Inventory of Supplementary Information

Complete sequence alignments for 78 archaea, 865 bacteria, and 187 eukarya are available in HTML format from our web site:

http://blanco.biomol.uci.edu/download/Bondar_SecYE_align.zip

save date: 9 Apr 2010, 06:35 *Structure* D-09-00288R

Supplementary Information

Dynamics of SecY Translocons with Translocation-Defective Mutations

Ana-Nicoleta Bondar[1,2], Coral Muñoz del Val[3], J. Alfredo Freites[1,2,4],

Douglas J. Tobias[2,4] and Stephen H. White[1,2]

[1]Department of Physiology and Biophysics and the [2]Center for Biomembrane Systems,

University of California at Irvine, Irvine, CA 92697-4560

[3]Department of Computer Science and Artificial Intelligence

University of Granada, Granada, E-18071, Spain

[4]Department of Chemistry and [5]Institute for Surface and Interface Science, University of California at Irvine
Irvine, CA 92697-2025

address correspondence to stephen.white@uci.edu

Contents

Complete sequence alignments for 63 archaea, 865 bacteria, and 187 eukarya are available in HTML format from our web site:

http://blanco.biomol.uci.edu/download/Bondar_SecYE_align.zip

Figure S1. Comparison of the SecYEG wild-type and mutant structures referred to in Figure 1. Overlap between the SecY mutant structures and wild-type SecY, and rmsd of the mutant structures relative to the starting crystal structure (Cα rmsd, in Å) for K250E (A; Sim2), T72V/T80V/R104A (B; Sim3), L406K (C, Sim4), and E336R (D, Sim5). Wild-type SecYEG is shown with SecY green, SecE purple, and Secβ blue. For the mutant translocons, SecE and Secβ are shown in transparent purple and transparent blue, respectively. Mutant SecY is depicted as follows: K250E - pink, T72V/T80V/R104A – cyan, L406K – orange, and E336R – gray. Mutated amino-acids are shown as surfaces. The numbers under 'Last 10ns' give the average rmsd ± standard deviation for SecY (green), SecE (cyan), SecB (blue) and the TM region of SecY (dark orange), in Å. In the case of the wild-type Sim1 the rmsd ± standard deviation values for SecY, SecE, SecB, and the TM region of SecY are 3.2 ± 0.1 Å, 2.4 ± 0.2 Å, 1.9 ± 0.2 Å, and 1.9 ± 0.1 Å, respectively (see Figure 1C).

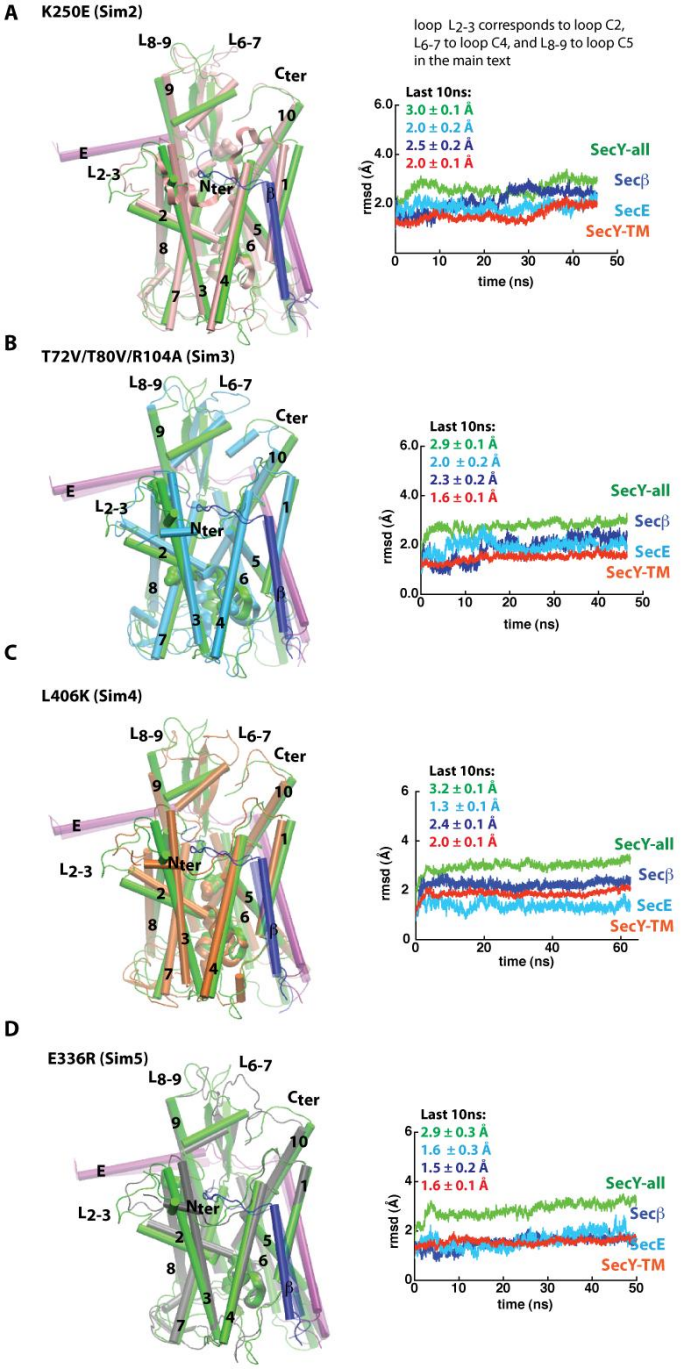


Figure S2. Location of H-Bonding Clusters referred to Figure 2. H-bonding amino acids of wild-type SecYEG were grouped into clusters located largely in the cytoplasmic (CP-1 to CP-6; panels A-F) and extracellular (EC-1 to EC-4; panels G-J) halves of the translocon. SecY is depicted in green, SecE in purple, and Secβ in iceblue. H-bonding amino acids are depicted as bonds with carbon atoms in cyan, oxygen red, and nitrogen blue. For simplicity, only the backbone is shown for non-polar amino acids whose backbone groups participate in H bonding. The figures were prepared using a snapshot from the simulation on the wild-type translocon (Sim1) after ~35ns of unconstrained dynamics.

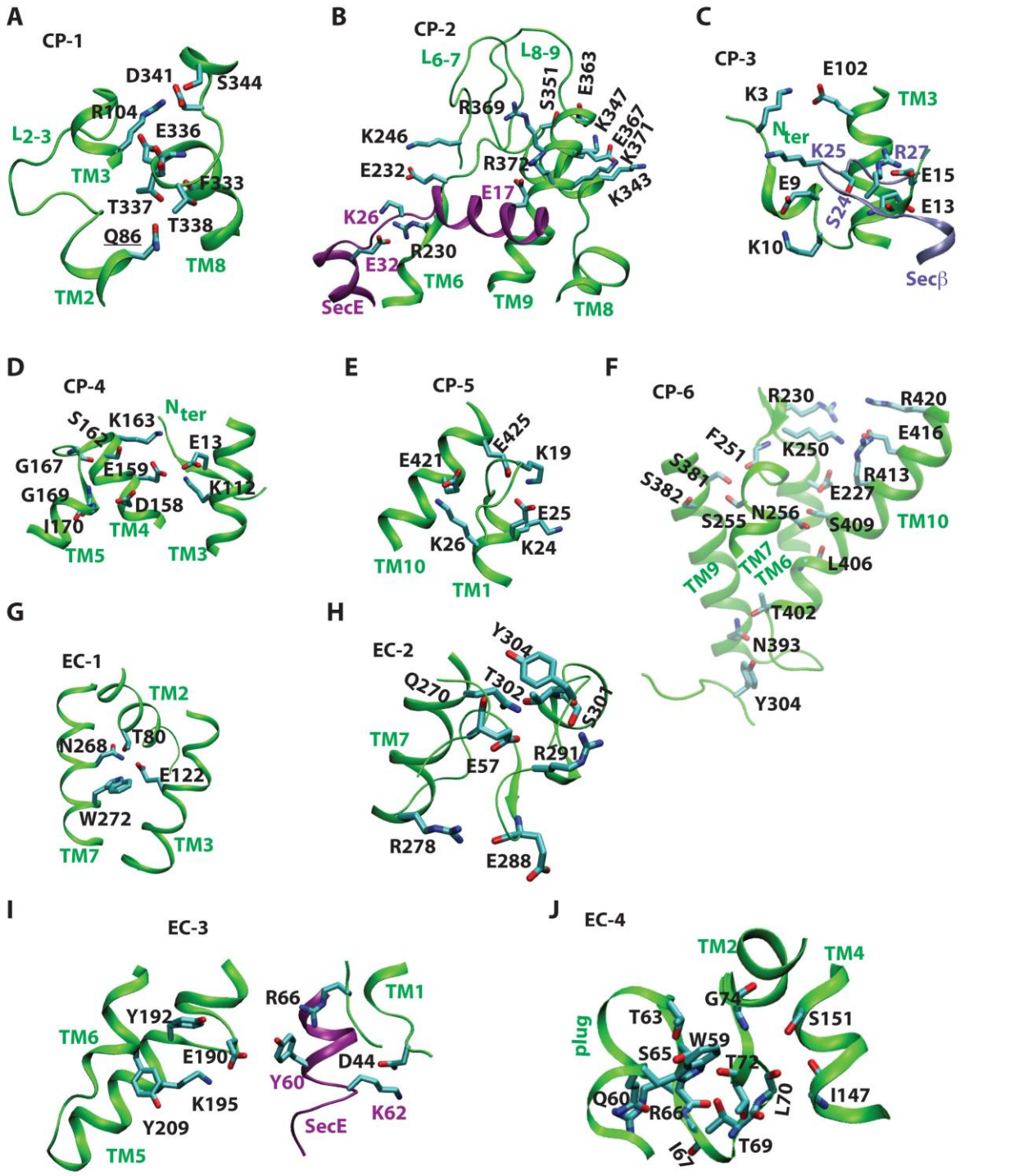


Figure S3. Examples of H-bond interactions in *T. maritima* corresponding to those of *M. jannaschii* referred to in Figure 2. The H-bond interactions were determined from the sequence alignments of Figure S5. The coordinates were taken from the SecYEG/SecA crystal structure 3DIN of Zimmer et al. (2008). (**A**) Amino acids T87 and Q131 (corresponding to *M. jannaschii* T80 and E122, respectively; Figure S1G) are within H-bonding distance of 3.4 Å. T83 is within H-bonding distance from S76 (3.0 Å); TM3-T124 H bonds to the carbonyl group of P84 (3.2 Å). (**B**) The TM7 amino acids S277 and S281 are not within H-bonding distance of TM2 or TM3. T168 corresponds to *M. jannaschii* S151 (see Figure S1J). T168 H bonds with the carbonyl groups of M80 and M164 (distances are 2.7 Å and 2.6 Å, respectively). The distance between the hydroxyl oxygen atoms of T79 and S163 is 3.8 Å. (**C**) On the cytoplasmic side of the *M. jannaschii* SecY, R104 and D341 H bond during the simulation on the wild type protein (see Table S1). The corresponding amino acids in *T. maritima* (R113 and D327) are located far apart from each other (the distance between the R113-CZ and D227-Cγ atoms is 16.1 Å)), but the distance between R107-NH2 and E330-Oε2 is significantly shorter at 4.2 Å. Q93-Oε_1, corresponding to *M. jannaschii* Q86 (Figure S1A), is within 5.4 Å from T318-Oγ_1. Assuming that the relative orientation of D327 and E352 is correct in the 4.5 Å resolution structure of Zimmer et al. (2008), the 2.9 Å distance between the carboxyl oxygens of these two acidic amino acids could be interpreted to suggest that D337 or E352 are protonated. In *M. jannaschii*, amino acids of the N terminus participate in H-bonding interactions with TM3, TM4, and Secβ (Figure S1C-D). In the structure of the SecYEG/SecA, these H bonds are not possible due to the N terminus being oriented towards TM10. (**D**) D158 and E159 of *M. jannaschii* participate in cluster CP-4 that also includes amino acids of TM3 and the N terminus (Figure S1D). D158 is highly conserved as Asp in archaea and eukarya (Figures S6, S8), and the D168A mutation in yeast (corresponding to *M. jannaschii* D158) affects topogenesis (Junne et al, 2007). In contrast, most bacteria have Gly at this position in the sequence (Figure S6). E159 is highly conserved in all organisms (Figures S6-S8). The *T. maritima* E176 (corresponding to *M. jannaschii* E159) could H bond with TM3-R121 -- the distance between E176-Oε_2 and R121-NH_2 is 4.2 Å. The distances between E176-Oε_1 and T179-OG1, and between Y85-OH and L172-O, are also longer than for a H bond (5.0 Å and 4.1 Å, respectively). (**E**) The crowded cluster of H bonds CP-6 of the *M. jannaschii* SecY involves amino acids of TM6, TM7, TM9, and TM10 (Figure S1F). *M. jannaschii* K250, whose mutation causes a *prl* phenotype in yeast (Junne et al, 2007), H bonds with TM6-E228 and TM10-E416 (Table S1). The interactions between these amino acids are greatly changed in the open structure of the *T. maritima* SecY. The distance between K264-Nζ and Q407-Oε_1 (corresponding to *M. jannaschii* K250 and E416, respectively) is 18.9 Å; likewise, the distance between K264-Nζ and Q234-Oε_1 (Q234 corresponds to *M. jannaschii* E227) is 13.1 Å. K264 H bonds instead with E237 (3.5 Å distance). It is not clear whether TM7-S277 and TM10-T393 could H bond (the distance between their hydroxyl oxygen atoms is 4.4 Å).

