Microbiology (2002), **148**, 3335–3346 Printed in Great Britain

REVIEW ARTICLE

Protein transport in the halophilic archaeon *Halobacterium* sp. NRC-1: a major role for the twin-arginine translocation pathway?

Albert Bolhuis

Tel: +44 24 76528360. Fax: +44 24 76523701. e-mail: abolhuis@bio.warwick.ac.uk

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

Keywords: archaea, Sec pathway, Tat pathway, signal peptide, lipoprotein

Background

Trafficking of proteins, which is an essential process for all living organisms, has been extensively studied in bacteria and eukarya. Very little, however, is known about protein transport in the third domain of life, the *Archaea*. These organisms are prokaryotes, thus have the same cellular organization as bacteria, but phylogenetic analyses showed that archaea are evolutionarily distinct from both bacteria and eukarya (Woese *et al*., 1990). A unique aspect of archaea is the composition of their cytoplasmic membrane: whereas bacterial and eukaryal membranes are composed of fatty acids linked to glycerol by an ester bond, archaeal phospholipids are composed of branched isoprene units linked to glycerol by an ether group. These lipids play an important role in surviving the extreme environments in which archaea are frequently found, as they are less permeable to ions and protons and more resistant to extreme temperatures or high salt concentrations (van de Vossenberg *et al*., 1998). Because of the unique cytoplasmic membrane and the extreme environment in which many archaea thrive, it is likely that protein translocation machineries of archaea contain specific adaptations or even completely novel components that are important for protein transport.

Systems for the transport of proteins across membranes function by common principles in all domains of life (Schatz & Dobberstein, 1996). Proteins destined for export are usually synthesized as pre-proteins with an amino-terminal signal peptide that is required for targeting to the membrane and initiation of protein translocation (von Heijne, 1990). At the membrane, pre-proteins are transported through a proteinaceous translocation channel in a process that is driven by the binding and hydrolysis of nucleoside triphosphates and/or the proton motive force. During or shortly after translocation, the signal peptide is removed from the pre-protein by signal peptidases (SPases), and the mature protein is released at the *trans* side of the membrane. Systems of protein transport that function according to these basic principles include protein import into mitochondria, chloroplasts and the lumen of the endoplasmic reticulum (ER), and protein export to extracytoplasmic compartments of prokaryotes, such as the periplasm and outer membrane (Gram-negative bacteria), the cell wall (Gram-positive bacteria and archaea) and the growth medium.

Halobacterium species are excellent model organisms among the archaea. These organisms are true extremophiles that thrive in an environment that is nearly saturated with salt (4–5 M NaCl). In contrast to several other archaea, halobacteria are easy to culture as they usually grow aerobically between 30 and 40 °C. Furthermore, they are genetically amenable organisms for which several genetic tools are available, such as transformation, shuttle vectors and gene knockout strategies (Sowers & Schreier, 1999; Peck *et al*., 2000). As with other archaea, little is known about the transport of proteins in these organisms. Nevertheless, the availability of the genomic sequences, including that of *Halobacterium* sp. NRC-1 (Ng *et al*., 2000), provides a wealth of information that can be used to make predictions about various cellular processes. This review will focus primarily on the two major protein transport routes found in archaea, i.e. the Sec and the Tat pathways, and the identification of proteins in *Halobacterium* sp. NRC-1 that may use these pathways to reach their extracytoplasmic destination. The analyses described in this review show that, in contrast to most other organisms, the majority of secretory proteins of *Halobacterium* sp. NRC-1 use the Tat pathway for export. In addition, there is the unexpected finding of many putative lipoproteins in this archaeon, which, as discussed below, indicates the presence of a completely novel mechanism for lipo-modification and processing in archaea.

The Sec pathway

One of the major pathways for protein translocation is the Sec pathway, which is conserved in all domains of

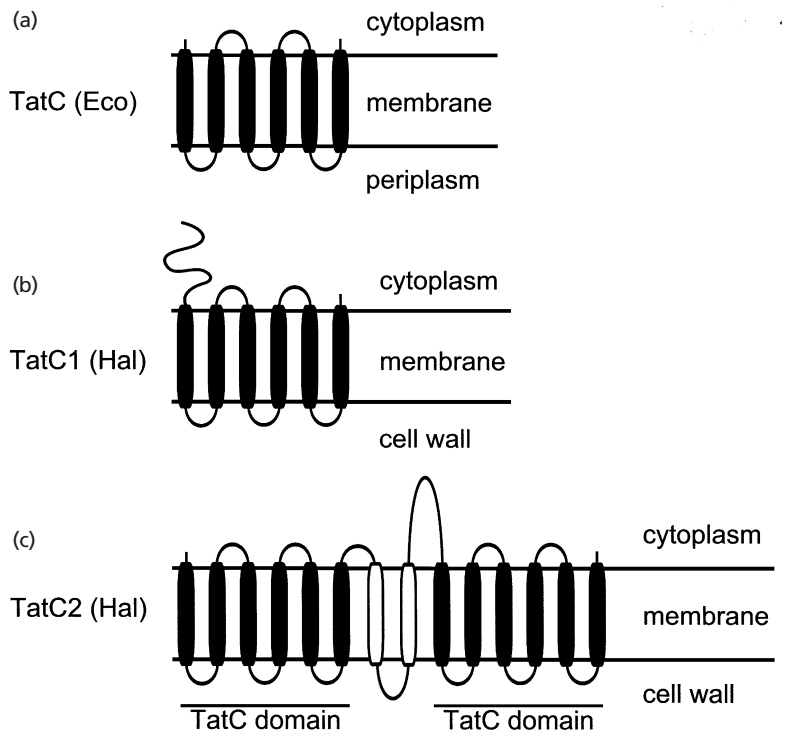

Fig. 1. Predicted membrane topologies of *E. coli* TatC, *Halobacterium* sp. NRC-1 TatC1 and *Halobacterium* sp. NRC-1 TatC2. The topologies were determined using the TMHMM2.0 server at www.cbs.dtu.dk/services. Transmembrane helices of the TatC proteins or domains are filled, and transmembrane helices of the linker domain in TatC2 are open.

life. In eukarya, most proteins destined for export to the lumen of the ER are translocated co-translationally via this pathway (Johnson & Waes, 1999). As soon as the signal peptide emerges from the ribosome, translation is arrested through binding of the signal recognition particle (SRP) to the nascent chain. In mammalian cells, SRP comprises six protein subunits (SRP9, SPR14, SRP19, SRP54, SRP68 and SRP72), which are bound to a 7S RNA scaffold (Keenan *et al.*, 2001). The entire complex of the ribosome, the nascent chain and the SRP is targeted to the ER membrane with the aid of the SRP receptor subunits SRα and SRβ. At the membrane, the SRP is released from the nascent chain in a process that depends on the binding of GTP, and translation is resumed. This occurs at the site of the translocase, and protein synthesis by the ribosome pushes the protein through the translocation channel. The channel is formed by the Sec61 complex, which comprises Sec61α, Sec61β and Sec61γ (Görlich & Rapoport, 1993). In yeast, in particular, several proteins are translocated post-translationally. In that case, transport is driven by a pulling action from Kar2p (called BiP in mammalian cells), which is an Hsp70 that resides in the lumen of the ER (Nguyen *et al.*, 1991). Accessory components that are important for protein transport include the Sec62/63 and Sec71/72 complexes, both of which are involved in post-translational translocation (Johnson & Waes, 1999), and TRAM, which is important for the translocation and integration of most secretory proteins (Voigt *et al.*, 1996). The signal peptidase complex (SPC; Evans *et al.*, 1986) removes the signal peptide at the *trans* side of the membrane.

