Heliyon 5 (2019) e01587

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.heliyon.com

Analysis of haloarchaeal twin-arginine translocase pathway reveals the diversity of the machineries

Deepanjan Ghosh [a,b], Debjyoti Boral [a,b], Koteswara Rao Vankudoth [a,b], Sureshkumar Ramasamy [a,*]

[a] Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune 411008, India
[b] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad- 201002, India

ARTICLE INFO

Keyword:
Microbiology

ABSTRACT

The twin-arginine translocase (Tat) pathway transports folded proteins across the plasma membrane and plays a critical role in protein transport in haloarchaea. Computational analysis and previous experimental evidence suggested that the Tat pathway transports almost the entire secretome in haloarchaea. The TatC, receptor component of this pathway shows greater variation in membrane topology in haloarchaea than in other organisms. The presence of a unique fourteen-transmembrane TatC homolog ($TatC_t$) in haloarchaea, over and above the expected TatC topological variants, indicates a strong correlation between the additional homologs and the large number of substrates transported via the haloarchaeal Tat pathway. Various combinations of TatC homologs with different topologies—$TatC_o$, $TatC_t$, $TatC_n$, and $TatC_x$ have been observed in haloarchaea. In this report, on the basis of these combinations we have segregated all haloarchaeal Tat substrates into two groups. The first group consists of substrates that are transported by $TatC_t$ alone, whereas the second group consists of substrates that are transported by the other TatC homologs ($TatC_o$, $TatC_n$, and $TatC_x$). The various haloarchaea TatA components also shows the possible segregation towards the substrates. We have also identified the possible homologs for Tat substrate chaperones, which act as a quality-control mechanism for proper protein folding. Further sequence analysis implies that the two TatC domains of $TatC_t$ complement each other's functionally. Substrate analysis also revealed subtle differences between the substrates being transported by various homologs: further experimental analysis is therefore required for better understanding of the complexities of the haloarchaeal Tat pathway.

1. Introduction

The twin-arginine translocase (Tat) pathway exclusively transports folded proteins across the plasma membrane [1, 2, 3, 4], unlike the Sec pathway, which transports proteins in an unfolded form [5] and requires ATP for the protein translocation. In almost all prokaryotes, the Tat pathway is responsible for transporting fewer proteins across the membrane than the Sec pathway. Archaea are known to be evolutionarily distinct from both bacteria and eukarya in many of their cellular characteristics. In the case of halophilic archaea, the Tat pathway has a decisive role in transporting almost 50% of secretome [6, 7, 8]. This might be because of the high cytoplasmic salt concentration in halophilic archaea, which these organisms maintain in order to balance the high sodium concentration of the environment. The high cytoplasmic salt concentration results in faster folding of proteins to prevent aggregation, so relatively few unfolded proteins are transported. The transport of folded proteins also ensures that the proteins are not folded in the extracellular environment, where chaperones might not be available and the chance of misfolding is therefore much higher [9].

Tat translocation is carried out by integral membrane proteins in the TatA and TatC families [1, 10]. In some organisms, such as archaea and Gram-positive bacteria with low guanine-cytosine (GC) genomic content, single TatA and TatC components are sufficient to mediate transport through the Tat pathway (the TatAC system). For example, *Bacillus subtilis* contains two *TatC* genes (denoted as T*atCd* and T*atCy*), which coordinate functions with their respective *TatA* partners (*TatAd* and *TatAy*). However, in other cases, Tat-mediated translocation involves another member of the TatA family [11], termed TatB. The Tat pathway in *Escherichia coli* consists of three distinct membrane-localized proteins (TatA, TatB, and TatC) and is the most studied model in which all the three components are essential for Tat functioning. TatABC systems are present in Gram-negative bacteria (including *E. coli*) [12, 13], in

* Corresponding author.
E-mail address: s.ramasamy@ncl.res.in (S. Ramasamy).

https://doi.org/10.1016/j.heliyon.2019.e01587
Received 1 November 2018; Received in revised form 23 March 2019; Accepted 24 April 2019

Gram-positive bacteria with high GC-content genomes, and in plant chloroplasts [14]. TatC is the receptor component that recognizes the characteristic RR-signal in the substrate N-terminal [15, 16]. TatC or the TatBC complex interacts with the substrate and recruits TatA. TatA forms an appropriately sized pore through which the substrate is transported across the membrane [17]. In plants, the thylakoid Tat system is required for the assembly of many essential components, such as photosystem II and the cytochrome *b6f* complex [18, 19]. The recently solved crystal structures of TatC from the thermophilic bacteria *Aquifex aeolicus* [20, 21] confirmed the presence of six transmembrane (TM) domains and the orientation of the amino and carboxy termini towards the cytoplasmic side of the membrane [12, 22, 23]. The structure of TatAd from *Bacillus subtilis* has been determined by NMR to be composed of a single TM domain and a cytosolically located amphipathic helix [24]. Common Tat substrates include respiratory redox enzymes, bacterial virulence factors, lipoproteins [25, 26, 27], and proteins involved in maintaining cell-wall integrity. These are mostly complexes containing cofactors, and many form oligomeric assemblies. Delivery of these proteins is monitored by a special class of cytoplasmic chaperones that bind specifically to the Tat signal of their substrate protein masking the signal sequence thereby ensuring proper maturation and co-factor loading. These chaperones, dubbed redox-enzyme maturation proteins (REMP), prevent the futile export of immature protein [28, 29, 30].

In this study, we examined 20 halophilic archaea whose complete proteomes were available in the Universal Protein Resource (UniProt) database. One or two TatA homologs were found in each organism. In contrast, many of the organisms contained two to four TatC homologs. We investigated the implications of the presence of these TatC homologs in these organisms. Experiments have shown that an atypical TatC homolog with fourteen TM domains ($TatC_t$) is present in halophilic archaea [31]. Unlike other organisms, the archaeal cell membrane phospholipid is composed of branched isoprene units linked by ether groups to glycerol. It may have some advantages to possess the novel protein transport components such as $TatC_t$ may be the result of the unusual membrane phospholipid structure. We believe that the variety in TatC homologs would have some significance for the substrates being transported [26]. To assess this, we analyzed the topology, horizontal gene transfer, and substrate diversity for the different TatC homologs. We also analyzed the TatA components with respect to *E. coli* and *A. aeolicus* TatA.

2. Materials and methods

2.1. Distribution of Tat components and topology assignment

All data on Tat pathway components was obtained via a query search in the UniProt database and was tabulated by organism. The amino acid sequences for TatC in these organisms were retrieved and were analyzed with the TMHMM Server v. 2.0 [32], which segregated the TatC homologs into different topologies. Different TM prediction servers were used to reinforce these results: the DAS-TM filter server, the HMMTOP server, and SCAMPI [33, 34, 35].

