SUPPORTING INFORMATION FOR:

Solution Structure of 4'-Phosphopantetheine - GmACP3 from *Geobacter metallireducens*: A Specialized Acyl Carrier Protein with Atypical Structural Features and a Putative Role in Lipopolysaccharide Biosynthesis[†]

Theresa A. Ramelot[1], Matthew J. Smola[1], Hsiau-Wei Lee[2], Colleen Ciccosanti[3], Keith Hamilton[3], Thomas B. Acton[3], Rong Xiao[3], John K. Everett[3], James H. Prestegard[2], Gaetano T. Montelione[3,4], and Michael A. Kennedy[1*]

[1]Department of Chemistry and Biochemistry, Miami University, Oxford, Ohio 45056, United States and the Northeast Structural Genomics Consortium, [2]Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, United States and the Northeast Structural Genomics Consortium, [3]Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States and the Northeast Structural Genomics Consortium, and [4]Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey, 08854, United States

Contents of Supplementary Material

Figure S1. LC-MS Electrospray Mass Spectrum of GmACP3.

Figure S2 Overlayed ^{1}H-^{15}N HSQC spectra of holo-GmACP3 and a mixture of holo- and apo-GmACP3.

Figure S3. $^{15}N\{^{1}H\}$ heteronuclear NOE data for holo-GmACP3 recorded at 20°C.

Figure S4. The 600 MHz Fast ^{1}H-^{15}N HSQC NMR spectrum of 5%-^{13}C,100%-^{15}N *holo*-GmACP3.

Table S1. GmACP3 homologs from Blast search of the KEGG database.

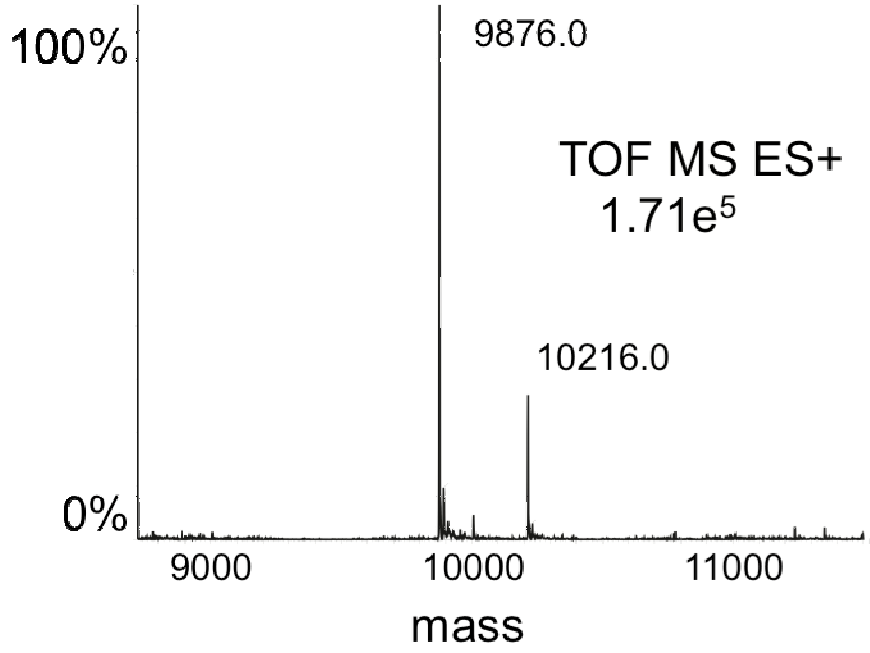
Table. S2. NOEs between 4'PP and protein resonances in holo-GmACP3.

Table. S3. DALI hits against holo-GmACP3 NMR structure.

Table S4. RMSDs between holo-GMACP3 and ACPs structural homologs.

Text S1. Additional discussion of possible functions of Gmet_2337 and Gmet_2340.

Figure S1. LC-MS Electrospray Mass Spectrum of GmACP3. Spectrum was collected at the Ohio State University, Mass Spectrometry and Proteomics Facility (http://www.ccic.ohio-state.edu/MS). Deconvoluted ESI data in on top and raw data is on the bottom.