In bacteria, most extracytoplasmic proteins are targeted post-translationally, with the aid of a targeting factor such as SecB, to the translocase, which comprises SecY, SecE and SecG (for a review, see Manting & Driessen, 2000). SecY and SecE are related to the eukaryotic Sec61α and Sec61γ, respectively. SecG and Sec61β are not related, but may have the same role in protein translocation. A pivotal role in bacterial protein transport is played by the peripheral membrane protein SecA, which is an ATPase that pushes the pre-protein through the translocase channel (Economou, 1998). Protein translocation is assisted by accessory components such as SecD and SecF. During, or shortly after, translocation, the signal peptide is removed by a type I SPase (Dalbey *et al.*, 1997) or, in case of lipoproteins (see below), by a type II SPase (Hayashi & Wu, 1990).

Several membrane proteins that use the Sec machinery for integration into the membrane are targeted by an SRP-like complex that comprises a homologue of SRP54 (Ffh) and a 4·5S RNA molecule (Seluanov & Bibi, 1997; Ulbrandt *et al.*, 1997). Targeting is mediated by a homologue of the eukaryotic SRα called FtsY. For certain proteins SecA is still required for SRP-dependent translocation (Neumann-Haefelin *et al.*, 2000), indicating that their translocation is not driven by the ribosome. The integral membrane protein YidC is also vital for the insertion of membrane proteins in *Escherichia coli* (Dalbey & Kuhn, 2000). This protein, which interacts with the SecYEG complex (Beck *et al.*, 2001; Scotti *et al.*, 2000), is also involved in the insertion of Sec-independent membrane proteins (Samuelson *et al.*, 2000).

Archaea lack the chaperone SecB, but do contain an SRP pathway (Eichler & Moll, 2001). The archaeal components of this pathway are SRP54, SRP19 (which is absent in bacteria), 7S RNA, and a homologue of

SRα/FtsY (denoted Dpa in *Halobacterium* sp. NRC-1). Analysis of the SRP pathway in the halophilic archaeon *Haloferax volcanii* showed that SRP54 is essential for viability and that it interacts with SRP19 and 7S RNA (Rose & Pohlschröder, 2002). All archaea contain homologues of SecY/Sec61α and SecE/Sec61γ but, interestingly, these are more similar to their eukaryotic counterparts (Pohlschröder *et al*., 1997; Eichler, 2000). A SecG homologue is not present in archaea, but a recent study suggested the presence of a homologue of the eukaryotic protein Sec61β (Kinch *et al*., 2002). Strikingly, archaea lack the bacterial translocation motor SecA. Thus archaea may use an alternative unique form of motor protein, or use an entirely different translocation mechanism that is possibly similar to the eukaryotic co-translational translocation mechanism. Since the archaeal translocase and SRP show more similarity to those found in eukaryotes, the latter possibility seems to be the most likely one. It is interesting to note, however, that of the accessory components only homologues of the bacterial SecD and SecF proteins are present. The role of these proteins is not clear, but they have been proposed to be involved in the membrane cycling of SecA (Duong & Wickner, 1997). Thus it is conceivable that SecD and SecF are involved in the cycling of a protein with a function similar to SecA. It is also possible that SecD and SecF serve a completely different role in archaea, or that the effect of SecD and SecF on SecA cycling is only indirect and that they have a different function in bacterial protein translocation as well. Strikingly, other roles have also been proposed for SecD/F in bacteria, such as involvement in the release of the mature secretory protein from the translocase (Matsuyama *et al*., 1993), gating or assembly of the translocase channel (Pohlschröder *et al*., 1997), or removal of signal peptides from the translocase (Bolhuis *et al*., 1998). The latter function is compatible with the observation that SecD and SecF show similarity to secondary solute transporters (Bolhuis *et al*., 1998; Tseng *et al*., 1999). Notably, the previously suggested function of SecD and SecF in maintenance of the proton motive force (Arkowitz & Wickner, 1994) was recently refuted (Nouwen *et al*., 2001). Archaea lack a number of components that are pivotal for protein transport in bacteria or eukarya, including homologues of the bacterial YidC protein and the eukaryotic TRAM and Sec63 proteins. Therefore, it is very likely that novel components will be involved in the archaeal translocation process.

The Tat pathway

Recently, a novel Sec-independent system for protein export was discovered that is present in most prokaryotes and thylakoid membranes of chloroplasts. Proteins using this pathway contain a typical twin-arginine motif in their signal peptide (Chaddock *et al*., 1995; Berks, 1996) and, therefore, it was denoted the twin-arginine translocation (Tat) pathway. In contrast to the Sec pathway, the Tat system has the unique ability to transport fully folded proteins and does not seem to depend on the presence of nucleoside triphosphates (for a review, see Robinson & Bolhuis, 2001). This pathway is involved in the export of proteins that either have to fold before translocation, such as certain co-factor containing proteins, or just fold too quickly. These folded proteins are incompatible with the Sec machinery and can only be exported via the Tat system.

Bacteria and thylakoid membranes of chloroplasts usually contain three major membrane-bound components that are required for Tat-dependent export (for reviews, see Berks *et al*., 2000; Robinson & Bolhuis, 2001). The core of the *E. coli* Tat-translocase is probably formed by TatB and TatC, both of which are present in a strict 1:1 ratio (Bolhuis *et al*., 2001). These proteins function together with a third component, TatA. Similarly, chloroplast Hcf106 (a TatB homologue) and TatC are tightly associated with each other and interact only with Tha4 (a TatA homologue) in the presence of precursor protein (Mori & Cline, 2002). TatA and TatB are similar in structure and contain one transmembrane helix at the N-terminus followed by a cytoplasmic domain. TatC is an integral membrane protein with six transmembrane helices (Fig. 1a).

Most archaea contain components for the Tat machinery, as a BLAST search against available genomic sequences (http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/genom_table_cgi) showed that 9 out of 15 archaea of which the genome has been fully sequenced contain at least one or two TatC-like proteins (also see Yen *et al*., 2002). Several of these also seem to contain one or two TatA and/or TatB-like proteins but, as these are not very similar to each other in general (Robinson & Bolhuis, 2001), the role of these proteins in translocation needs to be determined experimentally. *Halobacterium* sp. NRC-1 contains two TatC proteins, denoted TatC1 and TatC2 (Ng *et al*., 2000). TatC1, which has six predicted membrane-spanning domains, resembles other prokaryotic TatC-like proteins, except for its large N-terminal cytoplasmic domain (approx. 100 residues), which is absent in other TatC proteins (Fig. 1b). The TatC2 protein has a very unusual topology as it comprises two domains, each of which is similar to TatC proteins. These two domains, which share 45% identical and conserved residues with each other, are joined together by a linker region with two membrane-spanning domains (Fig. 1c). Thus the TatC2 protein contains 14 membrane-spanning domains in total. Interestingly, the cytoplasmic domain of the linker region is homologous to the N-terminal cytoplasmic domain of TatC1 (40% identical and conserved residues). The only other organism containing a TatC-like protein with a similar topology is the archaeon *Hf. volcanii* (http://wit-scranton.mbi.scranton.edu/Haloferax/), which is a halophile that is closely related to *Halobacterium* species.

Simple BLAST searches did not immediately show the presence of TatA- or TatB-like proteins in *Halobacterium* sp. NRC-1, but with PSI-BLAST (Altschul *et al*., 1997) a protein was identified showing similarity to

```
Vng0801 H. sp. MFTSTPLFIGGLPGGMEMAVVLLIAILLFGANKIPKLARSSGEAIGEFQK 50
TatA E. coli           MGGIS-IWQLLIIAVIVVLLFGTKKLGSIGSDLGASIKGFKK 41
                       :**:     : :: :* :****: *:  :    * :*  * *

Vng0801 H. sp. GR--EEVEQELQEIKSESAPDASADATADTTSDTTTDATSSADDTATN 96
TatA E. coli   AMSDDEPKQDKTSQDADFTAKTIADKQADTNQEQAKTEDAKRHDKEQV 89
                   :*  *:     :: :  : **  ***  : :    :   *
```

Fig. 2. Comparison of Vng0801 of *Halobacterium* sp. NRC-1 with TatA of *E. coli*. Identical (*) and conservative (:) replacements are marked. The transmembrane helix is indicated in grey shading, and the FG motif at the end of the transmembrane helix, which is diagnostic for proteins belonging to the TatA family, is indicated in bold letters.