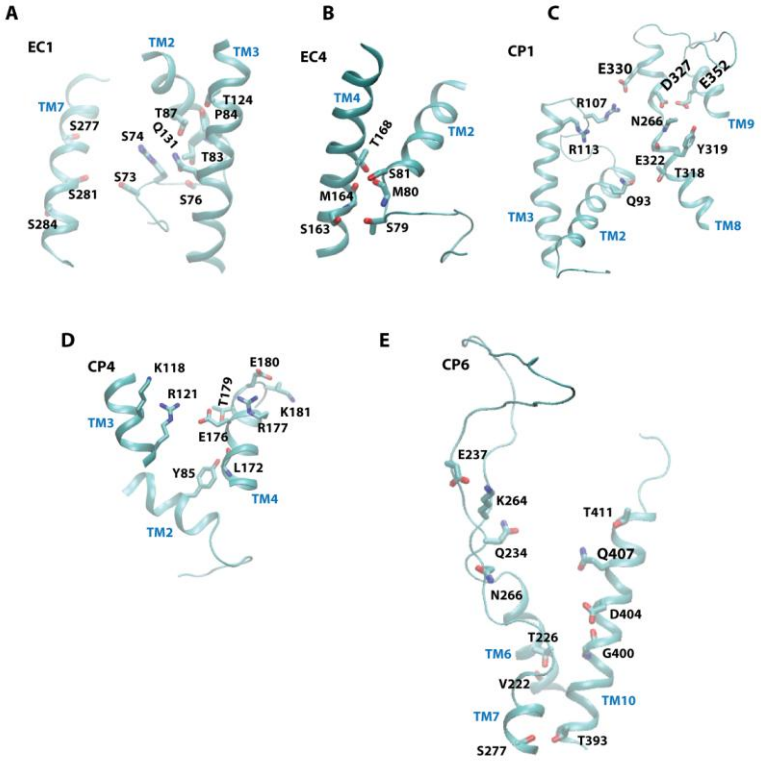


Figure S4. Examples of the hydrogen-bond interactions in *T. thermophilus* corresponding to those of *M. jannaschii* referred to in Figure 2. The H-bond interactions were determined from the sequence alignments of Figure S5. The coordinates were taken from the 3.2 Å SecYE-Fab crystal structure 2ZJS of Tsukazaki et al. (2008). (**A**) Similar to *T. maritima*, in *T. thermophilus* E122 is replaced by Gln (Q126). The distances between T82-Oγ_1 (T82 corresponds to *M. jannaschii* T80; Figure S1G) and Q126-Nε_2 and Q126-Oε_1 atoms are 4.7 Å and 5.1 Å, respectively. TM7-Q282 is also not within H-bonding distance from T80 and Q126 (the distance between T82-Oγ_1 and Q282-Nε_2 is 7.9 Å). (**B**) On the cytoplasmic side, Q86 and R108, which correspond *to M. jannaschii* Q86 and R104 (Figure S1A), are not involved in H bonding with other amino acid sidechains. The distance between R108-CZ and D332-Cγ (D332 corresponds to D341 in *M. jannaschii*) is 15.2 Å. K334 (corresponds to *M. jannaschii* K343, Figure S1 B) could be part of a H-bonding cluster with E354 (E363 in *M. jannaschii*), K358, E361. (**C**) E173 (*M. jannaschii* E159, Figure S1D) could be part of a H-bonding cluster involving N112, Q113, R109, R116, R174, E177, Y178. The distance between E173-Oε_2 and N112-N$\delta 2$ is 3.5 Å. Because in the crystal structure of the *T. thermophilus* SecY the N terminus points away from TM3, R116 (K112 in *M. jannaschii*, Figure S1D) cannot H bond to amino acids of the N terminus; the distance between E13-Cδ (*M. jannaschii* E13, Figure S1D) and R116-Cζ is 28.7 Å. E13 of the N terminus could instead salt-bridge to R422 (the distance between E13-Oε_1 and R422-NH1 is 4.2 Å). (**D**) TM7-K265 (K250 in *M. jannaschii*, Figure S1F) is not involved in H bonding. The distances between K265-Nζ and E238-Oε_2, and between K265-Nζ and E416-Oε_2 are 5.1 Å and 10.7 Å, respectively. The distance between Q235- (E227 in *M. jannaschii*) and E416-Oε_1 is 7.9 Å.

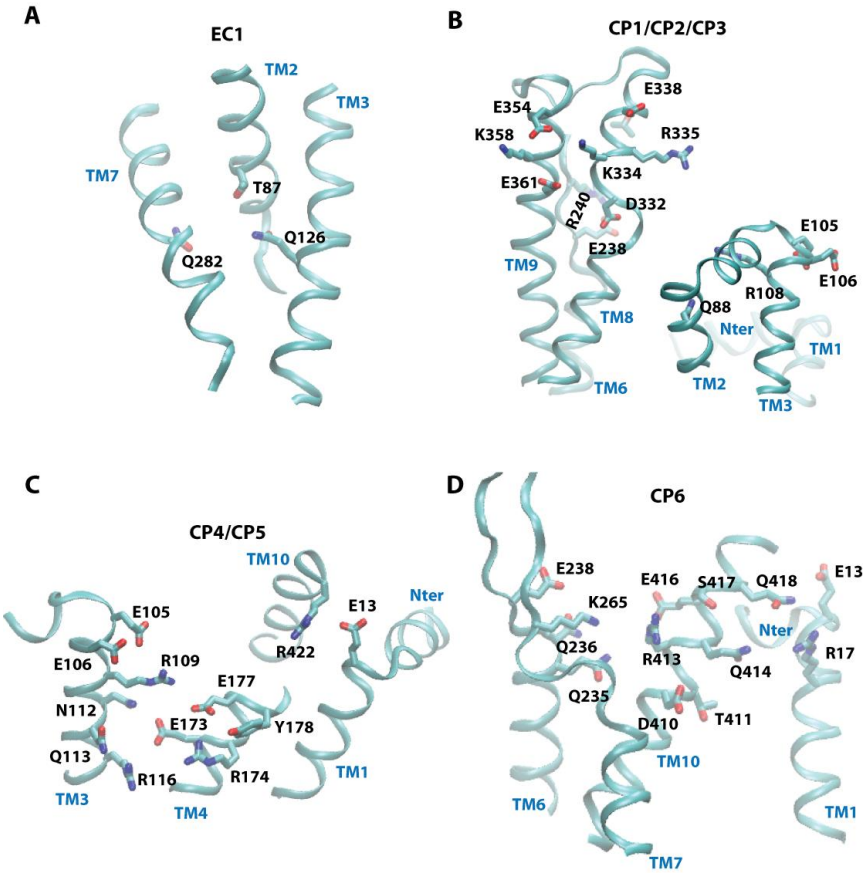


Figure S5. Alignment of Sec61p/SecY sequences from *T. thermophilus*, *M. jannaschii*, *E. coli*, *T. maritima*, and *S. cerevisiae* (referred to in Figure 3). We aligned the SecY sequences whose X-ray crystal structures have been solved (*M. jannaschii*, van den Berg 2003; *T. thermophilus*, Tsukazaki 2008; *T. maritima*, Zimmer 2008), and the sequences of SecY/Sec61 from *E. coli*/*S. cerevisiae*.

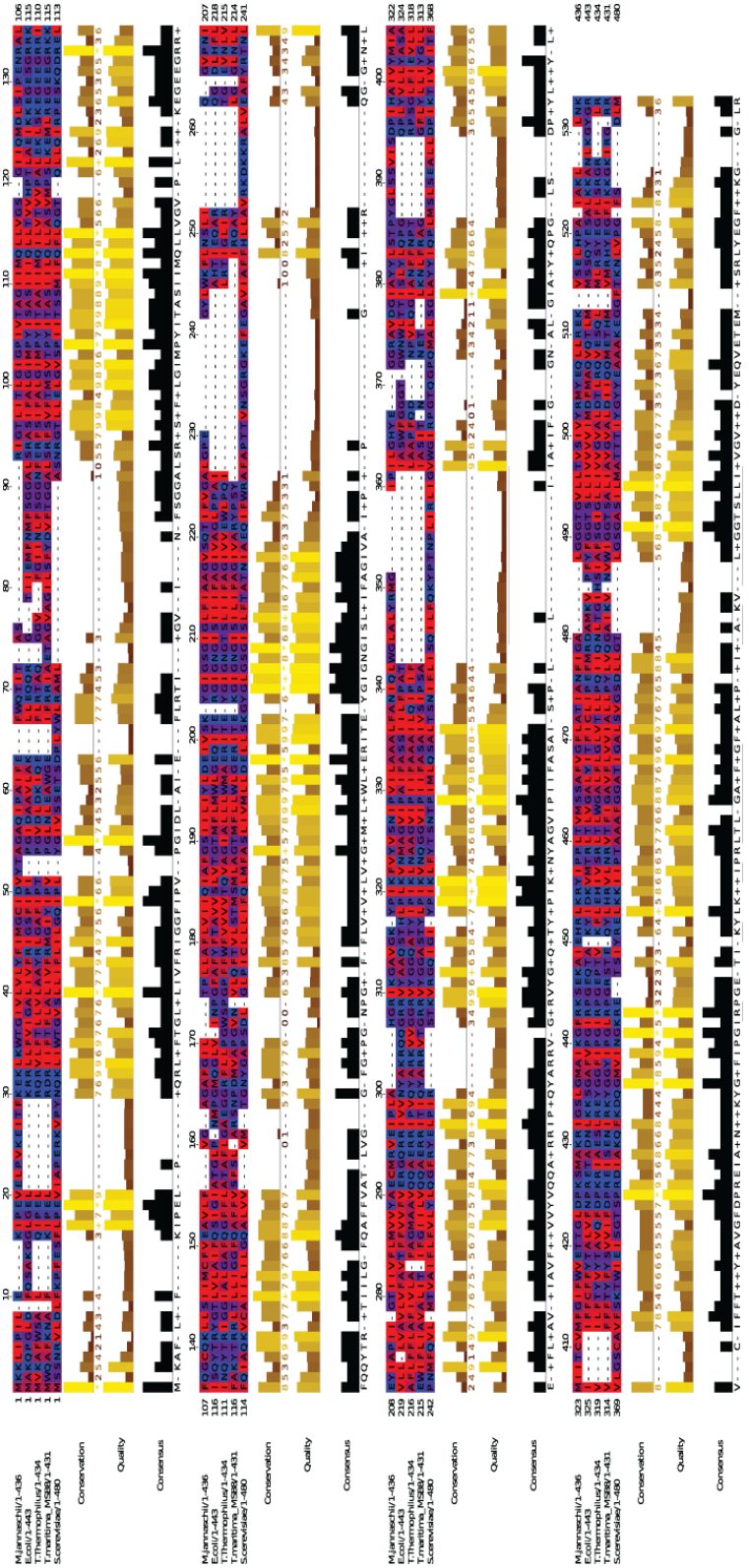


Figure S6. Frequency of analysis for selected H-bonding amino acids in SecY from archaea referred to in Figure 3. The frequency of SecY H-bonding amino-acids from Tables S1-S2 was analyzed as described in 'Protocol for sequence analysis', below. See Figure S2 for the location of H-bonding amino acids, and Tables S1-S2 for the H-bonding dynamics.

