BIOINFORMATION

open access

Supplementary material:

Table 1: Physicochemical properties of hypothetical proteins by Protparam tool

Sequence ID	No of aa	MW	pI	(-) R	(+ R)	EC	II	AI	GRAVY
gi|166409299	97	10407.8	10.11	1	12	25440	43.32	65.36	-0.182
gi|166409303	129	15748.3	10.14	13	31	22920	41.52	83.8	-0.877
gi|166409302	208	23392.4	9.29	27	34	7450	22.22	115.72	-0.148
gi|166409301	103	12163.4	9.73	11	20	11920	22.96	113.4	-0.287
gi|166409300	644	75501.5	9.14	60	73	77825	35.81	119.1	0.128
gi|390516769	31	3677.5	9.3	3	5	1490	7.98	138.39	0.726
gi|166409293	139	15938.3	9.37	12	18	13075	36.4	95.4	-0.443
gi|390516759	209	24226.7	9.25	23	32	19495	23.33	91.87	-0.612
gi|390516760	80	9250.4	4.76	15	11	4470	64.89	101.12	-0.611
gi|166409294	323	35653	6.25	36	34	40340	30.12	105.82	0.004

Table 2: Identification of domains by CDD-BLAST

Sequence ID	Domains
gi|166409299	Lactococcin_972 super family
gi|166409303	Mob_Pre super family
gi|166409302	ABC_MJ0796_Lo1CDE_FtsE, L_ocin_972_ABC
gi|166409301	DUF1093 super family
gi|166409300	DUF1430 super family, COG4652
gi|166409293	HTH_MARR, HTH super family
gi|166409294	MDR_yhdh_yhfp, oxido_YhdH

Table 3: Functional description of superfamilies of hypothetical proteins

Superfamily	Description
Lactococcin_972	Represent bacteriocins related to lactococcin. Associates with a seven transmembrane putative immunity protein.
Mob_Pre	With some plasmids, recombination can occur in a site specific manner that is independent of RecA. In such cases, the recombination event requires another protein called Pre. Pre is a plasmid recombination enzyme. This protein is also known as Mob (conjugative mobilisation).
L_ocin_972_ABC	putative bacteriocin export ABC transporter, lactococcin 972 group; A gene pair with a fairly wide distribution consists of a polypeptide related to the lactococcin 972 and multiple-membrane-spanning putative immunity protein. This model represents a small clade within the ABC transporters that regularly are found adjacent to these bacteriocin system gene pairs and are likely serve as export proteins.
ABC_MJ0796_Lo1CDE_FtsE	This family is comprised of MJ0796 ATP-binding cassette, macrolide-specific ABC-type efflux carrier (MacAB), and proteins involved in cell division (FtsE), and release of liporoteins from the cytoplasmic membrane (LolCDE). They are clustered together phylogenetically. MacAB is an exporter that confers resistance to macrolides, while the LolCDE system is not a transporter at all. An FtsE null mutants showed filamentous growth and appeared viable on high salt medium only, indicating a role for FtsE in cell division and/or salt transport. The LolCDE complex catalyses the release of lipoproteins from the cytoplasmic membrane prior to their targeting to the outer membrane.
DUF1093	Proteins of unknown function
DUF1430	This family represents the C-terminus (approximately 120 residues) of a number of hypothetical bacterial proteins of unknown function. These are possibly membrane proteins involved in immunity.
COG4652	Uncharacterized protein conserved in bacteria [Function unknown]
HTH	Helix-turn-helix domains; A large family of mostly alpha-helical protein domains with a characteristic fold; most members function as sequence-specific DNA binding domains, such as in transcription regulators. This superfamily also includes the winged helix-turn-helix domains.
HTH_MARR	helix_turn_helix multiple antibiotic resistance protein
oxido_YhdH	putative quinone oxidoreductase, YhdH/YhfP family; a superfamily in which some members are zinc-binding medium-chain alcohol dehydrogenases while others are quinone oxidoreductases with no bound zinc. This subfamily includes proteins studied crystallographically for insight into function: YhdH from Escherichia coli and YhfP from Bacillus subtilis. Members bind NADPH or NAD, but not zinc.
MDR_yhdh_yhfp	Yhdh and yhfp-like putative quinone oxidoreductases; QOR catalyzes the conversion of a quinone + NAD(P)H to a hydroquinone + NAD(P)+. Quinones are cyclic diones derived from aromatic compounds. Membrane bound QOR actin the respiratory chains of bacteria and mitochondria, while soluble QOR acts to protect from toxic quinones (e.g. DT-diaphorase). QOR reduces quinones through a semi-quinone intermediate via a NAD(P)H-dependent single electron transfer. QOR is a member of the medium chain dehydrogenase/reductase family, but lacks the zinc-binding sites of the prototypical alcohol dehydrogenases of this group. NAD(P)(H)-dependent oxidoreductases are the major enzymes in the interconversion of alcohols and aldehydes, or ketones.

