ELSEVIER

Biochimica et Biophysica Acta 1609 (2003) 115–125

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

www.bba-direct.com

The general protein secretory pathway: phylogenetic analyses leading to evolutionary conclusions

Thien B. Cao, Milton H. Saier Jr.*

Division of Biology, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0116, USA

Received 15 August 2002; received in revised form 21 October 2002; accepted 31 October 2002

Abstract

We have identified all homologues in the current databases of the ubiquitous protein constituents of the general secretory (Sec) pathway. These prokaryotic/eukaryotic proteins include (1) SecY/Sec61α, (2) SecE/Sec61γ, (3) SecG/Sec61β, (4) Ffh/SRP54 and (5) FtsY/SRP receptor subunit-α. Phylogenetic and sequence analyses lead to major conclusions concerning (1) the ubiquity of these proteins in living organisms, (2) the topological uniformity of some but not other Sec constituents, (3) the orthologous nature of almost all of them, (4) a total lack of paralogues in almost all organisms for which complete genome sequences are available, (5) the occurrence of two or even three paralogues in a few bacteria, plants, and yeast, depending on the Sec constituent, and (6) a tremendous degree of sequence divergence in bacteria compared with that in archaea or eukaryotes. The phylogenetic analyses lead to the conclusion that with a few possible exceptions, the five families of Sec constituents analyzed generally underwent sequence divergence in parallel but at different characteristic rates. The results provide evolutionary insights as well as guides for future functional studies. Because every organism with a fully sequenced genome exhibits at least one orthologue of each of these Sec proteins, we conclude that all living organisms have relied on the Sec system as their primary protein secretory/membrane insertion system. Because most prokaryotes and many eukaryotes encode within their genomes only one of each constituent, we also conclude that strong evolutionary pressure has minimized gene duplication events leading to the establishment of Sec paralogues. Finally, the sequence diversity of bacterial proteins as compared with their archaeal and eukaryotic counterparts is in agreement with the suggestion that bacteria were the evolutionary predecessors of archaea and eukaryotes.

Keywords: Protein secretion; Membrane insertion; General secretory pathway; SecYEG; SRP; FtsY; Ffh

1. Introduction

Protein complexes of the general secretory (Sec) pathway (TC #3.A.5) are found in bacteria, archaea and eukaryotes [1–3]. Each translocase minimally consists of three integral inner membrane proteins, SecYEG/Sec61αγβ [4–6]. Direct contact between the *Escherichia coli* SecY, SecE and SecG proteins has been documented [4,6–8]. The heterotrimeric SecYEG complex can exist dynamically in monomeric, dimeric, tetrameric and possibly higher oligomeric states [9,10]. A low-resolution (8 Å) structure of the dimeric SecYEG complex is available [4]. A total of 30 transmembrane α-helical segments (TMSs) per dimer of trimers (15 TMSs per heterotrimer) has been deduced in this three-dimensional reconstruction [4].

The cytoplasmic ATPase/helicase, SecA [11,12], is normally a homodimer that may dissociate to a monomer as part of its catalytic cycle [13,14]. SecA has been reported to recruit SecYEG complexes to form an active secretory translocation channel, an assembly that may consist of a SecA homodimer and four monomeric SecYEG complexes [15]. The C-terminal domain of SecY interacts with SecA [16,17]. SecA homologues have been identified only in prokaryotes and plant/algal chloroplasts but not in archaea or in nonplastidic eukaryotic cells [3]. SecE/Sec61γ and SecG/Sec61β homologues were believed to be lacking in some organisms with completely sequenced genomes [3], but recent evidence suggests that all three components of the SecYEG complex are universal (Ref. [5] and this report).

Two auxiliary proteins, SecD and SecF in *E. coli*, which are fused in a single protein in *Bacillus subtilis* [18], are homologous to members of the RND superfamily of proton motive force (pmf)-driven transporters (TC #2.A.6) [12,19]. Another protein, YajC of *E. coli*, forms a complex with

* Corresponding author. Tel.: +1-858-534-4084; fax: +1-858-534-7108.

E-mail address: msaier@ucsd.edu (M.H. Saier).

0005-2736/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0005-2736(02)00662-4

Table 1
Sequenced members of the SecY/Sec61α protein translocon family[a]

Abbreviation	Organism (alphabetical order)	Size[b] (no. of amino acids)	Genbank gi no.[c]
1. Archaea		(477 ± 28)	
SecY Ape	*Aeropyrum pernix*	494	gi|12230620 (sp|Q9YDDO)
Sec61α Afu	*Archaeoglobus fulgidus*	493	gi|7446502
SecY Hma	*Haloarcula marismortui*	487	gi|134414 (sp|P28542)
SecY Hsp	*Halobacterium* sp.	491	gi|10581182
SecY Mth	*Methanobacterium thermoautotrophicum*	465	gi|7446501
SecY Mja	*Methanococcus jannaschii*	440	gi|2129214
SecY Mva	*Methanococcus vannielii*	438	gi|134416 (sp|P28541)
SecY Pab	*Pyrococcus abyssi*	468	gi|7446504
SecY Pho	*Pyrococcus horikoshii*	468	gi|7446505
SecY Sac	*Sulfolobus acidocaldarius*	463	gi|1711369 (sp|P49978)
SecY Sso	*Sulfolobus solfataricus*	469	gi|11134755 (sp|Q9UX84)
Sec61α Tac	*Thermoplasma acidophilum*	535	gi|10640594
Sec61α Tvo	*Thermoplasma volcanium*	565	gi|14324567
2. Eukaryotes		(474 ± 13)	
Sec61α Aae	*Aedes aegypti*	476	gi|13173171
SecY Ath3	*Arabidopsis thaliana*	475	gi|3834321
Sec61α Cal	*Candida albicans*	479	gi|7710957
Orf Cel	*Caenorhabditis elegans*	473	gi|7510283
Sec61α Cfa	*Canis familiaris*	476	gi|585957 (sp|P38377)
Sec61α Dme	*Drosophila melanogaster*	423	gi|7297121
Sec61α Hro	*Halocynthia roretzi*	475	gi|2500736 (sp|Q25147)
Sec61α Hsa1	*Homo sapiens*	476	gi|7019415
Sec61α Hsa2	*Homo sapiens*	476	gi|7705736
SecY Mmu	*Mus musculus*	476	gi|7673003
Sec61αA Omy	*Oncorhynchus mykiss*	476	gi|13517985
Sec61αB Omy	*Oncorhynchus mykiss*	476	gi|13517987
Sec61α Pvu	*Phaseolus vulgaris*	476	gi|6581004
Sec61α Pfa	*Plasmodium falciparum*	472	gi|3057044
Sec61α Psal	*Pyrenomonas salina*	494	gi|585958 (sp|P38379)
Sec61α Sce1 (Sec61p)	*Saccharomyces cerevisiae*	480	gi|6323411
Sec61α Spo	*Schizosaccharomyces pombe*	479	gi|2500734 (sp|P79088)
Sec61α Tae	*Triticum aestivum*	475	gi|8886324
Sec61α Yli	*Yarrowia lipolytica*	471	gi|2500735 (sp|P78979)
3. Eukaryotes (divergent)		(483 ± 11)	
Sec61α Ath2	*Arabidopsis thaliana*	475	gi|3834321
Sec61α Sce2 (Ssh1p)	*Saccharomyces cerevisiae*	490	gi|6319760
4. Mycoplasma		(480 ± 11)	
SecY Mca	*Mycoplasma capricolum*	482	gi|134417 (sp|P10250)
SecY Mga	*Mycoplasma gallisepticum*	498	gi|622613 (sp|O52351)
SecY Mge	*Mycoplasma genitalium*	475	gi|1351060 (sp|P47416)
SecY Mpn	*Mycoplasma pneumoniae*	477	gi|2500726 (sp|Q59548)
SecY Uur	*Ureaplasma urealyticum*	471	gi|6899220
5. Low G+C Gram-positive bacteria		(432 ± 3)	
SecY Bha	*Bacillus halodurans*	431	gi|585985 (sp|P38375)
SecY Bli	*Bacillus licheniformis*	431	gi|464752 (sp|Q05207)
SecY Bsu	*Bacillus subtilis*	431	gi|134409 (sp|P16336)
SecY Lla	*Lactococcus lactis*	439	gi|134415 (sp|P27148)
SecY Sau	*Staphylococcus aureus*	430	gi|3122859 (sp|O08387)
SecY Sca	*Staphylococcus carnosus*	430	gi|464753 (sp|Q05217)
SecY Spy	*Streptococcus pyogenes*	434	gi|13621388
6. High G+C Gram-positive bacteria		(440 ± 7)	
SecY Cgl	*Corynebacterium glutamicum*	460	gi|585986 (sp|P38376)
SecY Mlu	*Micrococcus luteus*	436	gi|417764 (sp|P33108)
SecY Mbo	*Mycobacterium bovis*	441	gi|3024605 (sp|P94926)
SecY Mle	*Mycobacterium leprae*	438	gi|2344857
SecY Msm	*Mycobacterium smegmatis*	438	gi|2911815
SecY Sgb	*Streptomyces galbus*	437	gi|2500728 (sp|Q59912)
SecY Sco	*Streptomyces coelicolor*	437	gi|6094267 (sp|P46785)
SecY Sgr	*Streptomyces griseus*	437	gi|2500729 (sp|Q59916)
SecY Sli	*Streptomyces lividans*	437	gi|1711368 (sp|P49977)
SecY Ssc	*Streptomyces scabies*	437	gi|1173421 (sp|P43416)