TatA of *E. coli* (45 % identical and conservative replacements; Fig. 2). This protein, denoted Vng0801, is predicted to contain an N-terminal membrane-spanning domain that ends on the FG motif that is diagnostic for proteins belonging to the TatA family (Robinson & Bolhuis, 2001). Like other TatA- and TatB-like proteins (Settles *et al*., 1997), the membrane-spanning domain of Vng0801 is followed by a stretch of approximately 30 residues that may form an amphipathic helix. A protein belonging to the TatB family was not identified in *Halobacterium* sp. NRC-1.

Signal peptides

Based upon cleavage by SPases, signal peptides can be classified in four distinct groups (Tjalsma *et al*., 2000). Class I signal peptides are typical Sec-type signal peptides that are cleaved by type I SPases. These signal peptides are usually 18–35 amino acids long and do not contain a strict consensus sequence. They share, however, a tripartite structure. The amino-terminal N-domain, which is usually two to eight residues, contains one or more positively charged residues. The N-domain is important for interaction with the protein translocation machinery (Akita *et al*., 1990) and negatively charged lipid head groups on the cytoplasmic face of the lipid bilayer (de Vrije *et al*., 1990; Phoenix *et al*., 1993). The hydrophobic H-domain that follows the N-domain varies in length from 8 to 15 residues. This region has been proposed to form an α-helical conformation in the membrane (Briggs *et al*., 1986). The third domain (C-domain) of the signal peptide contains the cleavage site for SPase. The residues at positions −3 and −1 (relative to the start of the mature protein) are usually those with small neutral side chains, such as alanine, glycine, serine and threonine (von Heijne, 1984). A subgroup of Class I signal peptides directs proteins to the Tat pathway. These signal peptides are very similar to Sec-type signal peptides but contain, in addition, a typical twin-arginine motif just before the H-domain. The consensus sequence of this motif is (S/T)RRx$\phi\phi$, where the arginines are invariant, x is any residue, and ϕ is a hydrophobic residue (Berks, 1996; Cristóbal *et al*., 1999). In Gram-negative bacteria, in particular, the hydrophobic residues (often FL) are frequently followed by a lysine residue (Berks, 1996). Surprisingly, the *E. coli* Tat substrate pre-SufI is still exported (albeit inefficiently) by the Tat pathway when one of the two arginine residues is replaced by a lysine (Stanley *et al*., 2000). There even exists a natural Tat-dependent substrate in *Salmonella enterica* that contains a 'KR' motif instead of 'RR' (Hinsley *et al*., 2001). Such examples are, however, still very rare and may be more the exception than the rule.

Class II signal peptides are found in bacterial lipoproteins. These signal peptides, which contain similar N- and H-domains to Class I signal peptides, are characterized by a conserved lipobox in the C-domain with the consensus sequence L(A/S)(G/A)C (von Heijne, 1989; Hayashi & Wu, 1990). The invariable cysteine in this lipobox is lipid-modified by a diacylglyceryl transferase and becomes the first residue of the mature protein after cleavage by a lipoprotein-specific type II SPase. Due to the lipid-modified cysteine, the protein remains anchored to the cytoplasmic membrane. Most Class II signal peptides are predicted to be Sec-dependent, since twin-arginine motifs are only rarely found in these signal peptides (Tjalsma *et al*., 2000).

The third class of signal peptides are present in prepilin-like proteins. These signal peptides also have a positively charged N-domain and a hydrophobic H-domain. In contrast to other signal peptides, prepilin signal peptides are cleaved just before the H-domain by a specific SPase that has its active site on the cytoplasmic face of the membrane (Lory, 1994). The H-domain, which remains attached to the mature protein, plays an important role in the assembly of the prepilin-like structures (Forest & Tainer, 1997).

The fourth class of signal peptides are found in certain bacteriocins and pheromones that are exported by ABC transporters. These lack a hydrophobic domain, and only some of them contain a double glycine motif at the −1 and −2 positions relative to the cleavage site (Michiels *et al*., 2001).

Only few extracellular or wall-bound proteins from archaea have been described. Through genomic analysis, the eurarchaeon *Methanococcus jannaschii* was predicted to contain at least 34 secretory proteins (Nielsen *et al*., 1999), but the criteria used for identification of proteins containing signal peptides were rather conservative and it is likely that several were not identified. Recently, a more extensive survey was performed on the genome of *Sulfolobus solfataricus*, a member of the crenarchaeotes (Albers & Driessen, 2002). In that case,

131 putative secretory proteins were identified. The main conclusions from these studies were that archaeal class I signal peptides have a bacterial-like charge distribution and a eukaryotic-like cleavage site. Very recently, Rose *et al*. (2002) developed a software program called 'TATFIND' that is able to identify Tat-dependent substrates in any sequenced genome. They identified several Tat-dependent substrates in *Halobacterium* sp. NRC-1 and other archaea, and concluded that haloarchaea use the Tat pathway extensively.

Signal peptide predictions in *Halobacterium* sp. NRC-1

The availability of the genome sequence of *Halobacterium* sp. NRC-1 facilitates the identification of secretory proteins using SignalP2.0 (Nielsen *et al*., 1999; www.cbs.dtu.dk/services/SignalP-2.0/). This is a neural network that, due to the lack of data, has not been trained on archaeal proteins. The complete proteome of *Halobacterium* sp. NRC-1 (www.ebi.ac.uk/

Table 1. Predicted Sec-type signal peptides of *Halobacterium* sp. NRC-1

Positively charged residues are indicated in bold letters, the H-domain is indicated in grey shading, and the residues at positions – 3 to – 1 relative to the SPase I cleavage site are underlined. Since the archaeal SPase is more similar to its eukaryotic counterpart, the cleavage sites (indicated with an arrow) were predicted using the eukaryotic dataset of SignalP. Hypothetical indicates similarity to other proteins with unknown function; nh, no homologues identified.

