Supplementary Table S1. Detailed Information on the Cell Surface Biogenesis Components.[1]

Phylum	Bacteria	Eukaryotes	Euryarchaeota							Crenarchaeota			Micrarchaeota	Nanoarchaeota	Nanohaloarchaeota	Korarchaeota	Parvarchaeota	Thaumarchaeota	Lokiarchaeota
Species[2]	*E. coli*	*S. cerevisia*	*H. volcanii*	*M. maripaludis*	*M. mazei*	*T. kodakarensis*	*M. kandleri*	*M. thermoautotr.*	*A. fulgidus*	*M. jannaschii*	*S. acidocaldarius*	*A. pernix*	*T. tenax*	*Micrarchaeum*	*Nanoarchaeum*	*Haloredivivus*	*Korarchaeum*	*Parvarchaeum*	*Nitrosopumilus*	*Lokiarchaeum*
SRP RNA	4.5S	7S	7S	7S	7S	7S	7S	7S	7S	7S	7S	7S	-	-	-	-	7S	-	7S	-
Ffh/SRP54	b2610	YPR088C	0123	1551	1267	1486	1612	1321	0622	0101	1310	1735	0615	00530	-	02176	1590	0759	0531	47630
	P0AGD7	P20424	Q977V2	Q6LX03	Q8PXF3	Q5JJC8	Q8TUY9	O27376	O29633	Q57565	O07853	Q9YB62	G4RNY1	A0A1L9GXP0	-	H0AC09	B1L7A6	D6GW84	A9A214	A0A0F8VYB3
SRP19	-	human[3]	1109	1107	1557	1330	1551	165	1258	1034	0759	0772	2083	-	-	-	1508	-	0519	41090, 32750
	-	P09132	D4GW40	Q6LY84	Q8PWM7	Q5JGS5	Q8TV49	O26267	O29010	Q58440	Q4JAP4	Q9YDZ7	G4RML9	-	-	-	B1L725	-	A9A202	A0A0F8XZV0, A0A0F8XKW5
FtsY/SRα	b3464	YDR292C	0120	0683	0810	0653	1613	1608	2062	0291	1462	1081,1	1809	02405	-	00416	1330	1106	0396	13700
	P10121	P32916	D4GYW6	Q6LZE6	Q8PYQ0	Q5JF74	Q8TUY8	O27645	O28217	Q57739	P27414	Q9YD31	G4RLI5	A0A1L9GW55	-	H0AD11	B1L6J7	D6GW44	A9A593	A0A0F8VKN8
SRβ	-	YKL154W	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	P36057	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SecB	b3609	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	P0AG86	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SecA	b0098	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	P10408	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YidC/DUF106	b3705	-	2495, A0169	1030	2152	1516	0024	28	1901	0480	-	-	-	04185	-	-	1574	0996	0404	10200
	P25714		D4GTQ2, D4GQJ8	Q6LYG1	Q8PV22	Q5JJH3	Q8TZB1	O26136	O28378	Q57904				A0A1L9GU27			B1L790	D6GWX5	A9A5A1	A0A0F8YB98
SecY/Sec61α	b3300	YLR378C	2541, A0174	1422	2147	1518	0026	26	1902	0478	0574	0983.1	1416	00495	168	03514	1573	0997	0402	00020
	P0AGA2	P32915	Q977V3, D4GQK3	Q6LXD0	Q8PV27	Q5JJH1	Q8TZA9	O26134	O28377	Q60175	P49978	Q9YDD0	G4RKF3	A0A1L9GXQ7	Q74NB9	H0AF96	B1L789	D6GWX6	A9A599	A0A0F8VQL6
SecE/Sec61γ	b3981	YDR086C	0718	1435	1009	1420	0822	1677	0536	0371	1461	2176a	1808	04985	375a	-	-	0482	notAnno	-
	P0AG96	P35179	D4GT81	Q6LXB7	Q8PY55	Q5JH32	Q8TX54	O27713	O29714	Q57817	P27340	P58198	G4RLI4	A0A1L9GT68	notAnno	-	-	D6GVG9	notAnno	-
SecG	b3175	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	P0AG99	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sec61β	-	YER087C-B, YER019C-A	0437	1428	1372	1952	0208	1536.1	1776	0912.1	0729	0650a.1	1720	-	-	01324	-	0425	notAnno	43700
		P52870, P52871	D4GRX8	Q6LXC4	Q8PX50	Q5JDK7	Q8TYT4	P60461	O28498	P60460	P60464	Q9YEC1	G4RL96	-	-	H0A9J8	-	D6GVB3	notAnno	A0A0F8VAE6
SecD	b0408	-	1976	1523	1425	1594	1009	849	-	0111	-	-	-	04965	436	02800	-	0324	-	-
	P0AG90	-	D4GTK5	Q6LX30	Q8PWZ7	Q5JDS5	Q8TWM4	O26937	-	Q57575	-	-	-	A0A1L9GT97	Q74M77	H0ADH4	-	D6GV23	-	-
SecF	b0409	-	1975	1524	1424	1593	1008	848	-	1253	-	-	-	04970	437	02805	-	0325	-	-
	P0AG93	-	D4GTK4	Q6LX29	Q8PWZ8	Q5JDS6	Q8TWM5	O26936	-	Q58650	-	-	-	A0A1L9GT55	Q74M76	H0ADH5	-	D6GV24	-	-
YajC	b0407	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	P0ADZ7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bip	-	YJL034W	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	P16474	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