Table 1 (*continued*)

Abbreviation	Organism (alphabetical order)	Size[b] (no. of amino acids)	Genbank gi no.[c]
7. Chlamydia		(459 ± 3)	
SecY Cmu	*Chlamydia muridarum*	457	gi|7190825
SecY Ctr	*Chlamydia trachomatis*	457	gi|7404464 (sp|P28539)
SecY Cpn	*Chlamydophila pneumoniae*	462	gi|7189052
8. Primitive bacteria		(440 ± 14)	
SecY Aae	*Aquifex aeolicus*	429	gi|6226133 (sp|O66491)
SecY Dra	*Deinococcus radiodurans*	439	gi|7473361
SecY Lin	*Leptospira interrogans*	460	gi|5163224
SecY Tma	*Thermotoga maritima*	431	gi|7444763
9. Spirochetes		(442 ± 11)	
SecY Tpa	*Treponema pallidum*	450	gi|7444758
SecY Bbu	*Borrelia burgdorferi*	434	gi|3914976 (sp|O51451)
10. Gram-negative proteobacteria		(433 ± 24)	
SecY Ako	*Acyrthosiphon kondoi*	356	gi|1711366 (sp|P49976)
SecY Bsp	*Buchnera* sp.	437	gi|10039163
SecY Cje	*Campylobacter jejuni*	421	gi|6969103
SecY Ccr	*Caulobacter crescentus*	440	gi|13422601
SecY Eco	*Escherichia coli*	443	gi|134413 (sp|P03844)
SecY Hin	*Haemophilus influenzae*	441	gi|1173419 (sp|P43804)
SecY Hpy	*Helicobacter pylori*	420	gi|10720271 (sp|O25879)
SecY Mlo	*Mesorhizobium loti*	446	gi|13470571
SecY Nme	*Neisseria meningitidis*	436	gi|7225380
SecY Pmu	*Pasteurella multocida*	441	gi|12721770
SecY Pae	*Pseudomonas aeruginosa*	442	gi|9950460
SecY Rpr	*Rickettsia prowazekii*	433	gi|6226135 (sp|Q9ZCS5)
SecY Vch	*Vibrio cholerae*	444	gi|9657163
SecY Xfa	*Xylella fastidiosa*	457	gi|9106138
11. Chloroplasts, plastids, and bacteria		(461 ± 59)	
SecY Asp	*Antithamnion* sp.	405	gi|2500733 (sp|Q37143)
SecY Ath1	*Arabidopsis thaliana*	551	gi|4185137
SecY Cca	*Cyanidium caldarium*	410	gi|6466342
SecY Cpa	*Cyanophora paradoxa*	492	gi|1351059 (sp|P25014)
SecY Gth	*Guillardia theta*	420	gi|1351058 (sp|P28527)
SecY Osi	*Odontella sinensis*	425	gi|1351061 (sp|P49461)
SecY Plu	*Pavlova lutheri*	419	gi|1351062 (sp|P28540)
SecY Psat	*Pisum sativum*	527	gi|4929291
SecY Ppu	*Porphyra purpurea*	411	gi|1711367 (sp|P51297)
SecY Psal	*Pyrenomonas salina*	412	gi|585988
SecY Sol	*Spinacia oleracea*	545	gi|7484685
SecY Syn1	*Synechocystis* PCC6301	439	gi|401077 (sp|P31159)
SecY Syn2	*Synechocystis* PCC6803	442	gi|2500730 (sp|P77964)
SecY Zma	*Zea mays*	553	gi|7489840

[a] Proteins are grouped according to the phylogenetic group in which they are found (see Fig. 1).
[b] The average size ± S.D. is provided for each of the phylogenetic clusters (1–11).
[c] SwissProt/Tremble accession numbers are provided in parentheses.

SecD–SecF both independently of and in complexation with SecYEG [20–22]. The SecDF–YajC complex is not essential for secretion and is not found in all prokaryotes [19], but in *E. coli*, it stimulates the secretory process up to 10-fold under many conditions, particularly at lower temperatures [23]. This complex has never been identified in a eukaryotic cell [19]. The mechanistic role of this auxiliary complex is poorly understood, but it may facilitate the ATP-driven cycle of SecA membrane insertion and de-insertion at various stages in the translocation process [20,21]. Alternatively, or in addition, it may facilitate interaction of the essential Sec protein constituents with each other [22,24].

In *E. coli*, SecY is a 10 TMS protein, about 450 amino acyl residues long, that together with SecE (3 TMSs) and SecG (2 TMSs), is believed to form the protein-translocating channel [4,25,26]. The two small proteins, SecE and SecG, are each variable in length and topology, depending on the organismal source, but they are maximally of about 140 amino acyl residues in length, and contain one to three and one or two TMSs, respectively [3,7]. Translocation can be driven by ATP hydrolysis catalyzed by SecA, although the pmf is stimulatory [27]. Both energy sources may be required for efficient translocation with each acting at different steps [3,27]. ATP appears to be essential under all or most conditions. Both SecY and SecA are known to directly contact the substrate protein [28].

In eukaryotes, the heterotrimeric Sec61αβγ protein complex in the endoplasmic reticulum (ER) serves as the

channel for matrix protein transport and integral membrane protein insertion by either a cotranslational or a posttranslational mechanism [29–31]. In cotranslational export, directionality is determined by binding of the translating ribosome to the Sec61 complex. The channels in the ribosome and membrane are aligned so the lumenal end of the channel is the only exit site available to the elongating polypeptide chain [32]. By contrast, in posttranslational transport, the Sec61 complex associates with the tetrameric Sec62/63 complex, the resultant Sec complex binds the signal sequence of the translocation substrate, and translocation is energized by BiP (Kar2), a soluble, lumenal Hsp70 ATPase that hydrolyzes ATP to energize polypeptide translocation [33,34]. Translocation requires that BiP interacts with the Sec61 complex via a luminal domain of Sec63, the J domain. Thus, BiP may "pull" the protein through the channel and/or act as a "molecular ratchet", preventing backward movement. Although both mechanisms may be operative, the ratchet mechanism has been best documented under certain experimental conditions [34,35].

