Supporting Information

Porphyrins from a metagenomic library of the marine sponge *Discodermia calyx*

Rui He,[a,c] Toshiyuki Wakimoto,*[a] Yuya Takeshige,[a] Yoko Egami,[a] Hiromichi Kenmoku,[b] Takuya Ito,[b] Bochu Wang,*[c] Yoshinori Asakawa,[b] and Ikuro Abe*[a]

[a] *Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. Fax: +81 3 5841 4744; Tel: +81 3 5841 4741; E-mail: abei@mol.f.u-tokyo.ac.jp*
[b] *Faculty of Pharmaceutical Sciences, Tokushima Bunri University, Yamashiro-cho, Tokushima 770-8514, Japan. Fax: 81 88 655 3051; Tel: 81 88 622 9611; E-mail: asakawa@ph.bunri-u.ac.jp*
[c] *Bioengineering College, Chongqing University, No. 174, Shanpingba Main Street, Chongqing, 400030, P. R. China. Fax: 86 23 65112877; Tel: 86 23 65112840; E-mail: wangbc2000@126.com*

Comparative data of porphyrin production in the clones

Negative control means the *E. coli* carrying void vector. After each broths of three clones were extracted with HP-20, and then subjected to DAD-HPLC analysis. HPLC analysis was performed on ODS column (Cosmosil 5C18 PAQ waters, 4.6 x 250 mm) with a mixture of H_2O and MeCN, both containing 0.1% acetic acid: 0–5 min, 5% MeCN; 5–25 min, 5–100% MeCN; 25–30 min, 100% MeCN; and 30–38 min, 5% MeCN. 0.8mL/min. DAD profile were measured with an Shimadzu HPLC System: LC-20AD and SPD-20A Prominence Diode Array Detector.

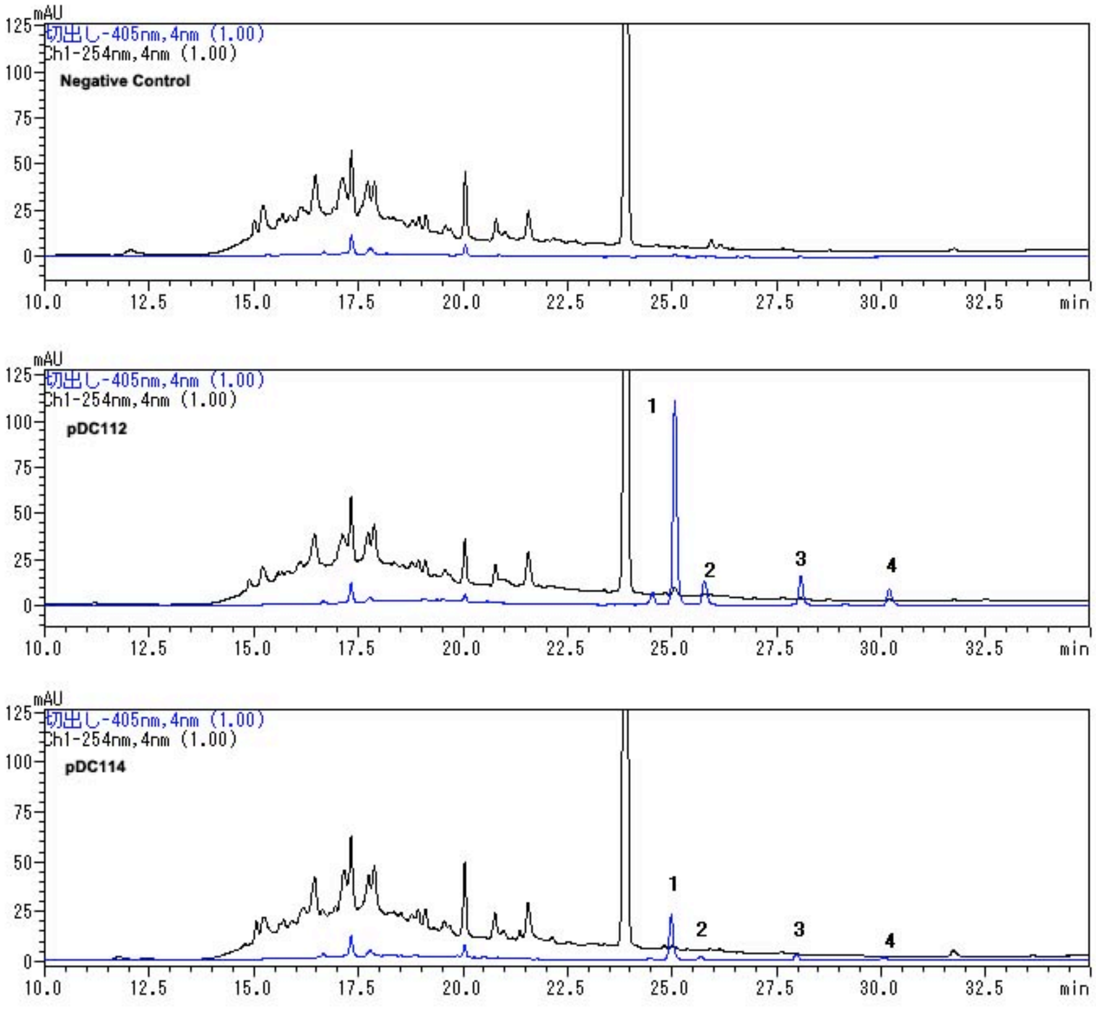
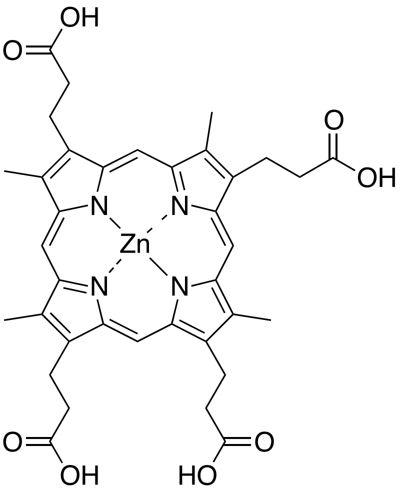