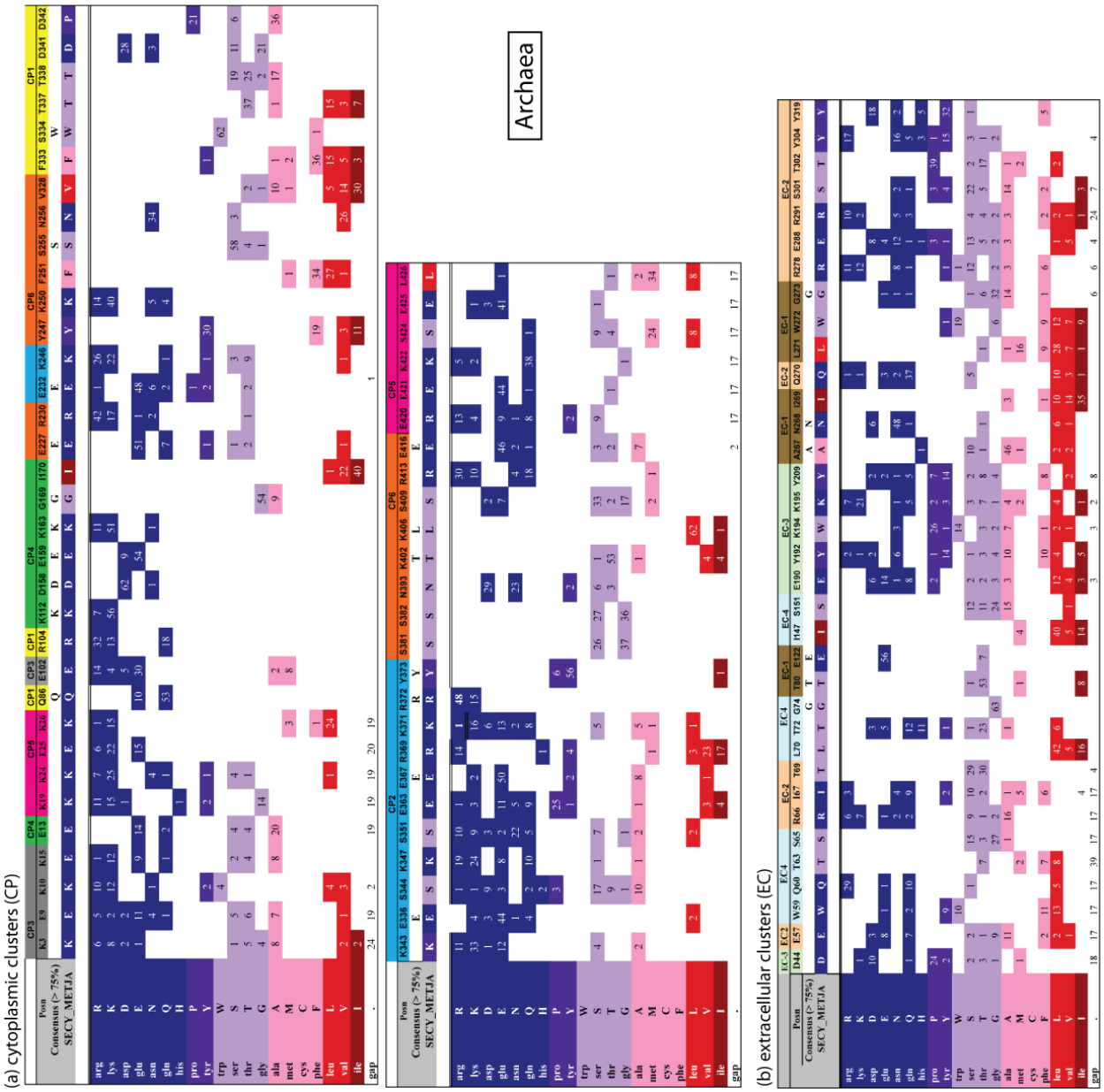


Figure S7. Frequency of analysis for selected H-bonding amino acids in SecY from bacteria referred to in Figure 3. The frequency of SecY H-bonding amino-acids from Tables S1-S2 was analyzed as described in 'Protocol for sequence analysis', below. See Figure S2 for the location of H-bonding amino acids, and Tables S1-S2 for the H-bonding dynamics.

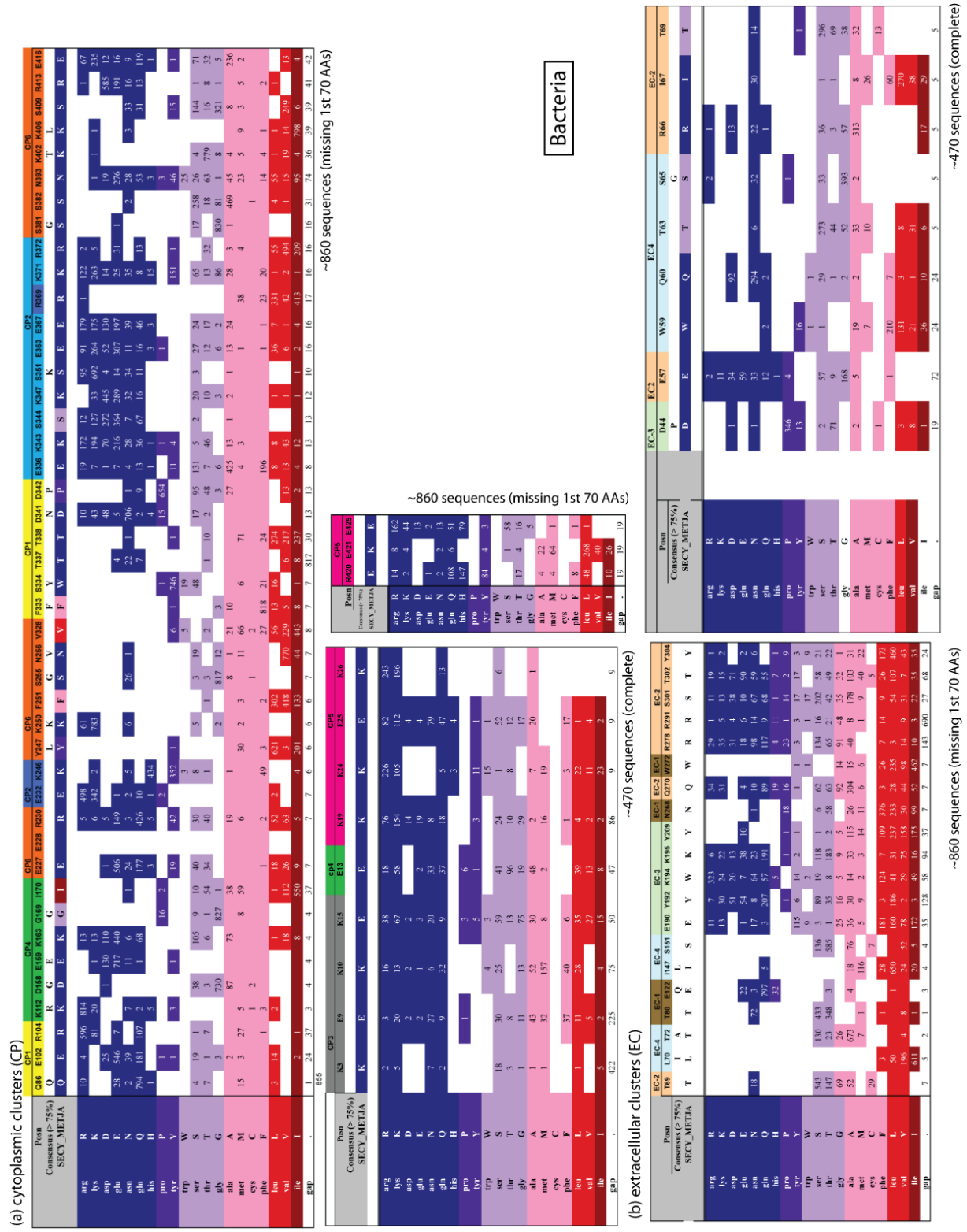


Figure S8. Frequency of analysis for selected H-bonding amino acids in SecY from eukarya referred to in Figure 3. The frequency of SecY H-bonding amino-acids from Tables S1-S2 was analyzed as described in 'Protocol for sequence analysis', below. See Figure S2 for the location of H-bonding amino acids, and Tables S1-S2 for the H-bonding dynamics.

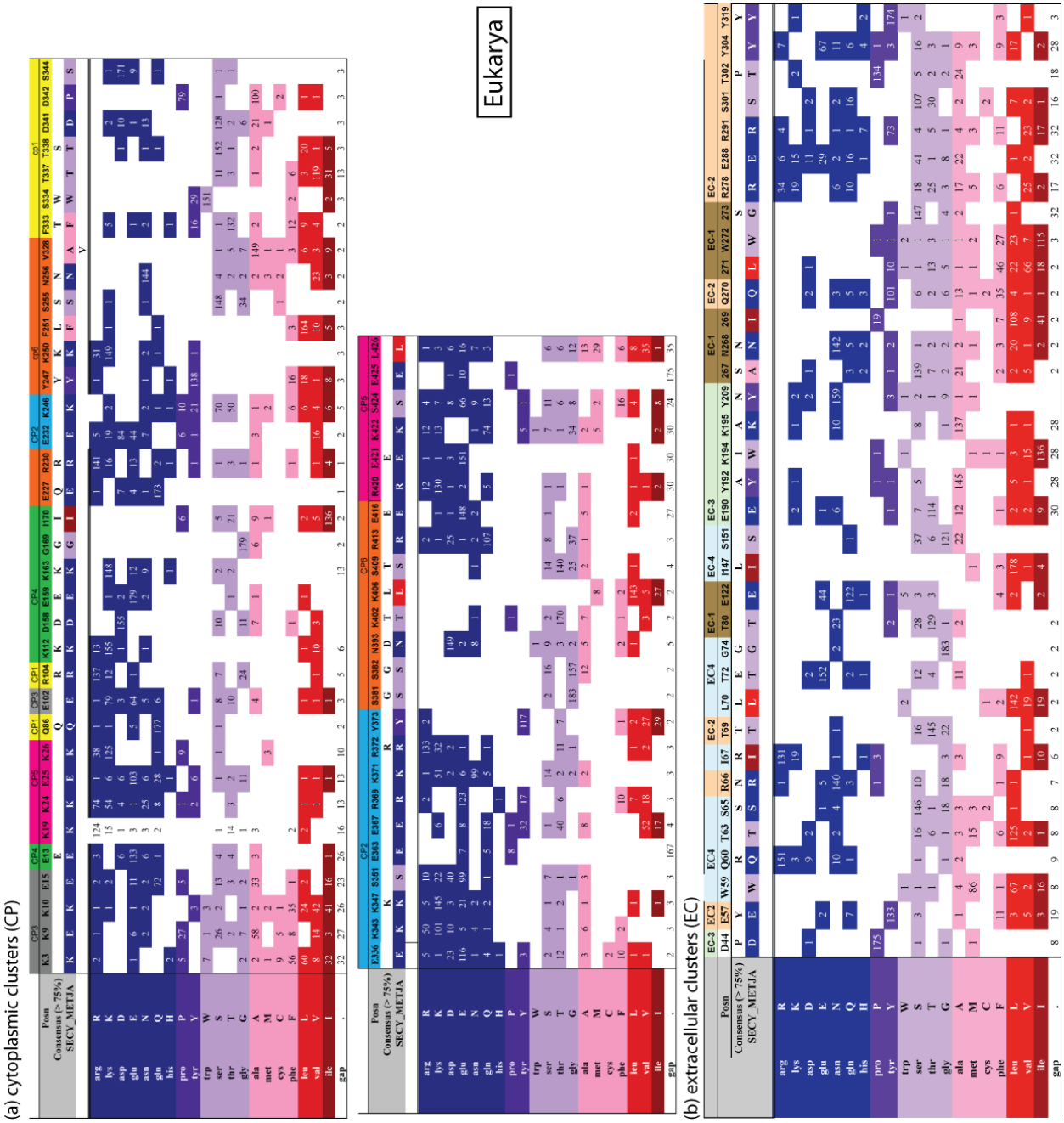
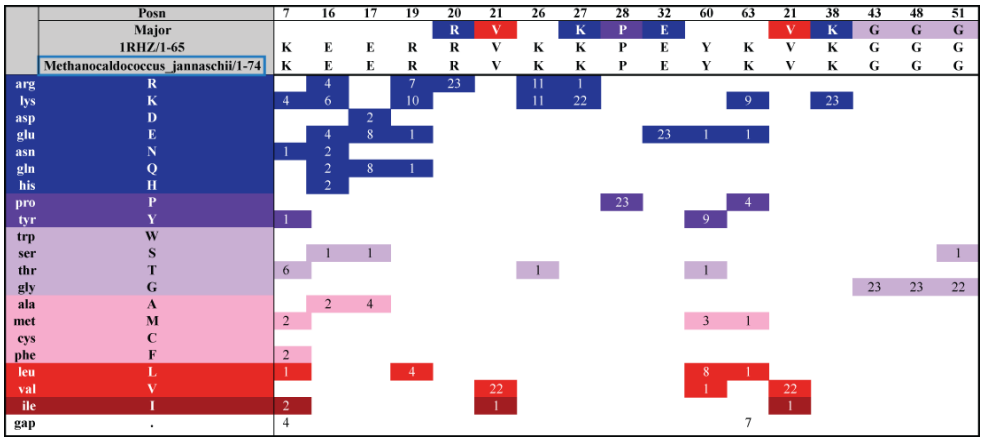
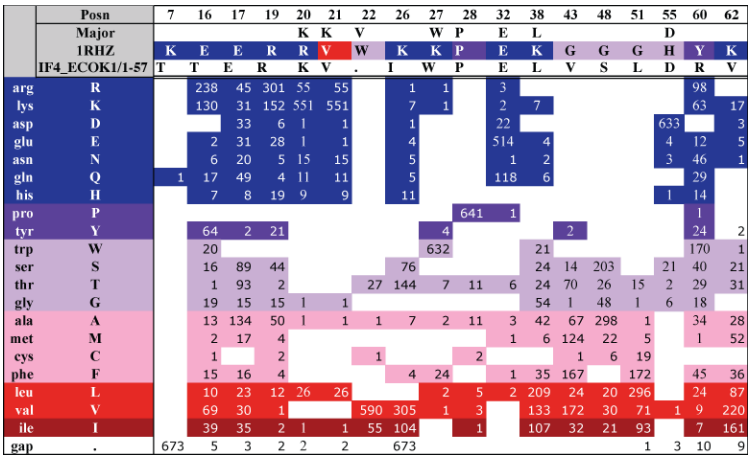


Figure S9. Frequency analysis for selected H-bonding amino acids in SecE from archaea referred to in Figure 3. The frequency of SecE H-bonding amino-acids from Tables S1-S2 was analyzed as described in 'Protocol for sequence analysis', below.