Protein	Signal peptide	↓		Predicted function
CbiN	MN**R**WLAAGGILLGALVVFSFVSAGAWG		GAD	Co-transport
Csg*	MTDTTG**K**L**R**AVLLTALMVGSVIGAGVAFTGGAAA		ANA	S-layer
Edp	MTNPID**K**HA**K**TLATSWTFIIVVAAIVGGVGGVA		LQS	Signal peptide peptidase
HemV1	MH**R**G**R**FATLVIVALAVTMTAPAGA		LAP	Iron-binding protein
Vng0198	MTQ**R**GQFVLAGAAVAALALASVAVA		YLQ	NH
Vng0200	MQ**R**GQANLVALVVGVLLVGAAVTLA		VGA	NH
Vng0204	MT**R**GQSAVVGVAVLVAATVVAVAA		LTA	NH
Vng0222	MVAE**R**TH**R**W**R**GAVALSLAVIAAGATT**R**VPAVLVLGVLGVAGA		GYA	Hypothetical
Vng0262	MLPEVGLLAP**R**SLAVLAGVAA		AVG	Hypothetical
Vng0483	MFEG**R**LWT**R**LLVGGLVAVAVLA		GGA	Hypothetical
Vng0539	MN**R**T**R**IVVVAASVLAVLTVIAAVVPA		GTV	NH
Vng0953	MGFSVSGSAAILFIAAFVSVGILYSAA		FNG	Hypothetical
Vng1355	ME**R**LDA**R**VVVVWLW**R**VLVFAVVLAGAAG		IGS	NH
Vng1372	MG**R**TAAI**R**LAGIAVVVTAAVVAS		QPA	NH
Vng1427	MVAA**R**LATAAVGVAGSLAVSVVAWQVFDLPVFVLA		VPF	NH
Vng1598	MFV**R**AA**R**SMALMFASVVPS**R**SAIS		CWV	NH
Vng1657	MD**R**E**R**LLAVGIAVVVLASAVTAVVIPGVVA		DAD	Hypothetical
Vng1787	MVDGGSA**R**VLLLAVSAAFA		GDD	NH
Vng1953	MTDHTPTT**K**L**R**AVALAALLVLSVLAGATTFTGAAA		AAA	NH
Vng2098	MVAA**R**TTDTLLWGCIAALAFLVLAQA		YVL	NH
Vng2148	MS**R**APLAVAALVVVATAMAATGLA		APG	NH
Vng2149	M**R**QALLLLVALACLVAVPVGAG		GSG	Hypothetical
Vng2236	MVVGVAVVAW**K**APVVAAGLTAVLVGLALVA**R**G		AIQ	NH
Vng2461	MPPVYAT**R**ALVAVLCELAADA		DPD	NH
Vng2567	MASVSSAHLVLFIAAILFSAALAGALT		DSA	Hypothetical
Vng2589	MNA**R**STLSVCAVAAVLVVAGIAGATA		LGM	Hypothetical
Vng6208	M**K**SLQL**K**VGAVVGAFVLVTVLVGGGLA		WNP	Hypothetical
Vng6209	MAG**R**LIVFGLLWLGAIALVTVGA		F**R**Y	NH

*The first methionine encoded by the mRNA of the *csg* gene is, most likely, the first residue of the precursor protein (Lechner & Sumper, 1987), and not the second methionine as indicated in the databases (see GenBank accession no. AAG20702).

Table 2. Predicted Tat-type signal peptides of *Halobacterium* sp. NRC-1

Positively charged residues are indicated in bold letters, twin-arginines are black-boxed, the H-domain is indicated in grey shading, and the residues at positions − 3 to − 1 relative to the SPase I cleavage site are underlined. Since the archaeal SPase is more similar to its eukaryotic counterpart, the cleavage sites (indicated with an arrow) were predicted using the eukaryotic dataset of SignalP. Hypothetical indicates similarity to other proteins with unknown function; NH, no homologues identified.

Protein	Signal peptide	↓		Predicted function
Aph	MPTPHTTESPSVD**RR**TFLAGLSGAVAGGAVA		SQF	Alkaline phosphatase
Chi	MPHD**RR**SYL**R**TSSAVIASLLAASTPTSA		ADT	Chitinase
DmsA	MSDTDLNAT**RR**DVL**K**SGAVAAVGLSGGGLLST		LQE	DMSO reductase
HcpB	MT**R**LDDTALS**RR**GVL**R**AAAGTATAVAAGTAATGAAAAQA		YDG	Halocyanin-like protein
HcpD	MTDSDSAVT**RR**VLQGSAGAGAAAAGIGGFAAGGAAQS		ASI	Halocyanin-like protein
Hly	MADNTNVT**RR**SFLTATGAAAGSVALVGVSA		GES	Halolysin
Sub	M**R**QT**RR**TFM**K**TAAAAIGGLSATTQPVTA		EGV	Subtilisin
Vng0199	MPE**R**G**RR**GQASLPAVEAAIGVFVILAVAATFTVG		VPG	NH
Vng0615	MN**K**WQ**RR**TAAYSLGLLGLMAAFAVG		YYY	Potassium transport
Vng0818	MPHD**RR**SYL**R**TSSAVIASLLAASTPTSA		ADT	Chitinase
Vng0819	MN**RR**TYIQ**R**GALTVGALLGASASATAA		SDS	Chitin-binding protein
Vng1036	MELT**RR**DVLAALAAGGVAA		GAG	NH
Vng1268	MPS**RR**LL**R**TVGALAAVPLVTS		APA	NH
Vng1952	MFDSMDTI**R**PTI**RR**TAVVVTAVAVIAALAAAGSTA		AAA	NH
Vng6162	MVS**K**QN**RR**TFL**K**TVGASGTTAIALSSAATVSA		EDK	NH
Vng6266	MCIMS**R**Q**R**T**RR**DII**R**TAGTTAAGLAVVSGSAVA		SEN	NH
Vng6296	M**RR**WGM**K**L**R**QIAGSLAAATGVVA		ATN	Hypothetical
Vng6323	MPSG**K**S**RR**VLAGAGSVILGGSVA		AFG	NH

proteome/index.html) was, therefore, analysed with the available datasets (eukaryotic, Gram-positive and Gram-negative). This resulted in a dataset containing proteins with putative signal peptides. These proteins were further analysed for the presence of additional membrane-spanning domains, and those predicted to be polytopic membrane proteins were excluded. The signal peptides in the final dataset were screened for the presence of an RR motif, a lipobox, or a cleavage site for prepilin SPase. Class IV signal peptides were not identified using this method.

Sec-type signal peptides

The analysis resulted in the identification of 28 proteins that are putative substrates for the Sec pathway and cleaved by type I SPase (Table 1). This represents 1 % of the proteome, which is relatively low as compared with for instance *Bacillus subtilis* (4 % ; Tjalsma *et al*., 2000), *E. coli* (6 % ; unpublished results) or *S. solfataricus* (4 % ; Albers & Driessen, 2002). The N-regions of these signal peptides usually contain one or two positively charged residues, with a strong bias to arginines (18 % Lys, 82 % Arg). This bias, however, does not seem to be limited to N-regions of signal peptides, as a similar ratio is found in the entire proteome of *Halobacterium* sp. NRC-1 (21 % Lys, 79 % Arg). In contrast to predicted signal peptides of *M. jannaschii* and *S. solfataricus* (Nielsen *et al*., 1999; Albers & Driessen, 2002), there is not an increased isoleucine content in the H-domain. Because archaeal SPases are more closely related to their eukaryotic counterparts (Tjalsma *et al*., 1998), the cleavage sites indicated in Table 1 were based upon the predictions of the eukaryotic dataset in SignalP. The cleavage site of the cell-surface glycoprotein Csg, which is one of the very few that has been determined experimentally (Lechner & Sumper, 1987), was correctly predicted.

A well-known protein of *Halobacterium* sp. NRC-1 is bacteriorhodopsin, which is a light-driven proton pump. It was shown that this membrane protein inserts co-translationally (Dale *et al*., 2000), suggesting that bacteriorhodopsin is SRP-dependent. The signal peptide of this protein, which is 13 residues in length, is very unusual as it does not contain a hydrophobic core or any positive charges (Seehra & Khorana, 1984). Since the signal peptide is not required for membrane insertion (Xu *et al*., 1995), it is unlikely that it is involved in SRP-dependent targeting and may not even be a true signal peptide.

Table 3. Predicted Sec-type lipoprotein signal peptides of *Halobacterium* sp. NRC-1

Positively charged residues are indicated in bold letters, and the H-domain is indicated in grey shading. Residues identical to the bacterial motif of the lipobox [L(AS)(GA)C; at postions − 3 to + 1] are bold and underlined. The positions of the cleavage sites are indicated with with an arrow. Hypothetical indicates similarity to other proteins with unknown function; NH, no homologues identified.

