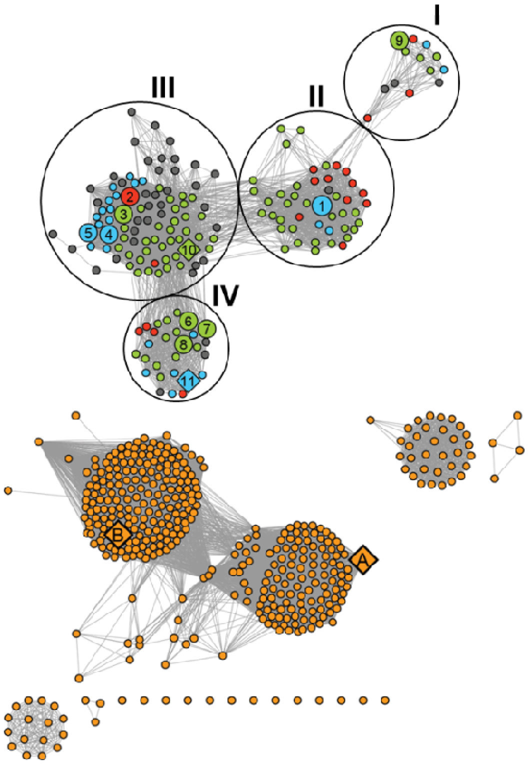

Figure S1: Sequence network diagram of cog 1387 (same as figure 2), with HPP enzymes and close homologs grouped into four arbitrary zones I, II, III, and IV. The details of each node (protein) are provided in Tables S1-S4.

Table S1: List of genes represented by blue nodes in **Figures 2 and S1**. These are authentic HPP enzymes which are found in the vicinity of the L-histidine biosynthesis operon on the genome of the organism.

Organism	**Locus tag**	**GI number**	**Sequence identity to L37351**	**Zone**
Bacillus anthracis str. Ames	BA_1432	30261511	22%	III
Bacillus anthracis str. 'Ames Ancestor'	GBAA_1432	47526706	22%	III
Bacillus anthracis str. Sterne	BAS1324	49184341	22%	III
Bacillus cereus 03BB102	BCA_1468	225863376	23%	III
Bacillus cereus AH187	BCAH187_A1572	217958990	21%	III
Bacillus cereus AH820	BCAH820_1504	218902621	22%	III
Bacillus cereus ATCC 10987	BCE_1533	42780607	21%	III
Bacillus cereus ATCC 14579	BC1413	161511209	21%	III
Bacillus cereus B4264	BCB4264_A1466	218234061	21%	III
Bacillus cereus E33L	BCZK1298	52143930	23%	III
Bacillus cereus G9842	BCG9842_B3878	218896444	20%	III
Bacillus cereus Q1	BCQ_1486	222095146	21%	III
Bacillus thuringiensis serovar konkukian str. 97-27	BT9727_1297	49477223	25%	III
Bacillus thuringiensis str. Al Hakam	BALH_1268	118476974	23%	III
Bacillus weihenstephanensis KBAB4	**BcerKBAB4_1335**	163939326	24%	III
Clostridium botulinum E3 str. Alaska E43	CLH_2622	188588114	31%	II
Clostridium kluyveri DSM 555	CKL_1305	153953930	31%	II
Exiguobacterium sibiricum 255-15	Exig_2369	172058376	24%	IV
Lactobacillus fermentum IFO 3956	LAF_0754	184155230	22%	IV
Lactococcus lactis subsp. cremoris MG1363	llmg_1288	125624110	80%	II
Lactococcus lactis subsp. cremoris SK11	LACR_1325	116512055	79%	II
***Lactococcus lactis* subsp. lactis Il1403**	**L37351**	**15673198**	**100%**	**II**
Listeria innocua Clip11262	LMHCC_2060	16799654	26%	IV

Listeria monocytogenes EGD-e	lmo0570	16802613	25%	IV
Listeria monocytogenes HCC23	LMHCC_2060	217965335	26%	IV
Listeria monocytogenes serotype 4b str. F2365	**LMOf2365_0599**	46906815	25%	IV
Listeria welshimeri serovar 6b str. SLCC5334	lwe0536	116871956	36%	IV
Natranaerobius thermophilus JW/NM-WN-LF	Nther_1995	188586605	25%	III
Oceanobacillus iheyensis HTE831	OB0554	23098009	33%	II
Streptococcus gordonii str. Challis substr. CH1	SGO_1401	157150565	26%	I
Streptococcus sanguinis SK36	SSA_1439	125718249	25%	I

Table S2: List of genes represented by green nodes in **figures 2 and S1**. These represent authentic HPP enzymes identified based on their primary sequence. All the genes necessary for L-histidine biosynthesis are not found clustered together in these organisms.