Considerable evidence suggests that the ER translocon can function as a "retrotranslocon" to transport improperly folded proteins from the lumen of the ER back into the cytoplasm where degradation occurs in proteasomes [36–39]. Thus, ER lumen proteins that are stalled at some point in their folding/assembly, and possibly integral membrane proteins that do not properly fold, may be recognized by specific chaperone proteins and targeted for retrotranslocation. Indeed, cytoplasmic entry of cholera toxin has been shown to depend on the lumenal disulfide reductase which presumably acts as an unfoldase [40]. The process of retrotranslocation, which requires energy, is likely to prove to be universal. Although the equivalent of retrotranslocation is known to occur in bacteria [41,42], the apparatus and mechanism involved have not yet been identified.

Proper folding and insertion of integral inner membrane proteins in bacteria is dependent on a complex of proteins resembling the eukaryotic signal receptor protein (SRP)–RNA complex [3]. These proteins, Ffh (an SRP54-like protein) and FtsY (an SRP receptor, subunit α-like protein), probably act as GTP hydrolysis-dependent chaperones, feeding the substrate protein into the SecYEG channel [1]. SecA is not required for insertion of polytopic integral membrane proteins [43–47], but insertion shows an absolute dependency on Ffh and FtsY as well as the SecYEG channel complex [3,45,48,49]. In some cases, the Oxa1 homologue in *E. coli*, YidC, may replace or function in conjunction with this protein complex [50–52]. YidC, an integral membrane *E. coli* protein that is abundant relative to the other Sec complex components, is itself inserted in a process requiring the Sec translocation complex [45,47]. It interacts with the SecYEG complex [24] as well as the SecDF–YajC complex [22]. It has been proposed that YidC plays a dual role both in an early step as a receptor and in a late step by facilitating partitioning of the substrate protein from the SecYEG complex to the lipid bilayer [24].

Although extensive biochemical and molecular genetic experiments have been conducted on Sec protein complex constituents, comprehensive phylogenetic analyses have not been reported. In this article, we correct this deficiency, identifying available sequenced homologues of the known essential constituents of the Sec protein export/insertion apparatus as of January 2002. These constituents include SecY/Sec61α, SecE/Sec61γ, SecG/Sec61β, Ffh/SRP54 and FtsY/SRP receptor-α found in prokaryotes/eukaryotes. We multiply align the sequences of all full-length homologues, characterize their distribution in the living world, identify common regions of conservation, hydropathy and amphipathicity, and construct phylogenetic trees in preparation for detailed evolutionary analyses. These studies reveal that the Sec apparatus is truly universal, and that all of the constituents analyzed were probably transmitted to present-day organisms by vertical transmission without loss, and with almost no gene duplication (for the few known prokaryotic exceptions, see Ref. [53]). The results provide a clear picture of the process of evolutionary divergence whereby the protein constituents of the Sec system became distributed among current-day organisms. Although only representative data are presented in this report, a complete set of tables, listing homologues and their properties together with the multiple alignments and phylogenetic trees, can be found on our ALIGN web site (www-biology.ucsd.edu/~msaier/transport/). Average hydropathy, amphipathicity and similarity plots, based on these multiple alignments, are also presented. Computer methods used were as described previously [54].

2. The SecY/Sec61α family

SecY–Sec61α homologues included in this study are listed in Table 1, and the phylogenetic tree for these proteins is shown in Fig. 1. The SecY–Sec61α phylogenetic tree generally reveals 11 clusters according to organismal phylogeny as follows: (1) archaea; (2) eukaryotes; (3) eukaryotes (divergent plant and yeast proteins); (4) mycoplasmas; (5 and 6) low G+C and high G+C Gram-positive bacteria, respectively; (7–11) five clusters of Gram-negative bacteria including (7) chlamydia, (8) primitive bacteria, (9) spirochetes (10) proteobacteria, and (11) cyanobacteria together with eukaryotic chloroplasts. Conclusions resulting from the phylogenetic analysis of SecY homologues are summarized in Table 2. These conclusions are based exclusively on the tree shown in Fig. 1 and are discussed below.

Every living organism examined so far, including all organisms with a completely sequenced genome, has a SecY homologue. This fact shows that by this criterion, the Sec system is truly ubiquitous. Surprisingly, only a few Gram-positive bacteria and no known Gram-negative bacterium have/has more than one SecY homologue. In certain Streptococci and Staphylococci, the "extra" SecY paralogue seems to be specialized for the export of very

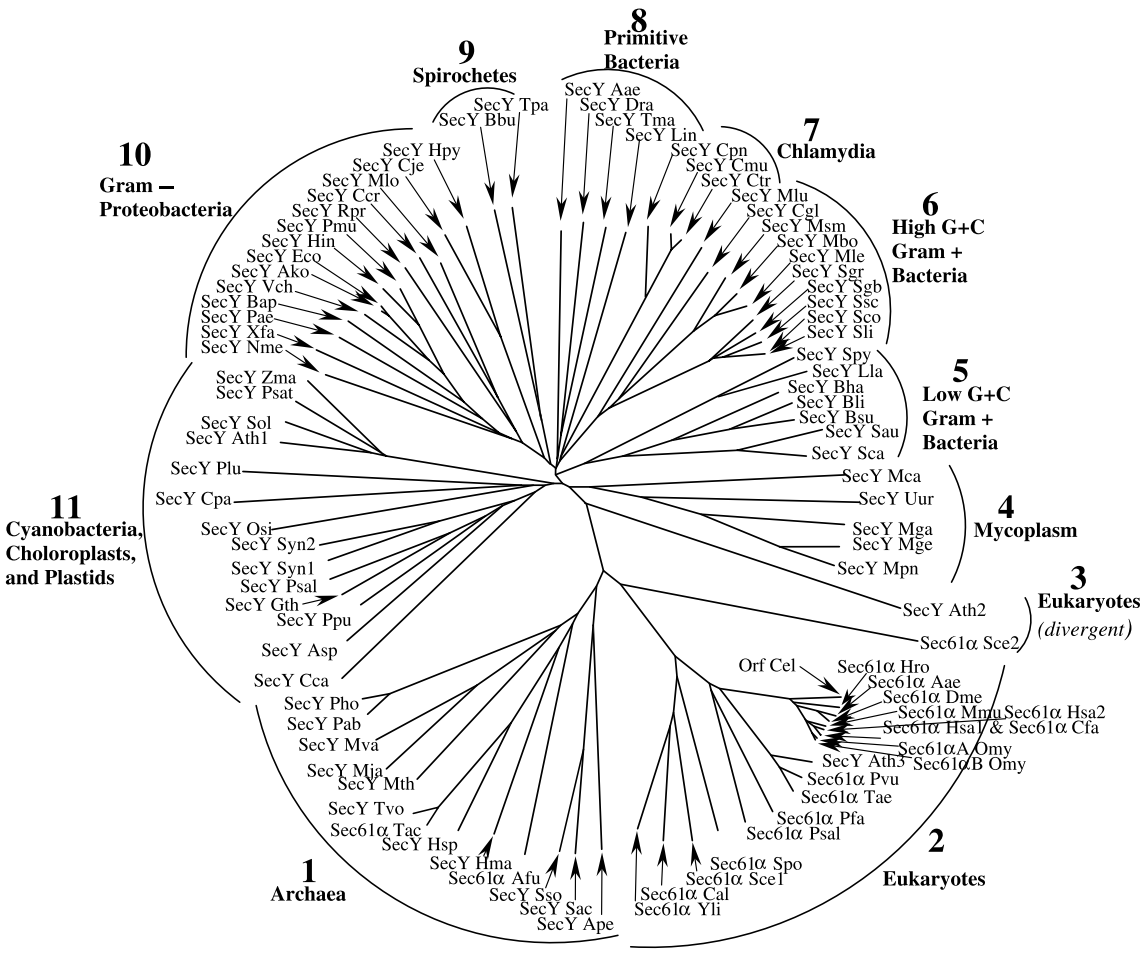

Fig. 1. Phylogenetic tree for sequenced homologues of SecY/Sec61α as of January 2002. Protein and organismal abbreviations are as recorded in Table 1. The Clustal X program was used to generate the multiple alignment upon which the tree was based. This alignment as well as the tabulated family members, alignments and trees for all secretory pathway constituents discussed in this review can be viewed on our ALIGN web site (see http://www-biology.ucsd.edu/~msaier/transport/) and downloaded for analytical purposes.

large serine-rich repeat proteins involved in virulence [53]. Why gene duplication giving rise to multiple SecY paralogues in prokaryotes has occurred so rarely is at present a mystery.

