Arch Microbiol (2002) 177 : 441–450
DOI 10.1007/s00203-002-0408-4

MINI-REVIEW

Ming-Ren Yen · Yi-Hsiung Tseng · Erin H. Nguyen
Long-Fe Wu · Milton H. Saier Jr.

Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system

Received: 8 October 2001 / Revised: 30 January 2002 / Accepted: 4 February 2002 / Published online: 21 March 2002

Abstract Twin-arginine targeting (Tat) protein secretion systems consist of two protein types, members of the TatA and TatC families. Homologues of these proteins are found in many archaea, bacteria, chloroplasts and mitochondria. Every prokaryotic organism with a fully sequenced genome exhibits either neither family member, or between one and three paralogues of these two family members. The *Arabidopsis thaliana* genome encodes three of each. Although many mitochondrially encoded TatC homologues have been identified, corresponding TatA homologues have not been found in this organelle. Phylogenetic analyses reveal that most prokaryotic Tat systems consist of one TatC homologue and two sequence-divergent TatA homologues (TatA and TatB). When only one TatA homologue is present, TatB is missing, and when three TatA homologues are present, the third one arose by duplication of TatA, not TatB. Further, homologues most resembling TatB are more sequence-divergent than those more closely resembling TatA. In contrast to the TatA family, the TatC family shows phylogenetic clustering in strict accordance with organismal type. These results are discussed in terms of their probable structural, functional and evolutionary significance.

Keywords Twin-arginine targeting · Protein secretion · Evolution · Phylogeny

M.-R. Yen · E.H. Nguyen · M.H. Saier Jr. (✉)
Department of Biology, University of California at San Diego,
La Jolla, CA 92093–0116 USA
e-mail: msaier@ucsd.edu,
Tel.: +1-858-5344084, Fax: +1-858-5347108

M.-R. Yen · Y.-H. Tseng
Institute of Molecular Biology, National Chung Hsing University,
Taichung, 402, Taiwan, Republic of China

L.-F. Wu
Laboratoire de Chimie Bactérienne, UPR9043 CNRS,
Institut de Biologie Structurale et Microbiologie,
31 chemin Joseph Aiguier, 13402 Marseille cedex 20, France

Introduction

Numerous protein secretion systems have been identified in prokaryotes, and a few of these are found in eukaryotes as well (Saier 2000a, b). One of the more recently discovered systems is the so-called *t*win-*a*rginine *t*argeting/*t*ranslocation (Tat) system (Berks 1996; Bogsch et al. 1998; Santini et al. 1998; Sargent et al. 1998), also referred to as the *m*embrane *t*argeting and *t*ranslocation (Mtt) system (Weiner et al. 1998). In *Escherichia coli*, the Tat system consists of four proteins, TatA, B, C and E (Wexler et al. 2000). Of these, the first three are encoded within a constitutively expressed operon, while TatE is encoded elsewhere on the bacterial chromosome (Berks et al. 2000a,b; Jack et al. 2001). The Tat system is structurally and mechanistically similar to the chloroplast, ΔpH-driven, thylakoidal protein import pathway (Dalbey and Robinson 1999; Keegstra and Cline 1999; Settles and Martienssen 1998). The signal peptides of the proteins translocated by the Tat/ΔpH pathway contain a conserved twin-arginine motif; these signal peptides are functionally interchangeable between the bacterial and chloroplast systems (Halbig et al. 1999; Mori and Cline 1998; Settles and Martienssen 1998; Voordouw 2000; Wexler et al. 1998). Complexes containing TatA-TatB (Sargent et al. 2001) or TatB-TatC (Bolhuis et al. 2001) have recently been purified from *E. coli*, and a functional Tat system has been reconstituted in vitro from cells overproducing TatABC (Yahr and Wickner 2001). Several models have been presented for the operation of these systems. Each has suggested that a different component serves as the precursor receptor, i.e., TatB and TatA because of their receptor-like topology and their different involvement in the translocation of various proteins (Chanal et al. 1999; Settles et al. 1997), and TatC because of its high degree of conservation among different organisms (Berks et al. 2000a,b; Jongbloed et al. 2000).