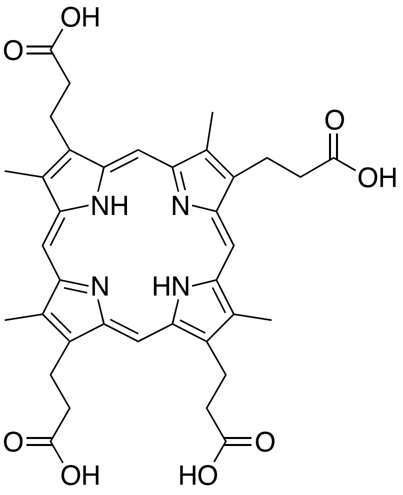
Figure S1. ODS HPLC profile of each extract of negative control, pDC112 and pDC114.

LC-MS data of coproporphyrin III

LC-MS (Agilent 1100 series-Bruker esquire 4000) analysis was performed on ODS column (TSK-Gel ODS-80Ts, 4.6 x 150 mm) with a mixture of H_2O and MeCN, both containing 0.1% acetic acid: 0–30 min, 50-90% MeCN; 30-40 min, 90-100% MeCN; 40-50 min, 100% MeCN. 0.2mL/min. Detected wavelength: 400nm. Positive ESI.



Zn-coproporphyrin III (**1**)



Coproporphyrin III (**2**)