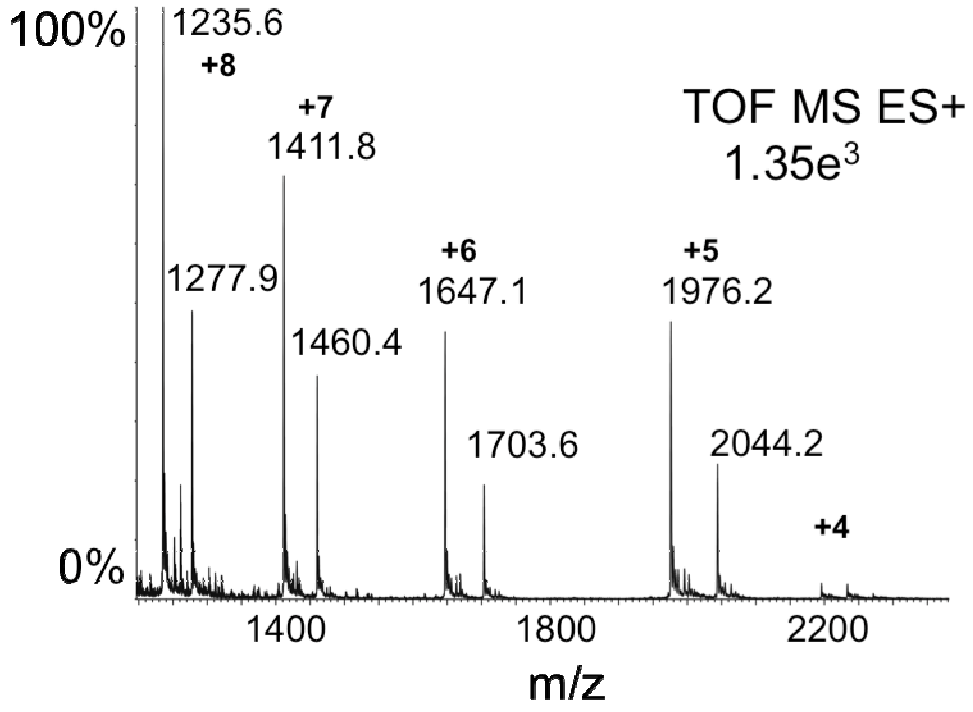


Figure S2. Overlayed ^{1}H-^{15}N HSQC spectra of holo-GmACP3 and a mixture of holo- and apo-GmACP3. Red is > 90% holo-GmACP3 NC5 NMR sample. Blue is mixture of apo- and [*U*-^{13}C, ^{15}N] holo-GmACP3 NMR sample. R12 N-Hε, R24 N-Hε, and H39 N-Hε2 peaks are folded in the 15N dimension from 86.4, 80.2, and 162.5 ppm, respectively. The W34 N-Hε is cropped for clarity (12.92 ppm, 133.9 ppm).