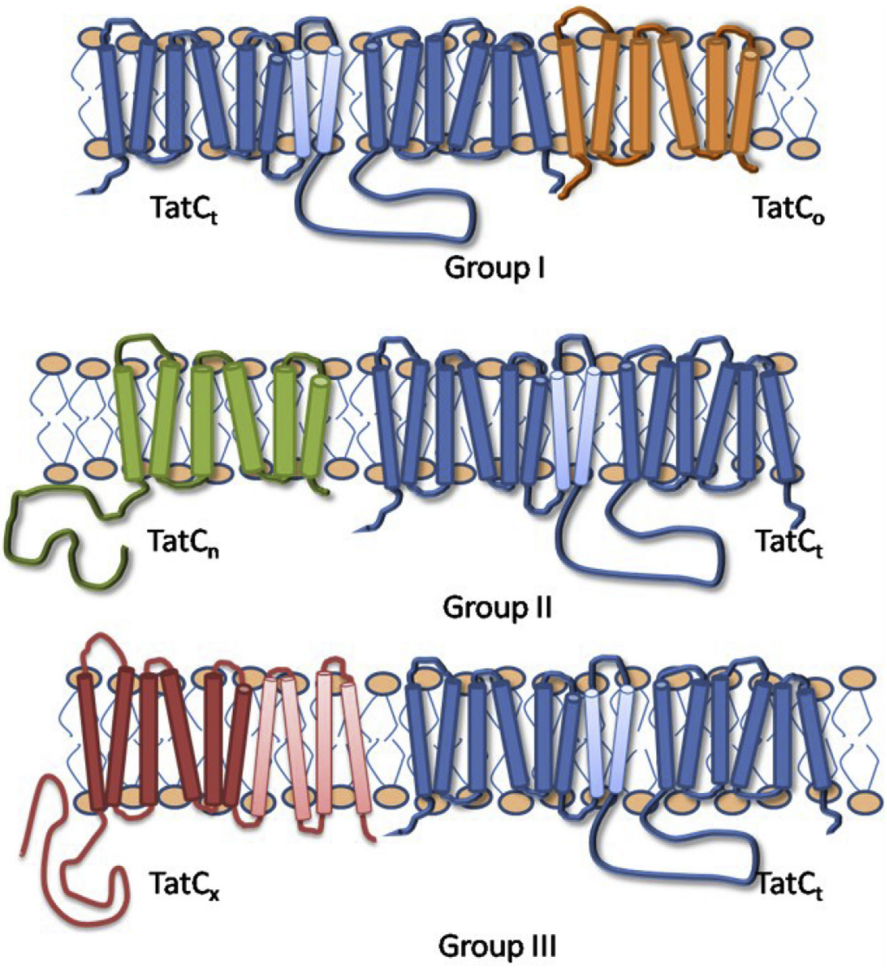

Fig. 1. Schematic representation of different TatC topology and grouped in to different groups based on combinations. $TatC_t + TatC_o$, $TatC_t + TatC_n$ and $TatC_t + TatC_x$ which are denoted in group 1, Group 2 and group 3, respectively.

2.2. Sequence analysis and GC-content analysis

All TatC homologs were aligned using ClustalW and a neighbor-joining tree with a bootstrap value of 10,000 was generated. The sequence analysis was performed by Clustal-Omega-generated multiple sequence alignment and the relative positions of secondary structures were marked according to the TMHMM data. The GC content of the TatC gene along with fifty upstream and fifty downstream genes were plotted as box plots. For comparison of TatC versus genomic GC content, the entire gene chromosome sequence was used to obtain the GC content. The two datasets were compared using a two tailed *z*-test.

2.3. Substrate analysis

The complete proteomes for all twenty haloarchaea were downloaded from the UniProt database. The Tat substrates were identified using the TatFind server. To segregate the substrates, a bidirectional BLAST was performed against each other using standalone BLAST version 2.2.29. An E-value cutoff of 0.5 was imposed; all hits with less than 50% query coverage were removed. From these hits only the bidirectional hits (BETs) were considered for analysis (Fig. S2).

2.4. Signal sequence analysis

Multiple sequence alignment files for the RR signal in each dataset were generated from the TatFind data with ClustalW. TheWebLogo was generated at http://weblogo.berkeley.edu/logo.cgi.

2.5. Pfam domain analysis

The Pfam hidden Markov models (HMMs) present in each substrate were identified through a Pfam batch sequence search. The unique and common HMMs were then identified in the different groups. Venn diagrams were generated with http://bioinfogp.cnb.csic.es/tools/venny/.

2.6. Chaperon identification

Sequences for the known tat related chaperons were obtained from UniProt and blasted against the proteomes of these organisms using standalone BLAST (blast-2.2.29+) with an E-value cutoff of 1. Sequences were blasted and homologs for each of them were obtained. Standard BLASTp output provided pairwise sequence alignment using which the signature sequence conservation was checked.

3. Results and discussion

The archaeal Tat system generally consists of TatA and TatC, but is devoid of TatB. The Tat pathway is present in about half of the Euryarchaeota members and several of the Crenarchaeota species that have been sequenced. The number of substrates transported by the Tat system in archaea is generally comparable to that for bacteria, with many archaea, such as *Sulfolobus tokodaii* and *Archaeoglobus fulgidus*, encoding only a few known Tat substrates, primarily cofactor-containing redox proteins [36, 37]. But, in the case of haloarchaea, a substantially greater number of substrates are translocated via the Tat pathway. Sequenced and annotated data for twenty haloarchaeal species available in UniProt were used for this analysis.

3.1. Membrane topological variants of TatC

The amino acid sequences for TatC in haloarchaeal were retrieved and using TMHMM server v. 2.0 [32] the TMs were assigned. The homologs were segregated into different topologies. This analysis segregated the TatC homologs into four classes TatC with 6-TM helices ($TatC_o$), TatC with 6-TM helices and possessing a long N-terminal cytoplasmic loop ($TatC_n$), TatC with 10-TM helices ($TatC_x$) and TatC with 14-TMhelices ($TatC_t$) based on the major differences in membrane topology, specifically the number of TM helices and the length of the N-terminal cytoplasmic region. Different TM-prediction servers were

Table 1
Analysis of Tat Pathway receptor component and number of substrates present in Haloarchaea that were sequenced and annotated.

Organism	No. of tat substrates	No. oftatC	TatCt	$TatC_x$	$TatC_n$	TatC0	tatC (id)
Halophilic							
Halalkalicoccus jeotgali	103/4212	2	1			1	D8J5Y4 (6),D8J5Y5 (14)
Haloarcula hispanica	122/3860	2	1	1			G0HQX6 (10),G0HQX5 (14)
Haloarcula marismortui	148/4237	2	1	1			Q5UYP5(10),Q5UYP6(14)
Halobacterium salinarum (*strain ATCC 29341*)	72/2578	2	1			1	B0R7G7 (6),B0R7G6 (14)
Halobacterium salinarum (*strain ATCC 700922*)	63/2426	2	1			1	B0R7G7 (6),B0R7G6 (14)
Haloferax mediterranei	171/4712	3	1		2		I3R116 (6L),I3R115 (14),M0IS38(6L)
Haloferax volcanii	143/4813	3	2		2[a]		L9UI27(6L),D4GZD0 (6L),D4GZC9 (14),L9UGM9 (14)
Halogeometricum borinquense	161/4438	4	1		2	1	L9UQ35 (6L),E4NRH6(14),E4NRH7(6L),E4NPB6(6)
Halomicrobium mukohataei	125/3343	2	1	1			C7P1B7(14),C7P1B8(10)
Halopiger xanaduensis	194/4221	2	1		1		F8D3F7 (6L),F8D3F6 (14)
Haloquadratum walsbyi (strain DSM 16790)	64/2546	2	1		1		Q18E63 (6L),Q18E62 (14)
Haloquadratum walsbyi (strain DSM 16854)	59/2638	2	1		1		G0LFQ9 (14),G0LFQ8 (6L)
Halorhabdus utahensis	100/3001	2	1	1			C7NVD1 (10),C7NVD2 (14)
Halorubrum lacusprofundi	98/3497	2	1		1		B9LTY5 (14),B9LTY6 (6L)
Haloterrigena turkmenica	230/5116	2	1		1		D2RTM5 (14),D2RTM6 (6L)
Halovivax ruber	132/3099	2	1		1		L0IGT4 (6L),L0ID04 (14)
Natrialba magadii	228/4936	2	1		1		D3SVL3 (6L),D3SVL2 (14)
Natrinemapellirubrum	171/5026	2	1		1		L0JIB8(6L),L0JEW5 (14)
Natrinema sp.	142/4296	2	1		1		I7CEK6 (6L),I7CPH1(14)
Natronobacterium gregoryi	138/3624	2	1		1		L0AHR8 (6L),L0AFL9 (14)
Natronomonas moolapensis	69/2721	2	1			1	M1XNN2(6),M1XZE0 (14)
Natronomonas pharaonis	106/2764	2	1			1	Q3ISX0 (6),Q3ISW9(14)
Others							
Aquifex aeolicus	15/1553	1				1	O67305 (6)
Acidianus hospitalis	2/2329	2				2	F4B4X0 (6),F4B6B8(6)
Desulfurococcus kamchatkensis	0/1470	0					
Sulfobus ilandicus	3/2661	0					
Archaeoglobus veneficus	7/2065	2				2	F2KP54(6),F2KQ26 (6)
Methanobrevibacter smithii	0/1783	0					
Picrophilus torridus	2/1535	1				1	Q6KZY6 (6)