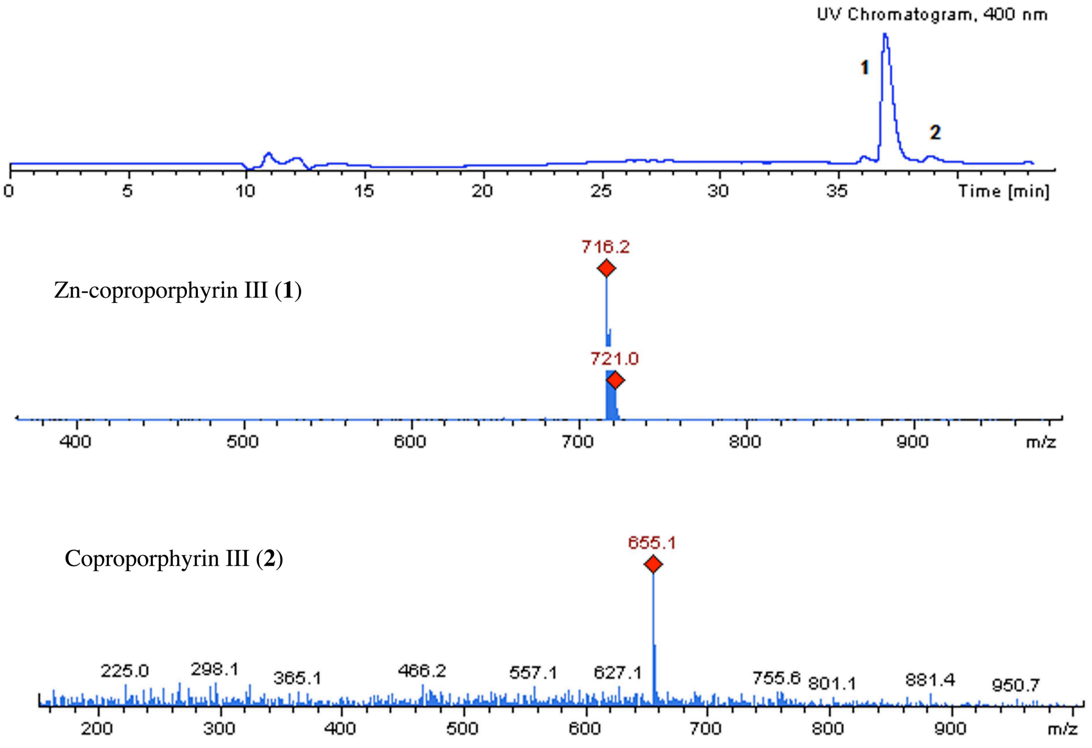
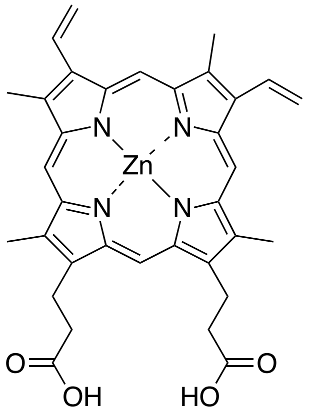


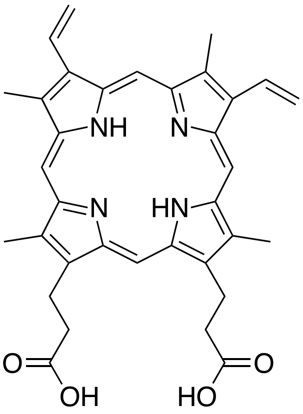
Figure S2. LC-MS data of Zn-coproporphyrin III and coproporphyrin III.

LC-MS data of protoporphyrin IX

Zn-protoporphyrin IX (**3**): HRESITOFMS *m/z* 625.1804 [M+H]+ (calcd. for $C_{34}H_{33}N_4O_4{}^{64}Zn$, 625.1793).

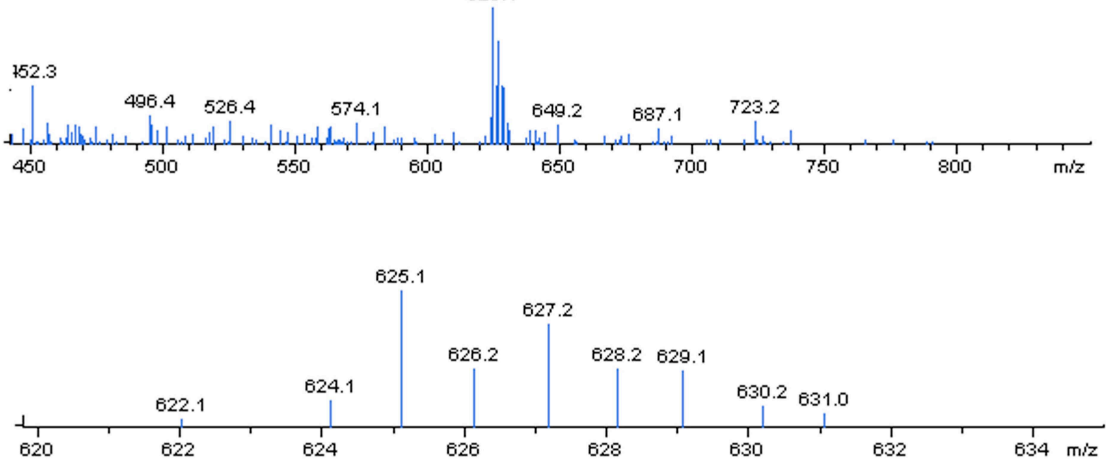


Zn-protoporphyrin IX (**3**)