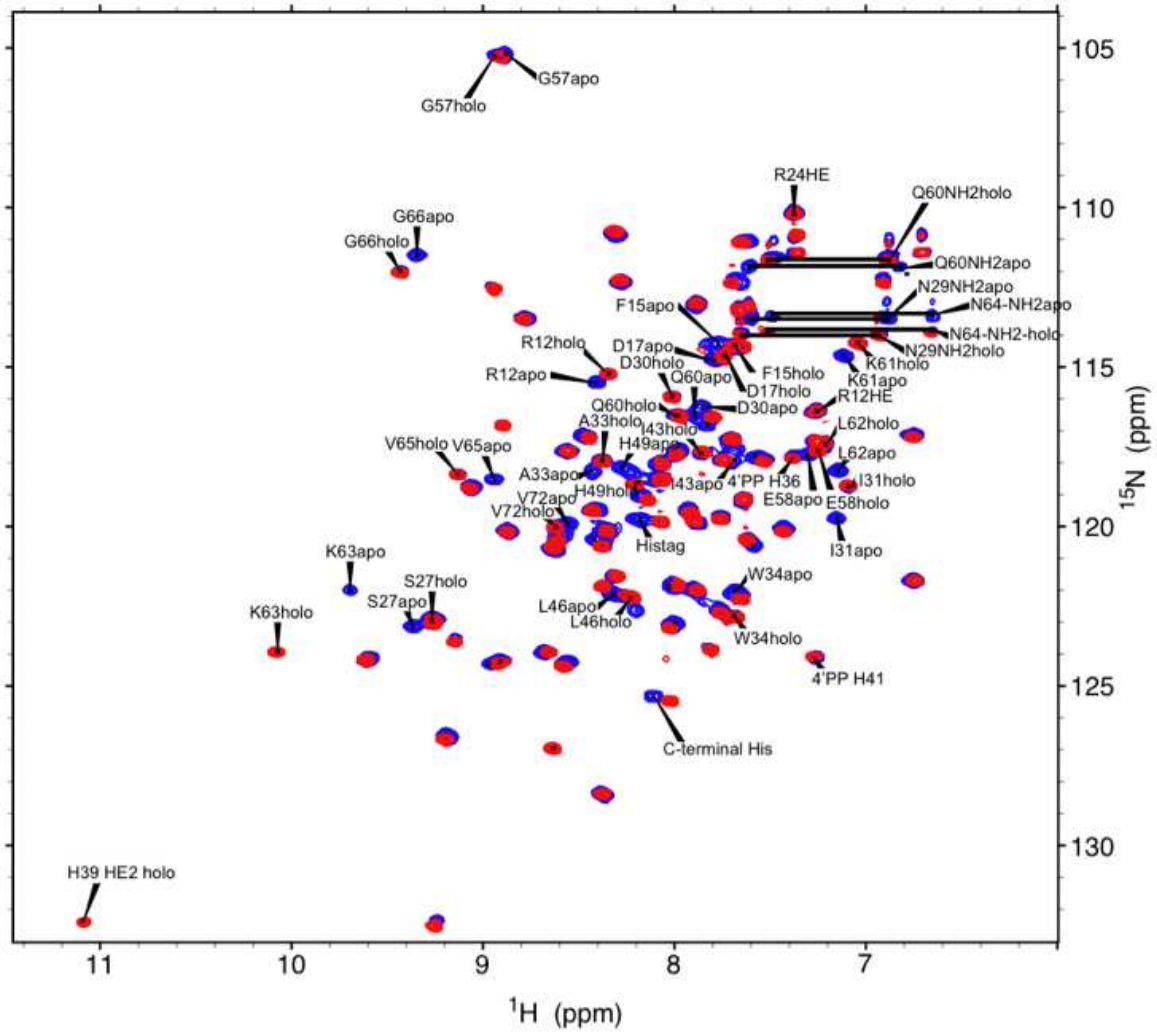


Figure S3. $^{15}N\{^{1}H\}$ heteronuclear NOE data for holo-GmACP3 recorded at 20°C.