TatC	b3839	-	0185, 0186	-	2586, 2585	-	-	-	1512, 1344	-	0254	2154,1	1059	-	-	-	0591	-	0774	-
	P69423	-	D4GZC9, D4GZD0	-	Q8PTX2, Q8PTX3	-	-	-	O28760, O28925	-	Q4JC08	Q9Y9Y5	G4RJF8	-	-	-	B1L4G5	-	A9A5F9	-
TatA	b3837, b0627	-	1027, 1162	-	2584, 2583	-	1201	-	2056	-	0256	2154a, 2139a	2052	-	-	-	1126	-	0773, 1318	-
	P69428, P0A843	-	D4GVK4, D4GWC8	-	Q8PTX4, Q8PTX5	-	Q8TW36	-	O28223	-	Q4JC06	Q05DY3, Q05DY4	G4RM70	-	-	-	B1L5Z3	-	A9A5F8, A9A2H5	-
TatB	b3838	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	P69425	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SPI	b2568	YIR022W	0002, 2603	0387	1344	2037, 1703	1031	1448	1791, 1655, 1657, 2078	0260	1795, 1380	1796.1	1710	04470	432	-	1465	0740	0208	-
	P00803	P15367	D4GYJ8, D4GUC5	Q6M084	Q8PX78	Q5JDG1, Q5JJ10	Q8TWK2	O27497	O28483, O28616, O28618, O28201	Q57708	Q4J7Y0, Q4J917	Q9YAZ9	G4RL86	A0A1L9GU74	Q74MX5	-	B1L6Y2	D6GW64	A9A2X6	-
sortase	EP54_00580[4]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Q9S446	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Xrt; archaeosortase[5] (ArtA, ArtC, ArtE)	NE1797[6]	-	0915(A)	1125(E)	1401(A), 2114(C)	-	-	-	2046(A)	O28233(E)	-	-	-	-	-	-	-	-	-	-
	Q82TS7	-	D4GUZ4	Q6LY66	Q8PX21, Q8PV60	-	-	-	O28233	Q60336	-	-	-	-	-	-	-	-	-	-
Lgt	b2828	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	P60955	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lnt	b0657	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	P23930	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SPII[7]	b0027	-	n.i.	n.i.	n.i.	n.i.	-	n.i.	n.i.	n.i.	-	-	n.i.	-	-	n.i.	-	-	-	n.i.
	P00804	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AglB/PglB/STT3	CJE1267[8]	YGL022W	1530	1424	0646, 0647, 2210	0810, 1718	-	1623, 1898, 1906, 420	0380, 0329, 0040	1525	1274	-	0519	04060	155	02810	1056	0254	0075	28460
	Q5HTX8	P39005	D4GYH4	Q6LXC8	Q8PZ48, Q8PZ47, Q8PUW8	Q5JH70, Q5JJ26	-	O27660, O27920, O27928, O26520	O29867, O29918, O30195	Q58920	Q4J9B4	-	G4RNP0	A0A1L9GU13	Q74MN3	H0ADH6	B1L5S3	D6GUV6	A9A114	A0A0F8Y4P4
PilD/PibD/EppA	b0108	-	2993	0555, 0232	1678	0053	0703	421	0936	0902, 1282.1, 0835.2	-	-	-	-	-	00391	-	0088	-	-
	P36647	-	D4GY85	Q6LZR9, Q6M0N8	Q8PWB1	Q5JED9	Q8TXH1	O26521	O29326	Q58312, P81318, P81324	-	-	-	-	-	H0AD06	-	D6GUF4	-	-
ArlH (FlaH)	-	-	1216	1674	0318, 0414	0047	-	-	1050	0899	1174	1898	-	-	-	-	-	-	-	-
	-	-	D4GWY9	Q6LWN5	Q8Q020, Q8PZS8	Q5JEE5	-	-	O29212	Q58309	Q4J9K9	Q9YAP5	-	-	-	-	-	-	-	-
ArlI[9] (FlaI)	-	-	1217	1675	0317, 0413	0048	-	-	1049	0900	1173	1896.1	-	-	-	-	-	-	-	-
	-	-	D4GWZ0	Q6LWN4	Q8Q021, Q8PZS9	Q5JEE4	-	-	O29213	Q58310	Q4J9L0	Q9YAP7	-	-	-	-	-	-	-	-
ArlJ[10] (FlaJ)	-	-	1218	1676	0316, 0412	0049	-	-	1048	0901	1172	1895.1	-	-	-	-	-	-	-	-
	-	-	D4GWZ2	Q6LWN3	Q8Q022, Q8PZT0	Q5JEE3	-	-	O29214	Q58311	Q4J9L1	Q9YAP8	-	-	-	-	-	-	-	-
PilB[9]	b0107, b3326	-	0620, 0748, 1034, 1160, 2385	0040	0938, 1516, 2298	1853	0707	1702	0659, 0996	0781, 1287, 1288	1494, 2317	2208.1	0895, 1134	02725	169, 425	01701, 03986	-	0031, 0601, 1030	-	-
	P36645, P45759	-	D4GSL6, D4GTC5,	Q6M177	Q8PYC5, Q8PWR0,	Q5JEN0	Q8TXG7	O27737	O29598, O29266	Q58191, Q58683,	Q4J8S4, Q4J6H9	Q9Y9T0	G4RPQ4, G4RJM9	A0A1L9GVK6	Q74NB8, Q74MY0	H0AA38, H0A9H4	-	D6GUA6, D6GVT4,	-	-