Protoporphyrin IX (**4**)

Zn-protoporphyrin IX (**3**)



Protoporphyrin IX (**4**)

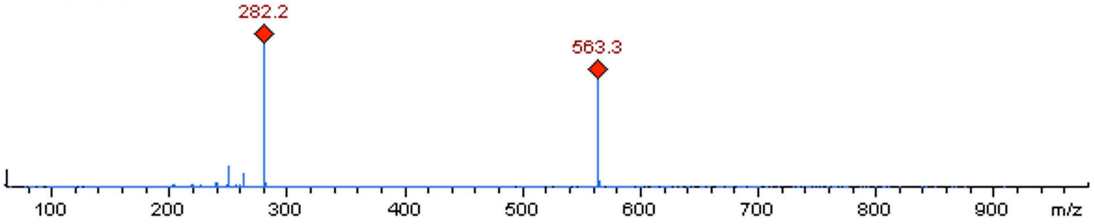


Figure S3. LC-MS data of Zn-protoporphyrin IX and protoporphyrin IX.

NMR data of Zn-coproporphyrin III

Blue and purple values indicate ^{1}H and ^{13}C chemical shifts, respectively.

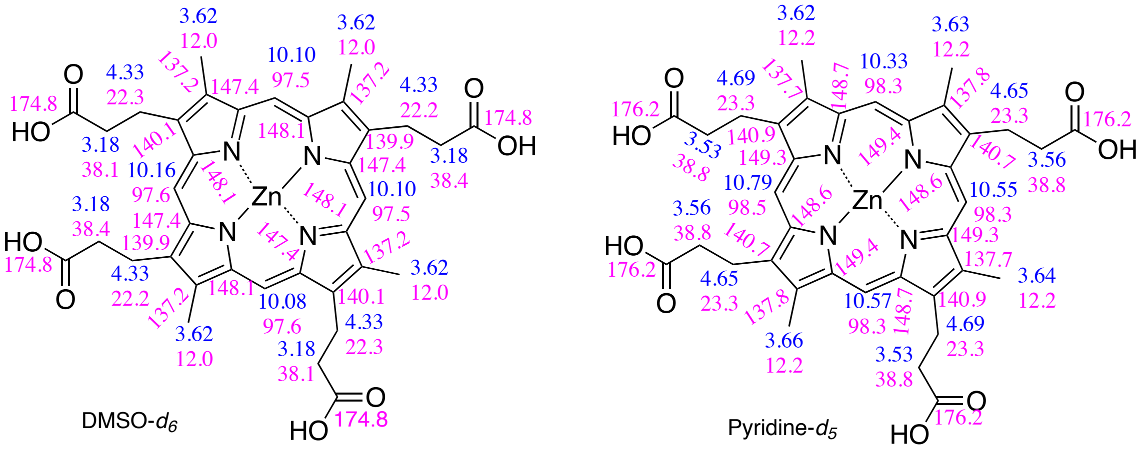


Figure S4. ^{1}H and ^{13}C chemical shifts of Zn-coproporphyrin III.

Table S1. Proposed ORFs in pDC112 (40.639kb).