In both *E. coli* and the thylakoid membrane of chloroplasts, Tat-dependent protein transport is energized by the proton motive force (pmf) rather than ATP in a process

Table 1 Proteins of the TatC family

Abbreviation	Organism	Size	GenBank index number
Archaea			
Ape	*Aeropyrum pernix*	255	7516974
Afu1	*Archaeoglobus fulgidus*	250	11499107
Afu2	*Archaeoglobus fulgidus*	292	11498942
Hsp1	*Halobacterium* sp.	344	10581680
Hsp2	*Halobacterium* sp.	724	10581679
Sso1	*Sulfolobus solfataricus*	294	13813641
Sso2	*Sulfolobus solfataricus*	289	13816528
Tac	*Thermoplasma acidophilum*	254	10640535
Tvo	*Thermoplasma volcanium*	272	13541201
Gram-negative bacteria			
Aae	*Aquifex aeolicus*	240	6226418
Ach	*Azotobacter chroococcum*	255	9988063
Cje	*Campylobacter jejuni*	245	9979039
Ccr	*Caulobacter crescentus*	300	13423470
TatC Eco	*Escherichia coli*	258	2851441
Hin	*Haemophilus influenzae*	256	1176320
Hpy	*Helicobacter pylori*	253	9978991
Lpn	*Legionella pneumophila*	238	13277315
Mlo	*Mesorhizobium loti*	291	13471181
Nme	*Neisseria meningitidis*	256	11279151
Pmu	*Pasteurella multocida*	245	12722099
Pae	*Pseudomonas aeruginosa*	267	11352718
Pst	*Pseudomonas stutzeri*	267	12831976
Rpr	*Rickettsia prowazekii*	251	9979045
Sty	*Salmonella typhimurium*	259	6960228
Syn-0	*Synechococcus* sp.	246	4590513
Syn-y	*Synechocystis* sp.	254	1723125
Vch	*Vibrio cholerae*	250	11279148
Xfa	*Xylella fastidiosa*	246	11279149
Gram-positive bacteria			
Bha1	Bacillus halodurans	253	9979492
Bha2	*Bacillus halodurans*	219	10176531
Bsu1	*Bacillus subtilis*	254	2811036
Bsu2	*Bacillus subtilis*	245	3183575
Dra	*Deinococcus radiodurans*	270	7473667
Mle	*Mycobacterium leprae*	310	12644313
Mtu	*Mycobacterium tuberculosis*	308	1731351
Rer	*Rhodococcus erythropolis*	55	1666186
Sau	Staphylococcus aureus	220	13700260
Sco	*Streptomyces coelicolor*	316	6119674
Sli	*Streptomyces lividans*	301	11323290
Chloroplasts			
Ath1	*Arabidopsis thaliana*	340	11279150
Cca	*Cyanidium caldarium* (red alga)	239	11465415
Gth	*Guillardia theta* (cryptomonad alga)	290	11467698
Osi	*Odontella sinensis* (centric diatom; alga)	263	11467559
Osa	*Oryza sativa*	359	13442818
Ppu1	*Porphyra purpurea* (red alga)	254	11465730
Mitochondria			
Ath2	*Arabidopsis thaliana*	277	3980178
Ath3	*Arabidopsis thaliana*	277	625980
Bvu	*Beta vulgaris* (sugar beet)	276	9838436
Ccri	*Chondrus crispus* (red alga)	262	7024427
Csy	*Chrysodidymus synuroideus* (golden alga)	241	11466478
Cme	*Cyanidioschyzon merolae* (red alga)	267	8954387

Table 1 (continued)

Abbreviation	Organism	Size	GenBank index number
Mpo	*Marchantia polymorpha* (liverwort)	244	586764
Nol	*Nephroselmis olivacea* (green alga)	247	11135954
Obe	*Oenothera berteriana* (flowering plant)	295	287897
Ovi	*Oenothera villaricae* (flowering plant)	295	1076558
Ppu2	*Porphyra purpurea* (red alga)	244	11465646
Pwi	*Prototheca wickerhamii* (green alga)	234	11497475
Mja	*Malawimonas jakobiformis* (jakobid)	251	11466665
Rsa	*Rhodomonas salina* (marine microalga)	240	11466594
Pin	*Phytophthora infestans* (pseudofungus)	248	9695398
Ram	*Reclinomonas americana* (jakobid)	260	11466529
Tau	*Thraustochytrium aureum* (marine fungus)	248	10802923

Protein sizes are given as number of amino acyl residues

dependent on the twin-arginine consensus motif, S/T R R X F L K (Angelini et al. 2001; Blaudeck et al. 2001; Stanley et al. 2000; Summer et al. 2000). This leader may play multiple roles in targeting, translocating and assembling folded proteins and protein complexes (Sambasivarao et al. 2000; Santini et al. 2001). Although it is clear that Tat can function independently of the general secretory pathway (GSP or Sec), the assembly of a functional periplasmic electron-transfer chain may depend on both systems (Cristóbal et al. 1999; Heikkilä et al. 2001). Numerous redox proteins are known to be exported by this system, but pleiotropic effects of *tat* mutations and the recent discovery of Tat-leader binding proteins suggest that much is yet to be learned about the biological importance of the system and its mechanism of action (Angelini et al. 2001; Bernhard et al. 2000; Oresnik et al. 2001; Stanley et al. 2001). Excellent reviews of Tat-dependent protein export have appeared (Berks et al. 2000, b; Settles and Martienssen 1998; Wu et al. 2000).