Some eukaryotes such as *C. elegans* and *D. melanogaster* have only one SecY homologue, and a second homologue has not been reported for any animal species. The yeast, *S. cerevisiae*, has two very distant paralogues (Sec61p and Ssh1p), and the plant, *A. thaliana*, has a chloroplast (cyanobacterial-like) homologue as well as two sequence divergent paralogues. Strikingly, all of the archaeal proteins (cluster 1) and all of the orthologous eukaryotic proteins (cluster 2) represent only a small portion of the phylogenetic tree, suggesting that at the molecular level, the many bacterial kingdoms comprise a large majority of the biological diversity found on Earth. Among the constituents of the major eukaryotic cluster, separate subclusters for the homologues from (a) plants, (b) animals, (c) yeast plus fungi and (d) protozoans are evident. Outside of this primary eukaryotic cluster, however, are the two sequence divergent yeast and plant paralogues (cluster 3 in Fig. 1) as well as the chloroplast paralogues of plants (cluster 11 in Fig. 1). The sequence divergent yeast paralogue, Ssh1p, when genetically deleted, has been shown to give rise to defects in both SRP-dependent and SRP-independent protein translocation as well as in retrotranslocation of misfolded ER proteins [55]. Ssh1p recognizes a subset of signal sequences and may be a component of a second yeast ER membrane protein translocon [55,56].

Table 2
Phylogeny of SecY/Sec61α homologues: major conclusions

1. Protein phylogeny follows that of the organisms (16S rRNAs).
2. All organisms with fully sequenced genomes have at least one homologue.
3. Very few bacteria and no archaea have more than one homologue.
4. Some eukaryotes have just one sequenced homologue.
5. *S. cerevisiae* has two very distant paralogues.
6. *A. thaliana* has two very distant paralogues plus a chloroplast (cyanobacterial-like) homologue.
7. Most of the phylogenetic divergence occurs within the bacterial domain, with the archaeal and eukaryotic domains representing two relatively tight clusters.

3. Uniform topology of SecY homologues

The *E. coli* SecY protein has been shown to possess 10 TMSs with its N- and C-termini in the cytoplasm [57]. Using the AveHAS program [58], average hydropathy, amphipathicity and similarity plots were generated for the complete SecY family (Fig. 2). Following a poorly conserved hydrophilic region present in several plant homologues, 10 peaks of hydrophobicity were observed. These are labeled 1 through 10 in Fig. 2A. Corresponding to each hydrophobic peak is a peak of similarity, showing that the TMSs are better conserved than the loop regions. None of the 10 TMSs is strongly amphipathic, but large peaks of amphipathicity are found in loop 2–3 (between TMSs 2 and 3) and in loop 8–9. Smaller peaks of amphipathicity occur before TMS1, after TMSs 5 and 7, and between TMSs 9 and 10 (Fig. 2). Regarding the conservation of inter-TMS loops, loops 2–3, 4–5, 6–7 and 8–9, which are localized to the cytoplasmic side of the membrane, are all well conserved, whereas loops 1–2, 3–4, 5–6, 7–8 and the short 9–10 loop, which are localized to the extracytoplasmic side, are less well conserved. It therefore appears that the cytoplasmic loops (with the possible exception of loop 6–7 which is less well conserved) have strongly conserved functions that are universal for all or most members of the SecY family. This function might deal with protein–protein recognition [16,17]. Based on this same criterion, the extracytoplasmic loops do not have a universally conserved function.

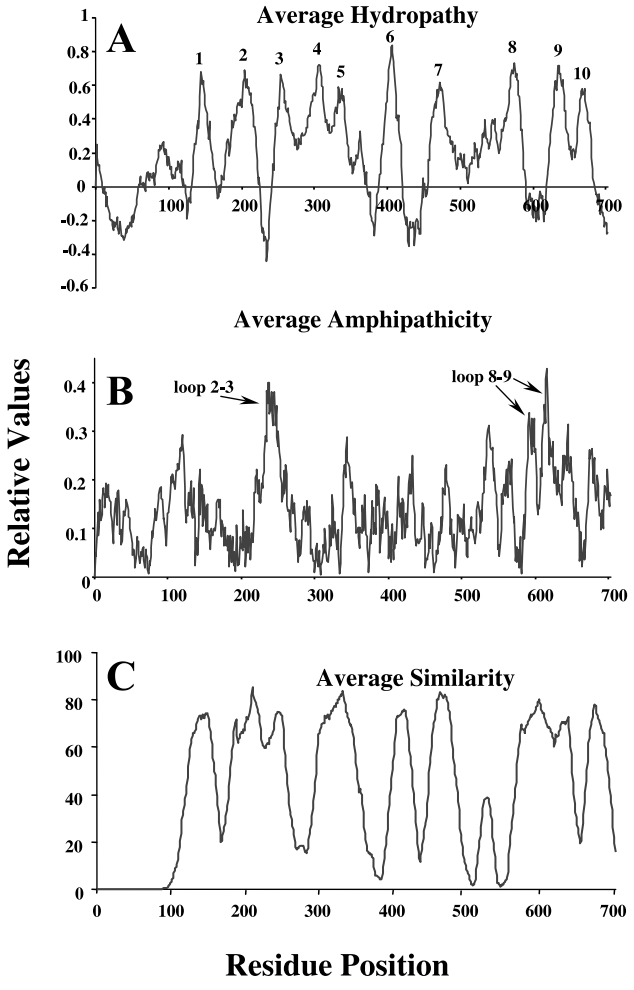

Fig. 2. Average hydropathy, amphipathicity and similarity plots derived from the multiple alignment of all SecY/Sec61α homologues listed in Table 1. The multiple alignment was generated with the Clustal X program, and the plots shown were obtained with the AveHAS program [58]. The length of the plot (700 residue positions) in part reflects the occurrence of N-terminal extensions of some of the plant homologues and inter-TMS insertions in many of the SecY/Sec61α homologues relative to others. Most of these proteins are 400–500 residues in length (see our ALIGN web site).

The AveHAS program was used to generate corresponding plots for the 11 phylogenetic clusters of proteins depicted in Fig. 1 (data not shown). All 11 plots could be interpreted in terms of a conserved 10 TMS topology; the features observed in Fig. 2 were generally observed for all 11 subfamilies. Thus, in all cases, the cytoplasmic loops were better conserved than the extracytoplasmic loops, with loop 6–7 sometimes exhibiting poorer conservation than the other cytoplasmic loops. Moreover, loops 2–3 and 8–9 consistently exhibited strong amphipathic character. The latter loop exhibited two closely spaced amphipathic peaks as shown in Fig. 2 for all clusters except cluster 3 proteins which exhibited a single such peak. These plots also revealed that the long, poorly conserved, N-terminal, hydrophilic extensions (Fig. 2) are present only in clusters 3 and 11, both of eukaryotic origin. Because only some of the proteins in these two clusters have these extensions (see Table 1 and our ALIGN web site), it cannot be concluded that this is a general characteristic of either of these two subfamilies. The occurrence of extra domains not found in prokaryotic homologues is a characteristic of eukaryotic proteins [59].