[a] Same type.

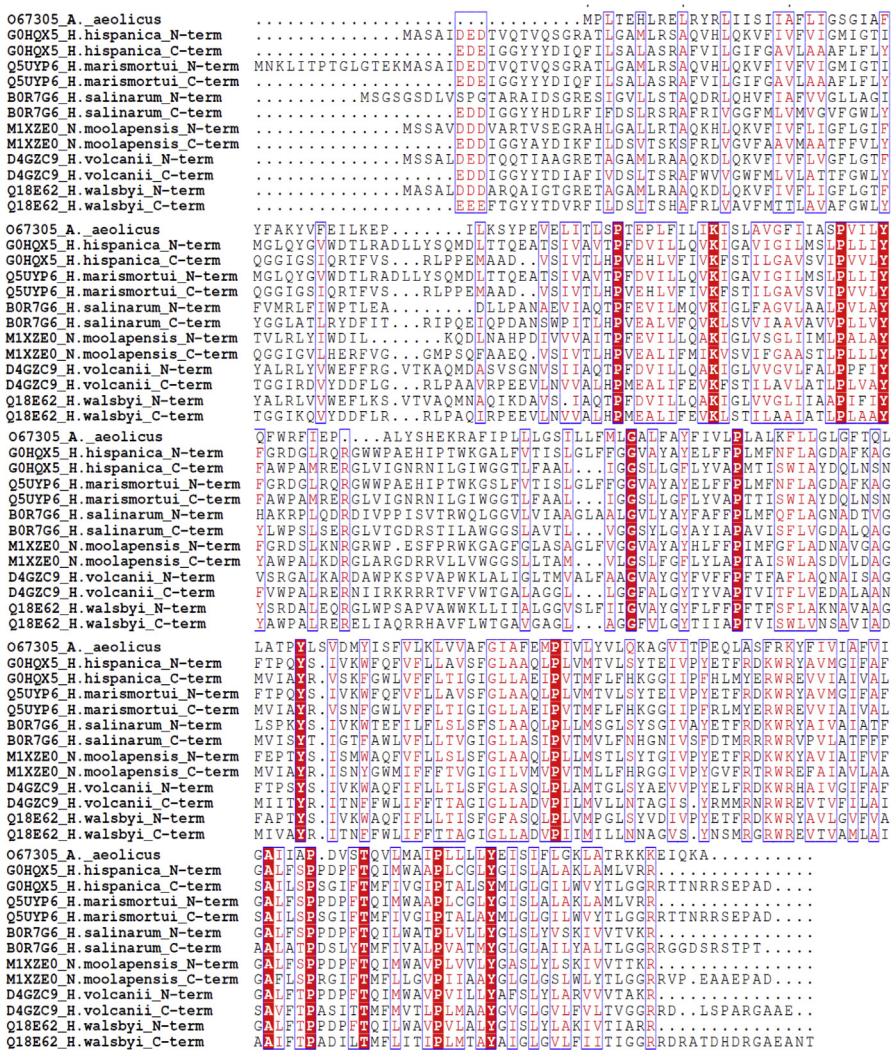

Fig. 2. The alignment between N and C terminal TatC domain of $TatC_t$.

Table 2
Percentage similarity between N-terminal and C-terminal TatC domains of $TatC_t$ of Holoarchaea.

Organism	Percentage similarity
Halalkalicoccus jeotgali_D8J5Y5	20
Haloarcula hispanica_G0HQX5	21
Haloarcula marismortui_Q5UYP6	21
Halobacterium salinarum_B0R7G6	23
Haloferax mediterranei_I3R115	23
Haloferax volcanii_D4GZC9	22
Haloferax volcanii_L9UGM9	26
Halogeometricum borinquense_E4NRH6	22
Halomicrobium mukohataei_C7P1B7	23
Haloplger xanaduensis_F8D3F6	21
Haloquadratum walsbyi_Q18E62	22
Haloquadratum walsbyi_G0LFQ9	22
Halorhabdus utahensis_C7NVD2	22
Halorubrum lacusprofundi_B9LTY5	23
Haloterrigena turkmenica_D2RTM5	22
Halovivax ruber_L0ID04	22
Natrialba magadi_D3SVL2	22
Natrinema pellirubrum_L0JEW5	20
Natrinema sp._I7CPH1	22
Natronobacterium gregoryi_L0AFL9	24
Natronomonas moolapensis_M1XZE0	23
Natronomonas pharaonis_Q3ISW9	25

used to validate these results: the DAS-TM filter server [34], the HMMTOP server [33], and SCAMPI [35]. The results of the analyses from these servers confirmed the results we obtained from the TMHMM server, except in few cases and on an average all the proteins fell into one of the topological categories described above (Fig. 1 and Table 1) [23].

The expected 6-TM topology of TatC has been confirmed through analysis of its crystal structure [20, 21] and the existence of a 14-TM homolog has been demonstrated experimentally [26]. Multiple sequence alignment additionally confirmed that the TM predictions were appropriate (data not shown). The consistent prediction by all tools of a TatC homolog with 10-TM helices strongly suggests the existence of a 10-TM TatC homolog. The existence of this $TatC_x$ topological class is further supported by previous studies in *H. marismortui* [26]. Topology similar to that of the $TatC_n$ category has been commonly observed in TatC from chloroplasts [38].

The $TatC_t$ topological class is noteworthy because it is unique to haloarchaea. Further investigation via BLASTp analysis indicated that $TatC_t$ comprises two putative TatC domains. Multiple sequence analysis and TMHMM data confirmed that the first six TM helices at the N-terminal and the last six helices at the C-terminal had significant homology with the normal TatC domains (Fig. 2 and Table 2). Two TM helices and a large cytoplasmic loop separate these two homologous domains. No homologs could be identified for the middle two TM helices. For $TatC_x$, the first six N-terminal TM helices form a characteristic TatC domain. No sequence homology was observed for the last four TM helices with any of

the sequences. We then examined the distribution of these four TatC topology classes among the haloarchaea species. It was observed that all the species encompasses atleast one copy of $TatC_t$ and combination with $TatC_o$, $TatC_n$ or $TatC_x$ (Table 1). Thus, we defined three main groups (Fig. 1): $TatC_t + TatC_o$ (Group I), $TatC_t + TatC_n$ (Group II) and $TatC_t + TatC_x$ (Group III).