Protein	Signal peptide ↓	Predicted function
HcpC	MLKSTGAAVTATLLAG ↓ CSGG	Halocyanin-like protein
Vng0121	MRWTVLVVAAFVVGAG ↓ CGGP	Hypothetical
Vng0346	MLISGYTKTGFGATLAAGTGTGWLLAG ↓ CLSV	NH
Vng0759	MRALWVALAVAALAVTAG ↓ CAGF	NH
Vng1094	MTQICDAMVSVRLCLILALVVLGG ↓ CVFV	Hypothetical
Vng1250	MVRDLHVPRTALLAVGLAVLLVAAG ↓ CTGM	NH
Vng1890	MVRSATIPLVLIALLAAAG ↓ CLGS	NH

Tat-type signal peptides

The signal peptide searches revealed 18 putative substrates for the Tat pathway (Table 2). This is 40 % of the Class I signal peptides, which is surprisingly high as compared to for instance *B. subtilis* (8 % ; Tjalsma *et al.*, 2000), *E. coli* (9 % ; unpublished data) or *S. solfataricus* (2·5 % ; Albers & Driessen, 2002). Several of the putative Tat-type signal peptides also contain the remainder of the bacterial RR motif, including the signal peptide of DmsA (DMSO reductase). The latter protein, which is predicted to bind a molybdopterin cofactor, is a typical Tat substrate that is involved in anaerobic respiration. Notably, *Halobacterium* species have, to my knowledge, never been reported to grow anaerobically using DMSO as the terminal electron acceptor. Several proteins that contain Tat-signal peptides are normally Sec-dependent in bacteria and do not bind cofactors, such as the subtilisin-like protease Sub and the alkaline phosphatase Aph. Interestingly, alkaline phosphatase from the thermophilic bacterium *Thermus thermophilus* was recently shown to be a Tat-dependent substrate (Angelini *et al.*, 2001). In this study, it was speculated that the Tat pathway could be a more suitable transport route for thermophilic enzymes because they are more rigidly folded.

Lipoproteins

Surprisingly, several signal peptides that were identified contained a conserved cysteine together with other residues that are diagnostic of bacterial lipoproteins. The H-domains of lipoprotein signal peptides are usually shorter than those of secretory proteins and are not always recognized by the SignalP server (Nielsen *et al.*, 1997; Tjalsma *et al.*, 2000). Therefore, all proteins from *Halobacterium* sp. NRC-1 were screened for the presence of the lipobox in the amino-terminal portion of proteins (using 'pattern search' in the Pedant database at http://pedant.gsf.de/). Positive scoring proteins were analysed for the presence of a hydrophobic domain, which resulted in several additional putative lipoproteins and bringing the total number of putative lipoproteins in *Halobacterium* sp. NRC-1 to 51 (Tables 3 and 4). More than 80 % of these (44) contained a twin-arginine motif, suggesting that most lipoproteins are exported by the Tat pathway. This is very remarkable for a number of reasons. Firstly, the presence of twin-arginine motifs in lipoproteins in bacteria is very rare as, for example, none of the lipoproteins from *E. coli* (unpublished data) and only 4 % of the lipoproteins from *B. subtilis* (Tjalsma *et al.*, 2000) contain a twin-arginine motif. Secondly, archaea do not contain homologues of any of the proteins involved in bacterial lipomodification and processing, and there is only limited evidence for the presence of lipoproteins in archaea. One study showed that halocyanin, a blue copper protein from the haloalkaliphilic archaeon *Natronobacterium pharaonis*, contains a typical lipoprotein signal peptide and that the protein is modified at the cysteine with, most likely, a diphytanyl (glycerol)diether (Mattar *et al.*, 1994). Interestingly, *Halobacterium* sp. NRC-1 encodes at least seven extracellular halocyanin, plastocyanin, and other copper-binding proteins (see Tables 2, 3 and 4). Five of these are predicted to be lipoproteins, and all but one (HcpC) are predicted to be Tat-dependent. Notably, halocyanin from *N. pharaonis* also contains a twin-arginine motif in its signal peptide. Several putative lipoproteins are also found in other archaea, such as *M. jannaschii* and *Archaeoglobus fulgidus* (Tjalsma *et al.*, 2001). Since archaea do not contain homologues of bacterial enzymes involved in modification and processing of lipoproteins, it is very likely that they contain a completely novel mechanism for this process. The presence of a twin-arginine motif in more than 80 % of putative lipoproteins may be specific to *Halobacterium* sp. NRC-1 and other halophiles. A preliminary survey of the partially sequenced halophilic archaeon *Hf. volcanii* genome (http://wit-scranton.mbi.scranton.

Table 4. Predicted Tat-type lipoprotein signal peptides of *Halobacterium* sp. NRC-1

Positively charged residues are indicated in bold letters, twin-arginines are black-boxed, and the H-domain is indicated in grey shading. Residues identical to the bacterial motif of the lipobox [L(AS)(GA)C; at postions – 3 to + 1] are bold and underlined. The positions of the cleavage sites are indicated with with an arrow. Hypothetical indicates similarity to other proteins with unknown function; NH, no homologues identified; BP, binding protein.

Protein	Signal peptide	↓	Predicted function
AppA	MAN**R**DNHA**R**PS**R**PNTDQSTSEAHS**RRR**VLKALGATGVAAGV**AG**	**C**LGG	Oligopeptide BP
BasB	MHSTT**RR**EWLGAIGATAATG**LAG**	**C**AGV	Amino acid BP
CosB	MMDTPEHASTSS**RR**QLLGMLAAGGTTAV**AG**	**C**TTI	BP
ComA	MSL**RR**LLAVFAVVGMVVI**AG**	**C**AGG	Competence protein
DppD	MF**K**YAIVPYS**R**MSDDTVS**RR**GFLKAAGAATVVATST**AG**	**C**TDS	Dipeptide BP
Fbr	MQ**RRR**EFLQATGAALAAVG**LAG**	**C**SDS	Plastocyanin-like
HcpA	MSAMG**R**APD**RR**TFL**R**SAVAGGLAAI**AG**	**C**TDR	Halocyanin-like
H0844	MNDNTNSPSHS**R**CID**RR**TVLSGLSGALVA**LAG**	**C**TDN	NH
HtrA	MPS**RR**DVL**R**LGAGVLAAGT**AG**	**C**TDT	Protease
Ibp	MSEESGF**RRR**TFLAATGAATLSG**LAG**	**C**SGM	Iron BP
Imp	MPTEHT**RRR**FLQATGATSIAA**LAG**	**C**AGG	Immunogenic protein
Pan1	MTDQIGAPGLGIS**RR**DFVAATAGVGTAA**IAG**	**C**TAP	Copper BP
PhoX	MPADDAE**R**TT**R**T**RR**QVLAGMGATGAAA**LAG**	**C**QST	Phosphate BP
Pcy	MH**RR**AFLAGGTTLSVGV**LAG**	**C**IGP	Plastocyanin-like
PotD	MTDHTDLS**RR**QYV**K**TAALSAAATTA**LAG**	**C**MGG	Spermidine/putrescine BP
PrrC	M**RRR**TYLSLVGSAAAAGT**AG**	**C**LGV	Hypothetical
Thb	MTS**RRR**FVAAVGSATAASLG**LAG**	**C**VGD	Thiamin BP
UgpB	MAPTTET**RR**ATAT**RR**AVLAGAGTAGLTAF**AG**	**C**GAL	Glycerol-3-phosphate BP
Vng0212	MD**RR**TYLATSVAL**LAG**	**C**STS	Hypothetical
Vng0404	M**K**R**R**AFLATAAVLTTT**G**	**C**LG**R**	NH
Vng0496	MD**RR**TLLG**R**AVAGLTACVSLSVG**AG**	**C**LGD	NH
Vng0903	MADFD**RR**EFL**K**LAGGTVGASLV**AG**	**C**SSG	Hypothetical
Vng0798	MNSDQ**R**GVP**RR**EFL**K**AAVAIGGASA**LSA**	**C**LG**R**	NH
Vng0914	MN**RR**AFLTASAGLGSTAA**LAG**	**C**LGA	NH
Vng1039	MMACT**RR**KALAAVGTTLS**LSG**	**C**A**R**I	NH
Vng1538	M**RR**QLAALLVVALVAV**AG**	**C**AGG	NH
Vng1589	M**R**P**R**IVLIAALVVSSIA**LAG**	**C**SGT	NH
Vng1595	MTGTAPVS**RR**QYLGTAGAIIGTT**AG**	**C**LTG	Hypothetical
Vng1720	MAGEAT**RR**RCLGHVAAGAAL**LAG**	**C**TSH	Hypothetical
Vng1869	MSP**R**QPDASDADAPADTTSDAGHPHSTVG**RR**SFLAATGAAASATT**LAG**	**C**LGG	Spermidine/putrescine BP
Vng2101	MS**R**A**R**T**RR**LLSSVALAVVAG**LTG**	**C**STH	NH
Vng2152	M**R**GQPVH**RR**SVLALVGGGAVSA**LAG**	**C**TDT	Hypothetical
Vng2165	MH**RR**ALLGGVAAAAAAS**LTG**	**C**MGA	NH
Vng2205	M**RRR**TVLVGVTGALFGVGG**G**	**C**LTM	NH
Vng2299	M**R**T**RR**QFLATTTSLTTVGL**LAG**	**C**ARS	NH
Vng2539	MA**RR**LLAVAVVCLVV**LAG**	**C**QGG	Hypothetical
Vng2549	MQ**RR**AFL**K**AGSAATLAGL**LAG**	**C**SSP	Hypothetical
Vng2562	MA**RRR**QILAGGASLIAAS**LAG**	**C**TSS	Hypothetical
Vng2637	MH**RR**PFLS**R**LCLGAVAAT**AG**	**C**LS**R**	NH
Vng6157	MNG**R**GNNM**R**QWFGDDMP**RR**VLL**R**SIGVCSAVGI**AG**	**C**LSQ	NH
Vng6297	MVDSICT**RR**LLAAVGATSVSAV**AG**	**C**SST	Cell-surface protein
Vng6432	MVPS**RR**AFLRNAGIAAIGG**LAG**	**C**LTD	NH
YcdH	MDEQTHTHLS**RR**T**R**TLTASAGVLSAG**LAG**	**C**ITS	Zinc BP
YqgG	MHSDPDDGASGPVS**RR**AFVAATGTAGVAA**LAG**	**C**ANS	Phosphate BP