Recently, comprehensive phylogenetic analyses of Tat systems have been reported (Wu et al.2000). In this review, we update and expand upon the studies of Wu et al. We identify all currently recognizable homologues of TatA/B/E and of TatC in the current databases using the PSI-BLAST program with iterations (Altschul et al. 1997) and analyze their organismal distribution and isoform occurrence. The sequences are multiply aligned using the CLUSTAL X program (Thompson et al. 1994) in preparation for phylogenetic and structural analyses. The resulting information provides guides for future structure/function experiments as well as insight into the evolutionary pathways that gave rise to the two families of interacting proteins that comprise Tat secretory systems.

Homologues of TatA and TatC

The TatABCE system of *E. coli*, has been both molecularly and functionally characterized. This system forms a large (~600 kDa) complex which interacts with fully folded substrate redox proteins that have the N-terminal S/T R R X F L K twin-arginine leader motif. About two dozen *E. coli* proteins may be synthesized with a putative twin-arginine leader (Stanley et al. 2001), and about half of them, including nitrate reductase (NapA), trimethylamine *N*-oxide reductase (TorA), hydrogenases, formate dehydrogenases and DMSO reductase, have been shown to be assembled and exported via the Tat pathway (see Berks et al.2000). Most of these proteins associate with their cofactors in the cell cytoplasm before translocation.

The *E. coli* TatA, TatB, TatC, and TatE proteins have 1, 1, 6, and 1 putative transmembrane α-helical spanners (TMSs), respectively (sizes of 98, 171, 258, and 67 amino acyl residues). TatA, B and E are homologous, and TatA and TatE can partially substitute for each other functionally.

Homologues of *E. coli* TatC (Table 1) and TatA, B and E (Table 2) are found in a variety of gram-negative and gram-positive bacteria, archaea, thylakoid membranes of plant chloroplasts and mitochondria as noted above (Table 3). Homologues are not demonstrable, for example, in mycoplasmas, *Thermatoga maritima*, *Methanococcus jannaschii*, yeast, and animals. Thus, by this criterion, the system is not ubiquitous. While no gram-negative bacterium has more than one TatC homologue, gram-positive bacteria may have two, and archaea may have two or three (Table 3). There is no good correlation between the number of TatC isoforms and the number of TatA paralogues in an organism, as a bacterium with either one or two TatC homologues may have one, two or three TatA homologues (Table 3). At present, we do not know if there is a correlation between the numbers of *tat* genes and the numbers and/or properties of the proteins translocated by this pathway.

Phylogeny of the TatC family

The phylogenetic tree for all recognized TatC homologues is shown in Fig. 1. All of the low G+C gram-positive bacterial homologues cluster together as do the high G+C gram-positive bacterial homologues. Most of the gram-