3.2. Sequence and phylogenetic analysis of TatC

A neighbor-joining tree generated from aligned TatC homologs was analyzed for all unique $TatC_t$ homologs, to assess the clustering of organisms in the three main groups outlined above. We found that the organisms with $TatC_t + TatC_x$ and $TatC_t + TatC_o$ were formed the distinct

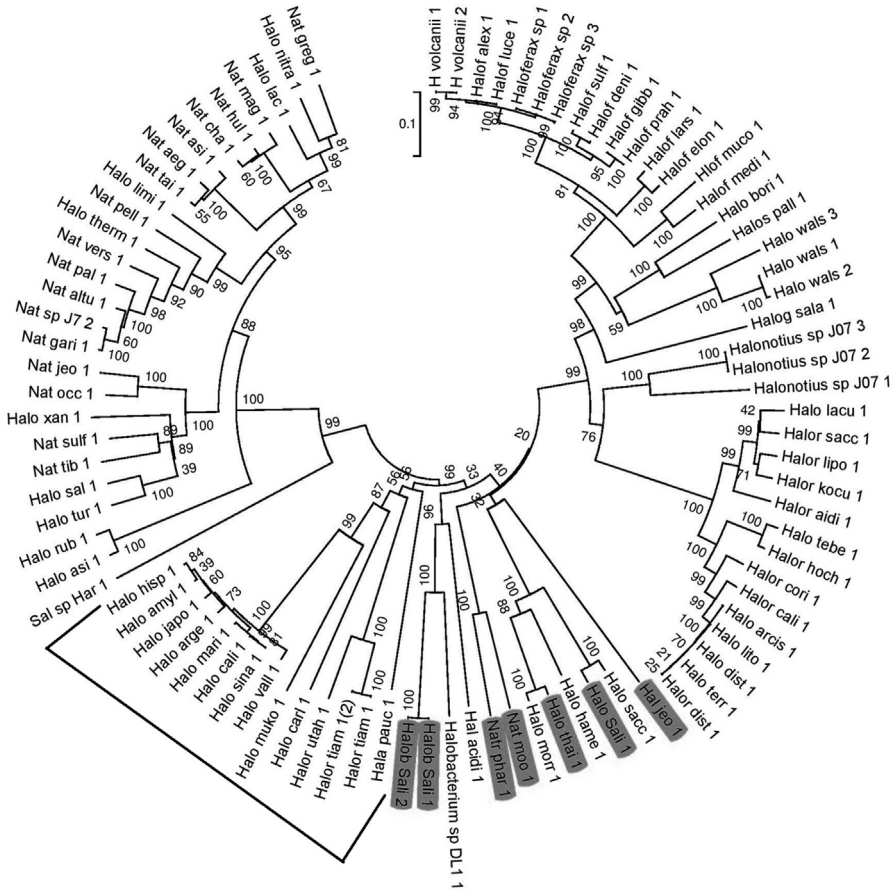

Fig. 3. Phylogenetic analysis of all $TatC_t$ from haloarchaea. Group 3 $TatC_t$ shown in bracket and group 1 members are highlighted; the remaining members from group 2 group.

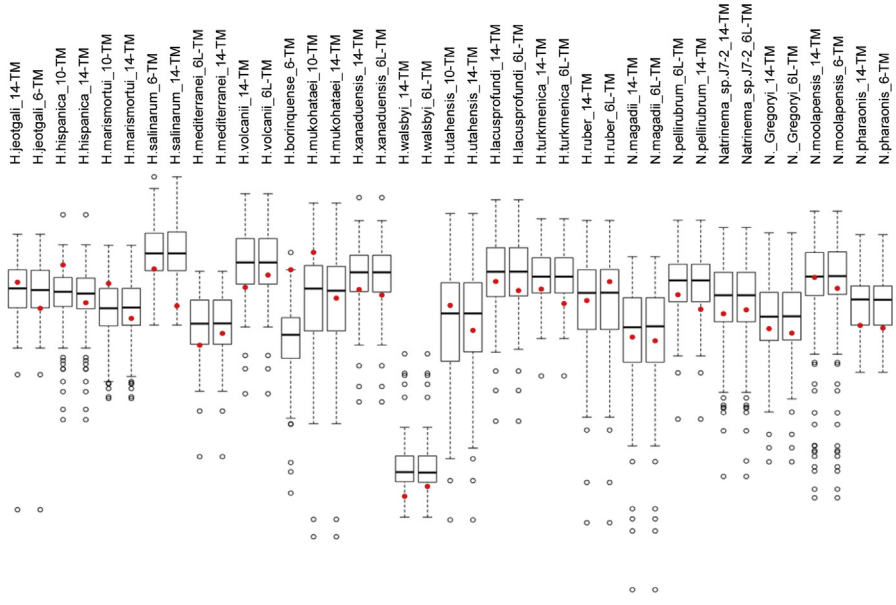

Fig. 4. The whisker plot representation of GC content of fifty genes flanking either side of the TatC (represented as filled dot) of respective organism.

cluster (Fig. 3). The organisms with $TatC_t$ + $TatC_n$ did not cluster together but were widely distributed.

The sequences were further analyzed by multiple sequence alignment against *E. coli* and *A. aeolicus*, for which crystal structure is available, to determine the crucial residues that were actively involved in Tat signal interaction and dimer formation, and explore whether these residues were conserved across sequences. This analysis provides an insight into the probable oligomerization state of the different classes of TatC in archaea (Fig S1 and Table ST1). The TM regions were marked to ensure that the sequences conserved in the analysis were in proximity to the reported regions of the protein. *A. aeolicus* Glu165, which is exposed at the center of the concave face and thus places an ionizable group in the hydrophobic interior of the bilayer [20, 21], was replaced by glutamine and was conserved across $TatC_o$, $TatC_n$, $TatC_x$ and the N-terminal domain of $TatC_t$. However, in the C-terminal $TatC_t$ domain, Glu165 was replaced with asparagine or aspartic acid. This forms hydrogen-bond networks with Ser107 and Tyr85 [39]. Both these residues were conserved across $TatC_o$, $TatC_n$ and $TatC_x$; they were only partially conserved in the two TatC domains of $TatC_t$ - Ser107 in the N-terminal $TatC_t$ domain and Tyr85 in the C-terminal $TatC_t$ domain. This evidence indicates that the difference in membrane properties in the haloarchaea Tat pathway might be due to the conservation of glutamine instead of glutamic acid. In addition, the differential conservation of the three residues Glu165, Ser107, and Tyr85 in the two Tat domains of $TatC_t$ might enable complementary functioning of the two domains for active transport.

Glu96 is another critical residue for Tat transport. This residue was conserved across $TatC_o$, $TatC_n$, $TatC_x$ and the C-terminal Tat domain of $TatC_t$. A similar distribution was found even in case of the residues involved in signal binding in *Aquifex* [40]. The Pro42, Phe87, Pro90 and Leu92 residues were conserved in $TatC_o$, $TatC_n$, $TatC_x$ and C-terminal Tat domain of $TatC_t$ but only Pro90 was conserved in $TatC_t$. Pro57 and Pro48 were conserved in the N-terminal Tat domain of $TatC_t$ instead of Pro42. We have also investigated the residues involved in TatC dimerization and interaction with other Tat pathway membrane components. While the Asp205 was conserved across all TatC classes except for the C-terminal TatC domain of $TatC_t$, the Tyr30 residue was not conserved in $TatC_o$, $TatC_n$ or the N-terminal TatC domain of $TatC_t$.