Organism	**Locus tag**	**GI number**	**Sequence identity to L37351**	**Zone**
Akkermansia muciniphila ATCC BAA-835	Amuc_1427	187735918	22%	III
Alkaliphilus metalliredigens QYMF	Amet_3162	150390911	33%	II
Alkaliphilus metalliredigens QYMF	Amet_3545	150391281	29%	II
Anoxybacillus flavithermus WK1	Aflv_0468	212638317		IV
Arcobacter butzleri RM4018	Abu_1904	157738118	25%	III
Bacillus clausii KSM-K16	ABC0751	56962525	22%	II
Bacillus clausii KSM-K16	ABC2752	161349990	26%	IV
Bacillus halodurans C-125	**BH3206**	15615768	28%	IV
Bacillus licheniformis DSM 13 = ATCC 14580	BLi03106	52786832	29%	IV
Bacillus pumilus SAFR-032	BPUM_2607	157693363	25%	IV
Bacillus subtilis subsp. subtilis str. 168	**BSU29620**	16080014	29%	IV
Brevibacillus brevis NBRC 100599	**BBR47_00270**	226309614	22%	III
Caldicellulosiruptor bescii DSM 6725	Athe_1711	222529696	33%	II
Caldicellulosiruptor saccharolyticus DSM 8903	Csac_2251	146297249	33%	II
Campylobacter concisus 13826	CCC13826_1445	157165198	25%	III
Campylobacter curvus 525.92	CCV52592_0277	154174464	25%	III
Campylobacter fetus subsp. fetus 82-40	CFF8240_1053	118475300	24%	III
Campylobacter hominis ATCC BAA-381	CHAB381_1502	154148198	26%	III
Candidatus Desulforudis audaxviator MP104C	Daud_1449	169831605	23%	III
Carboxydothermus hydrogenoformans Z-2901	CHY_1234	78043027	28%	II
Carboxydothermus	CHY_2084	78043205	27%	III

hydrogenoformans Z-2901				
Clostridium acetobutylicum ATCC 824	CA_C2727	15025762	30%	II
Clostridium beijerinckii NCIMB 8052	Cbei_0975	150015861	34%	II
Clostridium beijerinckii NCIMB 8052	Cbei_2106	150016975	28%	IV
Clostridium botulinum A str. ATCC 3502	CBO2571	148380526	34%	II
Clostridium botulinum A str. Hall	CLC_2443	153937403	34%	II
Clostridium botulinum B1 str. Okra	CLD_1994	170757551	35%	II
Clostridium botulinum E3 str. Alaska E43	CLH_2953	188590530	33%	II
Clostridium botulinum F str. Langeland	CLI_2634	153940040	35%	II
Clostridium cellulolyticum H10	Ccel_1927	220929346	37%	II
Clostridium difficile 630	CD630_07370	126698317	24%	IV
Clostridium difficile 630	CD630_21910	126699807	29%	II
Clostridium kluyveri DSM 555	CKL_0268	153952905	33%	II
Clostridium kluyveri NBRC 12016	CKR_0224	219853567	34%	II
Clostridium kluyveri NBRC 12016	CKR_1201	219854544	31%	II
Clostridium novyi NT	NT01CX_0001	118444162	28%	II
Clostridium phytofermentans ISDg	Cphy_2903	160881032	30%	II
Clostridium thermocellum ATCC 27405	Cthe_0724	125973242	33%	II
Deinococcus deserti VCD115	Deide_07040	226355575	25%	III
Deinococcus geothermalis DSM 11300	Dgeo_1608	94985708	26%	III
Deinococcus radiodurans R1	DR_0470	15805497	26%	III
Desulfitobacterium hafniense DCB-2	Dhaf_2109	219668146	30%	II
Desulfitobacterium hafniense DCB-2	Dhaf_4086	219670096	27%	III
Desulfitobacterium hafniense Y51	DSY1025	89893771	29%	II
Desulfitobacterium hafniense Y51	DSY2929	89895675	27%	III
Desulfotalea psychrophila LSv54	DP2888	51246740	25%	IV
Desulfotomaculum reducens MI-1	Dred_0665	134298533	24%	III
Desulfovibrio alaskensis G20	Dde_1057	78356104	24%	II
Desulfovibrio desulfuricans subsp.	Ddes_1985	220905247	22%	III