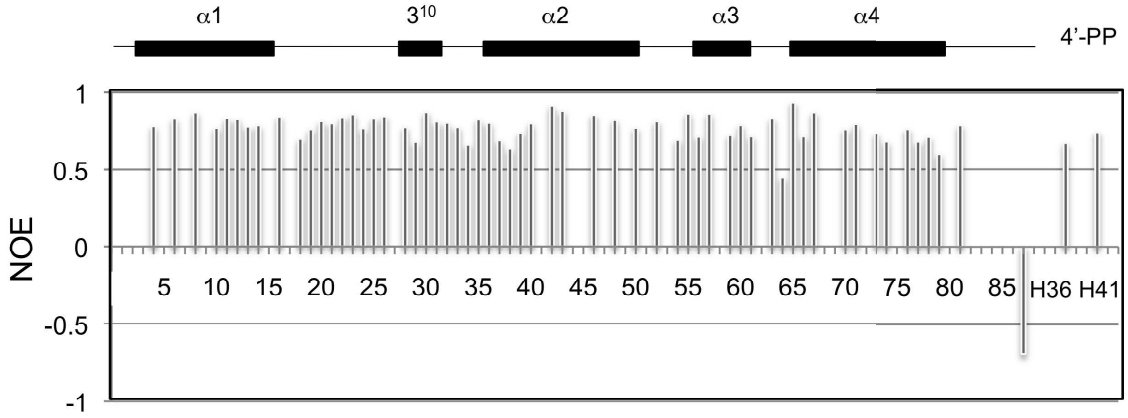
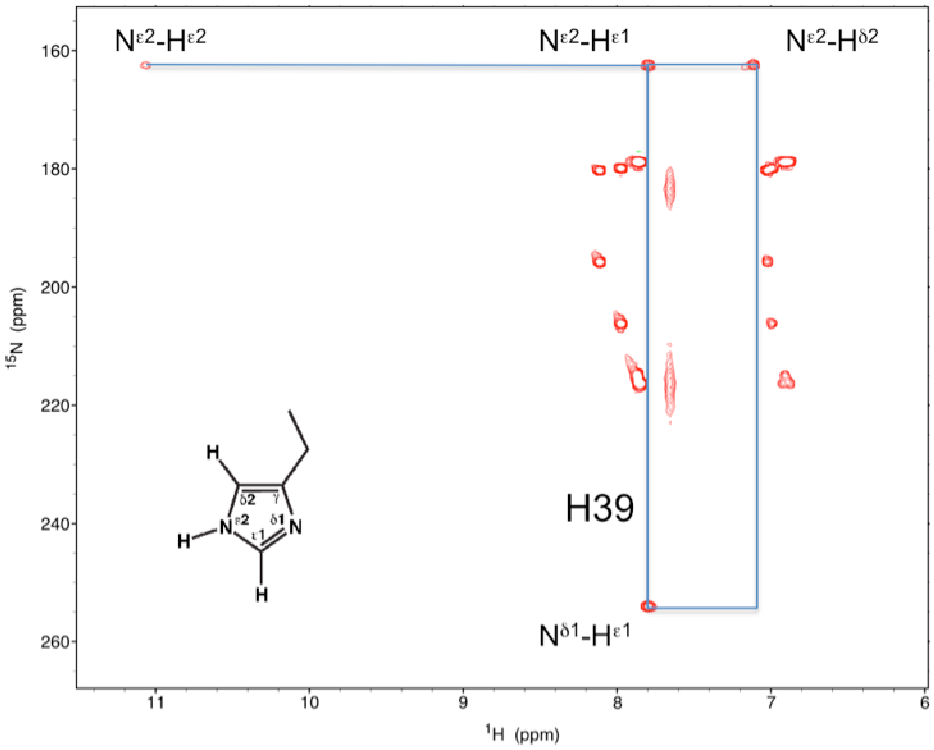


Figure S4. Top: The 600 MHz Fast ^{1}H-^{15}N HSQC NMR spectrum of 5%-^{13}C,100%-^{15}N *holo*-GmACP3 in pH 6.5 MES buffer at 20°C. Acquisition parameters: ^{15}N carrier at 198 ppm, J^{NH} = 22.8 Hz, 1 s relaxation delay, 1024 x 256 complex points, 200 ppm sweep width in the ^{15}N dimension, and 256 transients per t1 increment. Assigned cross-peaks for H39 are labeled. The upside down L pattern is indicative of the $H^{\varepsilon 2}$ neutral tautomer (inset). Other peaks are from H49 and the C-terminal His_6 tag. Bottom: close up view of the hydrogen-bonded W34 and H39 in the hydrophobic cavity of GmACP3.



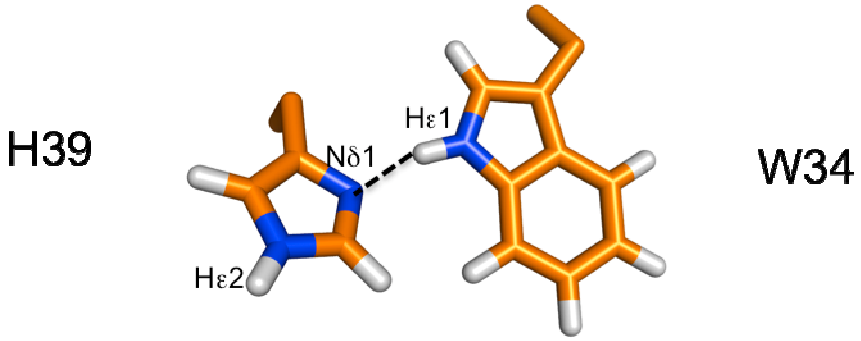


Table S1. GmACP3 homologs from Blast search of the KEGG database. The consensus motif is WDS-L/V/F-x-H/N. Everything below *Bacteroides fragilis* BF1532 has less than 40% identity. Genes with KEGG IDs marked with a star have a nearby homolog of *gmet_2338* that encodes a membrane-bound O-acyl transferase (MBOAT) protein.