We further surveyed GC-content for these TatC homologs in order to test the possibility of horizontal gene transfer to haloarchaea. The GC content of the TatC gene and 50 upstream and 50 downstream genes were plotted. For comparison of TatC versus genomic GC content, the entire gene chromosome sequence was used to obtain the GC content. The two datasets were compared using a two tailed z-test. It is clear from Fig. 4 that there is no horizontal gene transfer. No significant difference was observed between the individual GC content versus the genomic GC content for any of the TatC homologs ($z = 1.044$; two-tailed $p = 0.2965$), thereby indicating that these TatC gene homologs may have evolved within the haloarchaea and consequently the presence of a correlation between the different TatC homologs and the types of substrate transported by them.

3.3. Haloarchaeal TatA

The size of TatA components in the selected Haloarchaea ranges from 75 to 145 amino acids. Phylogenetic analysis of unique TatA sequences (aligned with *E. coli* TatA and TatE and *A. aeolicus* TatA1 and TatA2) established that unlike $TatC_t$, the TatA do not segregate according to the groups based on TatC combinations (Fig. 5). The alignment of *E. coli* TatA structure with haloarchaeal showed that the first TM region and the amphipathic helix region is highly conserved across all TatA components, but there was very little similarity in the n-terminal region (Fig. S2). TM residue Gln8 of *E. coli* TatA points inward, resulting in a short hydrophobic pore in the center of the complex. Different views were proposed based on simulations of the TatA complex in lipid bilayers indicate that the short TM domain distorts the membrane [41]. This residue is replaced with Glu in all haloarchaea. The *E. coli* Gly21 remains conserved in haloarchaea that keeps TM helix and amphipathic helix at right angles except in the case of TatA1 of *Halomicrobiu mmukohataei* (UniProt id

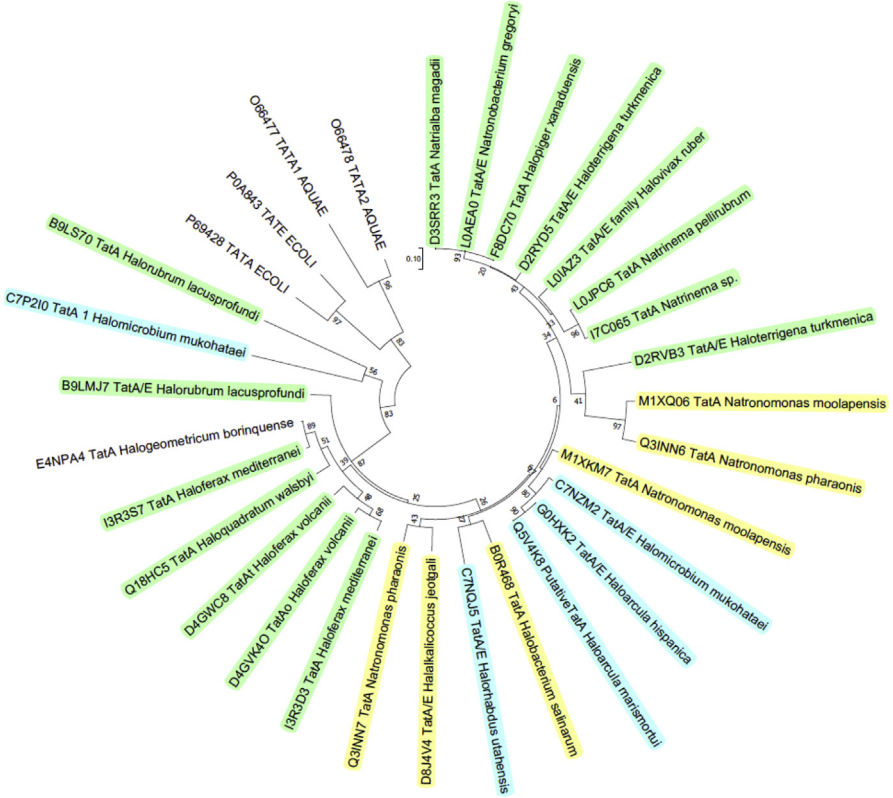

Fig. 5. Phylogenetic analysis of TatA/E from 20 haloarcheal species studies. Group 1 members are highlighted in yellow, group 2 in green and group 3 in blue.

C7P2I0). The movement of C-terminal portion of the amphipathic helix is considered important for TatA function and Phe39, mutation (F39A) of which causes TatA inactivation [42], is conserved across haloarchaea. *Halorubrum lacusprofundi* TatA (UniProt id B9LS70) is an exception which has a non-aromatic amino acid leucine. Phylogenetic analysis also showed that these two TatA homologs (C7P2I0 and B9LS70) segregate out, which may suggest different substrate interaction or inactive TatA homologs as these organisms contain another TatA homolog. Presence of a predicted coiled coil motif in the C-terminal of *Natronomonas moolapensis* TatA homolog (UniProt id M1XKM7) might play additional function in recognition.

3.4. Tat substrate distribution among the different groups of haloarchaea

As discussed above, the three groups of species had $TatC_t$ in common; thus we hypothesized that the Tat substrates in each group would be translocated either by $TatC_t$ or by $TatC_o$, $TatC_n$ or $TatC_x$. The Tat substrates were identified from proteome of all twenty haloarchaea using the TATFIND 1.4 server [6]. To segregate the substrates, a bidirectional BLAST was performed against each other using standalone BLAST version 2.2.29. From these hits, only the bidirectional hits (BETs) were considered for analysis. The BETs were considered as shared sets and the substrates which had no hits were termed as unique sets. Furthermore, the substrates translocated by $TatC_t$ might exhibit common characteristics across all three groups in terms of signal and functional domains. We conducted extensive sequence analysis of the Tat substrates for these haloarchaea to check for any unique features exhibited by each group of substrates: those potentially transported by $TatC_t$ and those transported by the other TatC homologs [26]. From a total of 1,287,495 proteins, 2670 substrates were identified for the Group II ($TatC_t$ + $TatC_n$), Group III ($TatC_t$ + $TatC_x$), and Group I datasets ($TatC_t$ + $TatC_o$) (Fig. S3).

3.5. Tat signal motif analysis

The Tat signal motif consists of three basic domains: a positively charged region at the N-terminal, a hydrophobic core and a more polar region that contains the cleavage site for a signal peptidase [6, 36, 43, 44]. The TATFIND server identifies Tat substrates and also provides information about the N-terminal signature RR signal and the middle hydrophobic region. The canonical RR signal is the region, which directly interacts with the TatC receptor [45, 46]; thorough analysis of this region was conducted for all Tat substrates in our dataset.