desulfuricans str. ATCC 27774				
Desulfovibrio vulgaris DP4	Dvul_0754	120601804	21%	III
Desulfovibrio vulgaris str. 'Miyazaki F'	DvMF_0940	218886042	22%	III
Desulfovibrio vulgaris str. Hildenborough	DVU2490	46580894	21%	III
Dictyoglomus turgidum DSM 6724	Dtur_0185	217966622	25%	III
Geobacillus kaustophilus HTA426	**GK2799**	56421334	26%	IV
Geobacillus thermodenitrificans NG80-2	GTNG_2701	138896338	24%	IV
Helicobacter hepaticus ATCC 51449	HH1714	32267213	26%	III
Heliobacterium modesticaldum Ice1	HM1_1034	167629052	30%	II
Leuconostoc mesenteroides subsp. mesenteroides ATCC 8293	LEUM_1558	116618652	26%	IV
Lysinibacillus sphaericus C3-41	Bsph_4155	169829588	24%	IV
Methylacidiphilum infernorum V4	Minf_0970	189218982	29%	II
Moorella thermoacetica ATCC 39073	Moth_0817	83589668	26%	III
Nautilia profundicola AmH	NAMH_1464	224373484	25%	II
Nitratiruptor sp. SB155-2	NIS_1518	152991260	21%	III
Oceanobacillus iheyensis HTE831	OB2759	23100214	28%	II
Pelotomaculum thermopropionicum SI	PTH_1872	147678207	25%	III
Rubrobacter xylanophilus DSM 9941	Rxyl_2572	108805364	21%	III
Streptococcus agalactiae A909	SAK_1303	76563046	27%	I
Streptococcus agalactiae NEM316	gbs1289	25011338	27%	I
Streptococcus mutans UA159	**SMU_1486c**	24379879	28%	I
Streptococcus thermophilus CNRZ1066	str1232	55823151	25%	I
Streptococcus thermophilus LMD-9	STER_1212	116627978	26%	I
Sulfurimonas denitrificans DSM 1251	Suden_1866	78778060	25%	III
Sulfurovum sp. NBC37-1	SUN_0354	152991950	24%	III
Symbiobacterium thermophilum IAM 14863	STH2574	51893712	24%	III
Thermotoga maritima MSB8	TM0804	15643567	27%	II
Thermotoga petrophila RKU-1	Tpet_0124	148269269	27%	II
Thermotoga sp. RQ2	TRQ2_0122	170287929	27%	II

Thermus thermophilus HB27	TTC1652	46199954	26%	III
Thermus thermophilus HB8	TTHA0331	55980300	26%	III
Wolinella succinogenes DSM 1740	WS0471	34556896	23%	III

Table S3: List of genes represented by red nodes in **figures 2 and S1**. These represent HPP homologs identified based on their primary sequence, which have all the sequence features necessary for HPP activity. However, the genomes of these organisms do not possess several genes necessary for L-histidine biosynthesis, based on the current available information.

Organism	**Locus tag**	**GI number**	**Sequence identity to L37351**	**Zone**
Clostridium botulinum A3 str. Loch Maree	CLK_1956	170758750	34%	II
Clostridium botulinum B str. Eklund 17B	CLL_A3211	187934801	32%	II
Clostridium botulinum B str. Eklund 17B	CLL_A2893	187935558	32%	II
Clostridium perfringens ATCC 13124	CPF_1312	110798699	33%	IV
Clostridium perfringens ATCC 13124	CPF_2137	110800838	30%	II
Clostridium perfringens SM101	CPR_1849	110802089	30%	II
Clostridium perfringens SM101	CPR_1128	110802533	33%	IV
Clostridium perfringens str. 13	CPE1056	18310038	32%	IV
Clostridium perfringens str. 13	CPE1883	18310865	30%	II
Clostridium tetani E88	CTC01046	28210748	33%	II
Coprothermobacter proteolyticus DSM 5265	COPRO5265_0239	206895927	30%	II
Fervidobacterium nodosum Rt17-B1	Fnod_0198	154248897	30%	II
Finegoldia magna ATCC 29328	FMG_0406	169824103	27%	I
Fusobacterium nucleatum subsp. nucleatum ATCC 25586	FN0428	19703770	26%	I
Lawsonia intracellularis PHE/MN1-00	LI0725	94987167	21%	III
Leuconostoc citreum KM20	LCK_00178	170016536	25%	IV
Macrococcus caseolyticus JCSC5402	**MCCL_0344**	222150594	27%	III
Streptococcus thermophilus LMG 18311	stu1232	55821239	26%	I
Thermosipho africanus TCF52B	THA_992	217077076	28%	II
Thermosipho melanesiensis BI429	Tmel_0785	150020678	28%	II