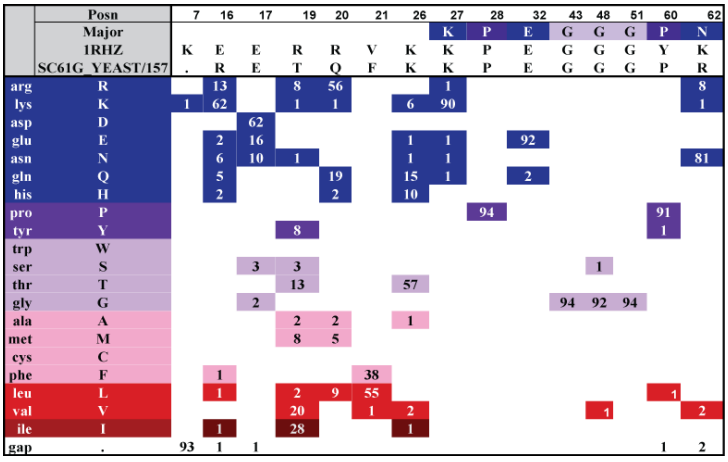
	Posn	7	16	17	19	20	21	26	27	28	32	60	63	21	38	43	48	51
	Major					R	V		K	P	E			V	K	G	G	G
	1RHZ/1-65	K	E	E	R	R	V	K	K	P	E	Y	K	V	K	G	G	G
	Methanocaldococcus_jannaschii/1-74	K	E	E	R	R	V	K	K	P	E	Y	K	V	K	G	G	G
arg	R		4		7	23		11	1									
lys	K	4	6		10			11	22				9		23			
asp	D			2														
glu	E		4	8	1						23	1	1					
asn	N	1	2															
gln	Q		2	8	1													
his	H		2															
pro	P									23			4					
tyr	Y	1										9						
trp	W																	
ser	S		1	1														1
thr	T	6						1				1						
gly	G															23	23	22
ala	A		2	4														
met	M	2										3	1					
cys	C																	
phe	F	2																
leu	L	1			4							8	1					
val	V						22					1		22				
ile	I	2					1							1				
gap	.	4											7					

Figure S10. Frequency analysis for selected H-bonding amino acids in SecE from bacteria referred to in Figure 3. The frequency of SecE H-bonding amino-acids from Tables S1-S2 was analyzed as described in 'Protocol for sequence analysis', below.



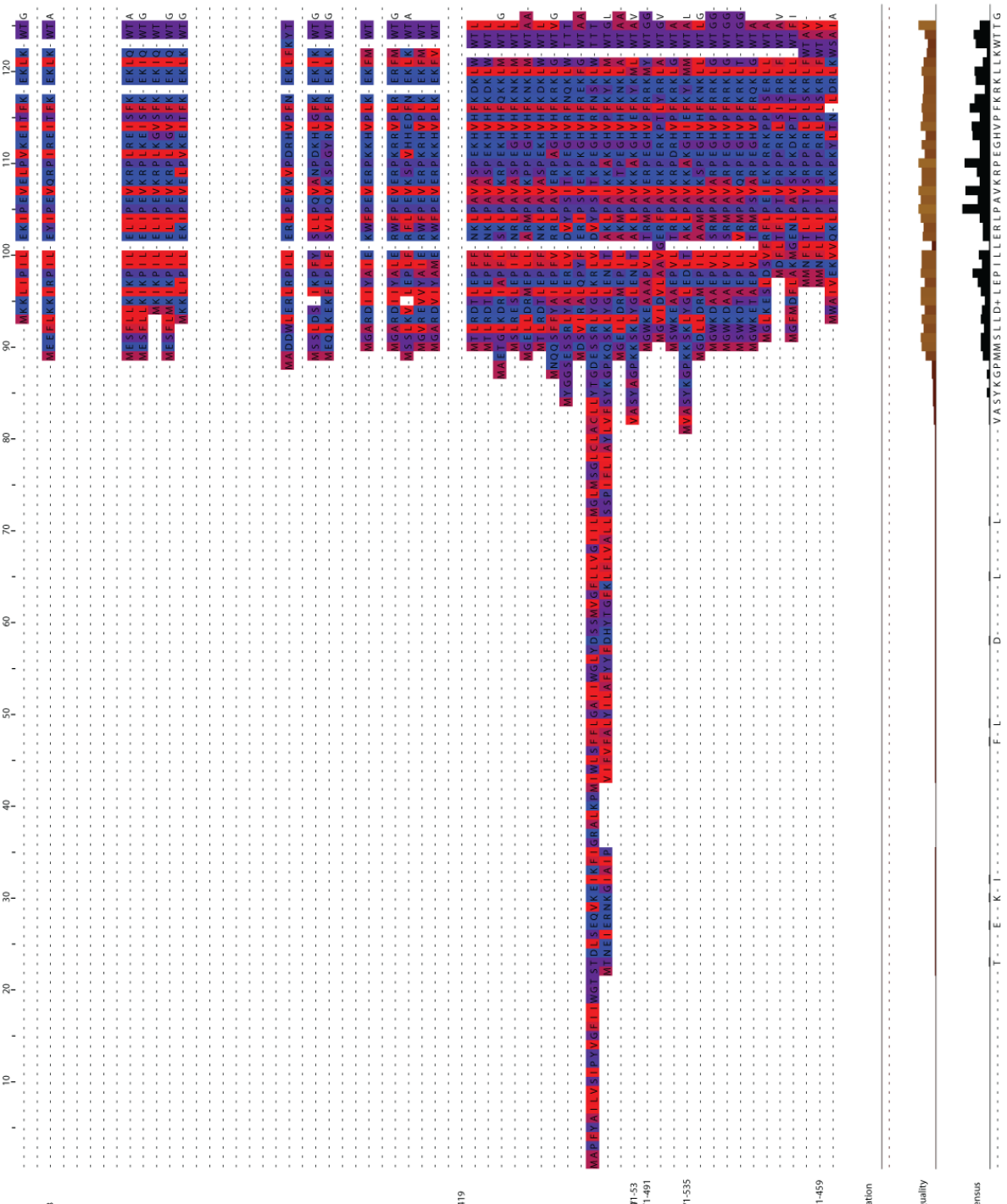
	Posn	7	16	17	19	20	21	22	26	27	28	32	38	43	48	51	55	60	62
	Major					K	K	V		W	P	E	L				D		
	1RHZ	K	E	E	R	R	V	W	K	K	P	E	K	G	G	G	H	Y	K
	IF4_ECOK1/1-57	T	T	E	R	K	V	.	I	W	P	E	L	V	S	L	D	R	V
arg	R		238	45	301	55	55		1	1		3						98	
lys	K		130	31	152	551	551		7	1		2	7					63	17
asp	D			33	6	1	1		1			22					633		3
glu	E		2	31	28	1	1		4			514	4				4	12	5
asn	N		6	20	5	15	15		5			1	2				3	46	1
gln	Q	1	17	49	4	11	11		5			118	6					29	
his	H		7	8	19	9	9		11								1	14	
pro	P										641	1						1	
tyr	Y		64	2	21					4			2					24	2
trp	W		20							632			21					170	1
ser	S		16	89	44				76			24	14	203			21	40	21
thr	T		1	93	2			27	144	7	11	6	24	70	26	15	2	29	31
gly	G		19	15	15	1	1						54	1	48	1	6	18	
ala	A		13	134	50	1	1	1	7	2	11	3	42	67	298	1		34	28
met	M		2	17	4							1	6	124	22	5		1	52
cys	C		1		2			1			2			1	6	19			
phe	F		15	16	4				4	24		1	35	167		172		45	36
leu	L		10	23	12	26	26			2	5	2	209	24	20	296		24	87
val	V		69	30	1			590	305	1	3		133	172	30	71	1	9	220
ile	I		39	35	2	1	1	55	104		1		107	32	21	93		7	161
gap	.	673	5	3	2	2	2		673							1	3	10	9

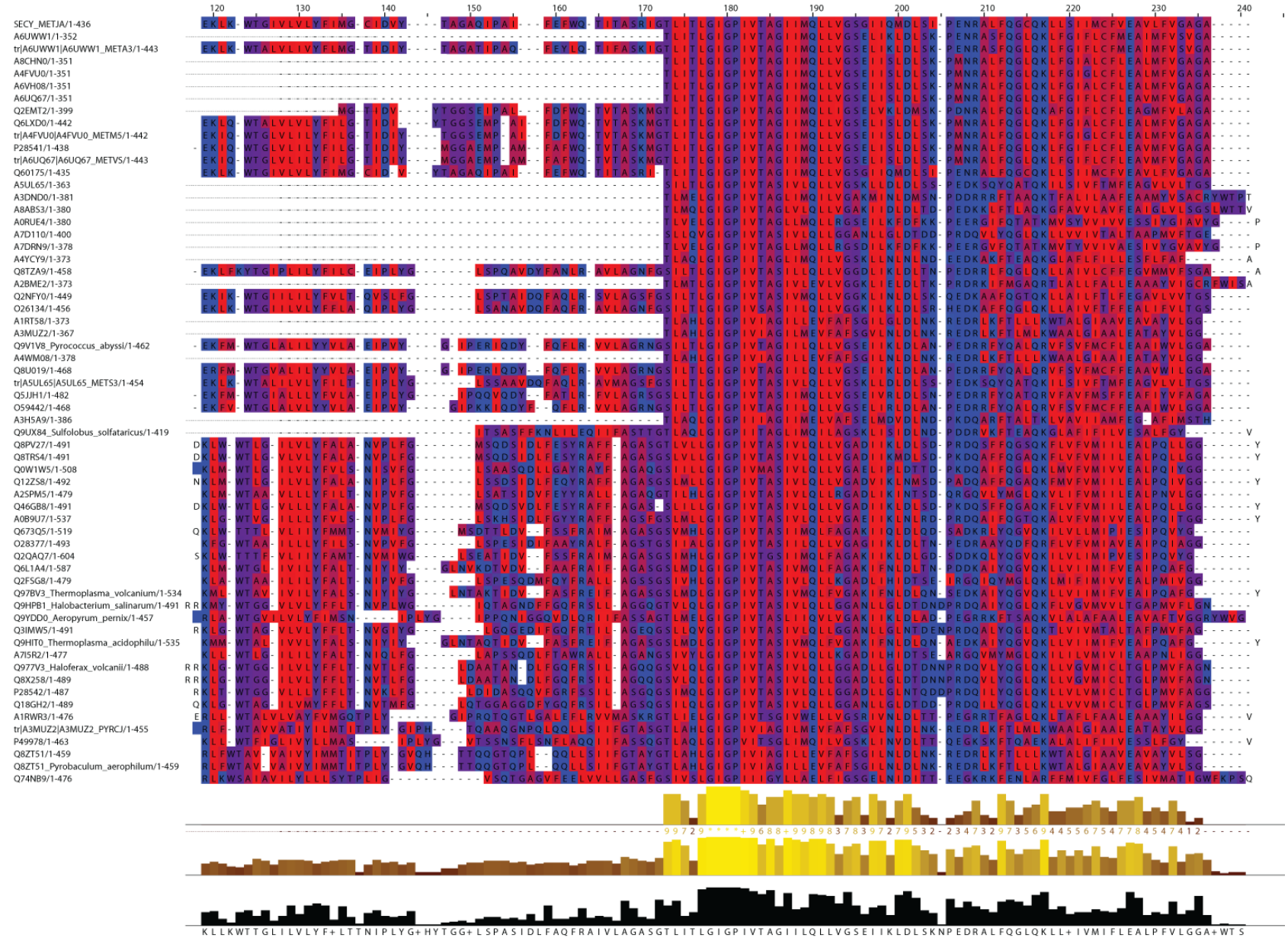
Figure S11. Frequency analysis for selected H-bonding amino acids in SecE from eukarya referred to in Figure 3. The frequency of SecE H-bonding amino-acids from Tables S1-S2 was analyzed as described in 'Protocol for sequence analysis', below.