edu/Haloferax/) indicates that also this halophilic archaeon contains several putative lipoproteins, most of which are putative Tat substrates (data not shown). As mentioned before, *Hf. volcanii* also contains a TatC-like protein with the same topology as TatC2 (Fig. 1c). Thus it is conceivable that halophiles contain specially adapted Tat pathways that enable the export of a large number of Tat-dependent (lipo)proteins.

Table 5. Flagellin proteins of *Halobacterium* sp. NRC-1

Flagellin proteins with prepilin-like signal peptides. Postively charged residues are indicated in bold letters and the H-domain is indicated in grey shading. The positions of the cleavage sites are indicated with an arrow. The starting codons of the genes encoding FlaA2 (Genbank accession no. AAG19425), FlaB2 (GenBank accession no. NP–279904) and FlaB3 (GenBank accession no. NP–279905)were assigned to the condon corresponding to the Met residue that is located in the middle of the H-domain. Because of the similarity between these signal peptides, and the identity to the flagellin proteins of H. salinarum (see Thomas et al., 2001), the amino-termini of these proteins have been corrected as shown.

Protein	↓ Signal peptide
FlaA1a	MFEFITDEDE**R**G QVGIGTLIVFIAMVLVAAIAAGVLINTAGF
FlaA1b	MESM**RR**G QVGIGTLVVFMAMILVAAMAASTLVDIGGM
FlaA2	MFEFITDEDE**R**G QVGIGTLIVFIAMVLVAAIAAGVLINTAGF
FlaB1	MFEFITDEDE**R**G QVGIGTLIVFIAMVLVAAIAAGVLINTAGY
FlaB2	MFEFITDEDE**R**G QVGIGTLIVFIAMVLVAAIAAGVLINTAGY
FlaB3	MFEFITDEDE**R**G QVGIGTLIVFIAMVLVAAIAAGVLINTAGY

Interestingly, lipid modification of archaeal proteins such as the cell surface glycoprotein (Csg) of *Hf. volcanii* and *Halobacterium salinarum* has been reported before (Kikuchi *et al*., 1999; Konrad & Eichler, 2002). However, the modification described in these studies appears to be of a different type than signal peptide modification since it involved the modification of the C-terminal domain of the mature protein (Konrad & Eichler, 2002). Whether the same components are involved in both types of modification remains an open question.

Type IV prepilin signal peptides

In archaea, flagellin proteins contain a signal peptide that is similar to that of type IV prepilins (see Thomas *et al*., 2001). It is not known how these signal peptides are cleaved, as archaea lack a prepilin-type SPase, or how these flagellin proteins are translocated across the cytoplasmic membrane (Thomas *et al*., 2001). Interestingly, the archaeon *S. solfataricus* contains several sugar-binding proteins with a similar signal peptide, and for some of these the cleavage site, which is located just before the hydrophobic domain, was experimentally verified (Albers & Driessen, 2002). Strikingly, most binding proteins of *Halobacterium* sp. NRC-1 seem to be retained at the membrane through lipo-modification and not through the hydrophobic domain of a prepilin-like signal peptide (Table 4). The genome of *Halobacterium* sp. NRC-1 does encode six flagellin proteins, all of which have very similar prepilin-like signal peptides (Table 5; Ng *et al*., 2000). Four genes that are closely located on the chromosome (*vng0198*, *vng0199*, *vng0200* and *vng0204*; see Tables 1 and 2) encode proteins containing the same residues as the flagellin proteins in their cleavage site (RGQ). The processing of these proteins should, however, be experimentally determined as they also contain a predicted cleavage site for type I SPase.

Concluding remarks

In total, 103 proteins with putative signal peptides were identified in *Halobacterium* sp. NRC-1, and more than 60% of these contain a twin-arginine motif. This is extremely high since in bacteria and non-halophilic archaea the majority of proteins (> 90%) appear to be Sec-dependent. Although it has to be proven experimentally that *Halobacterium* sp. NRC-1 proteins are indeed Tat-dependent, it suggests that the halobacterial Tat pathway plays a major role in protein translocation. Similar conclusions were reached by Rose *et al*. (2002) through *in silico* analyses and the observation that an α-amylase (which is usually a Sec substrate) from the alkalihalophilic archaeon *Natronococcus* sp. was not exported in *Hf. volcanii* when the RR motif in its signal peptide was replaced with a KK motif. The question remains why so many proteins in halophilic archaea are Tat-dependent. The reason for this may lie in the environment in which *Halobacterium* spp. and other haloarchaea thrive. They thrive in conditions with about 4–5 M NaCl, and to maintain an osmotic balance K^+ is accumulated inside the cell to a concentration that approximates that of the external Na^+ concentration (Lanyi, 1974). Proteins from non-halophilic organisms would just aggregate and precipitate under these conditions, but halobacterial proteins have increased negative surface charge that enables the binding of essential water molecules and prevents aggregation through electrostatic repulsion (Elcock & McCammon, 1998; Kennedy *et al*., 2001; Madern *et al*., 2000). Newly synthesized proteins will have to fold rapidly in their native conformation to prevent aggregation. Thus many secretory proteins may fold into their 3-dimensional conformation before they reach the membrane. This makes these proteins incompatible with the Sec system and necessitates export via the Tat system. The alternative solution to the problem of aggregation in the cytoplasm is co-translational trans-

location in which the secretory protein is synthesized at the site of translocation. Therefore, it may be that in *Halobacterium* spp. post-translational translocation occurs exclusively via the Tat pathway, leaving the Sec pathway with co-translational translocation only.