The multiple alignment of all SecY/Sec61α proteins revealed a high degree of sequence conservation although no single residue was fully conserved. The most conserved region occurs in cytoplasmic loop 4–5 and the neighboring TMS5. The consensus sequence for this region is:

G (L I V) G N G (L I V) S $(L I V)_3$ F (L I V A)

G $(L I V)_2$ $(S A T)_2$ (L I V) P.

(Alternative residues at a single position are indicated in parentheses.)

4. The SecE/Sec61γ family

All SecE/Sec61γ homologues identified are tabulated on our web site, and the corresponding phylogenetic tree is presented therein. These proteins are small (57–177 amino acyl residues [aas]), preventing high phylogenetic resolution. Consequently, the tree was expected to exhibit much

greater experimental error than observed for the SecY/Sec61α tree. In spite of this prediction, the tree was found to exhibit clustering patterns that are very similar to those observed for the SecY/Sec61α tree. Thus, all clusters (1–11) found in Fig. 1 are present in the SecE/Sec61γ tree with the sole exception of cluster 3 (the sequence divergent eukaryotic proteins) (see our ALIGN web site). Moreover, with the exception of *A. thaliana*, no prokaryotic or eukaryotic organism exhibits more than one SecE homologue. The redundancy found for SecY paralogues in *S. cerevisiae* was not observed for SecE. *Arabidopsis* exhibits three SecE paralogues, two close paralogues in the eukaryotic cluster (cluster 2) and one distant paralogue, probably a chloroplast protein. The three *A. thaliana* SecE paralogues thus correlate numerically with the three SecY paralogues found in this organism. However, the two SecE paralogues in cluster 2 are very similar in sequence, and thus probably arose by a recent gene duplication event, whereas all three *A. thaliana* SecY paralogues are very dissimilar, having either arisen from much earlier gene duplication events, or from horizontal gene transfer events.

The output of the AveHAS program [58] applied to the multiple alignment of all SecE and Sec61γ homologues revealed a single strong peak of hydrophobicity in the extreme C-terminal portion of the alignment (alignment positions 130–160), and this peak was preceded by an equally striking peak of amphipathicity (alignment positions 100–130). The poorly conserved N-terminal portions of the alignment (not represented in most homologues) exhibited typical hydrophilic properties (data not shown; see our web site). Thus, the three-TMS topology for the *E. coli* SecE [7,8] is not a general characteristic of its homologues. TMSs 1 and 2 are not essential in the *E. coli* SecE [7].

Examination of the SecE multiple alignment revealed that no residue was fully conserved although fairly good conservation was observed for the C-terminal regions corresponding to the highly tilted TMS 3 in the *E. coli* protein [8]. A consensus sequence encompassing the amphipathic helical region and the adjacent TMS is:

(L I V) $(R K)_2$ (L I V A) X W P (S A T) (R K)

X E X_6 $(L I V)_4$ (G F) $(L I V)_4$ (G A S) $(L I V)_2$

(X = any residue).

5. The SecG/Sec61β family

As for the SecY/Sec61α and the SecE/Sec61γ families, recognizable homologues of *E. coli* SecG and yeast Sec61β were retrieved from the databases. All organisms with a fully sequenced genome possess such a homologue, and no prokaryotic organism or animal species was found to exhibit more than one such homologue [5]. However, *S. cerevisiae* has two Sec61β paralogues, and *A. thaliana* has three, correlating with the numbers of SecY paralogues in these two organisms.

The phylogenetic tree for the SecG/Sec61β family revealed clustering in accordance with expectation based on the SecY/Sec61α and SecE/Sec61γ trees (see Fig. 1 and our web site). It is interesting to note that in contrast to the SecY/Sec61α tree, the two *S. cerevisiae* proteins are close paralogues, and the three *A. thaliana* proteins are even more closely related than are the two yeast proteins. Clustering of the eight bacterial subfamilies (clusters 4–11) is the same as for the SecY and SecE homologues within experimental error.

The output of the AveHAS program [58] applied to the multiple alignment of the SecG/Sec61β proteins revealed two well-conserved peaks of hydropathy centered at alignment positions 69 and 122 with peak 2 being less hydrophobic than peak 1. Examination of the multiple alignment revealed that hydrophobic peak 2 is present in all homologues, but hydrophobic peak 1 is present only in bacteria. From size and topological standpoints, the archaeal homologues more closely resemble the eukaryotic proteins than the bacterial proteins [5]. In most homologues, however, the last TMS is preceded by a glycine-rich region of about 20 residues, followed by a weakly amphipathic (α-helical) region of about 15 residues, just preceding the conserved TMS. A consensus sequence shared by both prokaryotic and eukaryotic homologues is: (L I V) (S T D N) (R P) (L I V) (T P) $(L I V A)_4$ (S T) $(L I V)_3$. The occurrence of this common consensus sequence suggests that all of these proteins share a common ancestry. However, due to extensive sequence divergence among members of this family, the sequence similarity between the bacterial, archaeal and eukaryotic sequences is insufficient to establish homology based on rigorous, established, statistical criteria [60].

6. The Ffh/SRP54 family

Every organism with a fully sequenced genome encodes an Ffh/SRP54 family member. With the exception of a single pair of paralogues in *A. thaliana* (one related to other eukaryotic homologues and one related to the cyanobacterial/chloroplast homologues), no organism was found to display more than one sequenced member of the Ffh/SRP54 family (see our web site). The phylogenetic tree for this family (Fig. 3) shows clustering essentially as observed for the SecY/Sec61α tree except that the sequence divergent eukaryotic branch (cluster 3 in Fig. 1) is not represented. Thus, there is just one archaeal cluster (cluster 1) and one eukaryotic cluster (cluster 2). In the latter cluster, the proteins segregate according to organismal source (animals, plants, fungi and protozoa) as observed for the SecY/Sec61α tree (Fig. 1). All of the bacterial kingdoms represented in Fig. 1 are also represented in Fig. 3.