Table 4: Prediction of Subcellular localization sites in hypothetical proteins by PSORTB

Sequence ID	Localization
gi|390516760	Cytoplasmic
gi|390516759	Unknown
gi|166409293	Cytoplasmic
gi|166409299	CytoplasmicMembrane
gi|166409303	Unknown
gi|166409302	CytoplasmicMembrane
gi|390516769	CytoplasmicMembrane
gi|166409301	Unknown
gi|166409300	CytoplasmicMembrane
gi|166409294	CytoplasmicMembrane

Table 5: Families found by PFAM database

Sequence ID	Pfam-A	Pfam-B	Domain
gi|166409294	ADH_zinc_N	-	ADH_N
gi|390516760	DUF885	-	YolD
gi|390516759	-	Pfam-B 7040	-
gi|166409293	MarR	Pfam-B_33	-
gi|390516769	DUF1430	Pfam-B_2627	
gi|166409300	DUF1430		
gi|166409301	Tmemb_9 DUF1312 DUF1093	Pfam-B_15664	
gi|166409302		Pfam-B_16450	ABC_tran
gi|166409303	Mob_Pre		
gi|166409299	Lactococcin_972		

Table 6: Descriptions of Pfam families of hypothetical proteins

Sequence ID	Description
gi|166409294	Alcohol dehydrogenase GroES-like domain Zinc-binding dehydrogenase family
gi|390516760	YolD-like protein
gi|166409293	MAR family; multiple antibiotic resistance
gi|166409301	TMEM9; widely expressed and localised to the late endosomes and lysosomes
gi|166409302	ABC transporters belong to the ATP-Binding Cassette superfamily

Table 7: SOSUI result of hypothetical proteins

Sequence ID	N-terminal	Transmembrane region	C-terminal	Type	Length
gi|166409299	42	FVSSCIASTILFGTLLGVTYKAE	64	PRIMARY	23
gi|166409301	41	TLIIVTIVLLIIILSLLLVRN	61	PRIMARY	21
gi|166409300	184	ILFFELVIDNDLLVPFIFLGVLY	206	SECONDARY	23
	242	KIIYWLVLTTIFFIANILIIHIA	264	PRIMARY	23
	279	VLLFVFACITTLLWLISYSLLLK	301	PRIMARY	23
	317	SIFMGTFTKCIMVLILSFLLAQN	339	SECONDARY	23
	575	IITIVTIIINVFILLIATVFEII	597	PRIMARY	23
	633	IMLAYTTHILFGSKVLLFIIMSI	655	PRIMARY	23

Table 8: Hypothetical proteins interacting with functionally important proteins

Sequence ID	Interacting proteins
gi|166409299	arsenate reductase
gi|166409302	elongation factor G 30S ribosomal protein S7 30S ribosomal protein S5
gi|166409301	ABC transporter ATP-binding protein
gi|166409300	cysteine protease precursor
gi|390516769	type I restriction-modification system endonuclease
gi|166409293	carboxy-terminal processing proteinase CtpA
gi|166409294	regulatory protein MarR

Table 9: Templates used by PS2 server for modeling

Sequence ID	Templates
gi|166409300	1rh1A
gi|166409294	1tt7A
gi|166409293	1z91A
gi|166409302	2it1A

Table 10: Residues involved in ligand binding sites predicted by QSITE finder

Sequence ID	Site volume	Residues
gi|166409293	209	SER 14, ARG 17, VAL 18, PHE 19, HIS 21, PHE 22, LEU 111, ASN 112, PHE 116, ALA 117, ASN 118, LEU 119.
gi|166409294	419	ILE 40, ASN 41 & 243, TYR 42, ALA 128, GLY 156, 159,160 & 162, 242, 313 & 317, VAL 161 & 244, THR 219, CYS 241, GLN 309 & 314, LEU 310 & 311, LYS 312, HIS 315, ARG 318, PHE 152, SER 153.
gi|166409300	169	GLY 56, LEU 57 & 58, LYS 60, ASN 61, ILE 64, TYR 90, ASN 148, VAL 149, PHE 152.
gi|166409302	241	GLU 157, THR 159, GLY 160, LEU 162, ASP 163, THR 164, GLY 167, LYS 168, ILE 171, THR 188, HIS 189, ASP 190, GLU 192, LEU 193.