Table 2 Proteins of the TatA family

Abbreviation	Organism	Size[1]	GenBank index number
Archaea			
Ape1	Aeropyrum pernix	71	14602210
Ape2	*Aeropyrum pernix*	59	14602211
Afu1	Archaeoglobus fulgidus	113	11499638
Afu2	*Archaeoglobus fulgidus*	81	11498390
Hsp	*Halobacterium* sp. NRC-1	96	10580374
Sso1	*Sulfolobus solfataricus*	112	13813639
Sso2	*Sulfolobus solfataricus*	91	13816527
Tva	*Thermoplasma acidophilum*	100	–
Tvo	*Thermoplasma volcanium*	100	13541202
Gram-negative bacteria			
Aae1	*Aquifex aeolicus*	77	9978997
Aae2	*Aquifex aeolicus*	59	9978994
Ach	*Azotobacter chroococcum*	192	1224007
Cje1	*Campylobacter jejuni*	79	9979042
Cje2	*Campylobacter jejuni*	138	9979036
Ccr1	*Caulobacter crescentus*	74	13423474
Ccr2	*Caulobacter crescentus*	200	13423471
TatA Eco	*Escherichia coli*	89	9988060
TatB Eco	*Escherichia coli*	171	12644081
TatE Eco	*Escherichia coli*	67	2506618
Hin1	*Haemophilus influenzae*	89	9988065
Hin2	*Haemophilus influenzae*	186	12643470
Hpy1	*Helicobacter pylori*	79	9978986
Hpy2	*Helicobacter pylori*	160	9978989
Lpn1	*Legionella pneumophila*	61	13277311
Lpn2	*Legionella pneumophila*	89	13277313
Mlo1	*Mesorhizobium loti*	73	13471183
Mlo2	*Mesorhizobium loti*	251	–
Nme1	*Neisseria meningitidis*	67	9979005
Nme2	*Neisseria meningitidis*	228	13431937
Pmu1	*Pasteurella multocida*	76	12722097
Pmu2	*Pasteurella multocida*	191	13431924
Pae1	*Pseudomonas aeruginosa*	82	11352708
Pae2	*Pseudomonas aeruginosa*	141	13431934
Pst1	*Pseudomonas stutzeri*	57	9979022
Pst2	*Pseudomonas stutzeri*	139	12831975
Pst3	*Pseudomonas stutzeri*	76	12831974
Rpr	*Rickettsia prowazekii*	54	9979048
Sty1	*Salmonella typhimurium*	84	9978999
Sty2	*Salmonella typhimurium*	182	9979001
Sty3	*Salmonella typhimurium*	67	9979008
Ssp1	*Synechocystis* sp.	75	7444817
Ssp2	*Synechocystis* sp.	126	7470329
Vch1	*Vibrio cholerae*	82	9979012
Vch2	*Vibrio cholerae*	133	9979015
Vch1–2	*Vibrio cholerae*	78	11356195
Xfa1	*Xylella fastidiosa*	71	9979030
Xfa2	*Xylella fastidiosa*	140	9979033
Gram-positive bacteria			
Bha1	*Bacillus halodurans*	65	4514350
Bha2	*Bacillus halodurans*	68	10176530
Bsu1	*Bacillus subtilis*	70	7444820
Bsu2	*Bacillus subtilis*	57	7444821
Bsu3	*Bacillus subtilis*	62	7444823
Dra1	*Deinococcus radiodurans*	132	7471368

Table 2 (continued)

Abbreviation	Organism	Size[1]	GenBank index number
Dra2	*Deinococcus radiodurans*	117	7471564
Mle	*Mycobacterium leprae*	88	1731352
	Mycobacterium leprae	120	13093079
Mtu1	*Mycobacterium tuberculosis*	83	1731353
Mtu2	*Mycobacterium tuberculosis*	131	7476688
Rer	*Rhodococcus erythropolis*	98	9979018
Sau	*Staphylococcus aureus*	71	13700261
Sco1	*Streptomyces coelicolor*	95	6119675
Sco2	*Streptomyces coelicolor*	161	9714435
Sli	*Streptomyces lividans*	158	12227261
Eukaryotes			
Ath1	*Arabidopsis thaliana*	260	4894914
Ath2	*Arabidopsis thaliana*	147	7682781
Ath3	*Arabidopsis thaliana*	272	10177936
Psa	*Pisum sativum*	137	4929305
Zma1	*Zea mays*	170	4877984
Zma2	*Zea mays*	169	4877986
Zma3	*Zea mays*	243	7489747
Zma4	*Zea mays*	238	7542514

Protein sizes are given as number of amino acyl residues: – accession number not available

negative bacterial proteins cluster into two distinct but adjacent clusters. Moreover, the archaeal homologues are found in a single diverse cluster. The eukaryotic proteins are found in two clusters, one probably representing the chloroplast homologues and one representing the mitochondrial homologues. The former cluster includes cyanobacterial homologues. The putative mitochondrial cluster is exceptionally divergent in sequence in spite of the close phylogenetic relationships of several of the source organisms. No animal, protozoan or yeast homologue is represented.

The TatC family multiple alignment

Almost all TatC homologues fall into the size range of 240–310 amino acyl residues regardless of organismal source. The pea chloroplast TatC (cpTatC) has a much longer N-terminal soluble domain than those predicted for bacterial and algal TatC proteins (Mori et al. 2001). TatC of *E. coli* probably has six TMSs, and the multiple alignment and derived average hydropathy plots suggest that this is true of all TatC homologues (see our ALIGN web site http://www-biology.ucsd.edu/~msaier/transport/). Recently, biochemical data have shown that cpTatC is an integral membrane protein with its N- and C-termini exposed to the stromal face of the membrane (Mori et al. 2001).