It was clear that the phenylalanine in the fifth position of the signal sequence was highly conserved in almost all the Tat substrates examined

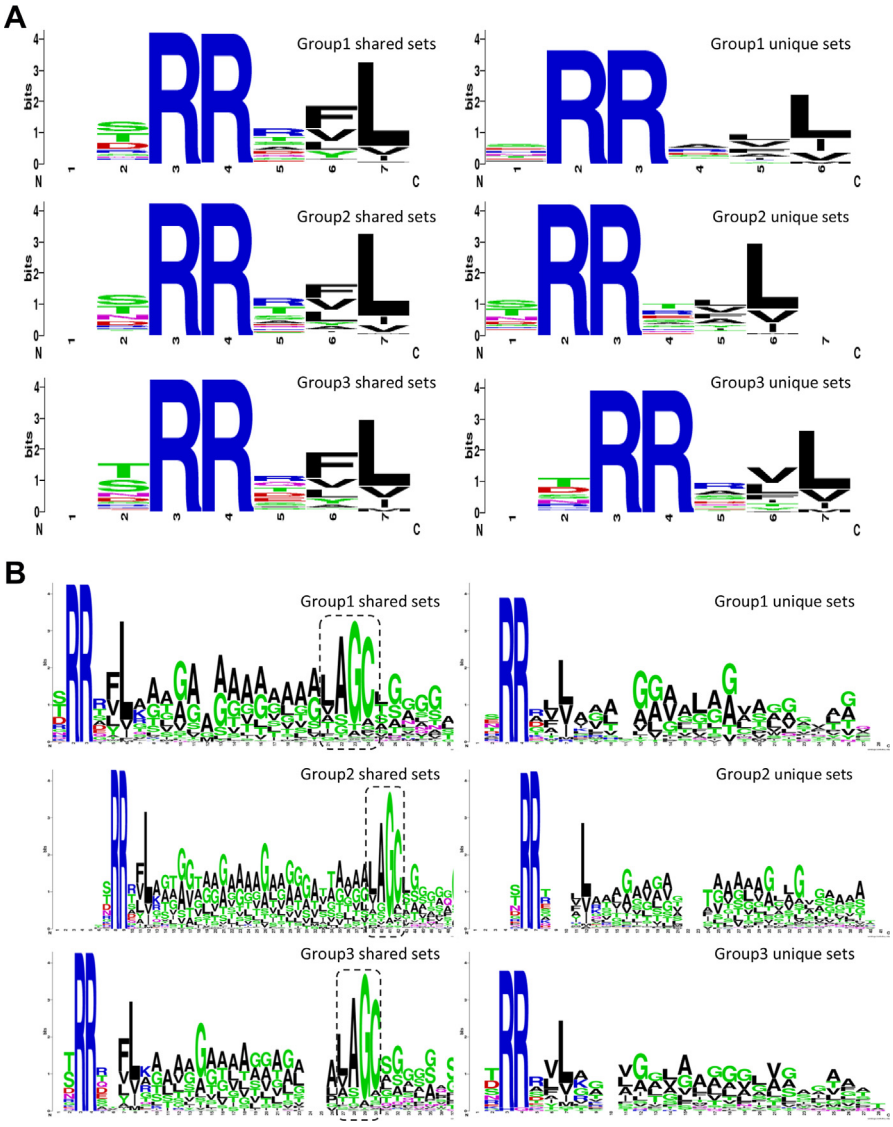

Fig. 6. a) Weblogo representation of Tat signal motifs in each class of substrates. b) Weblogo representation of Tat signal peptide regions from different classes of substrates showing the occurrence of lipobox motif (AGC) in the shared substrates. WebLogo was generated at http://weblogo.berkeley.edu/logo.cgi.

(Fig. 6a), however, there were also differences in the mean hydrophobicity (data not shown). These results indicate that the signal region clearly plays a role in recognition, but that recognition is not solely due to either the RR signal or the hydrophobicity score of the signal region alone. The multiple sequence alignments of fifty amino acids of the N-terminal clearly showed a high occurrence of AGC sequences at the C-terminal of the hydrophobic region in a shared set of Tat substrates, but this was not observed in a unique set of substrates. This AGC sequence forms a distinctive recognition site called the 'lipobox' [31] and was depicted via the WebLogos (Fig. 6b). In the mature lipoprotein, this cysteine residue is attached to a membrane-associated lipid anchor. Analysis using the TatLipo server [7] confirmed that most of the substrates in the shared set were lipobox positive and were probably lipoproteins. This feature therefore clearly distinguishes between the shared (exported by more than one class of TatC) substrate datasets and the unique (exported by one type of TatC) substrate datasets. It is also probable that $TatC_t$ mainly transports proteins from the shared set, which are probably lipoproteins, whereas the other TatC homolog classes may transport the remaining unique set of substrates, which may be secreted proteins.

Table 3
Distribution of PFAM families.

Datasets	Number of pfam families	Number of proteins with pfam families predicted
Group 1 shared sets	109	231
Group 1 unique sets	47	29
Group 2 shared sets	237	555
Group 2 unique sets	208	144
Group 3 shared sets	136	276
Group 3 unique sets	47	31

3.6. Functional analysis of the Tat pathway of haloarchaea

We assessed the substrates in our dataset for specific functional roles, but in haloarchaea only a few have been experimentally tested and most are uncharacterized. We analyzed each substrate using the Pfam database to identify the different domains, repeats and signature motifs. The unique and common HMMs were identified in the different Tat groups (Table 3). A total of 891 unique families were identified among the 2670 haloarchaea Tat substrates. From the analysis, it can be inferred that there is more domain diversity across the unique set and that the shared set has many domains in common (Fig. 7 and Table 3).

The Pfam families for the shared set of substrates were then compared across datasets. Fifty eight families were common to all the groups for the shared set, but only two families were common to all the groups for the unique set (Fig. 7). To assess the significance of the shared and unique sets of families, the distribution of these domains across the groups were analyzed by re-mapping the list of Pfam families to the substrates in each set and then mapping all the families present in these proteins. Although there were many unique Pfam families in both sets, most of the shared substrates contained one or more common Pfam families.

The 58 frequently occurring Pfam families found in all the shared sets were mostly metal-binding domains, periplasmic substrate-binding

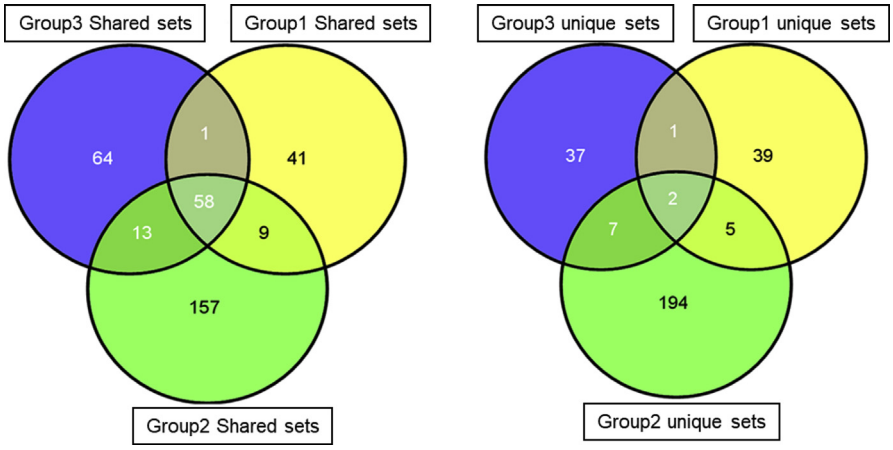

Fig. 7. Venn diagrammatic representation of the Pfam family distribution in shared and unique group of substrates. Venn diagrams were generated with http://bioinfogp.cnb.csic.es/tools/venny/.