Organism	KEGG org	KEGG ID	sequence
Geobacter metallireducens	gme	*gmet_2339	WDSLxH
Gemmatimonas aurantiaca	gau	*GAU_0573	WDSLxH
Solibacter usitatus	sus	*acid_6044	WDSLxH
Psychromonas ingrahamii	pin	*ping_1634	WDSLxH
Acidobacteria bacterium	aba	Acid345_0414	WDSLxH
Xanthobacter autotrophicus	xau	*Xaut_0883	WDSLxN
Pseudomonas stutzeri	psa	*PST_2875	WDSLxH
Rhodobacter sphaeroides KD131	rsk	*RSKD131_2926	WDSLxH
Rhodobacter sphaeroides 2.4.1	rsp	*RSP_1534	WDSFxH
Rhodobacter sphaeroides	rsh	*Rsph17029_0184	WDSFxH
Bacteroides fragilis YCH46	bfr	*BF3277	WDSLxH
Bacteroides fragilis YCH46	bfr	BF1532	WDSLxH
Sinorhizobium meliloti	sme	*SMc01553	WDSLxN
Sorangium cellulosum	scl	5812467_acpA	WDSLxN
Frankia alni	fal	FRAAL6258	WDSLxH
Frankia sp.	fra	Francci3_3942	WDSLxH
Clostridium acetobutylicum	cac	CAC3313	WDSLxH
Solibacter usitatus	sus	acid_7809	WDSVxH
Clost. beijerinckii	cbe	Cbei_2095	WDSLxH
Nostoc puntiforme	npu	Npun_F3159	WNSMxH
Clostridium thermocellum	cth	Cthe_2227	WDSLxH
Aeromonas hydrophia	aha	AHA_3187	WDSLxH
Campylobacter jejuni	cjr	CJE1489	WTSLxH
Acaryochloris marina	amr	AM1_2708	WDSMxN
Geobacillus sp.	gym	GYMC10_1448	WDSLxH
Methylbacterium nodulans	mno	Mnod_2747	WDSLxH
Nitrosospira multiformis	nmu	Nmul_A0283	WDSLxH
Aeromonas salmonicida	asa	ASA_0139	WDSLxH
Bacteroides fragilis	bfr	BF0825	WDSLxH
Vibrio cholerae	vch	VC0248	WDSLxH
Herpetosiphon aurantiacus	hau	Haur_2031	WDSLxH
Candidatus Methanoregular booneis	mbn	Mboo_1741	WDSLxH

Table S2. NOEs between 4'PP and protein resonances in holo-GmACP3. Atom names are in Xplor format.

Residue	proton	4'PP proton	Upper bound (A)
A28	HN	H31*	5.0
A28	HB*	H30*	2.5
A28	HB*	H31*	3.0
A28	HB*	H36	5.0
A28	HB*	H37*	4.0
A28	HB*	H38*	5.0
A28	HB*	H41	5.0
W34	HD1	H31*	5.0
D35	HA	H31*	3.5
S36	HA	H28*	4.0
S36	HA	H30*	5.0
S36	HA	H31*	3.5
S36	HN	H31*	6.0
S36	HB*	H28*	5.0
H39	HB*	H30*	6.0
H39	HB*	H31*	5.0
H39	HD2	H30*	5.0
H39	HE1	H30*	5.0
H39	HE1	H31*	5.0
K63	HA	H38*	3.0
K63	HA	H41	3.0
N64	HA	H42*	4.0
N64	HA	H431	3.0
N64	HA	H432	3.0