Several regions of TatC proved to be well conserved, particularly the six TMSs and adjacent regions. The consensus sequences for these six regions indicate a distinctive amphipathic character that may be indicative of membrane orientation and protein structure. These six consensus sequences are as follows:

- TMS1: H L X E L R X R (L I V)$_2$ X$_2$ (L I V)$_2$ (S T A G) (L I V)$_3$ (S T A G) (L I V F)$_2$ (S T A G C)$_2$
- TMS2: (S T A G) (L I V F)$_3$ (S T A G) (L I V F)$_3$ (S T A G) X P (L I V)$_3$ Y Q (L I V) W A F (L I V) (S T A G) P G L Y
- TMS3: E (R K)$_2$ (L I V)$_3$ P (L I V F)$_3$ (S T A G P)$_3$ (L I V) L F (L I V Y F) X G X$_2$ F (S T A G) (Y F)$_2$ (L I V F)$_3$ P (L I V F)$_3$ X (L I V F)$_3$
- TMS4: Y (L I V) (D E S) F (L I V)$_2$ X (L I V F)$_3$ (S T A G P) F G (L I V) (S T A G) F (E Q) (L I V) P (L I V A)$_2$ X (L I V F)$_3$ X$_3$ G
- TMS5: (R K) (R K P) (Y F) (L I V)$_3$ (S T A G)$_2$ F (L I V F)$_2$ (G A) X (L I V)$_2$ (T S) P P
- TMS6: Q (S T A G) (L I V)$_2$ (S T A G) (L I V) P (L I V M) X$_2$ L (F Y) E (L I V) (S T A G)$_2$ (L I V F)$_3$ (S T A G) (R K)

[Alternative residues at a single position are indicated in parentheses; X=any residue]

Attempts to identify an internal gene duplication or triplication event that might have given rise to these proteins failed. Thus, each TMS exhibits characteristic features that do not resemble those of any other.

Phylogeny of the TatA family

The TatA/B/E phylogenetic tree is presented in Fig. 2. Like the TatC family tree, clustering is generally in accordance with organismal type, but unlike the TatC tree, there are multiple clusters for most of the organismal types. For example, archaeal proteins are found in three clusters although one of these clusters includes most of them. When two archaeal paralogues are present, they can have arisen

Table 3 Organismal distribution of TatA and TatC homologues[a]

Organism	TatC	TatABE	Complete genome sequence
Archaea			
Aeropyrum pernix	1	2	+
Archaeoglobus fulgidus	2	2	+
Halobacterium sp.	3	1	+
Sulfolobus solfataricus	2	2	+
Thermoplasma acidophilum	1	1	+
Thermoplasma volcanium	1	1	+
Gram-negative bacteria			
Aquifex aeolicus	1	2	+
Azotobacter chroococcum	1	2	
Campylobacter jejuni	1	2	+
Caulobacter crescentus	1	2	+
Escherichia coli	1	3	+
Haemophilus influenzae	1	2	+
Helicobacter pylori	1	2	+
Legionella pneumophila	1	2	
Mesorhizobium loti	1	2	+
Neisseria meningitidis	1	2	+
Pasteurella multocida	1	2	+
Pseudomonas aeruginosa	1	2	+
Pseudomonas stutzeri	1	3	
Rickettsia prowazekii	1	1	+
Salmonella typhimurium	1	3	
Synechococcus sp.	1	0	
Synechocystis sp.	1	2	+
Vibrio cholerae	1	3	+
Xylella fastidiosa	1	3	+
Gram-positive bacteria			
Bacillus halodurans	2	2	+
Bacillus subtilis	2	3	+
Deinococcus radiodurans	1	2	+
Mycobacterium leprae	1	2	+
Mycobacterium tuberculosis	1	2	+
Rhodococcus erythropolis	1	1	
Staphylococcus aureus	1	1	+
Streptomyces coelicolor	1	2	
Streptomyces lividans	1	0	
Eukaryote[a]			
Arabidopsis thaliana	3	3	+

[a]See Tables 1 and 2 for eukaryotes with incompletely sequenced genomes

either by an early gene duplication event (i.e., Afu) or by a late duplication event (i.e., Sso and Ape). Note that several archaea with fully sequenced genomes are not represented in either Fig. 1 or Fig. 2, and they thus lack the Tat system.

The high G+C gram-positive bacterial homologues are found in three sequence-divergent clusters, but all of the low G+C gram-positive bacterial proteins are found in a single cluster that includes the *Deinococcus* homologue, Dra2. The other *Deinococcus* homologue, Dra1, is found loosely associated with one of the high G+C gram-positive bacterial proteins, Sco1. Since the single *Deinococcus* TatC homologue clustered with the high G+C gram-positive proteins, one can propose that Dra2 was obtained from a bacterium by lateral transfer.