	Size (aa)	Homology	E value	Identity (%)
1	159	unnamed protein product [*Dechlorosoma suillum* PS]	9e-31	52
		multi-sensor signal transduction histidine kinase	8e-22	43
2	492	unnamed protein product [*Methylomonas methanica* MC09]	1e-66	34
		diguanylate cyclase [*Nitrosococcus halophilus* Nc4]	3e-58	29
		hybrid signal transduction histidine kinase and diguanylate cyclase/phosphodiesterase [*Nitrosococcus oceani* ATCC 19707]	1e-54	31
3	552	multi-sensor hybrid histidine kinase	1e-135	45
4	195	putative transmembrane protein [NC10 bacterium 'Dutch sediment']	1e-31	39
		unnamed protein product [*Geobacter metallireducens* GS-15]	9e-27	36
		hypothetical protein N47_E47780 [uncultured *Desulfobacterium* sp.]	4e-25	37
5	560	unnamed protein product [*Geobacter metallireducens* GS-15]	0	53
		TonB-dependent receptor, plug [NC10 bacterium 'Dutch sediment']	2e-171	49
		similar to ferric enterobactin receptor FepA [*Candidatus Kuenenia stuttgartiensis*]	5e-157	51
6	971	hypothetical protein Acid345_1683 [*Candidatus Koribacter versatilis*	0	47
		Ellin345]	0	48
		conserved hypothetical protein, membrane [*Candidatus Poribacteria* sp.	0	46
		WGA-A3]	4e-74	57
		unnamed protein product [*Candidatus Solibacter usitatus* Ellin6076]		
		cytochrome c-like protein [uncultured bacterium]		
7	321	serine/threonine kinase [*Streptomyces griseoaurantiacus* M045]	6e-17	31
		conserved hypothetical protein [*Streptomyces ambofaciens* ATCC 23877]	2e-16	32
		sulfatase modifying factor [*Streptomyces sviceus* ATCC 29083]	2e-16	31
		hypothetical protein Sfla_0405 [*Streptomyces flavogriseus* ATCC 33331]	2e-15	31
		putative domain of unknown function (DUF323) [uncultured marine	9e-14	43
		crenarchaeote HF4000_APKG3K8]	93-13	45
		methyltransferase [*Cyanothece* sp. PCC 7425]		
8	583	unnamed protein product [*Desulfobacterium autotrophicum* HRM2]	3e-09	41
		hypothetical protein Hoch_5883 [*Haliangium ochraceum* DSM 14365]	3e-08	46
9	405	amidohydrolase [*Sebaldella termitidis* ATCC 33386]	3e-105	39
		possible cytosine deaminase [*Enterococcus faecalis* TX1322]	2e-100	40
		amidohydrolase family protein [*Enterococcus faecalis* TX0411]	3e-100	40
10	200	conserved hypothetical protein [*Oscillatoria* sp. PCC 6506]	5e-68	52
		hypothetical protein MicvaDRAFT_2891 [*Microcoleus vaginatus* FGP-2]	4e-67	54
		hypothetical protein MC7420_1683 [*Microcoleus chthonoplastes* PCC 7420]	7e-67	54
11	464	conserved protein of unknown function [NC10 bacterium 'Dutch sediment']	4e-179	56
		hypothetical protein NIDE3054 [*Candidatus Nitrospira defluvii*]	6e-161	51
12	236	putative proteasome A-type and B-type [NC10 bacterium 'Dutch sediment']	5e-94	63
		proteasome subunit alpha [*Candidatus Nitrospira defluvii*]	1e-85	59
13	226	proteasome subunit beta [*Candidatus Nitrospira defluvii*]	1e-93	59
		putative proteasome A-type and B-type [NC10 bacterium 'Dutch sediment']	2e-92	58
14	80	unnamed protein product [*Streptomyces scabiei* 87.22]	5e-16	54
		Prokaryotic ubiquitin-like protein Pup [*Streptomyces zincirESistens* K42]	2e-15	53
		unnamed protein product [*Streptosporangium roseum* DSM 43021]	3e-15	54
15	491	putative proteasome component [NC10 bacterium 'Dutch sediment']	0	54
		hypothetical protein Tbis_1836 [*Thermobispora bispora* DSM 43833]	2e-175	51
		Pup ligase/deamidase [*Frankia* sp. CN3]	3e-169	50
16	611	conserved protein of unknown function [NC10 bacterium 'Dutch sediment']	0	56
		AAA ATPase central domain-containing protein [*Thermomonospora curvata*	0	58
		DSM 43183]	0	58