			D4GVL0, D4GWC7, D4GWN3		Q8PUN4					Q58684								D6GX06		
PilC[9]	b0106, b3327	-	0619, 0749, 1033, 1159, 2386	0038, 0039	0937, 1515, 2297	1851, 1852	0705, 0706	1703	0658, 0995	0779, 0780, 1286	1495, 2318	2207.1, 2206.1	0894, 1133, 1135	02735, 02740	267, 268, 285	01696, 03976, 03981	-	0032, 0033, 0600, 1029	-	-
	P36646, P41441	-	D4GSL5, D4GTC7, D4GVK9, D4GWC6, D4GWN4	Q6M179, Q6M178	Q8PYC6, Q8PWR1, Q8PUN5	Q5JEN2, Q5JEN1	Q8TXG9, Q8TXG8	O27738	O29599, O29267	Q58189, Q58190, Q58682	Q4J8S3, Q4J6H8	Q9Y9T1, Q9Y9T2	G4RPQ3, G4RJM8, G4RJN0	A0A1L9GVF7, A0A1L9GVK2	Q74NG7, Q74NG6, Q74MT4	H0AA37, H0A9H2, H0A9H3	-	D6GUA7, D6GUA8, D6GVT3, D6GX05	-	-
SLG[11]	-	-	2072	0383	1976	0895	-	716	-	0822	2355 (SlaA) 2354 (SlaB)	0609.1	1887	-	300	-	-	-	-	-
	-	-	P25062[12]	Q6M088[13]	Q8PvI7[12]	Q5Ji58[14]	-	O26812[15]	-	Q58232[12]	Q4J6E5[12] Q4J6E6[12]	Q9YEG7[14]	G4RLQ9[14]	-	Q74MU7[15]	-	-	-	-	-

[1] Components were identified by an iterative BLASTp analysis against a database consisting of the proteomes represented in this table. For each component, the analysis was initiated by sequences from *Escherichia coli, Saccharomyces cerevisiae,* and *Haloferax volcanii* (if applicable). BLASTp hits better than e-3 were inspected and homologs were manually selected. Identified homologs were then used for subsequent BLASTp analyses until no additional homologs were identified. For each homolog, the locus tag (upper value) and the UniProt accession (lower values) is provided. To confirm absence of components from archaea outside the phyla Euryarchaeota and Crenarchaeota, text searches were performed in UniProt with a limited set of spelling variants (Feb-2018). It should be noted that the seeming absence of a component might be due to the draft nature of the underlying genome sequence or to an incomplete annotation. In few cases, an annotation incompleteness could be resolved by tBLASTn analyses against the nonredundant database with restriction to the corresponding organism ("notAnno"). The absence of SRP components in Nanoarchaeum equitans was already reported (Podar, Makarova et al. 2013). The assignment of the locus tag NEQ375a to the secE gene was also taken from the same publication.