The two high G+C gram-positive bacterial TatA paralogues, in *Mycobacterium tuberculosis*, *Mycobacterium leprae* and *Streptomyces coelicolor*, arose by early gene duplication events, but the low G+C gram-positive bacterial homologues in *Bacillus* species arose by more recent gene duplication events. Two of the three *Bacillus subtilis* paralogues have orthologues in *Bacillus halodurans*, but the third one does not. It is important to note that many low G+C gram-positive bacteria with fully sequenced genomes (i.e., mycoplasmas, ureaplasma, lactococci, streptococci, enterococci) lack recognizable TatA and TatC homologues and therefore lack the Tat system altogether. These organisms generally use fermentative pathways for energy generation and lack electron transfer complexes.

Cyanobacterial and chloroplast TatA homologues are found in one large cluster and one small cluster. A single *Synechocystis* protein is found in each cluster. Most of the proteins in these two clusters are from higher plants. It is interesting to note that no sequenced TatA homologue from the organisms possessing putative mitochondrial TatC homologues have yet been identified. Thus, in mitochondria, TatC homologues may function alone, in conjunction with one or more nuclear-encoded TatA homologue(s), or with one or more protein(s) non-homologous to TatA.

The gram-negative bacterial TatA homologues fall into one large cluster that includes both TatA and TatE of *E. coli*, and one moderately sized cluster that includes TatB of *E. coli*. However, gram-negative bacterial proteins are also found in four small clusters, one including both of the two *Aquifex aeolicus* paralogues. Interestingly, the apparently internally duplicated homologue from *Azotobacter chroococcum* has an N-terminal domain that clusters with the major gram-negative bacterial cluster, while the C-terminal domain clusters with the one of moderate size. It seems probable that the apparent internal duplication results from a sequencing error, but if not, a late gene fusion event is implied.

TatA isoform analyses

The configuration of the TatA tree suggests that a very early gene duplication events gave rise to TatA and TatB, that almost all gram-negative and high G+C gram-positive bacteria acquired both of these paralogues by vertical transmission, and that a few of the γ-proteobacteria (i.e., *E. coli*, *V. cholerae* and *S. typhimurium*) acquired a third paralogue by a late gene duplication event in which the TatA (but not the TatB) paralogue was duplicated.

It should be noted that the small size of many TatA homologues introduces substantial error in the phylogenetic tree and causes some proteins to appear more divergent in sequence than they actually are. However, the tighter clustering of TatA homologues as compared with TatB homo-