		ATPase [*Streptomyces* sp. C]	0	58
		ARC [*Streptomyces coelicolor* A3(2)]		
17	261	cytochrome c-type biogenesis protein CcsB [*Geobacter* sp. FRC-32]	3e-44	37
		cytochrome c-type biogenesis protein CcsB [*Desulfurivibrio alkaliphilus* AHT2]	2e-43	38
		unnamed protein product [*Geobacter metallireducens* GS-15]	7e-42	36
		ResC/HemX-like cytochrome c biogenesis membrane protein [*Geobacter bemidjiensis* Bem]	5e-38	35
18	420	hemA gene product [*Geobacter metallireducens* GS-15]	3e-133	50
		glutamyl-tRNA reductase [*Thermodesulfatator indicus* DSM 15286]	3e-132	49
19	322	hemC-porphobilinogen deaminase [endosymbiont of *Riftia pachyptila* (vent Ph05)]	5e-85	49
20	322	hemB-Porphobilinogen synthase [*Thermodesulfatator indicus* DSM 15286]	7e-156	65
21	155	RsmD family RNA methyltransferase [*Eubacterium infirmum* F0142]	2e-25	35
		unnamed protein product [*Clostridium clariflavum* DSM 19732]	2e-24	36
22	355	tRNA-specific 2-thiouridylase MnmA [*Geobacter* sp. FRC-32]	2e-129	54
		unnamed protein product [*Kyrpidia tusciae* DSM 2912]	4e-129	54
23	155	cytoplasmic protein[*Syntrophus aciditrophicus* SB]	8e-27	40
		protein of unknown function DUF1844[*Desulfovibrio fructosovorans* JJ]	7e-20	47
24	638	DNA mismatch repair protein	1e-131	39
25	499	inosine-5'-monophosphate dehydrogenase[*Deferribacter desulfuricans* SSM1]	0	67
26	1152	DNA polymerase III subunit alpha[*Geobacter sulfurreducens* PCA]	0	52
		unnamed protein product [*Geobacter metallireducens* GS-15]	0	51
27	236	transporter, MotA/TolQ/ExbB proton channel family, putative[marine gamma proteobacterium HTCC2148]	8e-36	40
28	145	Biopolymer transport protein ExbD/TolR[*Moritella* sp. PE36]	4e-17	36
29	192	hypothetical protein Sthe_0866 [*Sphaerobacter thermophilus* DSM 20745]	3e-07	33
		N-acetyltransferase GCN5	1e-06	40
30	436	group 1 glycosyl transferase	7e-119	48
31	214	unnamed protein product [*Azospirillum lipoferum* 4B]	4e-32	36
		beta-lactamase domain protein [*Arthrospira maxima* CS-328]	2e-30	39
		metallo-beta-lactamase [*Myxococcus xanthus* DK 1622]	2e-29	39

Table S2. Proposed ORFs in pDC114 (37.753kbp).