[2] The following species were analyzed (data in parenthesis are the locus tag prefix and UniProt proteome code): *Haloferax volcanii)* ("HVO_", UP000008243); *Methanococcus maripaludis* ("MMP", UP000000590); *Methanosarcina mazei* ("MM_", UP000000595); *Thermococcus kodakarensis* ("TK", UP000000536); *Methanopyrus kandleri* ("MK", UP000001826); *Methanothermobacter thermoautotrophicus* ("MTH_", UP000005223); *Archaeoglobus fulgidus* ("AF_", UP000002199); *Methanocaldococcus jannaschii* ("MJ", UP000000805); *Sulfolobus acidocaldarius* ("Saci_", UP000001018); *Aeropyrum pernix* ("APE_", UP000002518); *Thermoproteus tenax* ("TTX_", UP000002654); *Nitrosopumilus maritimus* ("Nmar_", UP000000792); Candidatus *Korarchaeum cryptofilum* ("Kcr_", UP000001686); *Nanoarchaeum equitans* ("NEQ", UP000000578); *Lokiarchaeum sp.* strain GC14_75 ("Lokiarch_", UP000034722); *Micrarchaeum acidophilum* ("BK997_", UP000184208); *Haloredivivus sp.* strain G17 ("HRED_", UP000003484); *Parvarchaeum acidophilus* ARMAN-5 ("BJBARM5_", UP000009376). "Bacteria" refers to *E.coli* (UP000000625); "Eukaryotes" refers to yeast (*Saccharomyces cerevisiae*) (UP000002311).

[3] *S. cerevisiae* SRP19 homolog (SRP65, YML105C) is significantly longer than human SRP19 homolog.

[4] Sortase, [6] exosortase (Xrt) and [8] PglB were not identified in *E. coli* but are present in bacteria like *S. aureus*, *Nitrosomonas europeae*, and *Campylobacter jejuni*, respectively.

[5] The subtype of archaeosortase is indicated in parenthesis after the locus tag (A for artA, C for artC, E for artE).

[7] Species which have lipobox proteins require signal peptidase II, which has not yet been identified ("n.i."); species lacking lipobox proteins are indicated as not having this enzyme; in case of very few lipobox predictions (up to 2), these are assumed to be false positives and the species is considered devoid of SPII.

[9] ArlI (FlaI) and ArlJ (FlaJ) are PilB and PilC homologs, respectively, which are genomically co-located with ArlH (FlaH) homologs.

[10] Only those PilB and PilC homologs are reported, which are genomic neighbors; an exception are those from *Nanoarchaeum equitans*; PilB homologs are those reported in the genome publication (Waters, Hohn et al. 2003); all PilC homologs are reported, independent of genomic vicinity.

[11] SLGs are important cell surface components but lack a specific biosynthesis machinery. Therefore, the structural units themselves have been included here. However, since SLGs lack global sequence similarity, assignments are based on [12] experimental characterization; [13] sequence similarity to characterized SLGs from related species, taken from [14] species-specific literature sources or from the [15] review by Claus *et al*. (Claus, Akca et al. 2005).

[12] Experimentally characterized SLGs are from *H.volcanii* (Sumper, Berg et al. 1990) *M. mazei* (Francoleon, Boontheung et al. 2009), M. *jannaschii* (Akca, Claus et al. 2002), *S. acidocaldarius* (Grogan 1996, Veith, Klingl et al. 2009, Peyfoon, Meyer et al. 2010)

[13] SLGs assigned by sequence similarity are from *M. maripaludis* (based on UniProt:Q8X235 (Akca, Claus et al. 2002).

[14] The SLG from *T. kodakarensis* is the 3rd listed for this species in *Claus et al*. (Claus, Akca et al. 2005) but its correct GenBank accession is BAD85084. This identification, *csg*, TK0895, was taken from Takemasa (Takemasa, Yokooji et al. 2011). The *T. tenax* SLG assignment is taken from Siebers *et al*. (Siebers, Zaparty et al. 2011) as the structural analysis of this species (Wildhaber and Baumeister 1987) did not identify the SLG. The *A. pernix* assignment is taken from Palmieri *et al*. (Palmieri, Cannio et al. 2009).