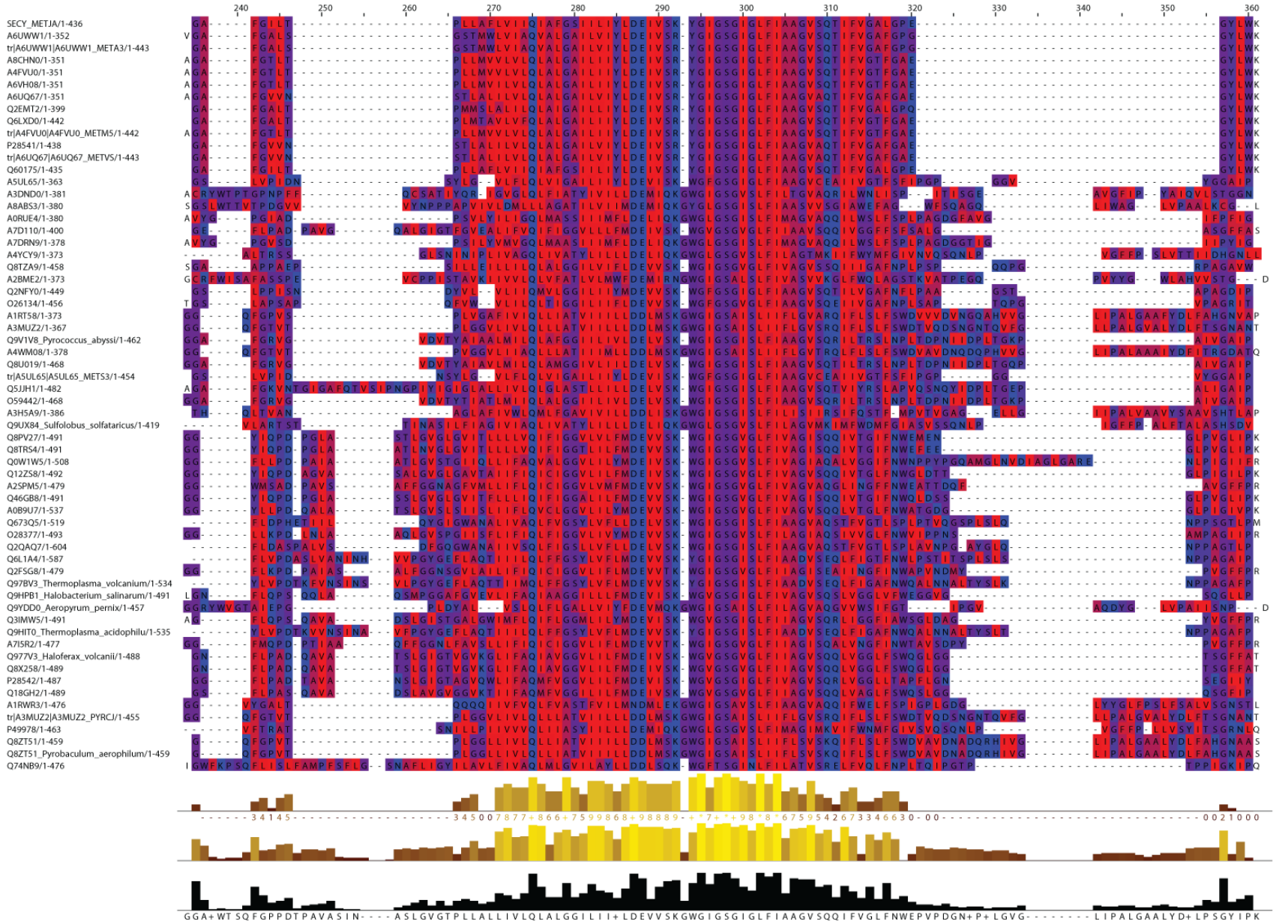


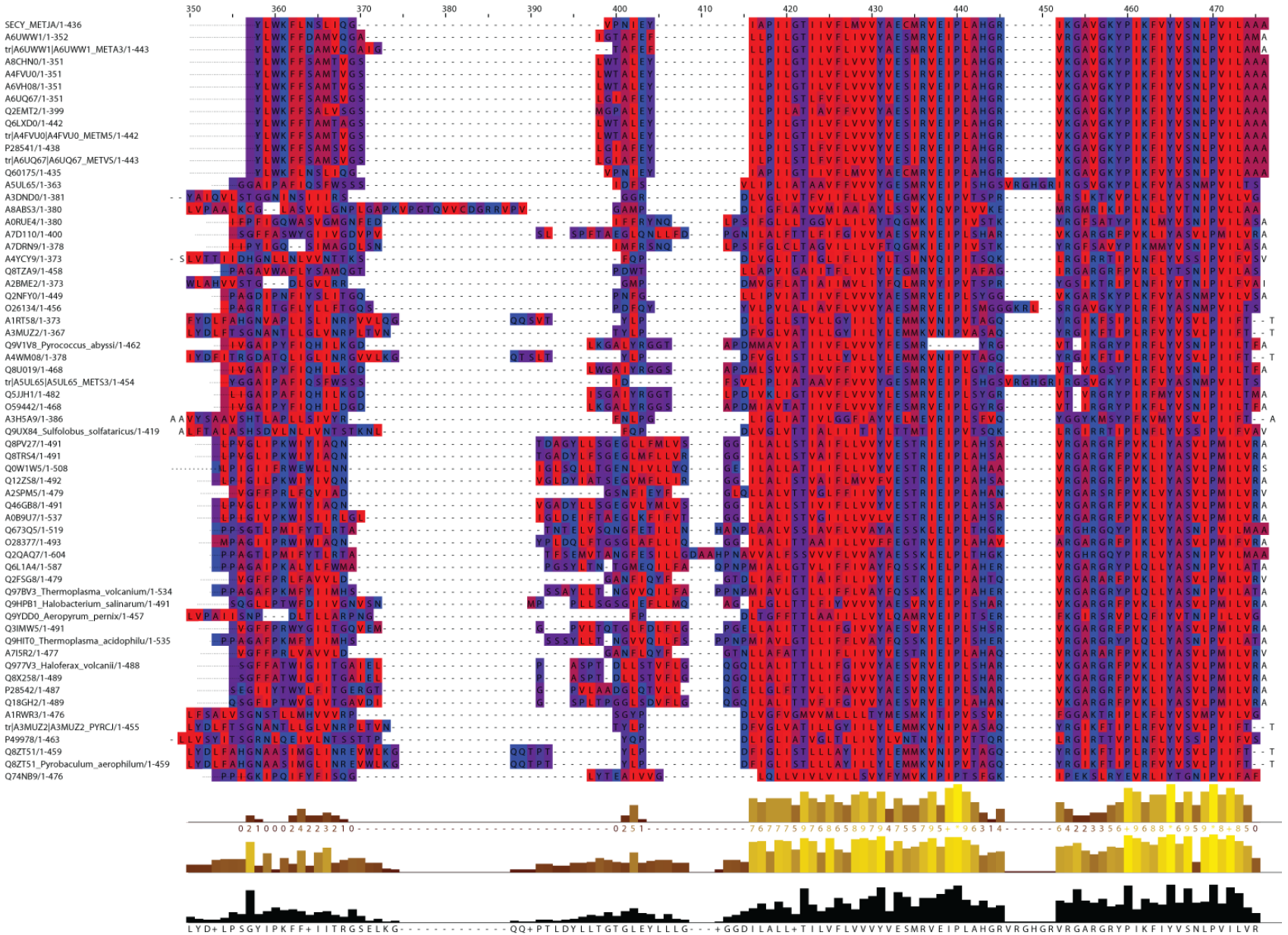
	Posn	7	16	17	19	20	21	26	27	28	32	43	48	51	60	62
	Major								K	P	E	G	G	G	P	N
	1RHZ	K	E	E	R	R	V	K	K	P	E	G	G	G	Y	K
	SC61G_YEAST/157	.	R	E	T	Q	F	K	K	P	E	G	G	G	P	R
arg	R		13		8	56			1							8
lys	K	1	62		1	1		6	90							1
asp	D			62												
glu	E		2	16				1	1		92					
asn	N		6	10	1			1	1							81
gln	Q		5			19		15	1		2					
his	H		2			2		10								
pro	P									94					91	
tyr	Y				8										1	
trp	W															
ser	S			3	3								1			
thr	T				13			57								
gly	G			2								94	92	94		
ala	A				2	2		1								
met	M				8	5										
cys	C															
phe	F		1				38									
leu	L		1		2	9	55								1	
val	V				20		1	2					1			2
ile	I		1		28			1								
gap	.	93	1	1											1	2

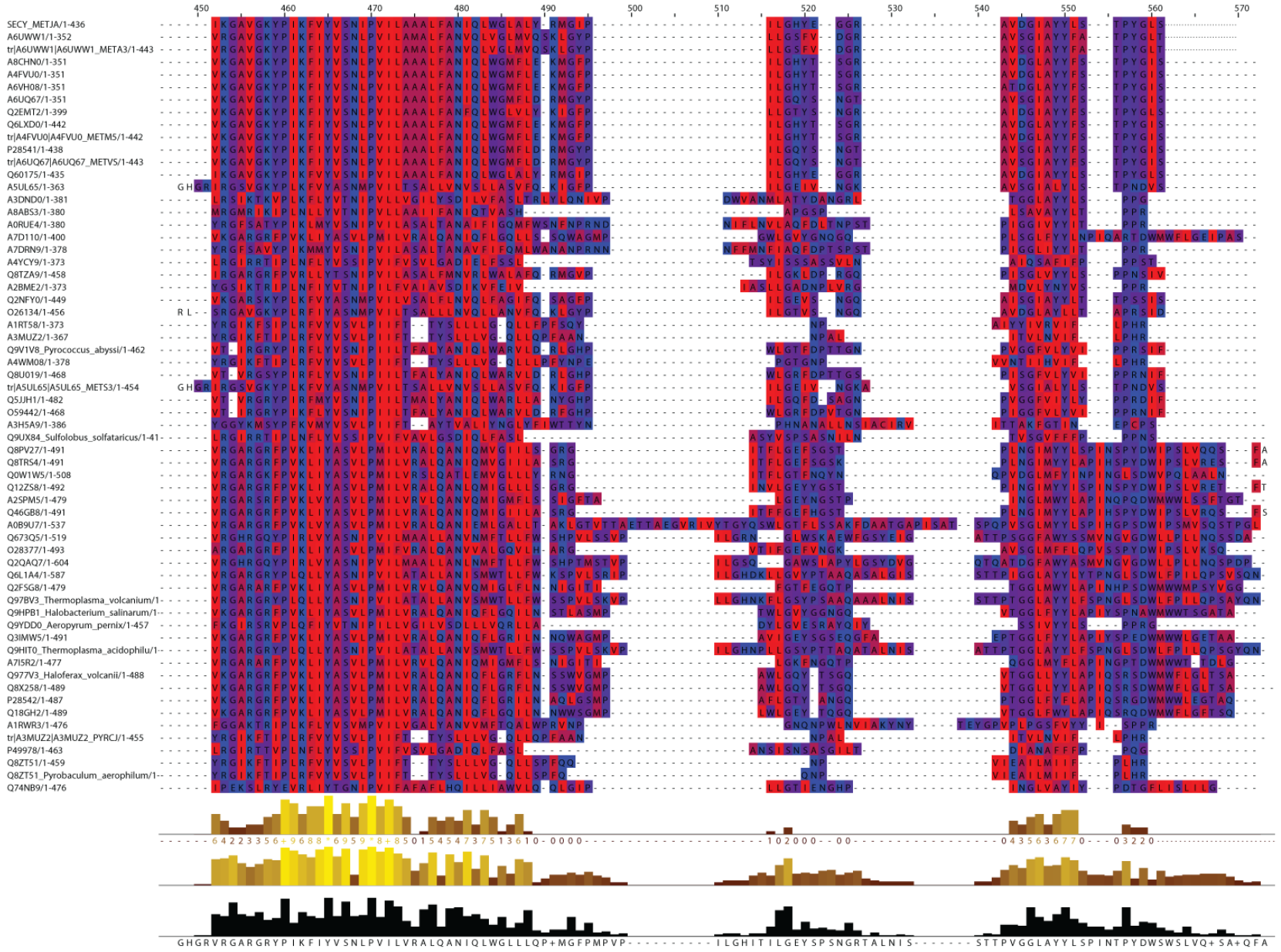
Figure S12. SecY sequence alignment for archaea referred to in Figure 3. The 63 sequences of SecY from archaea were aligned as described in 'Protocol for sequence analysis', below. The organism names and Pfam access codes for all SecY archaeal sequences are given in Table 4.