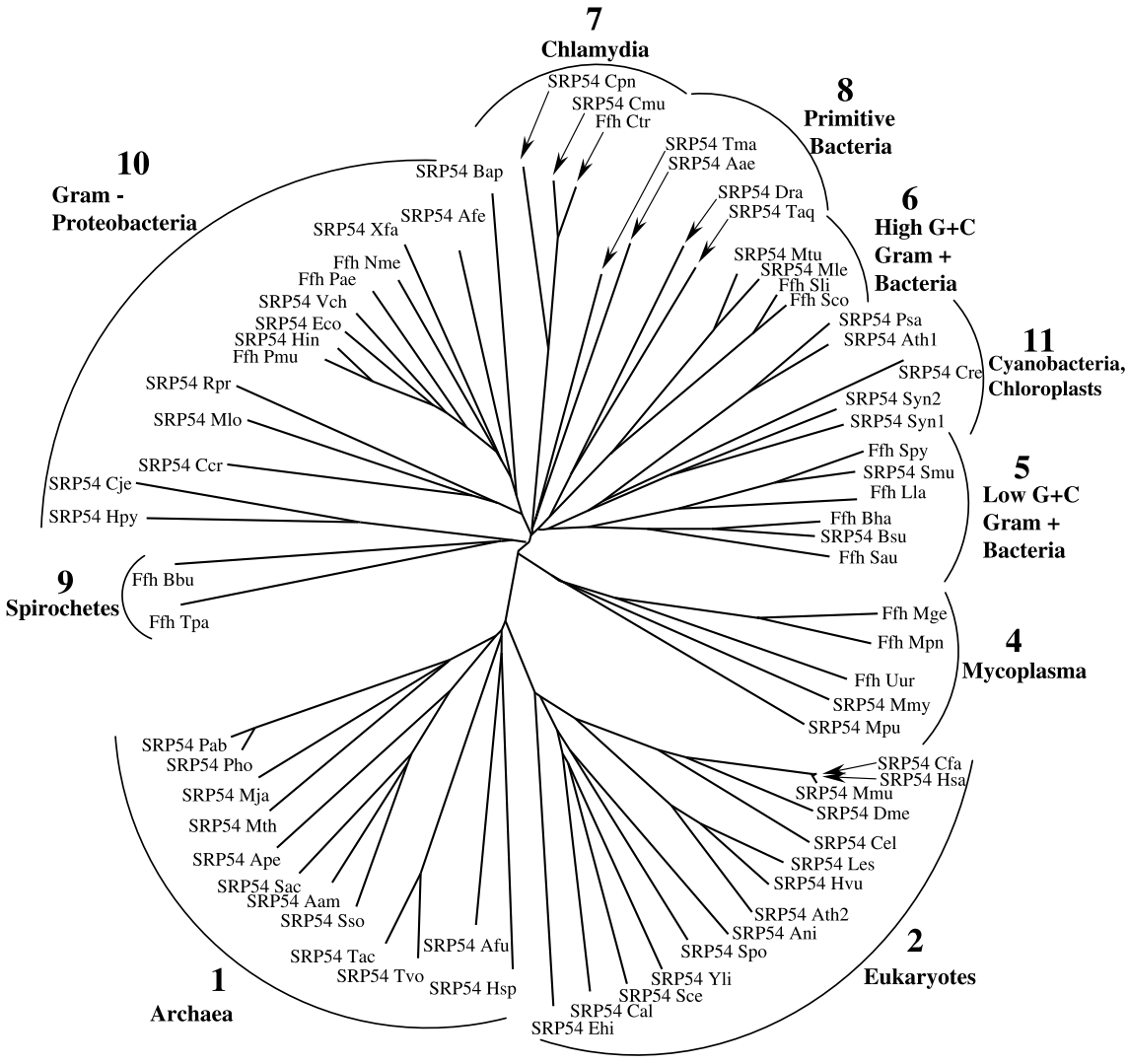

Fig. 3. Phylogenetic tree for the Ffh/SRP54 family. Proteins and organismal abbreviations are as presented in our ALIGN web site (see http://www-biology.ucsd.edu/~msaier/transport/). They generally correspond to those presented in Table 1 for the SecY/Sec61α family. Format of presentation and the computer programs used are as described in Fig. 1 and in Ref. [54].

Application of the AveHAS program to the multiple alignment of the Ffh/SRP54 family provided no evidence for a transmembrane segment. Several regions proved highly amphipathic when the angle was set at 100° as is appropriate for an α-helix.

The multiple alignment revealed 15 residues that were fully conserved. Most of these occurred in several distinct well-conserved motifs which (from N termini to C termini) were as follows:

1. Hy G Hy Q G (S A G T) G K T_2 (T S)
   X (S A G T) K Hy A (alignment positions 189 – 208)

2. D $(Hy)_4$ D T A G R (alignment positions 280 – 289)

3. P X E Hy_5 D (G A S T) X_2 G Q
   (alignment positions 311 – 324)

4. Hy_3 (S T) K Hy D G X_2 (R K) G G (A G)
   A L S X_6 P I X Hy_2 G X G E K Hy X P F X4 Hy_2
   S R Hy_2 G Hy G D (alignment positions 343 – 398).

(Hy, any hydrophobic residue; X = any residue; alternative residues at any one position are present in parentheses; an asterisk indicates a fully conserved residue.)

Motifs 1 and 2 are the Walker A and B motifs which comprise the GTP-binding sites in this family of proteins.

7. The FtsY/SRP receptor-α family

All currently sequenced protein members of the FtsY/SRP receptor-α family were tabulated as for the other

families, and as for the Ffh/SRP54 family, only a single member was identified in any one organism except for *A. thaliana* where two members were found. The phylogenetic tree revealed a single archaeal cluster and a single eukaryotic cluster (except for the second *Arabidopsis* protein that clustered loosely with the bacterial proteins). Clustering was usually, but not always, as expected based on the SecY/Sec61α tree. Differences included the *C. glutamicum* FtsY protein that clustered very loosely with the eukaryotic proteins, and the ε-proteobacterial proteins which did not cluster with the other proteobacterial proteins (as was observed for the Ffh/SRP54 family; see Fig. 3). Finally, FtsY from *A. thaliana* did not cluster with the cyanobacterial homologue.

Application of the AveHAS program to the multiple alignment revealed an N-terminal hydrophobic peak followed by a sharp amphipathic peak (alignment positions 1–100). This region is shared only by the eukaryotic proteins. A poorly conserved, strongly hydrophilic domain lacking amphipathicity when the angle was set at 100° was then observed (alignment positions 100–300) followed by a second much better conserved hydrophilic domain (alignment positions 300–430) that exhibits amphipathic character. Finally, the C-terminal region (alignment positions 430–650) displayed three or four hydrophobic peaks although the SRP receptor α-subunit exhibits biochemical characteristics of a soluble protein [3]. All of the 12 fully conserved residues were found within this domain, and all but one of these occurred in two well-conserved motifs as follows:

1. P Hy_6 G Hy N G V G K T_3 Hy (A G)
   K Hy A X_9 Hy_5 (A G) D T F R A (A G)

   A Hy E Q Hy (alignment positions 434 – 482)

2. D Hy_4 D T (A S) G R Hy X_5 L Hy X E L X K

   (alignment positions 543 – 572).

These two motifs correspond to the Walker A and B motifs (the GTP-binding motifs) and are similar to motifs 1 and 2 in the Ffh/SRP54 family (see Section 6).

8. Conclusions and perspectives

In this article, we report the identification of database sequences homologous to the known constituents of the protein secretion (Sec) systems found in both prokaryotes and eukaryotes. These sequences were multiply aligned, and their relative degrees of sequence conservation were evaluated. Because some of these proteins are long (e.g., SecY homologues), whereas others are short (e.g., SecE and SecG homologues), and because the degrees of sequence similarity are also very different, the levels of reliability of the different multiple alignments and phylogenetic trees differ. The SecY and Ffh family trees are probably the most reliable, whereas the SecE and SecG family trees are the least reliable. The analyses reported have led to major conclusions and postulates as summarized below.

1. All organisms with completely sequenced genomes exhibit homologues of all of the protein constituents of the Sec system that we have analyzed (SecY, SecG, SecE, Ffh and FtsY). All five proteins are therefore likely to be essential constituents of the secretory complex in all living organisms.
2. All but a few prokaryotic organisms [53], and some eukaryotes (i.e., *C. elegans* and *D. melanogaster*) exhibit only one homologue of each of these proteins although species of *Streptococcus* and *Staphylococcus* [53] as well as *S. cerevisiae* exhibit two paralogues of some of these proteins, whereas *A. thaliana* exhibits at least two, and sometimes three paralogues of these proteins.
3. The phylogenies of the Sec proteins qualitatively follow those of the 16S rRNAs from the same organisms with only a few exceptions. This observation strongly argues that all (or almost all) homologues are orthologues serving the same function in all organisms. The few exceptions are particularly interesting and worthy of note, but some of them may be artifactual (i.e., due to errors in the trees) rather than a consequence of lateral gene transfer or some other unexpected phenomenon that occurred during the evolutionary divergence of these proteins.
4. The phylogenetic trees shown in Figs. 1 and 3 reveal greater diversity within the bacterial domain than within either the archaeal or the eukaryotic domain. Because all but a few of the proteins included in these trees are orthologues, this surprising observation leads us to suggest that the protein phylogenies reflect those of the organisms, and that the bacterial domain is more diverse at the molecular level than either the archaeal or the eukaryotic domain. This interpretation would be consistent with the suggestion that bacteria were the evolutionary precursors of the archaea and eukaryotes [61]. Alternatively, one might argue that the extremely diverse environmental conditions encountered by bacteria, as compared with archaea and eukaryotes, account for the sequence diversity observed. In view of the diversity of archaeal environments, however, we find this second interpretation less likely.
5. The universality of the Sec protein constituents analyzed clearly argues that this pathway is the principal one operative in all living cells. It may be the only one operative in some membranous systems such as the endoplasmic reticulum of eukaryotic cells. Because of its ubiquity, we would also argue that it may be the most ancient of currently recognized protein secretion/membrane insertion systems. In an independent report [52],