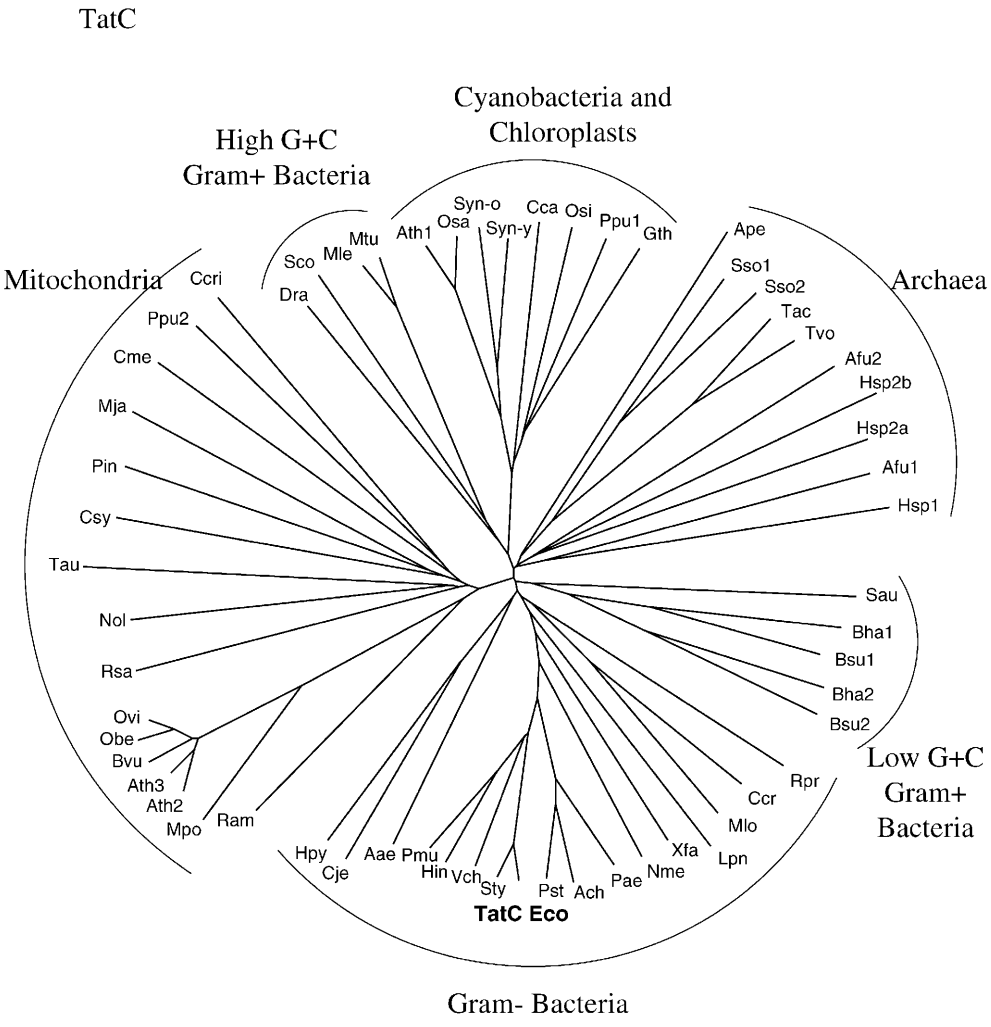

Fig. 1 Phylogenetic tree for the TatC family. Protein abbreviations are as presented in Table 1. The *E. coli* TatC protein is presented in *bold*. Incompletely sequenced (fragmentary) sequences (see Table 1) were not included. The organismal origins of the proteins in each phylogenetic cluster are indicated. This tree and the one shown in Fig. 2 were generated using the CLUSTAL X program

logues suggests that the latter have diverged in sequence more rapidly than the former. The only organisms with fully sequenced genomes that have only one TatA homologue are *Rickettsia prowazekii* and *Staphylococcus aureus*. The *R. prowazekii* homologue clusters with the major gram-negative bacterial TatA subfamily rather than with the moderately sized TatB subfamily. Additionally, TatA duplicated to give TatE, but TatB did not duplicate in any organism. All of these observations suggest that TatA is functionally more important than TatB and that the normally heterooligomeric TatA(E)/TatB complex can sometimes function as a homooligomeric complex, as in the cases of *R. prowazekii* and *S. aureus*. It should be noted that we screened the entire nucleotide sequences of both the *R. prowazekii* and the *S. aureus* genomes for a second paralogue of TatA, using the BLAST search tool, but none was found.

The TatA family multiple alignment

Only the single TMS and its neighboring hydrophilic regions were found to exhibit appreciable conservation in the TatA homologues. A consensus sequence for this region is:

- G Hy G Hy (S T A G P) (E Q) $(L\ I\ V)_{11}$ F G (G A S T P) X (R K) L P X (L I V) (G A) (S R K) (G A S T D) (L I V) G X_2 (L I V) (G K E) X F (R K).

[Hy=any hydrophobic residue; X=any residue]

Relative proportions of TatA vs TatC paralogues

Most prokaryotes with fully sequenced genomes that encode the Tat system have one TatC homologue and two TatA homologues. However, four have two, and one has three TatC homologues, while six have only one TatA homologue, and six have three TatA homologues (Table 3). Of the five prokaryotes with two TatC homologues, three are archaea, and two are *Bacillus* species. In these cases, the phylogenetic tree, which showed clustering according to organismal type, revealed that the gene duplication events occurred recently, both within the archaeal and *Bacillus* lineages. One archaeon (*Halobacterium* sp.) had an internally duplicated TatC homologue of twice the nor-