4'PP proton	4'PP proton	Upper bound (A)
H281	H30*	3.5
H282	H30*	3.5
H28*	H31*	4.0
H30*	H32	4.0
H30*	H421	4.0
H30*	H422	4.0
H30*	H43*	5.0
H31*	H32	3.5
H31*	H31*	5.5
H31*	H32	4.0
H31*	H371	3.0
H31*	H372	3.0
H31*	H38*	3.0
H36	H30*	3.0
H41	H30*	4.0
H41	H37*	5.0
H41	H38*	4.0
H41	H421	3.0
H41	H422	3.0
H41	H43	4.0
H42*	H43*	5.0

Table S3. DALI hits against holo-GmACP3 NMR structure. Z is the DALI Z-score. RMSD is the Cα root mean square deviation to holo-GmACP3 in Å. ALI is the number of aligned residues. LEN is the full length of the structural homolog. %I is the percent sequence identity with GmACP3.

	PDB	Z	RMSD	ALI	LEN	I%	Description
1	2kjs-A	12.4	1.4	78	87	100	>95% holo-GMACP3, self
2	2kci-A	11.6	1.8	78	87	100	25% holo-GmACP3
3	2fvf-A	9.5	1.8	76	82	21	NMR, spinach decanoyl-ACP
4	2cgq-A	9.3	1.7	72	74	25	1.8 Å, MtACP (Rv0033 from *M. tuberculosis*)
5	1vku-A	9.2	1.9	75	85	21	2.0 Å, TmACP (*Thermatoga maritima*)
6	2kgc-A	8.8	2.1	76	86	22	NMR, octanoyl-actACP
7	2eht-A	8.8	2.0	73	77	22	1.4 Å, apo-ACP *Aquifex aeolicus*
8	3gzl-A	8.7	2.0	74	81	22	2.6 Å, holo-PfACP dimer (*plasmodium falciparum*)
9	3ejb-E	8.7	2.0	72	78	19	2.0 Å, tetradecanoic- EcACP in P450Biol complex
10	3ejd-A	8.6	1.9	71	78	20	2.1 Å, hexadec-9Z-enoic- EcACP in P450Biol complex
11	3ejd-E	8.6	1.9	71	78	20	2.1 Å, hexadec-9Z-enoic EcACP- in P450Biol complex
12	3ejb-A	8.5	2.0	72	78	19	2.0 Å, tetradecanoic-EcACP in P450Biol complex
13	3eje-C	8.5	1.9	71	78	20	2.1 Å, octadec-9Z-enoic-EcACP in P450Biol complex
14	2ehs-A	8.5	1.9	72	75	22	1.3 Å, *Aquifex aeolicus*
15	3eje-A	8.4	1.9	71	78	20	2.1 Å, octadec-9Z-enoic-EcACP in P450Biol complex
16	1f80-E	8.4	1.8	70	74	23	2.3 Å, *Bacillus subtilis*, holo-ACP in AcpS complex
17	3gzm-A	8.3	1.9	72	78	22	1.8 Å, holoPfACP reduced monomer with malonate
18	2fae-A	8.3	2.2	73	77	21	1.6 Å, EcACP-decanoyl
19	2fad-A	8.3	2.2	73	77	21	1.6 Å, EcACP-heptanoyl
20	2fve-A	8.2	2.0	76	82	21	NMR, spinach decanoyl-ACP

Table S4. RMSDs between holo-GMACP3 and ACPs structural homologs. RMSD is the Cα RMSD in Å based on alignment of all four helices. This is residues: 3-15, 35-50, 56-60, and 65-73 of GmACP3. Helix α4 is longer (65-79), but only 65-73 line up with the shorter helix α4 in *E. coli* ACP.