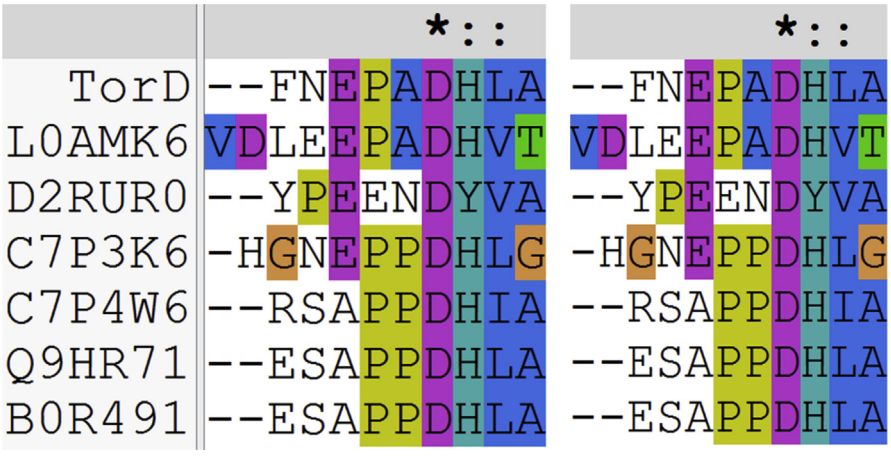

Fig. 8. REMP motifs *E. coli* TorD aligned to identified archaeal TorD homologs.

domains that are responsible for ion transport, glycosidases, dehydrogenases and domains responsible for the biosynthesis of pyrimidine and trypsin-2. There were also four domains of unknown function and four from PfamB. It was quite clear that the proteins in the shared substrates set were mostly membrane-associated proteins, which correlates with the presence of lipobox in the signal region. The unique substrates set generally had lyase or hydrolase-type domains. Interestingly, a group of unique substrates in the group II were identified as 7TM-GPCRs, PAZ domains, Cox II, and a few metal-binding domains. However, most of the Tat substrates present in the unique substrates set were secreted proteins.

3.7. Protein-folding quality control

The quality-control mechanism for substrates transported by the Tat machinery is well established. In the bacterial system, this process is taken care of by chaperones that referred as REMP [47], which mask the twin-arginine signal and ensure proper folding or substrate maturation with appropriate cofactor loading [48]. In the case of halophilic archaea, the presence of such a control mechanism may be even more vital because the Tat pathway transports almost the entire secretome. We therefore identified chaperone homologs from the proteomes of these organisms. Homologs for *E. coli* REMP such as DmsD, HyaE, HybE, NapD, NarJ, NarW, TorD and YcdY were identified. Pairwise sequence analysis showed that the key sequence motif shown in Fig. 8, Fig. S4 and Table ST2 was partially conserved and the protein sizes were comparable in most of the cases, indicating that the haloarchaeal Tat pathway uses chaperones similar to their bacterial counterparts.

4. Conclusions

Overall, these analyses provide insight into the diversity of the indispensable Tat pathway in the haloarchaeal system. The evolution of the Tat pathway itself suggests that there may be a fundamental conflict between the substrates transported and the Sec export mechanism in organisms such as haloarchaea that live in extreme habitats. The diversity and multiple unique topologies of the TatC receptor in haloarchaea indicate that they may exist specifically for transport of large numbers and a wide variety of Tat substrates. Conservation in the N-terminal region of TatA homologs to that of *E. coli* implies similar mechanism of transport but diversity in the C-terminal region could be useful in interaction with the wide range of substrates. This is a requirement for haloarchaeal proteins; they would otherwise aggregate after exiting the ribosome because haloarchaea have a high intracellular concentration of positive ions in order to maintain osmotic balance against the high extracellular sodium concentration. The quality of protein folding is a major concern for these organisms, and is positively facilitated by the sophisticated Tat pathway components. Unlike in other systems, many substrate groups are present here based on the class of receptor required for the translocation. This may be necessary for differential translocation dynamics and to ensure the stringent folding quality. Further *invivo* and *invitro* studies are required for dissecting the dynamics of the Tat pathway in haloarchaea.

Declarations

Author contribution statement

Deepanjan Ghosh: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Debjyoti Boral: Performed the experiments; Analyzed and interpreted the data.

Koteswara Rao Vankudoth: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Sureshkumar Ramasamy: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

Deepanjan Ghosh was supported by the Council of Scientific and Industrial Research (CSIR) New Delhi with a Senior Research Fellowship. Debjyoti Boral is supported by University Grant Commision (UGC), India.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2019.e01587.

References

[1] T. Palmer, B.C. Berks, The twin-arginine translocation (Tat) protein export pathway, Nat. Rev. Microbiol. 10 (2012) 483–496.
[2] B.C. Berks, F. Sargent, T. Palmer, The Tat protein export pathway, Mol. Microbiol. 35 (2000) 260–274.
[3] B.C. Berks, T. Palmer, F. Sargent, Protein targeting by the bacterial twin-arginine translocation (Tat) pathway, Curr. Opin. Microbiol. 8 (2005) 174–181.
[4] G.A. Sutherland, K.J. Grayson, N.B.P. Adams, D.M.J. Mermans, A.S. Jones, A.J. Robertson, D.B. Auman, A.A. Brindley, F. Sterpone, P. Tuffery, P. Derreumaux, P.L. Dutton, C. Robinson, A. Hitchcock, C.N. Hunter, Probing the quality control mechanism of the *Escherichia coli* twin-arginine translocase with folding variants of a de novo-designed heme protein, J. Biol. Chem. 293 (2018) 6672–6681.
[5] E. Park, T.A. Rapoport, Mechanisms of Sec61/SecY-mediated protein translocation across membranes, Annu. Rev. Biophys. 41 (2012) 21–40.
[6] R. Rose, T. Brüser, J.C. Kissinger, M. Pohlschröder, Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway, Mol. Microbiol. 45 (2002) 943–950.
[7] S. Stefanie, P. Friedhelm, D. Kieran, C.Z. Qiang, I. Saheed, P. Mechthild, Mutational and bioinformatic analysis of haloarchaeal lipobox-containing proteins, Archaea (2010).
[8] K. Dilks, R.W. Rose, E. Hartmann, M. Pohlschröder, Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey, J. Bacteriol. 185 (2003) 1478–1483.
[9] A. Bolhuis, D. Kwan, J.R. Thomas, Halophilic adaptations of proteins, Protein Adapt. Extremophiles (2008) 71–104.
[10] F. Alcock, P.J. Stansfeld, H. Basit, J. Habersetzer, M.A. Baker, T. Palmer, M.I. Wallace, B.C. Berks, Assembling the Tat protein translocase, eLife 5 (2016).
[11] O.V. Nolandt, T.H. Walther, S. Roth, J. Burck, A.S. Ulrich, Structure analysis of the membrane protein TatC(d) from the Tat system of *B. subtilis* by circular dichroism, Biochim. Biophys. Acta 1788 (2009) 2238–2244.
[12] F. Sargent, E.G. Bogsch, N.R. Stanley, M. Wexler, C. Robinson, B.C. Berks, T. Palmer, Overlapping functions of components of a bacterial Sec-independent protein export pathway, EMBO J. 17 (1998) 3640–3650.
[13] J.H. Weiner, P.T. Bilous, G.M. Shaw, S.P. Lubitz, L. Frost, G.H. Thomas, J.A. Cole, R.J. Turner, A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins, Cell 93 (1998) 93–101.
[14] E.G. Bogsch, F. Sargent, N.R. Stanley, B.C. Berks, C. Robinson, T. Palmer, An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria, J. Biol. Chem. 273 (1998) 18003–18006.
[15] J.D. Jongbloed, U. Martin, H. Antelmann, M. Hecker, H. Tjalsma, G. Venema, S. Bron, J.M. van Dijl, J. Muller, TatC is a specificity determinant for protein secretion via the twin-arginine translocation pathway, J. Biol. Chem. 275 (2000) 41350–41357.
[16] Q. Huang, T. Palmer, Signal peptide hydrophobicity modulates interaction with the twin-arginine translocase, mBio 8 (2017).
[17] D. Mangels, J. Mathers, A. Bolhuis, C. Robinson, The core TatABC complex of the twin-arginine translocase in Escherichia coli: TatC drives assembly whereas TatA is essential for stability, J. Mol. Biol. 345 (2005) 415–423.
[18] S. Molik, I. Karnauchov, C. Weidlich, R.G. Herrmann, R.B. Klösgen, The Rieske Fe/S protein of the cytochromeb 6/f complex in chloroplasts missing link in the evolution of protein transport pathways in chloroplasts? J. Biol. Chem. 276 (2001) 42761–42766.
[19] C.P. New, Q. Ma, C. Dabney-Smith, Routing of thylakoid lumen proteins by the chloroplast twin arginine transport pathway, Photosynth. Res. (2018).
[20] S. Ramasamy, R. Abrol, C.J. Suloway, W.M. Clemons Jr., The glove-like structure of the conserved membrane protein TatC provides insight into signal sequence recognition in twin-arginine translocation, Structure 21 (2013) 777–788.
[21] S.E. Rollauer, M.J. Tarry, J.E. Graham, M. Jääskeläinen, F. Jäger, S. Johnson, M. Krehenbrink, S.-M. Liu, M.J. Lukey, J. Marcoux, Structure of the TatC core of the twin-arginine protein transport system, Nature 492 (2012) 210–214.
[22] C. Punginelli, B. Maldonado, S. Grahl, R. Jack, M. Alami, J. Schröder, B.C. Berks, T. Palmer, Cysteine scanning mutagenesis and topological mapping of the *Escherichia coli* twin-arginine translocase TatC component, J. Bacteriol. 189 (2007) 5482–5494.