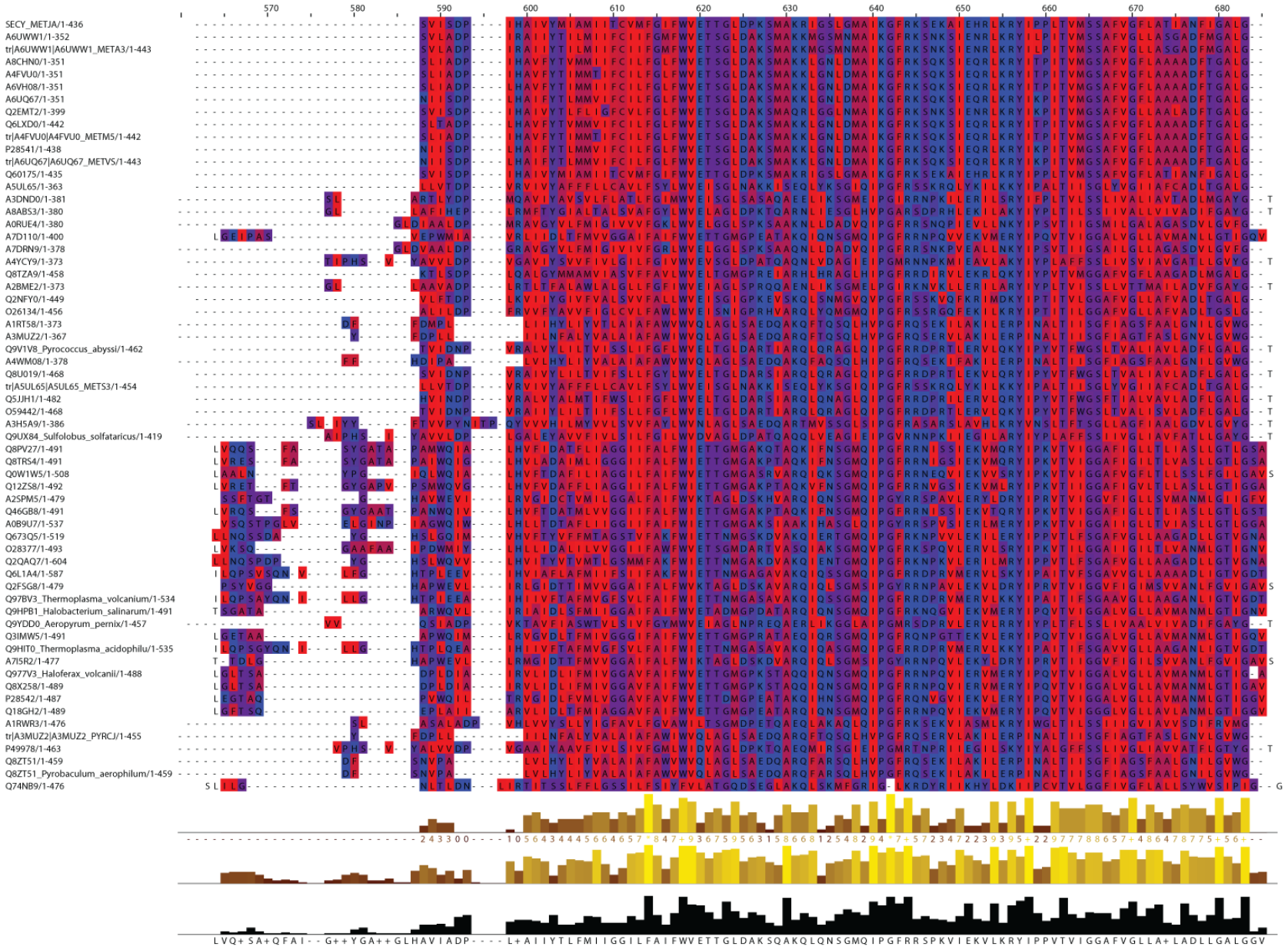



120 130 140 150 160 170 180 190 200 210 220 230 240
SECY_METJA/1-436
A6UWW1/1-352
tr|A6UWW1|A6UWW1_META3/1-443
A8CHN0/1-351
A4FVU0/1-351
A6VH08/1-351
A6UQ67/1-351
Q2EMT2/1-399
Q6LXD0/1-442
tr|A4FVU0|A4FVU0_METM5/1-442
P28541/1-438
tr|A6UQ67|A6UQ67_METVS/1-443
Q60175/1-435
A5UL65/1-363
A3DND0/1-381
A8ABS3/1-380
A0RUE4/1-380
A7D110/1-400
A7DRN9/1-378
A4YCY9/1-373
Q8TZA9/1-458
A2BME2/1-373
Q2NFY0/1-449
O26134/1-456
A1RT58/1-373
A3MUZ2/1-367
Q9V1V8_Pyrococcus_abyssi/1-462
A4WM08/1-378
Q8U019/1-468
tr|A5UL65|A5UL65_METS3/1-454
Q5JJH1/1-482
O59442/1-468
A3H5A9/1-386
Q9UX84_Sulfolobus_solfataricus/1-419
Q8PV27/1-491
Q8TRS4/1-491
Q0W1W5/1-508
Q12ZS8/1-492
A2SPM5/1-479
Q46GB8/1-491
A0B9U7/1-537
Q673Q5/1-519
O28377/1-493
Q2QAQ7/1-604
Q6L1A4/1-587
Q2FSG8/1-479
Q97BV3_Thermoplasma_volcanium/1-534
Q9HPB1_Halobacterium_salinarum/1-491
Q9YDD0_Aeropyrum_pernix/1-457
Q3IMW5/1-491
Q9HIT0_Thermoplasma_acidophilu/1-535
A7I5R2/1-477
Q977V3_Haloferax_volcanii/1-488
Q8X258/1-489
P28542/1-487
Q18GH2/1-489
A1RWR3/1-476
tr|A3MUZ2|A3MUZ2_PYRCJ/1-455
P49978/1-463
Q8ZT51/1-459
Q8ZT51_Pyrobaculum_aerophilum/1-459
Q74NB9/1-476
KLLKWTTGLILVLYF+LTTNIPLYG+HYTGG+LSPASIDLFAQFRAIVLAGASGTLITLGIGPIVTAGIILQLLVGSEIIKLDLSKNPEDRALFQGLQKLL+IVMIFLEALPFVLGGA+WTS


240
250
260
270
280
290
300
310
320
330
340
350
360
SECY_METJA/1-436
A6UWW1/1-352
tr|A6UWW1|A6UWW1_META3/1-443
A8CHN0/1-351
A4FVU0/1-351
A6VH08/1-351
A6UQ67/1-351
Q2EMT2/1-399
Q6LXD0/1-442
tr|A4FVU0|A4FVU0_METM5/1-442
P28541/1-438
tr|A6UQ67|A6UQ67_METVS/1-443
Q60175/1-435
A5UL65/1-363
A3DND0/1-381
A8AB53/1-380
A0RUE4/1-380
A7D110/1-400
A7DRN9/1-378
A4YCY9/1-373
Q8TZA9/1-458
A2BME2/1-373
Q2NFY0/1-449
O26134/1-456
A1RT58/1-373
A3MUZ2/1-367
Q9V1V8_Pyrococcus_abyssi/1-462
A4WM08/1-378
Q8U019/1-468
tr|A5UL65|A5UL65_METS3/1-454
Q5JJH1/1-482
O59442/1-468
A3H5A9/1-386
Q9UX84_Sulfolobus_solfataricus/1-419
Q8PV27/1-491
Q8TRS4/1-491
Q0W1W5/1-508
Q12ZS8/1-492
A2SPM5/1-479
Q46GB8/1-491
A0B9U7/1-537
Q673Q5/1-519
O28377/1-493
Q2QAQ7/1-604
Q6L1A4/1-587
Q2FSG8/1-479
Q97BV3_Thermoplasma_volcanium/1-534
Q9HPB1_Halobacterium_salinarum/1-491
Q9YDD0_Aeropyrum_pernix/1-457
Q3IMW5/1-491
Q9HIT0_Thermoplasma_acidophilu/1-535
A7I5R2/1-477
Q977V3_Haloferax_volcanii/1-488
Q8X258/1-489
P28542/1-487
Q18GH2/1-489
A1RWR3/1-476
tr|A3MUZ2|A3MUZ2_PYRCJ/1-455
P49978/1-463
Q8ZT51/1-459
Q8ZT51_Pyrobaculum_aerophilum/1-459
Q74NB9/1-476
GGA+WTSQFGPPDTPAVASIN----ASLGVGTPLLALLIVLQLALGGILII+LDEVVSKGWGIGSGIGLFIAAGVSQQIFVGLFNWEPVPDGN+P+LGVG--------LIPALGAALYD+LPSGYIPK


350
360
370
380
390
400
410
420
430
440
450
460
470
SECY_METJA/1-436
A6UWW1/1-352
tr|A6UWW1|A6UWW1_META3/1-443
A8CHN0/1-351
A4FVU0/1-351
A6VH08/1-351
A6UQ67/1-351
Q2EMT2/1-399
Q6LXD0/1-442
tr|A4FVU0|A4FVU0_METM5/1-442
P28541/1-438
tr|A6UQ67|A6UQ67_METVS/1-443
Q60175/1-435
A5UL65/1-363
A3DND0/1-381
A8AB53/1-380
A0RUE4/1-380
A7D110/1-400
A7DRN9/1-378
A4YCY9/1-373
Q8TZA9/1-458
A2BME2/1-373
Q2NFY0/1-449
O26134/1-456
A1RT58/1-373
A3MUZ2/1-367
Q9V1V8_Pyrococcus_abyssi/1-462
A4WM08/1-378
Q8U019/1-468
tr|A5UL65|A5UL65_METS3/1-454
Q5JJH1/1-482
O59442/1-468
A3H5A9/1-386
Q9UX84_Sulfolobus_solfataricus/1-419
Q8PV27/1-491
Q8TRS4/1-491
Q0W1W5/1-508
Q12ZS8/1-492
A2SPM5/1-479
Q46GB8/1-491
A0B9U7/1-537
Q673Q5/1-519
O28377/1-493
Q2QAQ7/1-604
Q6L1A4/1-587
Q2FSG8/1-479
Q97BV3_Thermoplasma_volcanium/1-534
Q9HPB1_Halobacterium_salinarum/1-491
Q9YDD0_Aeropyrum_pernix/1-457
Q3IMW5/1-491
Q9HIT0_Thermoplasma_acidophilu/1-535
A7I5R2/1-477
Q977V3_Haloferax_volcanii/1-488
Q8X258/1-489
P28542/1-487
Q18GH2/1-489
A1RWR3/1-476
tr|A3MUZ2|A3MUZ2_PYRCJ/1-455
P49978/1-463
Q8ZT51/1-459
Q8ZT51_Pyrobaculum_aerophilum/1-459
Q74NB9/1-476


450
460
470
480
490
500
510
520
530
540
550
560
570
SECY_METJA/1-436
A6UWW1/1-352
tr|A6UWW1|A6UWW1_META3/1-443
A8CHN0/1-351
A4FVU0/1-351
A6VH08/1-351
A6UQ67/1-351
Q2EMT2/1-399
Q6LXD0/1-442
tr|A4FVU0|A4FVU0_METM5/1-442
P28541/1-438
tr|A6UQ67|A6UQ67_METVS/1-443
Q60175/1-435
A5UL65/1-363
A3DND0/1-381
A8ABS3/1-380
A0RUE4/1-380
A7D110/1-400
A7DRN9/1-378
A4YCY9/1-373
Q8TZA9/1-458
A2BME2/1-373
Q2NFY0/1-449
O26134/1-456
A1RT58/1-373
A3MUZ2/1-367
Q9V1V8_Pyrococcus_abyssi/1-462
A4WM08/1-378
Q8U019/1-468
tr|A5UL65|A5UL65_METS3/1-454
Q5JJH1/1-482
O59442/1-468
A3H5A9/1-386
Q9UX84_Sulfolobus_solfataricus/1-41
Q8PV27/1-491
Q8TRS4/1-491
Q0W1W5/1-508
Q12ZS8/1-492
A2SPM5/1-479
Q46GB8/1-491
A0B9U7/1-537
Q673Q5/1-519
O28377/1-493
Q2QAQ7/1-604
Q6L1A4/1-587
Q2FSG8/1-479
Q97BV3_Thermoplasma_volcanium/1-
Q9HPB1_Halobacterium_salinarum/1-
Q9YDD0_Aeropyrum_pernix/1-457
Q3IMW5/1-491
Q9HIT0_Thermoplasma_acidophilu/1-
A7I5R2/1-477
Q977V3_Haloferax_volcanii/1-488
Q8X258/1-489
P28542/1-487
Q18GH2/1-489
A1RWR3/1-476
tr|A3MUZ2|A3MUZ2_PYRCJ/1-455
P49978/1-463
Q8ZT51/1-459
Q8ZT51_Pyrobaculum_aerophilum/1-
Q74NB9/1-476
GHGRVRGARGRYPIKFIYVSNLPVILVRALQANIQLWGLLLQP+MGFPMPVP
ILGHITILGEYSPSNGRTALNIS
STTPVGGLAYYLSPINTPYDWSWSLVQ+SA+QFA