PDB	RMSD	Description
1tk8	1.7	apo-EcACP
1loh-A	1.8	"apo"-EcACP from butyrl-ACP crystal
1loi-A	1.7	butryl-EcACP I62M mutant
2fac-A	1.6	hexanoyl-EcACP
2fad-A	1.6	heptanoyl-EcACP
2fad-B	1.6	heptanoyl-EcACP
2fae-A	1.6	decanoyl-EcACP
3ejb	1.4	tetradecanoyl-EcACP in $P450_{Biol}$ complex
3ejd	1.4	hexadec-9Z-anoyl EcACP- in $P450_{Biol}$ complex
3eje	1.4	octadec-9Z-anoyl-EcACP in $P450_{Biol}$ complex
2k92-1	2.6	apo-EcACP (NMR)
2k93-1	2.3	holo-EcACP (NMR)
2k94-1	2.3	butyrl-EcACP (NMR)
2k0y-1	2.2	apo-actACP (NMR)
2k0x-1	2.2	holo-actACP (NMR)
2kgd-1	2.7	oxybutyl-actACP (NMR)
2kge-1	2.6	dioxohexyl-actACP (NMR)
2kg6-1	2.5	acetyl-actACP (NMR)
2kg8-1	2.4	malonyl-actACP (NMR)
2kg9-1	2.1	butryl-actACP (NMR)
2kgA-1	1.5	hexanoyl-actACP (NMR)
2kgC-1	1.5	octanoyl-actACP (NMR)
2cgq-A	1.5	MtACP (Rv0033 from *M. tuberculosis*) "FDSLKL"
1vku-A	1.8	TmACP TmACP (*Thermatoga maritima*) – has longer h4 "ADSLDL"
3gzl	1.4	holo-PfACP dimer (*plasmodium falciparum*) "ADSLDL"
2eht	1.4	apo-ACP *Aquifex aeolicus* "ADSLDV"

Text S1. Additional discussion of possible functions of Gmet_2337 and Gmet_2340.

Gmet_2337 is annotated as a hypothetical protein and has 31% sequence identity to *E. coli* YiiM, annotated as a 6-N-hydroxylaminopurine resistance protein, and has 26% sequence identity to the C-terminal 100 residues of AlgJ. Gmet_2337 contains conserved GDSFS and DDTHW sequences that that are similar to those found in GDSL serine esterases/lipases and SGNH-hydrolase proteins as well as a potential S-D-H triad (*1, 2*). GDSL enzymes possess many functions, including hydrolysis and synthesis of lipids and esters. This family includes a subfamily of carbohydrate esterases that includes rhamnogalacturonan acetylesterase from *Aspergillus aculeatus*, which removes acetyl groups from the polysaccharide rhamnogalacturonan (*3*). The role of Gmet_2337 could be to hydrolyze incorrect thioesters from GmACP3, with a parallel function to the *Lactobacillus rhamnosus* thioesterase, DltD, that hydrolyzes mischarged D-alanyl-ACP, so that the correct D-alanyl-ACP will be synthesized (*4*).

The neighboring gene *gmet_2340* that has recently been annotated as an acetyl-CoA synthetase based on its 42% sequence identity with the *B. subtilis* enzyme (AcsA, BSU29680) (*5*). The action of this enzyme may allow *G. metallireducens* to metabolize acetate found in soil. It is possible that acetyl-CoA derived from enzymatic conversion could be a substrate for acylation of GmACP3. The resulting acetyl-ACP could then be an extender unit for further synthesis.

1. Upton, C., and Buckley, J. T. (1995) A new family of lipolytic enzymes? *Trends Biochem. Sci. 20*, 178-179.

2. Akoh, C. C., Lee, G. C., Liaw, Y. C., Huang, T. H., and Shaw, J. F. (2004) GDSL family of serine esterases/lipases. *Prog. Lipid Res. 43*, 534-552.

3. Molgaard, A., Kauppinen, S., and Larsen, S. (2000) Rhamnogalacturonan acetylesterase elucidates the structure and function of a new family of hydrolases. *Structure 8*, 373-383.

4. Kiriukhin, M. Y., and Neuhaus, F. C. (2001) D-alanylation of lipoteichoic acid: role of the D-alanyl carrier protein in acylation. *J. Bacteriol. 183*, 2051-2058.

5. Aklujkar, M., Krushkal, J., DiBartolo, G., Lapidus, A., Land, M. L., and Lovley, D. R. (2009) The genome sequence of *Geobacter metallireducens*: features of metabolism, physiology and regulation common and dissimilar to *Geobacter sulfurreducens*. *BMC Microbiol. 9*, 109.