[23] J. Behrendt, K. Standar, U. Lindenstrauß, T. Brüser, Topological studies on the twin-arginine translocase component TatC, FEMS Microbiol. Lett. 234 (2004) 303–308.
[24] Y. Hu, E. Zhao, H. Li, B. Xia, C. Jin, Solution NMR structure of the TatA component of the twin-arginine protein transport system from gram-positive bacterium *Bacillus subtilis*, J. Am. Chem. Soc. 132 (2010) 15942–15944.
[25] H. Shruthi, P. Anand, V. Murugan, K. Sankaran, Twin arginine translocase pathway and fast-folding lipoprotein biosynthesis in *E. coli*: interesting implications and applications, Mol. Biosyst. 6 (2010) 999–1007.
[26] K. Dilks, M.I. Giménez, M. Pohlschröder, Genetic and biochemical analysis of the twin-arginine translocation pathway in halophilic archaea, J. Bacteriol. 187 (2005) 8104–8113.
[27] M. Moser, S. Panahandeh, E. Holzapfel, M. Muller, In vitro analysis of the bacterial twin-arginine-dependent protein export, Methods Mol. Biol. 390 (2007) 63–79.
[28] R.J. Turner, A.L. Papish, F. Sargent, Sequence analysis of bacterial redox enzyme maturation proteins (REMPs), Can. J. Microbiol. 50 (2004) 225–238.
[29] S.K. Ramasamy, W.M. Clemons Jr., Structure of the twin-arginine signal-binding protein DmsD from *Escherichia coli*, Acta Crystallogr. Section F Struct. Biol. Cryst. Commun. 65 (2009) 746–750.
[30] A.L. Papish, C.L. Ladner, R.J. Turner, The twin-arginine leader-binding protein, DmsD, interacts with the TatB and TatC subunits of the *Escherichia coli* twin-arginine translocase, J. Biol. Chem. 278 (2003) 32501–32506.
[31] A. Bolhuis, Protein transport in the halophilic archaeon *Halobacterium* sp. NRC-1: a major role for the twin-arginine translocation pathway? Microbiology 148 (2002) 3335–3346.
[32] A. Krogh, B. Larsson, G. Von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J. Mol. Biol. 305 (2001) 567–580.
[33] G.E. Tusnady, I. Simon, The HMMTOP transmembrane topology prediction server, Bioinformatics 17 (2001) 849–850.
[34] M. Cserzö, F. Eisenhaber, B. Eisenhaber, I. Simon, On filtering false positive transmembrane protein predictions, Protein Eng. 15 (2002) 745–752.
[35] A. Bernsel, H. Viklund, J. Falk, E. Lindahl, G. von Heijne, A. Elofsson, Prediction of membrane-protein topology from first principles, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 7177–7181.
[36] M. Pohlschröder, M.I. Giménez, K.F. Jarrell, Protein transport in Archaea: sec and twin arginine translocation pathways, Curr. Opin. Microbiol. 8 (2005) 713–719.
[37] D. Kwan, A. Bolhuis, Analysis of the twin-arginine motif of a haloarchaeal Tat substrate, FEMS Microbiol. Lett. 308 (2010) 138–143.
[38] H. Mori, E.J. Summer, K. Cline, Chloroplast TatC plays a direct role in thylakoid (Delta)pH-dependent protein transport, FEBS Lett. 501 (2001) 65–68.
[39] E. Holzapfel, G. Eisner, M. Alami, C.M. Barrett, G. Buchanan, I. Lüke, J.-M. Betton, C. Robinson, T. Palmer, M. Moser, The entire N-terminal half of TatC is involved in twin-arginine precursor binding, Biochemistry 46 (2007) 2892–2898.
[40] G. Buchanan, E. d. Leeuw, N.R. Stanley, M. Wexler, B.C. Berks, F. Sargent, T. Palmer, Functional complexity of the twin-arginine translocase TatC component revealed by site-directed mutagenesis, Mol. Microbiol. 43 (2002) 1457–1470.
[41] F. Rodriguez, S.L. Rouse, C.E. Tait, J. Harmer, A. De Riso, C.R. Timmel, M.S. Sansom, B.C. Berks, J.R. Schnell, Structural model for the protein-translocating element of the twin-arginine transport system, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) E1092–E1101.
[42] M.G. Hicks, P.A. Lee, G. Georgiou, B.C. Berks, T. Palmer, Positive selection for loss-of-function tat mutations identifies critical residues required for TatA activity, J. Bacteriol. 187 (2005) 2920–2925.
[43] B.C. Berks, A common export pathway for proteins binding complex redox cofactors? Mol. Microbiol. 22 (1996) 393–404.
[44] A.M. Chaddock, A. Mant, I. Karnauchov, S. Brink, R.G. Herrmann, R. Klösgen, C. Robinson, 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 (1995) 2715.
[45] M. Alami, I. Lüke, S. Deitermann, G. Eisner, H.-G. Koch, J. Brunner, M. Müller, Differential interactions between a twin-arginine signal peptide and its translocase in *Escherichia coli*, Mol. Cell 12 (2003) 937–946.
[46] X. Ma, K. Cline, Mapping the signal peptide binding and oligomer contact sites of the core subunit of the pea twin arginine protein translocase, Plant Cell 25 (2013) 999–1015.
[47] C.S. Chan, L. Chang, K.L. Rommens, R.J. Turner, Differential interactions between Tat-specific redox enzyme peptides and their chaperones, J. Bacteriol. 191 (2009) 2091–2101.
[48] C.M. Stevens, T.M. Winstone, R.J. Turner, M. Paetzel, Structural analysis of a monomeric form of the twin-arginine leader peptide binding chaperone *Escherichia coli* DmsD, J. Mol. Biol. 389 (2009) 124–133.