570
580
590
600
610
620
630
640
650
660
670
680
SECY_METJA/1-436
A6UWW1/1-352
tr|A6UWW1|A6UWW1_META3/1-443
A8CHN0/1-351
A4FVU0/1-351
A6VH08/1-351
A6UQ67/1-351
Q2EMT2/1-399
Q6LXD0/1-442
tr|A4FVU0|A4FVU0_METM5/1-442
P28541/1-438
tr|A6UQ67|A6UQ67_METVS/1-443
Q60175/1-435
A5UL65/1-363
A3DND0/1-381
A8ABS3/1-380
A0RUE4/1-380
A7D110/1-400
A7DRN9/1-378
A4YCY9/1-373
Q8TZA9/1-458
A2BME2/1-373
Q2NFY0/1-449
O26134/1-456
A1RT58/1-373
A3MUZ2/1-367
Q9V1V8_Pyrococcus_abyssi/1-462
A4WM08/1-378
Q8U019/1-468
tr|A5UL65|A5UL65_METS3/1-454
Q5JJH1/1-482
O59442/1-468
A3H5A9/1-386
Q9UX84_Sulfolobus_solfataricus/1-419
Q8PV27/1-491
Q8TRS4/1-491
Q0W1W5/1-508
Q12ZS8/1-492
A2SPM5/1-479
Q46GB8/1-491
A0B9U7/1-537
Q673Q5/1-519
O28377/1-493
Q2QAQ7/1-604
Q6L1A4/1-587
Q2FSG8/1-479
Q97BV3_Thermoplasma_volcanium/1-534
Q9HPB1_Halobacterium_salinarum/1-491
Q9YDD0_Aeropyrum_pernix/1-457
Q3IMW5/1-491
Q9HIT0_Thermoplasma_acidophilu/1-535
A7I5R2/1-477
Q977V3_Haloferax_volcanii/1-488
Q8X258/1-489
P28542/1-487
Q18GH2/1-489
A1RWR3/1-476
tr|A3MUZ2|A3MUZ2_PYRCJ/1-455
P49978/1-463
Q8ZT51/1-459
Q8ZT51_Pyrobaculum_aerophilum/1-459
Q74NB9/1-476

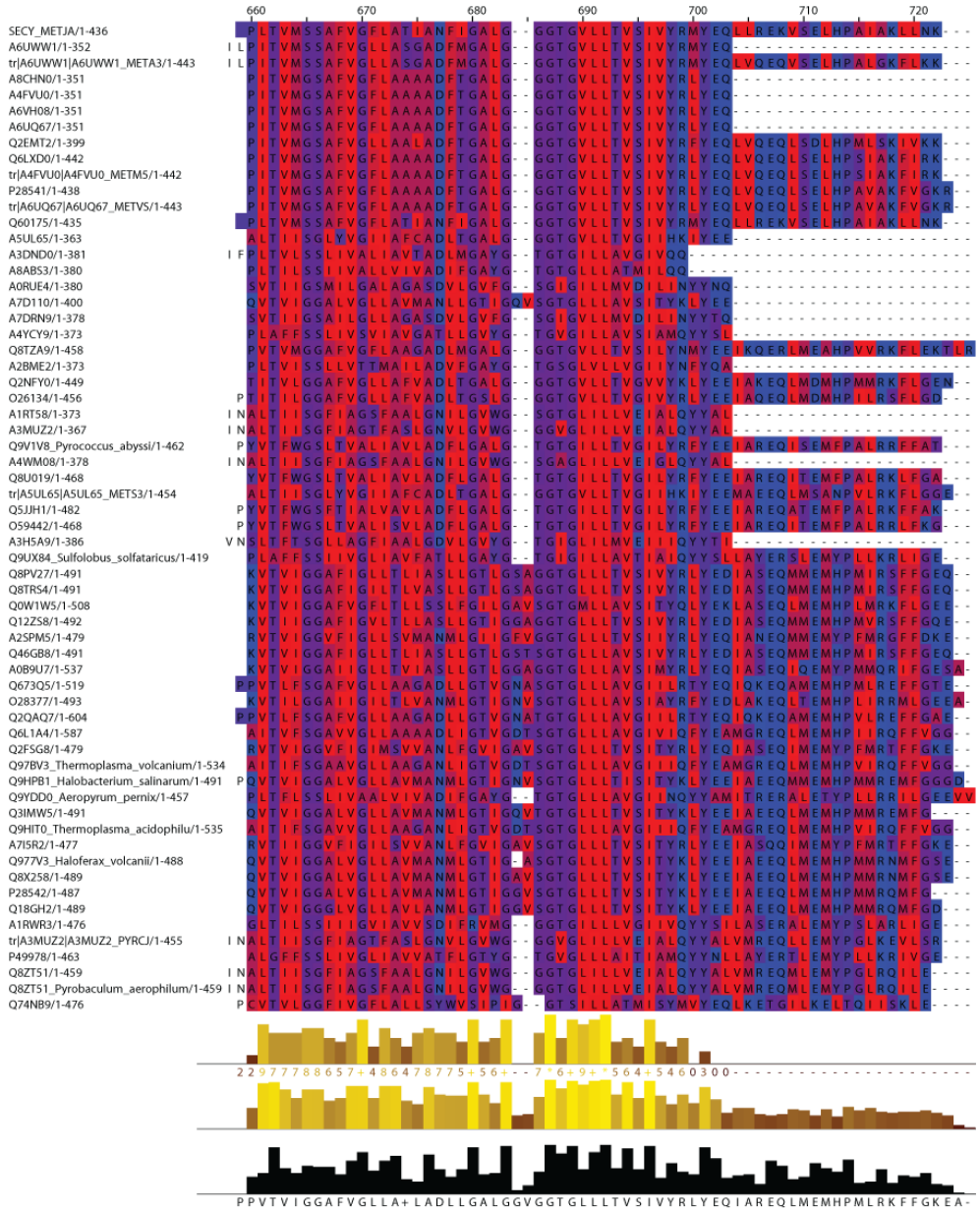

660
670
680
690
700
710
720
SECY_METJA/1-436
A6UWW1/1-352
tr|A6UWW1|A6UWW1_META3/1-443
A8CHN0/1-351
A4FVU0/1-351
A6VH08/1-351
A6UQ67/1-351
Q2EMT2/1-399
Q6LXD0/1-442
tr|A4FVU0|A4FVU0_METM5/1-442
P28541/1-438
tr|A6UQ67|A6UQ67_METVS/1-443
Q60175/1-435
A5UL65/1-363
A3DND0/1-381
A8ABS3/1-380
A0RUE4/1-380
A7D110/1-400
A7DRN9/1-378
A4YCY9/1-373
Q8TZA9/1-458
A2BME2/1-373
Q2NFY0/1-449
O26134/1-456
A1RT58/1-373
A3MUZ2/1-367
Q9V1V8_Pyrococcus_abyssi/1-462
A4WM08/1-378
Q8U019/1-468
tr|A5UL65|A5UL65_METS3/1-454
Q5JJH1/1-482
O59442/1-468
A3H5A9/1-386
Q9UX84_Sulfolobus_solfataricus/1-419
Q8PV27/1-491
Q8TRS4/1-491
Q0W1W5/1-508
Q12ZS8/1-492
A2SPM5/1-479
Q46GB8/1-491
A0B9U7/1-537
Q673Q5/1-519
O28377/1-493
Q2QAQ7/1-604
Q6L1A4/1-587
Q2FSG8/1-479
Q97BV3_Thermoplasma_volcanium/1-534
Q9HPB1_Halobacterium_salinarum/1-491
Q9YDD0_Aeropyrum_pernix/1-457
Q3IMW5/1-491
Q9HIT0_Thermoplasma_acidophilu/1-535
A7I5R2/1-477
Q977V3_Haloferax_volcanii/1-488
Q8X258/1-489
P28542/1-487
Q18GH2/1-489
A1RWR3/1-476
tr|A3MUZ2|A3MUZ2_PYRCJ/1-455
P49978/1-463
Q8ZT51/1-459
Q8ZT51_Pyrobaculum_aerophilum/1-459
Q74NB9/1-476
PPVTVIGGAFVGLLA+LADLLGALGGVGGTGLLLTVSIVYRLYEQIAREQLMEMHPMLRKFFGKEA-

Table S1. Summary of H-bonding analysis for the cytoplasmic half of wild-type SecY (Sim 1). Extent of hydrogen-bonding interactions observed in simulations of the SecYEG translocon. The extent is represented as percent of time hydrogen bonds were made. We analyzed the dynamics of distances for selected H-bonding amino acids in the wild-type translocon for the last 10 ns and 20 ns segments of Sim1 (Figure 1G). As the H-bonding criterion, we used a distance of less than 3.5 Å between the heavy atoms. See Figure S1 for the H-bonding clusters. Indicated in *italics* are amino acids whose mutation is known to cause translocation defects (see Table S3 for details).

Cluster	Hydrogen bond		Last 10 ns	Last 20 ns
CP1	S344-Oγ	D341-Oδ1	89.2	69.0
	S344-Oγ	D341-Oδ2	84.9	67.0
	S344-Oγ	R104-NH1	0	3.0
	R104-NH1	D341-Oδ1	0	9.2
	R104-NH1	D341-Oδ2	0	4.2
	R104-NH1	*E336*-O	22.3	13.8
	R104-NH2	*E336*-OE1	48.7	24.4
	T337-Oγ1	*Q86*-Nε2	0.1	6.2
	T338-Oγ1	*F333*-O	25.4	36.9
CP2	K343-NZ	E367-Oε1	44.9	73.9
	K343-NZ	E367-Oε2	85.8	73.2
	K347-NZ	E363-Oε1	27.5	15.7
	K347-NZ	E363-Oε2	32.4	18.0
	K371-NZ	E367-Oε1	7.6	12.0
	R369-NH1	L234-O	44.1	26.1
	R372-NH2	SecE_E17-Oε1	75.1	79.7
	R230-NH1	SecE_E32-Oε1	54.6	56.9
	E232-Oε1	SecE_K26-NZ	21.9	13.2
	E232-Oε1	K246-NZ	51.7	62.3
	E232-Oε2	K246-NZ	81.1	66.4
CP3	E102-Oε1	K3-NZ	12.4	40.3
	E102-Oε2	K3-NZ	12.0	45.5
	E9-Oε1	K10-NZ	65.0	46.8
	E9-Oε2	K10-NZ	4.3	22.0
	E13-N	Secβ_S25-OG	61.5	50.0
CP4	*K112*-NZ	E13-Oε2	28.8	61.7
	K112-NZ	*E159*-Oε2	33.5	35.3
	K163-NZ	*E159*-Oε2	0.3	0.7
	D158-Oδ1	I170-N	45.6	47.6
	D158-Oδ2	I170-N	26.9	30.2
	D158-Oδ1	G169-N	13.3	11.4
CP-5	E25-Oε2	K19-NZ	87.1	87.8
	E421-Oε2	K26-NZ	50.4	46.3
	E425-Oε2	K24-NZ	36.1	18.0
	E425-Oε1	K24-NZ	38.8	19.4
CP-6	S409-Oγ	*E227*-Oε2	99.9	96.6
	K250-NZ	*E416*-Oε2	38.2	49.3
	K250-NZ	*E227*-Oε1	5.9	3.9
	E416-Oε2	R413-NH1	10.5	41.3
	E416-Oε1	R420-NH2	2.4	10.1
	R419-NH1	L432-OT2	54.3	27.1
	S382-Oγ	S255-Oγ	10.5	9.6
	S381-Oγ	S255-Oγ	98.5	98.6
	S381-Oγ	F251-O	100.0	99.9
	S409-Oγ	*N256*-Nd2	0.3	0.3
	N393-Nδ2	T402-Oγ1	97.3	97.1

Table S2. Summary of H-bonding analysis for the periplasmic half of wild-type SecY (Sim 1). Extent of H-bonding interactions observed in simulations of the SecYEG translocon. The extent is represented as percent of time H bonds were made. We analyzed the dynamics of distances for selected H-bonding amino acids in the wild-type translocon for the last 10 ns and 20 ns segments of Sim1 (Figure 1G). As the H-bonding criterion, we used a distance of less than 3.5 Å between the heavy atoms. See Figure S1 for the H-bonding clusters. Indicated in *italics* are amino acids whose mutation is known to cause translocation defects (see Table S3 for details).

Cluster	Hydrogen bond		Last 10ns	Last 20 ns
EC-1	E122-Oε2	T80-Oγ1	23.1	35.6
	E122-Oε1	*W272*-Nε1	75.8	64.4
	E122-Oε1	N268-Nδ2	85.8	80.3
	T80-Oγ1	N268-Nδ2	95.0	94.9
EC-2	R278-NH2	E288-O	78.0	83.4
	R291-NH1	T302-O	100.0	100.0
	R291-NH1	S301-Oγ	21.4	19.9
	T302-Oγ1	Q270-Nε2	47.4	53.1
	E57-O	Y304-OH	7.0	17.1
EC-3	E190-Oε2	K195-NZ	37.1	44.0
	E190-Oε1	K195-NZ	49.8	40.2
	Y192-OH	E190-Oε1	0.4	1.0
	Y192-OH	E190-Oε2	0.6	1.6
	Y209-OH	K195-NZ	1.7	1.4
	E190-Oε1	SecE_Y60-OH	22.4	18.5
EC-4	*W59*-Nε1	*T72/V*-O	99.8	99.9
	T72-Oγ1	*T69*-N	97.9	96.6
	T72-Oγ1	*T69*-O	99.5	99.7
	S151-Oγ	I147-O	91.6	95.6
	S151-Oγ	G74-N	23.9	51.3
	L70-O	*T72*-N	77.4	81.1
	Q60-Oε1	*R66*-NH1	6.9	9.4
	T69-Oγ1	*R66*-O	100.0	100.0
	Q60-Oε1	I67-N	92.7	93.0
	S65-Oγ	T63-Oγ1	96.8	96.6

Table S3. Known mutation effects of H-bonding amino acids relevant to Figures 4-9. See Figure S5 and S12 for detailed sequence alignments. Summaries of the effects of specific mutations on the function of the *E. coli* and yeast translocons are given in Smith et al., 2005, and Junne et al., 2006, respectively.