Table S4: List of genes represented by gray nodes in **figures 2 and S1**. These represent enzymes which bear reasonable sequence identity to authentic HPP enzymes but lack the primary sequence features necessary for HPP activity, as determined in this study. Hence, these enzymes must have a different physiological function.

Organism	**Locus tag**	**GI number**	**Sequence identity to L37351**	**Zone**
Agrobacterium radiobacter K84	Arad_7540	222081247	22%	III
Alkaliphilus oremlandii OhILAs	Clos_0339	158319390	28%	III
Anoxybacillus flavithermus WK1	Aflv_1054	212638895	26%	III
Bacteroides fragilis NCTC 9343	BF0879	60680430	28%	III
Bacteroides fragilis NCTC 9343	BF1597	60681093	24%	III
Bacteroides fragilis YCH46	BF0961	53712253	28%	III
Bacteroides fragilis YCH46	BF1583	53712875	24%	III
Bacteroides thetaiotaomicron VPI-5482	BT_1478	29346888	24%	III
Bacteroides thetaiotaomicron VPI-5482	BT_4257	29349665	28%	III
Brachyspira hyodysenteriae WA1	BHWA1_00960	225619891	25%	III
Brevibacillus brevis NBRC 100599	BBR47_43860	226313971	22%	III
Clostridium novyi NT	NT01CX_0128	118444797	31%	II
Clostridium perfringens ATCC 13124	CPF_1620	110798872	29%	III
Clostridium perfringens ATCC 13124	CPF_0801	110799120	28%	IV
Clostridium perfringens ATCC 13124	CPF_1229	110801334	24%	III
Clostridium perfringens SM101	CPR_1040	110803006	24%	III
Clostridium perfringens SM101	CPR_0788	110803561	33%	IV
Clostridium perfringens str. 13	CPE0804	18309786	27%	IV
Clostridium perfringens str. 13	CPE0972	18309954	24%	III
Clostridium perfringens str. 13	CPE1370	18310352	29%	III
Clostridium phytofermentans ISDg	Cphy_2094	160880232	34%	III

Desulfobacterium autotrophicum HRM2	HRM2_30690	224370152	27%	III
Desulfotalea psychrophila LSv54	DP0093	51243945	35%	III
Desulfotomaculum reducens MI-1	Dred_2577	134300416	27%	III
Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	Ddes_0366	220903646	18%	III
Dictyoglomus thermophilum H-6-12	DICTH_1913	206900475	24%	III
Geobacillus kaustophilus HTA426	GK2248	56420783	24%	III
Geobacillus thermodenitrificans NG80-2	GTNG_2185	138895826	25%	III
Halorubrum lacusprofundi ATCC 49239	Hlac_1354	222479778	28%	III
Halothermothrix orenii H 168	Hore_11520	220931991	26%	III
Heliobacterium modesticaldum Ice1	HM1_2038	167630107	25%	III
Lactobacillus casei ATCC 334	LSEI_1206	116494700	25%	I
Lactobacillus casei BL23	LCABL_14250	191638198	25%	I
Lysinibacillus sphaericus C3-41	Bsph_1650	169827220	24%	III
Methylacidiphilum infernorum V4	Minf_0443	189218458	28%	III
Parabacteroides distasonis ATCC 8503	BDI_1238	150007878	32%	III
Pediococcus pentosaceus ATCC 25745	PEPE_0107	116491924	24%	I
Pseudomonas aeruginosa UCBPP-PA14	PA14_03330	116053983	23%	III
Syntrophomonas wolfei subsp. wolfei str. Goettingen	Swol_0653	114566197	27%	III
Thermoanaerobacter pseudethanolicus ATCC 33223	Teth39_0146	167036577	25%	III
Thermoanaerobacter sp. X514	Teth514_2305	167040919	25%	III
Treponema denticola ATCC 35405	TDE0796	42526308	24%	III