	Size (aa)	Homology	E value	Identity
1	175	precorrin-4 C(11)-methyltransferase [*Roseiflexus* sp. RS-1]	4e-35	55
		cobalt-precorrin-4 methyltransferase [*Lactobacillus reuteri*]	1e-31	46
2	391	unnamed protein product [*Kyrpidia tusciae* DSM 2912]	4e-85	41
		cobalamin biosynthesis protein [*Bacillus* sp. 2_A_57_CT2]	3e-82	38
		cbiG gene product [*Geobacillus thermodenitrificans* NG80-2]	3e-81	40
3	358	precorrin-3B methylase [*Cenarchaeum symbiosum* A]	5e-82	48
		precorrin methylase [*Lyngbya* sp. PCC 8106]	9e-82	56
		Cobalamin biosynthesis protein CbiG / Cobalt-precorrin-3b C17-methyltransferase [*Crocosphaera watsonii* WH 0003]	7e-81	52
4	463	hemA gene product [*Carboxydothermus hydrogenoformans* Z-2901]	1e-91	40
		glutamyl-tRNA reductase [*Thermodesulfovibrio yellowstonii* DSM 11347]	9e-90	36
		glutamyl-tRNA reductase [*Geobacter sulfurreducens* PCA]	9e-86	39
5	428	putative aminotransferase class-III [uncultured marine microorganism]	0	75
		glutamate-1-semialdehyde-2,1-aminomutase [*Desulfotomaculum nigrificans* DSM 574]	0	61
		glutamate-1-semialdehyde-2,1-aminomutase [*Moorella thermoacetica* ATCC 39073]	0	62
6	278	hypothetical protein Tter_1481 [*Thermobaculum terrenum* ATCC BAA-798]	2e-28	35
		hypothetical protein OSCT_2474 [*Oscillochloris trichoides* DG6]	4e-18	29
		type 11 methyltransferase [*Roseiflexus* sp. RS-1]	2e-15	34
7	467	amidase [*Pseudomonas syringae* pv. aceris str. M302273PT]	7e-131	45
		unnamed protein product [*Azorhizobium caulinodans* ORS 571]	3e-125	44
		amidase [*Roseiflexus castenholzii* DSM 13941]	3e-124	46
8	87	cytotoxic translational repressor of toxin-antitoxin stability system [*Microbacterium testaceum* StLB037]	1e-15	44
		Plasmid stabilization system [*Beggiatoa* sp. PS]	1e-14	35
		toxin-antitoxin system, toxin component, RelE family [*Actinomyces odontolyticus* F0309]	1e-12	39
9	363	hypothetical protein PM8797T_00417 [*Planctomyces maris* DSM 8797]	3e-91	42
		hypothetical protein DSM3645_21042 [*Blastopirellula marina* DSM 3645]	6e-83	44
		6-phosphogluconolactonase [*Herbaspirillum seropedicae* SmR1]	1e-66	37
10	492	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate	3e-111	43
		dehydrogenase complex (E2) [*Candidatus Poribacteria* sp. WGA-A3]	1e-98	41
		unnamed protein product [*Desulfovibrio desulfuricans* ND132]	3e-96	38
11	327	unnamed protein product [*Sphingomonas wittichii* RW1]	1e-116	57
		pyruvate dehydrogenase E1 component, beta subunit [*Phenylobacterium zucineum* HLK1]	2e-114	53
		transketolase central region [*Frankia* sp. EuI1c]	6e-109	55
12	352	unnamed protein product [*Sphingomonas wittichii* RW1]	2e-120	56
		pyruvate dehydrogenase (acetyl-transferring) [*Frankia* sp. EuI1c]	4e-119	54
		pyruvate dehydrogenase [*Halanaerobium hydrogeniformans*]	6e-119	53
13	148	Cupin 2 barrel domain-containing protein [*Acidovorax avenae* subsp. avenae ATCC 19860]	2e-10	34
		unnamed protein product [*Acidovorax citrulli* AAC00-1]	2e-10	36
		unnamed protein product [*Oceanithermus profundus* DSM 14977]	2e-10	35
14	408	Linalool 8-monooxygenase [*Mycobacterium rhodesiae* JS60]	7e-129	47
		Peroxidase [*Mycobacterium rhodesiae* JS60]	4e-126	47
		peroxidase [*Saccharomonospora azurea* SZMC 14600]	1e-79	38
15	533	unnamed protein product [*Thermomicrobium roseum* DSM 5159]	8e-165	48
		thiamine pyrophosphate domain-containing TPP-binding protein [*Sphaerobacter thermophilus* DSM 20745]	3e-164	49