Amino acid	Location	Figure	yeast[a] / *E. Coli*[b]	Effect of mutation	Reference	Revised asignment[3]
D44	TM1	S1 I	P40S[b]*	SecY100	Ito 1989 Smith 2005	
W59	plug	S1 J	L66N[a]	affects topology[1]	Junne 2006	L53[b] gap[c]
			F64C[b]	prlA300	Osborne 1993	
				prlA300:open-plug[2]	Smith 2005	
Q60	plug	S1 J	R67E[a]	affects topology	Junne 2007	
T63	plug	S1 J	L70N[a]	affects topology[1]	Junne 2007	
R66	plug	S1 I, J	A71D[b]	prlA302	Osborne 1993	R57[b] gap[c]
				prlA302:open-plug[2]	Smith 2005	
T69	plug	S1 J	S76F[b]	secY125	Taura 1994	
T72	TM2	S1 J	E79G[a]	affects topology[1]	Junne 2007	
Q86	TM2	S1 A	Q93R[a]	affects topology[1]	Junne 2007	
K112	TM3	S1 D	R121C	reduced functionality	Mori et al, 2004	
S151	TM4	S1 J	S161T[a]	affects topology[1]; prl	Junne 2007	
D158	TM4	S1 D	D168A[a]	affects topology[1]	Junne 2007	
			G175D[b]	SecY104	Taura 1994	
E159	TM4	S1 D	E176Q[b]	affects translocation	van der Sluis 2006	
			E176C[b]	reduced functionality	Mori et al, 2004	
E227	C4/TM6	S1 F	Q261R[a]	affects topology[1]	Junne 2007	
K250	TM7/C4	S1 F	K259E[a]	prl affects topology[1]	Junne 2007	
N256	TM7	S1 F	V274G[b]	prlA1	Emr 1981, Osborne 1993	
				prlA1: CS stable[2]	Smith 2005	
N268	TM7	S1 F	F286Y	restores translocation in I408N	Duong & Wickner 1999	
W272	TM7	S1 G	I290T[b]	secY121	Sako 1991	gap[b] I183[c]
F333	TM8	S1 A	T379I	affects topology[1]	Junne 2007	
E336	TM8	S1 A	E382R[a]	affects topology[1]	Junne 2007	
E416	TM10	S1 F	E460K	affects topology[1]	Junne 2007	

[1]June et al., 2007 assessed the effect of mutating specific residues on the membrane protein topology by measuring the efficiency of translocation of substrates with different charge distributions of the signal anchors.

[2]Smith et al., 2005 categorized the prl suppressor mutations with respect to their mechanism. Two classes of prlA suppressor mutations are open-ring stabilization (e.g., prlA300 and prlA302), and closed-state (CS) destabilization (e.g., prlA1). *sec* phenotypes are nonfunctional under restrictive conditions; in contrast, *prl* phenotypes expand the translocation function of SecY/Sec61 to substrates with mutant/absent signal peptides.

[3]Revised correspondence of amino acids from the sequences of *M. janaaschii*, *E. coli*, and *S. cerevisiae* based on the sequence alignment from Figure S6.

[a]The sequence of *S. cerevisiae*.

[b]The sequence of *E. coli*.

(*) The mutation effect is caused by a multiple mutation.

Table S4. The organism names and Pfam access codes for all SecY archaeal sequence alignments of Figure S12.

Accnr	Organism
SECY_METJA	*Methanococcus jannaschii:*
3din_F_PDB_sequence	*Methanococcus jannaschii:*
tr|A3MUZ2|A3MUZ2_PYRCJ	*Pyrobaculum calidifontis*
tr|A4FVU0|A4FVU0_METS5	*Methanococcus maripaludis*
tr|A5UL65|A5UL65_METS3	*Methanobrevibacter smithii*
tr|A6UQ67|A6UQ67_METVS	*Methanococcus vannielii*
tr|A6UWW1|A6UWW1_META3	*Methanococcus aeolicus*
SECY_METJA	*Methanococcus jannaschii:*
Q9YDD0	*Aeropyrum_pernix*
Q9V1V8	*Pyrococcus_abyssi*
Q9UX84	*Sulfolobus_solfataricus*
Q9HPB1	*Halobacterium_salinarum*
Q9HIT0	*Thermoplasma_acidophilum*
Q97BV3	*Thermoplasma_volcanium*
Q977V3	*Haloferax_volcanii*
Q8ZT51	*Pyrobaculum_aerophilum*
Q8X258	*Haloferax_volcanii*
Q8U019	*Pyrococcus_furiosus*
Q8TZA9	*Methanopyrus_kandleri*
Q8TRS4	*Methanosarcina_acetivorans*
Q8PV27	*Methanosarcina_mazei*
Q74NB9	*Nanoarchaeum_equitans*
Q6LXD0	*Methanococcus_maripaludis*
Q6L1A4	*Picrophilus_torridus*
Q673Q5	*uncultured_marine group II euryarchaeote DeepAnt-JyK*
Q60175	*Methanocaldococcus_jannaschii*
Q5JJH1	*Thermococcus_kodakarensis KOD1*
Q46GB8	*Methanosarcina_barkeri str. Fusaro*
Q3IMW5	*Natronomonas_pharaonis DSM 2160*
Q2QAQ7	*uncultured_marine group II euryarchaeote HF70_59C08*
Q2NFY0	*Methanosphaera_stadtmanae DSM 3091*
Q2FSG8	*Methanospirillum_hungatei JF-1*
Q2EMT2	*Methanococcus_voltae*
Q18GH2	*Haloquadratum_walsbyi DSM 16790*
Q12ZS8	*Methanococcoides_burtonii DSM 6242*
Q0W1W5	*uncultured_methanogenic archaeon*
P28541	*Methanococcus_vannielii*
P28542	*Haloarcula_marismortui*
P49978	*Sulfolobus_acidocaldarius*
O26134	*Methanothermobacter_thermautotrophicus str. Delta H*
O28377	*Archaeoglobus_fulgidus*
O59442	*Pyrococcus_horikoshii*
A8CHN0	*Methanococcus maripaludis*
A8ABS3	*Ignicoccus hospitalis*
A7I5R2	*Methanoregula boonei (strain 6A8)*
A7DRN9	*Candidatus Nitrosopumilus maritimus SCM1*
A7D110	*Halorubrum lacusprofundi ATCC 49239*
A6VH08	*Methanococcus maripaludis (strain C7 / ATCC BAA-1331) GN=MmarC7*
A6UWW1	*Methanococcus aeolicus (strain Nankai-3 / ATCC BAA-1280) GN=Maeo*
A6UQ67	*Methanococcus vannielii (strain SB / ATCC 35089 / DSM 1224) GN=Mevan*
A5UL65	*Methanobrevibacter smithii (strain PS / ATCC 35061 / DSM 861) GN=Msm*
A4YCY9	*Metallosphaera sedula*
A4WM08	*Pyrobaculum arsenaticum*
A4FVU0	*Methanococcus maripaludis (strain C5 / ATCC BAA-1333) GN=MmarC5*
A3MUZ2	*Pyrobaculum calidifontis (strain JCM 11548 / VA1) GN=Pcal*
A3H5A9	*Caldivirga maquilingensis IC-167*
A3DND0	*Staphylothermus marinus*
A2SPM5	*Methanocorpusculum labreanum*
A2BME2	*Hyperthermus butylicus*
A1RWR3	*Thermofilum pendens (strain Hrk 5) GN=Tpen*
A1RT58	*Pyrobaculum islandicum*
A0RUE4	*Cenarchaeum symbiosum*
A0B9U7	*Methanosaeta thermophila*

Protocol for sequence analysis.

The initial list of SecY proteins was compiled using the PFAM database (Finn et al., 2008) Hidden Markov Model (HHM) PF00344 for SecY, which has a coverage of 78 % of the full-length sequences.

SecY PFAM model offers a hand curated seed alignment of 17 reference sequences. The seed SecY alignment HMM (Durbin et al., 1999) is used to generate a full alignment, which are all related sequences with score higher than the manually set threshold values for the HMMs of a particular PFAM entry (Eddy et al, 2001), in this case the SecY family. We used the sequences selected by those models with the best scores for further analysis. The full alignment contained 1154 sequences from all Phyla and is not hand curated.

The sequences were then divided in three groups: archeas, bacteria and eukaryotes. To create the definitive full-length alignments we used T-coffee (Notredame et al., 2000), which allows aligning with good accuracy, profiles, structures and individual sequences. Resulting full-length analysis were manually inspected.

Conservation analysis, paying special attention to the conservation of hydrophobicity was generated following the color scheme from Kyte and Doolittle (Kyte and Doolittle, 1982). According to this scheme, the most hydrophobic residues are colored in red, and the most hydrophilic residues in blue. Tables with the frequencies of the amino acids were generated from the alignments for each important position in the alignment indicating with the Kite & Doolittle color scheme if the hydrophobic/hydrophilic properties for a certain position are conserved.

Alignment figures and modifications were made using Jalview (Clamp et al., 2004), a Java multiple alignment editor and analysis tool. The amino acid histogram representation for each position in the alignment was perform using a java implementation of LogoBar (Perez-Bercoff et al., 2006)

Supplementary References

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, J.S., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L., and Bateman, A. (2008). Nucleic Acids Research Database Issue 36: D281-D288.

Duong, F., and Wickner, W. (1999). The prlA and prlG phenotypes are caused by a loosened association among the translocase SecYEG subunits. EMBO J. 18: 3263-3270.

Durbin, R., Eddy, S.R., Krogh, A., and Mitchison, G. (1999). Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press. ISBN 0-521-62971-3.

Eddy, S.R. (2001). HMMER User's Guide: Biological sequence analysis using profile hidden Markov models, version 2.2. Washington University School of Medicine, 2001.

Emr, S. D., Hanley-Way, S., and Silhavy, T. J. (1981). Supressor mutations that restore export of a protein with a defective signal sequence. Cell 23: 79-88.

Notredame C, Higgins DG, Heringa J.T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol. 2000 Sep 8;302(1):205-17.

Kyte, J., and Doolittle, R.F., J. Mol. Biol. 1157, 105-132, 1982.

Pérez-Bercoff, Å., Koch, J. and Bürglin, T.R. (2006) LogoBar: bar graph visualization of protein logos with gaps. Bioinformatics, 22, 112-114.

Clamp, M., Cuff, J., Searle, S. M. and Barton, G. J. (2004), The Jalview Java Alignment Editor, Bioinformatics, 20, 426-7.

Ito, K., Hirota, Y., and Akiyama, Y. (1989). Temperature-sensitive *sec* mutants of *Escherichia coli*: inhibition of protein export at permissive temperatures. J. Bacteriol. 171: 1742-1743.

Junne, T., Schwede, T., Goder, V., and Spiess, M. (2007). Mutations in the Sec61p channel affecting signal sequence recognition and membrane topology. J. Biol. Chem. 282: 33201-33209.

Mori, H. Shimokawa, N., Satoh, Y., and Ito, K. (2004). Mutational analysis of transmembrane regions 3 and 4 of SecY, a central component of protein translocase. J. Bacteriol. 186: 3960-3969.

Osborne, R. S., and Silhavy, T. J. (1993). PrlA suppressor mutations cluster in regions corresponding to three distinct topological domains. EMBO J. 12: 3391-3398.

Sako, T. (1991). Novel *prlA* alleles defective in supporting staphylokinase processing in *Escherichia coli*.

Smith, M. A., Clemons, W. M., DeMars C. J., and Flower, A. M. (2005) Modeling the effects of prl mutations on the *Escherichia coli* SecY complex. J. Bacteriol. 187, 6454-6465

Taura, T., Akiyama, Y., and Ito, K. (1994). Genetic analysis of SecY: additional export-defective mutations and factors affecting their phenotype. Mol. Gen. Genet. 243: 261-269.

Tsukazaki, T., Mori, H., Fukai, S., Ishitani, R., Mori, T., Dohmae, N., Prederina, A., Sugita, Y., Vassylyev, D. G., Ito, K., and Nureki, O. (2008). Conformational transition of Sec machinery inferred from bacterial SecYE structures. Nature 455: 988-991.

Van den Berg, B., Clemons, W.M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapoport, T.A. (2004). X-ray structure of a protein-conducting channel. Nature *427*, 36-44.

Van der Sluis, E. O., Nouwen, N., Koch, J., de Keyzer, J., van der Does, C., Tampé, R., and Driessen, A. J. M. (2006). Identification of two interaction sites in SecY that are important for the functional interaction with SecA. J. Mol. Biol. 361: 839-849.

Zimmer, J., Nam, Y., and Rapoport, T. A. (2008). Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature 455: 936-943.