In conclusion, several novel aspects can be found in the protein transport pathways of *Halobacterium* sp. NRC-1, some of which may be specific for halophiles. Depending on the environment in which they thrive, specific adaptations may indeed be found in other archaea as well. Nevertheless, because of the genetic amenability and ease of culturing, *Halobacterium* species are very good model systems to study protein translocation in archaea and provide a fundamental insight into an important aspect of the cell biology of the third domain of life.

Acknowledgements

I thank Dr M. Pohlschröder for sharing information prior to publication, Dr V. DelVecchio and other members of the *H. volcanii* sequencing group for making sequences available, and Dr J. M. van Dijl for stimulating discussions. I also thank two anonymous referees for helpful suggestions and comments. This work is supported by a Royal Society University Research Fellowship.

References

Akita, M., Sasaki, S., Matsuyama, S. & Mizushima, S. (1990). SecA interacts with secretory proteins by recognizing the positive charge at the amino terminus of the signal peptide in *Escherichia coli*. *J Biol Chem* **265**, 8164–8169.

Albers, S. V. & Driessen, A. M. (2002). Signal peptides of secreted proteins of the archaeon *Sulfolobus solfataricus*: a genomic survey. *Arch Microbiol* **177**, 209–216.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Angelini, S., Moreno, R., Gouffi, K., Santini, C., Yamagishi, A., Berenguer, J. & Wu, L. (2001). Export of *Thermus thermophilus* alkaline phosphatase via the twin-arginine translocation pathway in *Escherichia coli*. *FEBS Lett* **506**, 103–107.

Arkowitz, R. A. & Wickner, W. (1994). SecD and SecF are required for the proton electrochemical gradient stimulation of preprotein translocation. *EMBO J* **13**, 954–963.

Beck, K., Eisner, G., Trescher, D., Dalbey, R. E., Brunner, J. & Muller, M. (2001). YidC, an assembly site for polytopic *Escherichia coli* membrane proteins located in immediate proximity to the SecYE translocon and lipids. *EMBO Rep* **2**, 709–714.

Berks, B. C. (1996). A common export pathway for proteins binding complex redox cofactors? *Mol Microbiol* **22**, 393–404.

Berks, B. C., Sargent, F. & Palmer, T. (2000). The Tat protein export pathway. *Mol Microbiol* **35**, 260–274.

Bolhuis, A., Broekhuizen, C. P., Sorokin, A., van Roosmalen, M. L., Venema, G., Bron, S., Quax, W. J. & van Dijl, J. M. (1998). SecDF of *Bacillus subtilis*, a molecular siamese twin required for the efficient secretion of proteins. *J Biol Chem* **273**, 21217–21224.

Bolhuis, A., Mathers, J. E., Thomas, J. D., Barrett, C. M. & Robinson, C. (2001). TatB and TatC form a functional and structural unit of the twin-arginine translocase from *Escherichia coli*. *J Biol Chem* **276**, 20213–20219.

Briggs, M. S., Cornell, D. G., Dluhy, R. A. & Gierasch, L. M. (1986). Conformations of signal peptides induced by lipids suggest initial steps in protein export. *Science* **233**, 206–208.

Chaddock, A. M., Mant, A., Karnauchov, I., Brink, S., Herrmann, R. G., Klosgen, R. B. & Robinson, C. (1995). A new type of signal peptide: central role of a twin-arginine motif in transfer signals for the delta pH-dependent thylakoidal protein translocase. *EMBO J* **14**, 2715–2722.

Cristóbal, S., de Gier, J. W., Nielsen, H. & von Heijne, G. (1999). Competition between Sec- and Tat-dependent protein translocation in *Escherichia coli*. *EMBO J* **18**, 2982–2990.

Dalbey, R. E. & Kuhn, A. (2000). Evolutionarily related insertion pathways of bacterial, mitochondrial, and thylakoid membrane proteins. *Annu Rev Cell Dev Biol* **16**, 51–87.

Dalbey, R. E., Lively, M. O., Bron, S. & van Dijl, J. M. (1997). The chemistry and enzymology of the type I signal peptidases. *Protein Sci* **6**, 1129–1138.

Dale, H., Angevine, C. M. & Krebs, M. P. (2000). Ordered membrane insertion of an archaeal opsin *in vivo*. *Proc Natl Acad Sci U S A* **97**, 7847–7852.

de Vrije, G. J., Batenburg, A. M., Killian, J. A. & de Kruijff, B. (1990). Lipid involvement in protein translocation in *Escherichia coli*. *Mol Microbiol* **4**, 143–150.

Duong, F. & Wickner, W. (1997). The SecDFyajC domain of preprotein translocase controls preprotein movement by regulating SecA membrane cycling. *EMBO J* **16**, 4871–4879.

Economou, A. (1998). Bacterial preprotein translocase: mechanism and conformational dynamics of a processive enzyme. *Mol Microbiol* **27**, 511–518.

Eichler, J. (2000). Archaeal protein translocation crossing membranes in the third domain of life. *Eur J Biochem* **267**, 3402–3412.

Eichler, J. & Moll, R. (2001). The signal recognition particle of Archaea. *Trends Microbiol* **9**, 130–136.

Elcock, A. H. & McCammon, J. A. (1998). Electrostatic contributions to the stability of halophilic proteins. *J Mol Biol* **280**, 731–748.

Evans, E. A., Gilmore, R. & Blobel, G. (1986). Purification of microsomal signal peptidase as a complex. *Proc Natl Acad Sci U S A* **83**, 581–585.

Forest, K. T. & Tainer, J. A. (1997). Type-4 pilus-structure: outside to inside and top to bottom – a minireview. *Gene* **192**, 165–169.

Görlich, D. & Rapoport, T. A. (1993). Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell* **75**, 615–630.

Hayashi, S. & Wu, H. C. (1990). Lipoproteins in bacteria. *J Bioenerg Biomembr* **22**, 451–471.

Hinsley, A. P., Stanley, N. R., Palmer, T. & Berks, B. C. (2001). A naturally occurring bacterial Tat signal peptide lacking one of the "invariant" arginine residues of the consensus targeting motif. *FEBS Lett* **497**, 45–49.

Johnson, A. E. & Waes, M. A. (1999). The translocon: a dynamic gateway at the ER membrane. *Annu Rev Cell Dev Biol* **15**, 799–842.

Keenan, R. J., Freymann, D. M., Stroud, R. M. & Walter, P. (2001). The signal recognition particle. *Annu Rev Biochem* **70**, 755–775.

Kennedy, S. P., Ng, W. V., Salzberg, S. L., Hood, L. & DasSarma, S. (2001). Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res* **11**, 1641–1650.

Kikuchi, A., Sagami, H. & Ogura, K. (1999). Evidence for covalent attachment of diphytanylglyceryl phosphate to the cell-surface glycoprotein of *Halobacterium halobium*. *J Biol Chem* **274**, 18011–18016.

Kinch, L. N., Saier, M. H., Jr & Grishin, N. V. (2002). Sec61β – a component of the archaeal protein secretory system. *Trends Biochem Sci* **27**, 170–171.

Konrad, Z. & Eichler, J. (2002). Lipid modification of proteins in archaea: attachment of a mevalonic acid-based lipid moiety to the S-layer glycoprotein of *Haloferax volcanii* follows protein translocation. *J Biol Chem* (in press).

Lanyi, J. K. (1974). Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol Rev* **38**, 272–290.

Lechner, J. & Sumper, M. (1987). The primary structure of a procaryotic glycoprotein. Cloning and sequencing of the cell surface glycoprotein gene of halobacteria. *J Biol Chem* **262**, 9724–9729.

Lory, S. (1994). Leader peptidase of the type IV prepilins and related proteins. In *Signal Peptidases*, pp. 17–29. Edited by G. von Heijne. Austin, TX: R. G. Landes.