		acetolactate synthase, large subunit protein (thiamine pyrophosphate-dependent enzyme) [alpha proteobacterium BAL199]	1e-136	44
16	251	glutamine amidotransferase [*Prosthecochloris aestuarii* DSM 271]	2e-22	33
		glutamine amidotransferase class-I [*Actinosynnema mirum* DSM 43827]	2e-22	39
17	431	TRAP dicarboxylate transporter subunit DctP [*Dinoroseobacter shibae* DFL 12]	2e-25	26
		unnamed protein product [*Shewanella woodyi* ATCC 51908]		
		extracellular solute-binding protein [*Pusillimonas* sp. T7-7]	1e-23	26
			2e-22	30
18	280	hypothetical protein STH42 [*Symbiobacterium thermophilum* IAM 14863]	8e-09	25
		hypothetical protein SSPSH_13162 [*Salinisphaera shabanensis* E1L3A]	1e-07	32
		hypothetical protein cje21_00447 [*Campylobacter jejun*i subsp. jejuni 1997-7]	3e-06	23
19	271	ABC transport system protein [uncultured candidate division OP1 bacterium]	2e-109	62
		ABC transporter [*Sphaerobacter thermophilus* DSM 20745]		
		putative ABC transporter [uncultured marine microorganism HF4000_007D16]	1e-100	59
			1e-91	52
20	190	unnamed protein product [*Chromohalobacter salexigens* DSM 3043]	4e-70	53
		branched-chain amino acid transport system ATP-binding protein [uncultured candidate division OP1 bacterium]	1e-68	59
		ABC transporter [*Crocosphaera watsonii* WH 8501]	1e-65	54
21	344	hydrophobic amino acid ABC transporter permease [uncultured candidate division OP1 bacterium]	7e-106	55
		hydrophobic amino acid ABC transporter permease [*Synechococcus* sp. JA-3-3Ab]	2e-71	46
22	341	unnamed protein product [*Chromohalobacter salexigens* DSM 3043]	6e-91	49
		hydrophobic amino acid ABC transporter permease [*Synechococcus* sp. JA-2-3B'a(2-13)]	6e-84	51
		putative amino acid or sugar ABC transport system, permease protein [*Labrenzia alexandrii* DFL-11]	3e-79	43
23	464	extracellular ligand-binding receptor [*Sphaerobacter thermophilus* DSM 20745]	8e-119	51
		hydrophobic amino acid ABC transporter amino acid-binding protein [*Synechococcus* sp. JA-3-3Ab]	4e-94	44
		hydrophobic amino acid ABC transporter periplasmic amino acid-binding protein [*Synechococcus* sp. JA-2-3B'a(2-13)]	5e-91	43
24	513	hydantoinase/oxoprolinase [*Thermobispora bispora* DSM 43833]	0	57
		hydantoinase/oxoprolinase [*Streptomyces* sp. AA4]	0	57
		hydantoinase/oxoprolinase [*Saccharopolyspora erythraea* NRRL 2338]	0	57
25	425	conserved hypothetical protein [*Octadecabacter antarcticus* 307]	1e-115	50
		unnamed protein product [*Thermomicrobium roseum* DSM 5159]	1e-110	53
		hypothetical protein Desca_2498 [*Desulfotomaculum carboxydivorans* CO-1-SRB]	2e-106	46
26	159	unnamed protein product [*Streptomyces cattleya* NRRL 8057]	4e-23	39
		glyoxalase/bleomycin resistance protein/dioxygenase [*Streptomyces cattleya* DSM 46488]	5e-23	39
		Glyoxalase/bleomycin resistance protein/dioxygenase [*Halobacterium* sp. DL1]	2e-18	41
27	486	4-hydroxyphenylacetate 3-monooxygenase, oxygenase subunit [*Sphaerobacter thermophilus* DSM 20745]	0	55
		unnamed protein product [*Kyrpidia tusciae* DSM 2912]	1e-170	49
		unnamed protein product [*Thermaerobacter marianensis* DSM 12885]	6e-165	49
28	524	putative GMC oxidoreductase [uncultured marine microorganism HF4000_ANIW93N21]	5e-151	46
		glucose-methanol-choline oxidoreductase [*Geodermatophilus obscurus* DSM 43160]	4e-93	40
29	163	unnamed protein product [*Cupriavidus necator* N-1]	6e-71	66

		hypothetical protein H16_B0885 [*Ralstonia eutropha* H16]	1e-69	64
		hypothetical protein RALTA_A1751 [*Cupriavidus taiwanensis* LMG 19424]	5e-64	67
30	60	unnamed protein product [*Sulfolobus islandicus* M.14.25]	4e-07	50
		haloacid dehalogenase-like hydrolase [*Sulfolobus acidocaldarius* DSM 639]	5e-07	50
		haloacetate dehalogenase [*Sulfolobus solfataricus* P2]	6e-07	48
31	498/1497	unnamed protein product [*Shewanella woodyi* ATCC 51908]	3e-47	28
		4-hydroxyphenylacetate 3-monooxygenase, oxygenase component [*Streptomyces* sp. AA4]	4e-44	28
		Chain A, Crystal Structure Of The Mutant Of Hpab (T198i, A276g, And R466h)	5e-39	30
32	270/813	unnamed protein product [*Haliangium ochraceum* DSM 14365]	3e-51	39
		aldo/keto reductase [*Methylobacterium nodulans* ORS 2060]	2e-47	40
		unnamed protein product [*Azotobacter vinelandii* DJ]	1e-44	37