we have found that the Oxa1/YidC family of putative protein exporters is also universal. Some evidence suggests that these proteins function in integral membrane insertion independently of the SecYEG translocon [62,63], whereas other evidence suggests that they function in conjunction with SecYEG [3,22,24,51,64]. Because one possibility does not exclude the other, it seems reasonable to suggest that Oxa1p homologues can function both independently of and cooperatively with the Sec protein translocase, depending on circumstances [50,52]. The ubiquity of Oxa1p homologues, and the occurrence of multiple Oxa1p paralogues in eukaryotes, leads to the suggestion that they may play essential roles as constituents of the Sec protein insertion apparatus [24]. This and many other interesting suggestions resulting from the phylogenetic analyses reported here should serve as guides for future research.

Acknowledgements

We thank Nick Grishin for collaboratively identifying the archaeal SecG homologues [5], and Mary Beth Hiller for assistance in the preparation of this manuscript. This work was supported by NIH grants GM55434-17 and GM 64368-24 from the National Institute of General Medical Sciences.

References

[1] P. Fekkes, A.J.M. Driessen, Protein targeting to the bacterial cytoplasmic membrane, Microbiol. Mol. Biol. Rev. 63 (1999) 161–173.
[2] E.H. Manting, A.J. Driessen, *Escherichia coli* translocase: the unravelling of a molecular machine, Mol. Microbiol. 37 (2000) 226–238.
[3] M. Müller, H.G. Koch, K. Beck, U. Schafer, Protein traffic in bacteria: multiple routes from the ribosome to and across the membrane, Prog. Nucleic Acid Res. Mol. Biol. 66 (2001) 107–157.
[4] C. Breyton, W. Haase, T.A. Rapoport, W. Kuhlbrandt, I. Collinson, Three-dimensional structure of the bacterial protein-translocation complex SecYEG, Nature 418 (2002) 662–665.
[5] L.N. Kinch, M.H. Saier Jr., N.V. Grishin, Sec61β — a component of the archaeal protein secretory system, Trends Biochem. Sci. 27 (2002) 170–171.
[6] E. van der Sluis, N. Nouwen, A. Driessen, SecY–SecY and SecY–SecG contacts revealed by site-specific crosslinking, FEBS Lett. 527 (2002) 159.
[7] K. Nishiyama, H. Suzuki, H. Tokuda, Role of the non-essential region encompassing the N-terminal two transmembrane stretches of *Escherichia coli* SecE, Biosci. Biotechnol. Biochem. 64 (2000) 2121–2127.
[8] A.K. Veenendaal, C. Van der Does, A.J. Driessen, The core of the bacterial translocase harbors a tilted transmembrane segment 3 of SecE, J. Biol. Chem. 277 (2002) 36640–36645.
[9] P. Bessonneau, V. Besson, I. Collinson, F. Duong, The SecYEG preprotein translocation channel is a conformationally dynamic and dimeric structure, EMBO J. 21 (2002) 995–1003.
[10] T.L. Yahr, W.T. Wickner, Evaluating the oligomeric state of SecYEG in preprotein translocase, EMBO J. 19 (2000) 4393–4401.
[11] M.H. Saier Jr., P. Werner, M. Müller, Insertion of proteins into bacterial membranes: mechanism, characteristics and comparisons with the eukaryotic process, Microbiol. Rev. 53 (1989) 333–366.
[12] K.H.M. van Wely, J. Swaving, R. Freudl, A.J.M. Driessen, Translocation of proteins across the cell envelope of Gram-positive bacteria, FEMS Microbiol. Rev. 25 (2001) 437–454.
[13] E. Or, A. Navon, T. Rapoport, Dissociation of the dimeric SecA ATPase during protein translocation across the bacterial membrane, EMBO J. 21 (2002) 4470–4479.
[14] R.L. Woodbury, S.J. Hardy, L.L. Randall, Complex behavior in solution of homodimeric SecA, Protein Sci. 11 (2002) 875–882.
[15] E.H. Manting, C. Van Der Does, H. Remigy, A. Engel, A.J. Driessen, SecYEG assembles into a tetramer to form the active protein translocation channel, EMBO J. 19 (2000) 852–861.
[16] K. Chiba, H. Mori, K. Ito, Roles of the C-terminal end of SecY in protein translocation and viability of *Eschericheria coli*, J. Bacteriol. 184 (2002) 2243–2250.
[17] H. Mori, Y. Shimizu, K. Ito, Super active *secY* variants that fulfill the essential translocation function with a reduced cellular quantity, J. Biol. Chem. (in press).
[18] A. Bolhuis, C.P. Broekhuizen, A. Sorokin, M.L. van Roosmalen, G. Venema, S. Bron, W.J. Quax, J.M. van Dijl, SecDF of *Bacillus subtilis*, a molecular Siamese twin required for the efficient secretion of proteins, J. Biol. Chem. 273 (1998) 21217–21224.
[19] T.-T. Tseng, K.S. Gratwick, J. Kollman, D. Park, D.H. Nies, A. Goffeau, M.H. Saier Jr., The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins, J. Mol. Microbiol. Biotechnol. 1 (1999) 107–125.
[20] F. Duong, W. Wickner, Distinct catalytic roles of the SecYE, SecG and SecDFYajC subunits of preprotein translocase holoenzyme, EMBO J. 16 (1997) 2756–2768.
[21] F. Duong, W. Wickner, The SecDFYajC domain of preprotein translocase controls preprotein movement by regulating SecA membrane cycling, EMBO J. 16 (1997) 4871–4879.
[22] N. Nouwen, A.J. Driessen, SecDFyajC forms a heterotetrameric complex with YidC, Mol. Microbiol. 44 (2002) 1397–1405.
[23] J.A. Pogliano, J. Beckwith, SecD and SecF facilitate protein export in *Escherichia coli*, EMBO J. 13 (1994) 554–561.
[24] E.N. Houben, M.L. Urbanus, M. Van Der Laan, C.M. Ten Hagen-Jongman, A.J. Driessen, J. Brunner, B. Oudega, J. Luirink, YidC and SecY mediate membrane insertion of a Type I transmembrane domain, J. Biol. Chem. 277 (2002) 35880–35886.
[25] K. Ito, SecY and integral membrane components of the *Escherichia coli* protein translocation system, Mol. Microbiol. 6 (1992) 2423–2428.
[26] T.H. Meyer, J.F. Ménétret, R. Breitling, K.R. Miller, C.W. Akey, T.A. Rapoport, The bacterial SecY/E translocation complex forms channel-like structures similar to those of the eukaryotic Sec61p complex, J. Mol. Biol. 285 (1999) 1789–1800.
[27] B.L. Geller, Energy requirements for protein translocation across the *Escherichia coli* inner membrane, Mol. Microbiol. 5 (1991) 2093–2098.
[28] M. Koster, W. Bitter, J. Tommassen, Protein secretion mechanisms in Gram-negative bacteria, Int. J. Med. Microbiol. 290 (2000) 325–331.
[29] S.U. Heinrich, W. Mothes, J. Brunner, T.A. Rapoport, The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain, Cell 102 (2000) 233–244.
[30] C. Stirling, Protein targeting to the endoplasmic reticulum in yeast, Microbiology 145 (1999) 991–998.
[31] B.M. Wilkinson, J.R. Tyson, P.J. Reid, C.J. Stirling, Distinct domains within yeast Sec61p involved in post-translational translocation and protein dislocation, J. Biol. Chem. 275 (2000) 521–529.
[32] T.A. Rapoport, B. Jungnickel, U. Kutay, Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes, Annu. Rev. Biochem. 65 (1996) 271–303.
[33] K.E. Matlack, W. Mothes, T.A. Rapoport, Protein translocation: tunnel vision, Cell 92 (1998) 381–390.
[34] K.E.S. Matlack, B. Misselwitz, K. Plath, T.A. Rapoport, BiP acts as a molecular ratchet during posttranslational transport of prepro-α factor across the ER membrane, Cell 97 (1999) 553–564.