References:

Akca, E., H. Claus, N. Schultz, G. Karbach, B. Schlott, T. Debaerdemaeker, J. P. Declercq and H. Konig (2002). "Genes and derived amino acid sequences of S-layer proteins from mesophilic, thermophilic, and extremely thermophilic methanococci." Extremophiles **6**(5): 351-358.

Claus, H., E. Akca, T. Debaerdemaeker, C. Evrard, J. P. Declercq, J. R. Harris, B. Schlott and H. Konig (2005). "Molecular organization of selected prokaryotic S-layer proteins." Can J Microbiol **51**(9): 731-743.

Francoleon, D. R., P. Boontheung, Y. Yang, U. Kin, A. J. Ytterberg, P. A. Denny, P. C. Denny, J. A. Loo, R. P. Gunsalus and R. R. Loo (2009). "S-layer, surface-accessible, and concanavalin A binding proteins of *Methanosarcina acetivorans* and *Methanosarcina mazei*." J Proteome Res **8**(4): 1972-1982.

Grogan, D. W. (1996). "Organization and interactions of cell envelope proteins of the extreme thermoacidophile *Sulfolobus acidocaldarius*." Canadian journal of microbiology **42**(11): 1163-1171.

Palmieri, G., R. Cannio, I. Fiume, M. Rossi and G. Pocsfalvi (2009). "Outside the unusual cell wall of the hyperthermophilic archaeon *Aeropyrum pernix* K1." Mol Cell Proteomics **8**(11): 2570-2581.

Peyfoon, E., B. Meyer, P. G. Hitchen, M. Panico, H. R. Morris, S. M. Haslam, S. V. Albers and A. Dell (2010). "The S-layer glycoprotein of the crenarchaeote Sulfolobus acidocaldarius is glycosylated at multiple sites with chitobiose-linked N-glycans." Archaea **2010**.

Podar, M., K. S. Makarova, D. E. Graham, Y. I. Wolf, E. V. Koonin and A. L. Reysenbach (2013). "Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park." Biology Direct **8**(1): 9.

Siebers, B., M. Zaparty, G. Raddatz, B. Tjaden, S. V. Albers, S. D. Bell, F. Blombach, A. Kletzin, N. Kyrpides, C. Lanz, A. Plagens, M. Rampp, A. Rosinus, M. von Jan, K. S. Makarova, H. P. Klenk, S. C. Schuster and R. Hensel (2011). "The complete genome sequence of *Thermoproteus tenax*: a physiologically versatile member of the Crenarchaeota." PLoS One **6**(10): e24222.

Sumper, M., E. Berg, R. Mengele and I. Strobel (1990). "Primary structure and glycosylation of the S-layer protein of *Haloferax volcanii*." J Bacteriol **172**(12): 7111-7118.

Takemasa, R., Y. Yokooji, A. Yamatsu, H. Atomi and T. Imanaka (2011). "*Thermococcus kodakarensis* as a Host for Gene Expression and Protein Secretion." Applied and Environmental Microbiology **77**(7): 2392-2398.

Veith, A., A. Klingl, B. Zolghadr, K. Lauber, R. Mentele, F. Lottspeich, R. Rachel, S. V. Albers and A. Kletzin (2009). "*Acidianus*, *Sulfolobus* and *Metallosphaera* surface layers: structure, composition and gene expression." Mol Microbiol **73**(1): 58-72.

Waters, E., M. J. Hohn, I. Ahel, D. E. Graham, M. D. Adams, M. Barnstead, K. Y. Beeson, L. Bibbs, R. Bolanos, M. Keller, K. Kretz, X. Y. Lin, E. Mathur, J. W. Ni, M. Podar, T. Richardson, G. G. Sutton, M. Simon, D. Soll, K. O. Stetter, J. M. Short and M. Noordewier (2003). "The genome of *Nanoarchaeum equitans*: Insights into early archaeal evolution and derived parasitism." Proceedings of the National Academy of Sciences of the United States of America **100**(22): 12984-12988.

Wildhaber, I. and W. Baumeister (1987). "The cell envelope of *Thermoproteus tenax*: three-dimensional structure of the surface layer and its role in shape maintenance." EMBO J **6**(5): 1475-1480.