Madern, D., Ebel, C. & Zaccai, G. (2000). Halophilic adaptation of enzymes. *Extremophiles* **4**, 91–98.

Manting, E. H. & Driessen, A. J. M. (2000). *Escherichia coli* translocase: unravelling of a molecular machine. *Mol Microbiol* **37**, 226–238.

Matsuyama, S., Fujita, Y. & Mizushima, S. (1993). SecD is involved in the release of translocated secretory proteins from the cytoplasmic membrane of *Escherichia coli*. *EMBO J* **12**, 265–270.

Mattar, S., Scharf, B., Kent, S. B., Rodewald, K., Oesterhelt, D. & Engelhard, M. (1994). The primary structure of halocyanin, an archaeal blue copper protein, predicts a lipid anchor for membrane fixation. *J Biol Chem* **269**, 14939–14945.

Michiels, J., Dirix, G., Vanderleyden, J. & Xi, C. (2001). Processing and export of peptide pheromones and bacteriocins in Gram-negative bacteria. *Trends Microbiol* **9**, 164–168.

Mori, H. & Cline, K. (2002). A twin-arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid ΔpH/Tat translocase. *J Cell Biol* **157**, 205–210.

Neumann-Haefelin, C., Schafer, U., Muller, M. & Koch, H. G. (2000). SRP-dependent co-translational targeting and SecA-dependent translocation analyzed as individual steps in the export of a bacterial protein. *EMBO J* **19**, 6419–6426.

Ng, W. V., Kennedy, S. P., Mahairas, G. G. & 40 other authors (2000). Genome sequence of *Halobacterium* species NRC-1. *Proc Natl Acad Sci U S A* **97**, 12176–12181.

Nguyen, T. H., Law, D. T. & Williams, D. B. (1991). Binding protein BiP is required for translocation of secretory proteins into the endoplasmic reticulum in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **88**, 1565–1569.

Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**, 1–6.

Nielsen, H., Brunak, S. & von Heijne, G. (1999). Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng* **12**, 3–9.

Nouwen, N., van der Laan, M. & Driessen, A. J. (2001). SecDFyajC is not required for the maintenance of the proton motive force. *FEBS Lett* **508**, 103–106.

Peck, R. F., Dassarma, S. & Krebs, M. P. (2000). Homologous gene knockout in the archaeon *Halobacterium salinarum* with *ura3* as a counterselectable marker. *Mol Microbiol* **35**, 667–676.

Phoenix, D. A., Kusters, R., Hikita, C., Mizushima, S. & de Kruijff, B. (1993). OmpF-Lpp signal sequence mutants with varying charge hydrophobicity ratios provide evidence for a phosphatidylglycerol-signal sequence interaction during protein translocation across the *Escherichia coli* inner membrane. *J Biol Chem* **268**, 17069–17073.

Pohlschröder, M., Prinz, W. A., Hartmann, E. & Beckwith, J. (1997). Protein translocation in the three domains of life: variations on a theme. *Cell* **91**, 563–566.

Robinson, C. & Bolhuis, A. (2001). Protein targeting by the twin-arginine translocation pathway. *Nat Rev Mol Cell Biol* **2**, 350–356.

Rose, R. W. & Pohlschröder, M. J. (2002). *In vivo* analysis of an essential archaeal signal recognition particle in its native host. *J Bacteriol* **84**, 3260–3267.

Rose, R. W., Brüser, T., Kissinger, J. C. & Pohlschöder, M. (2002). Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. *Mol Microbiol* **45**, 943–950.

Samuelson, J. C., Chen, M., Jiang, F., Moller, I., Wiedmann, M., Kuhn, A., Phillips, G. J. & Dalbey, R. E. (2000). YidC mediates membrane protein insertion in bacteria. *Nature* **406**, 637–641.

Schatz, G. & Dobberstein, B. (1996). Common principles of protein translocation across membranes. *Science* **271**, 1519–1526.

Scotti, P. A., Urbanus, M. L., Brunner, J., de Gier, J. W., von Heijne, G., van der Does, C., Driessen, A. J., Oudega, B. & Luirink, J. (2000). YidC, the *Escherichia coli* homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *EMBO J* **19**, 542–549.

Seehra, J. S. & Khorana, H. G. (1984). Bacteriorhodopsin precursor. Characterization and its integration into the purple membrane. *J Biol Chem* **259**, 4187–4193.

Seluanov, A. & Bibi, E. (1997). FtsY, the prokaryotic signal recognition particle receptor homologue, is essential for biogenesis of membrane proteins. *J Biol Chem* **272**, 2053–2055.

Settles, A. M., Yonetani, A., Baron, A., Bush, D. R., Cline, K. & Martienssen, R. (1997). Sec-independent protein translocation by the maize Hcf106 protein. *Science* **278**, 1467–1470.

Sowers, K. R. & Schreier, H. J. (1999). Gene transfer systems for the archaea. *Trends Microbiol* **7**, 212–219.

Stanley, N. R., Palmer, T. & Berks, B. C. (2000). The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in *Escherichia coli*. *J Biol Chem* **275**, 11591–11596.

Thomas, N. A., Bardy, S. L. & Jarell, K. F. (2001). The archaeal flagellum: a different kind of prokaryotic motility structure. *FEMS Microbiol Rev* **25**, 147–174.

Tjalsma, H., Bolhuis, A., van Roosmalen, M. L. & 7 other authors (1998). Functional analysis of the secretory precursor processing machinery of *Bacillus subtilis*; identification of a eubacterial homologue of archaeal and eukaryotic signal peptidases. *Genes Dev* **12**, 2318–2331.

Tjalsma, H., Bolhuis, A., Jongbloed, J. D., Bron, S. & van Dijl, J. M. (2000). Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol Mol Biol Rev* **64**, 515–547.

Tjalsma, H., Zanen, G., Bron, S. & van Dijl, J. M. (2001). The eubacterial lipoprotein-specific (type II) signal peptidase. In *The Enzymes*, vol. XXII, *Co- and Post Translational Proteolysis of*

Proteins, pp. 3–23. Edited by R. E. Dalbey & D. S. Sigman. San Diego, CA: Academic Press.

Tseng, T. T., Gratwick, K. S., Kollman, J., Park, D., Nies, D. H., Goffeau, A. & Saier, M. H., Jr (1999). The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. *J Mol Microbiol Biotechnol* **1**, 107–125.

Ulbrandt, N. D., Newitt, J. A. & Bernstein, H. D. (1997). The *E. coli* signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell* **88**, 187–196.

van de Vossenberg, J. L., Driessen, A. J. M. & Konings, W. N. (1998). The essence of being extremophilic: the role of the unique archaeal membrane lipids. *Extremophiles* **2**, 163–170.

Voigt, S., Jungnickel, B., Hartmann, E. & Rapoport, T. A. (1996). Signal sequence-dependent function of the TRAM protein during early phases of protein transport across the endoplasmic reticulum membrane. *J Cell Biol* **134**, 25–35.

von Heijne, G. (1984). How signal sequences maintain cleavage specificity. *J Mol Biol* **173**, 243–251.

von Heijne, G. (1989). The structure of signal peptides from bacterial lipoproteins. *Protein Eng* **2**, 531–534.

von Heijne, G. (1990). The signal peptide. *J Membr Biol* **115**, 195–201.

Woese, C. R., Kandler, O. & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**, 4576–4579.

Xu, Z. J., Moffett, D. B., Peters, T. R., Smith, L. D., Perry, B. P., Whitmer, J., Stokke, S. A. & Teintze, M. (1995). The role of the leader sequence coding region in expression and assembly of bacteriorhodopsin. *J Biol Chem* **270**, 24858–24863.

Yen, M. R., Tseng, Y. H., Nguyen, E. H., Wu, L. F. & Saier, M. H., Jr (2002). Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system. *Arch Microbiol* **177**, 441–450.