[35] B. Misselwitz, O. Staeck, T.A. Rapoport, J proteins catalytically activate Hsp70 molecules to trap a wide range of peptide sequences, Mol. Cell. 2 (1998) 593–603.
[36] A.E. Johnson, N.G. Haigh, The ER translocon and retrotranslocation: is the shift into reverse manual or automatic? Cell 102 (2000) 709–712.
[37] S.I. Nishikawa, S.W. Fewell, Y. Kato, J.L. Brodsky, T. Endo, Molecular chaperones in the yeast endoplasmic reticulum maintain the solubility of proteins for retrotranslocation and degradation, J. Cell Biol. 153 (2001) 1061–1070.
[38] B. Tsai, Y. Ye, T.A. Rapoport, Retro-translocation of proteins from the endoplasmic reticulum into the cytosol, Nat. Rev., Mol. Cell Biol. 3 (2002) 246–255.
[39] J. Wesche, Retrograde transport of ricin, Int. J. Med. Microbiol. 291 (2002) 517–521.
[40] B. Tsai, C. Rodighiero, W.I. Lencer, T.A. Rapoport, Protein disulfide isomerase acts as a redox-dependent chaperone to unfold cholera toxin, Cell 104 (2001) 937–948.
[41] A. Kihara, Y. Akiyama, K. Ito, Dislocation of membrane proteins in FtsH-mediated proteolysis, EMBO J. 18 (1999) 2970–2981.
[42] A.J. Romo, E.G. Ruby, M.H. Saier Jr., Effect of *Bdellovibrio bacteriovorus* infection on the phosphoenolpyruvate:sugar phosphotransferase system in *Escherichia coli*: evidence for activation of cytoplasmic proteolysis, Res. Microbiol. 143 (1992) 5–14.
[43] K. Beck, L.-F. Wu, J. Brunner, M. Müller, Discrimination between SRP- and SecA/SecB-dependent substrates involves selective recognition of nascent chains by SRP and trigger factor, EMBO J. 19 (2000) 134–143.
[44] H.G. Koch, T. Hengelage, C. Neumann-Haefelin, J. MacFarlane, H.K. Hoffschulte, K.L. Schimz, B. Mechler, M. Müller, In vitro studies with purified components reveal signal recognition particle (SRP) and SecA/SecB as constituents of two independent protein-targeting pathways of *Escherichia coli*, Mol. Biol. Cell 10 (1999) 2163–2173.
[45] H.G. Koch, M. Moser, K.L. Schimz, M. Müller, The integration of YidC into the cytoplasmic membrane of *Escherichia coli* requires the signal recognition particle, SecA and SecYEG, J. Biol. Chem. 277 (2002) 5715–5718.
[46] H.G. Koch, M. Müller, Dissecting the translocase and integrase functions of the *Escherichia coli* SecYEG translocon, J. Cell Biol. 150 (2000) 689–694.
[47] M.L. Urbanus, L. Fröderberg, D. Drew, P. Bjork, J.W. de Gier, J. Brunner, B. Oudega, J. Luirink, Targeting, insertion, and localization of *Escherichia coli* YidC, J. Biol. Chem. 277 (2002) 12718–12723.
[48] J. MacFarlane, M. Müller, The functional integration of a polytopic membrane protein of *Escherichia coli* is dependent on the bacterial signal-recognition particle, Eur. J. Biochem. 233 (1995) 766–771.
[49] N.D. Ulbrandt, J.A. Newitt, H.D. Bernstein, The *E. coli* signal recognition particle is required for the insertion of a subset of inner membrane proteins, Cell 88 (1997) 187–196.
[50] J. Luirink, T. Samuelsson, J.-W. de Gier, YidC/Oxa1p/Alb3: evolutionarily conserved mediators of membrane protein assembly, FEBS Lett. 501 (2001) 1–5.
[51] M.L. Urbanus, P.A. Scotti, L. Fröderberg, A. Sääf, J.W. de Gier, J. Brunner, J.C. Samuelson, R.E. Dalbey, B. Oudega, J. Luirink, Sec-dependent membrane protein insertion: sequential interaction of nascent FtsQ with SecY and YidC, EMBO Rep. 2 (2001) 524–529.
[52] M.-R. Yen, K.T. Harley, Y.-H. Tseng, M.H. Saier Jr., Phylogenetic and structural analyses of the Oxa1 family of protein translocases, FEMS Microbiol. Lett. 204 (2002) 223–231.
[53] A.B. Bensing, P.M. Sullam, An accessory *sec* locus of *Streptococcus gordonii* is required for export of the surface protein GspB and for normal levels of binding to human platelets, Mol. Microbiol. 44 (2002) 1081–1094.
[54] T.B. Cao, M.H. Saier Jr., Conjugal type IV macromolecular transfer systems of Gram-negative bacteria: organismal distribution, structural constraints and evolutionary conclusions, Microbiology 147 (2001) 3201–3214.
[55] B.M. Wilkinson, J.R. Tyson, C.J. Stirling, Ssh1p determines the translocation and dislocation capacities of the yeast endoplasmic reticulum, Dev. Cell 1 (2001) 401–409.
[56] S. Wittke, M. Dunnwald, M. Albertsen, N. Johnsson, Recognition of a subset of signal sequences by Ssh1p, a Sec61p-related protein in the membrane of endoplasmic reticulum of yeast *Saccharomyces cerevisiae*, Mol. Biol. Cell 13 (2002) 2223–2232.
[57] Y. Akiyama, K. Ito, Topology analysis of the SecY protein, an integral membrane protein involved in protein export in *Escherichia coli*, EMBO J. 6 (1987) 3465–3470.
[58] Y. Zhai, M.H. Saier Jr., A web-based program for the prediction of average hydropathy, average amphipathicity and average similarity of multiply aligned homologous proteins, J. Mol. Microbiol. Biotechnol. 3 (2001) 285–286.
[59] Y.J. Chung, C. Krueger, D. Metzgar, M.H. Saier Jr., Size comparisons among integral membrane transport protein homologues in bacteria, archaea, and eucarya, J. Bacteriol. 183 (2001) 1012–1021.
[60] M.H. Saier Jr., Computer-aided analyses of transport protein sequences: gleaning evidence concerning function, structure, biogenesis, and evolution, Microbiol. Rev. 58 (1994) 71–93.
[61] M.H. Saier Jr., Bacterial diversity and the evolution of differentiation, ASM News 66 (2000) 337–343.
[62] K. Hell, W. Neupert, R.A. Stuart, Oxa1p acts as a general membrane insertion machinery for proteins encoded by mitochondrial DNA, EMBO J. 20 (2001) 1281–1288.
[63] J.C. Samuelson, M. Chen, F. Jiang, I. Möller, M. Wiedmann, A. Kuhn, G.J. Phillips, R.E. Dalbey, YidC mediates membrane protein insertion in bacteria, Nature 406 (2000) 637–641.
[64] P.A. Scotti, M.L. Urbanus, J. Brunner, J.-W. de Gier, G. von Heijne, C. van der Does, A.J.M. Driessen, B. Oudega, J. Luirin, YidC, the *Escherichia coli* homologue of mitochondrial Oxa1p, is a component of the Sec translocase, EMBO J. 19 (2000) 542–549.