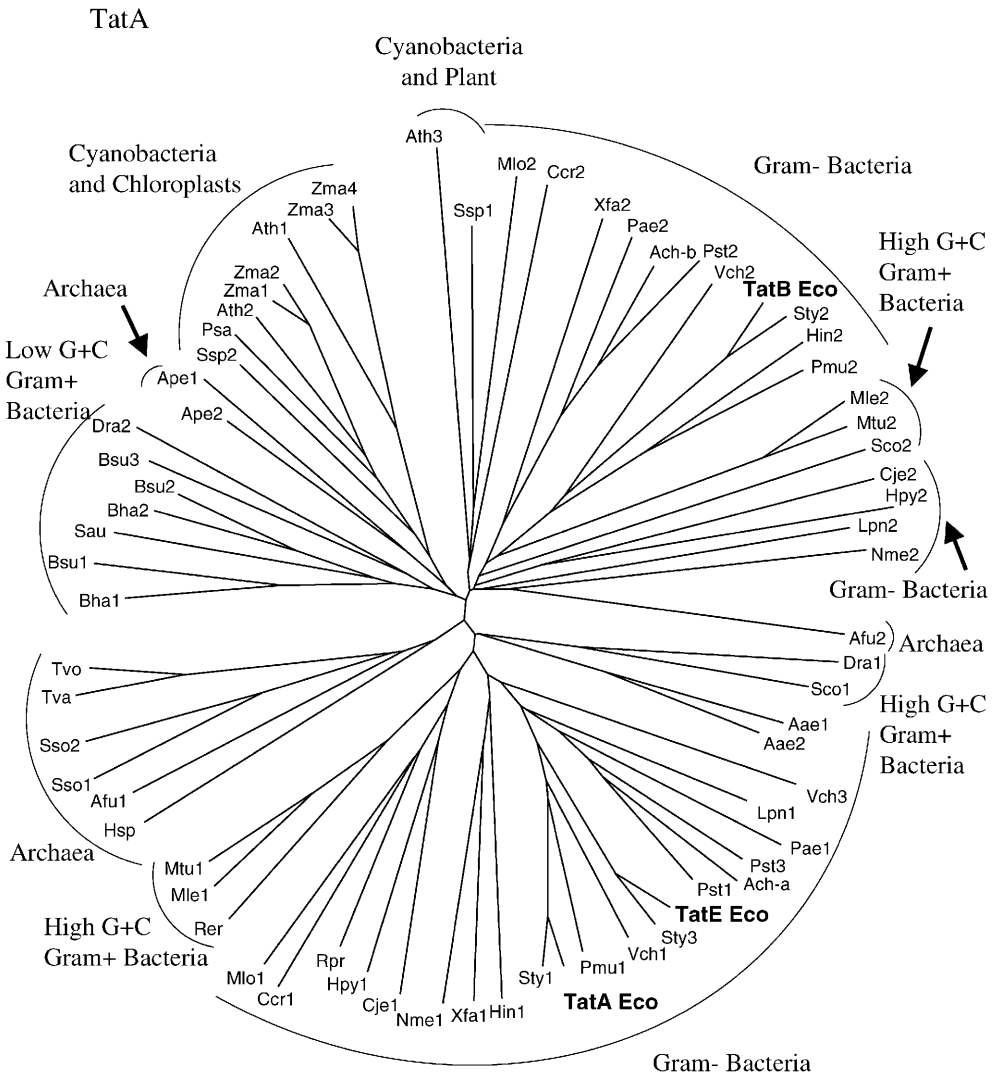

Fig. 2 Phylogenetic tree for the TatA family. Format of presentation is as for Fig. 1 with protein abbreviations as indicated in Table 2 and the *E. coli* homologues presented in *bold*

mal size, but unexpectedly, the two halves proved to be very divergent in sequence. Even more surprising, this organism, which exhibits the equivalent of three TatC homologues, has only one recognizable TatA homologue. In all other prokaryotes with two TatC paralogues, either two or three TatA paralogues were identified. The possibility that at least some organisms with pairs of both homologous proteins form two independently functioning complexes cannot be ruled out.

Evolutionary implications

While the duplicated TatC paralogues found in any one organism proved to have resulted from late-occurring gene duplication events (after the divergence of the major kingdoms of living organisms), the same was not true for the TatA paralogues. In some organisms (e.g., in the two *Bacillus* species and in two archaea), the *tatA* duplication event occurred late, after divergence of the kingdoms, but in gram-negative bacteria, in the high G+C gram-positive bacteria and in one archaeon (*A. fulgidus*), the gene duplication event occurred early. In a single protein, the TatA homologue of *Azotobacter chroococcum*, an internally duplicated TatA homologue was reported. Such a configuration would require that the two halves of this protein have opposite orientation in the membrane. Because sequence similarity usually implies common topology (Saier 2000b, 2001; Saier and Tseng 1999), and because TatA and TatB are usually encoded by adjacent genes within a single operon, we suggest that the apparent fusion is due to a sequencing error.

By using quantitative immunoblotting, it has been estimated that each pea chloroplast contains about 18,000 cpTatC, 95,000 Hcf106 (chloroplast TatB) and 140,000 Tha4 (chloroplast TatA) molecules (Mori and Cline 2001). cpTatC is present in approximately the same molar concentration as the number of active translocation sites, and Hcf106 and Tha4 are present in about five and eight copies per translocation site, respectively. A complex of about 700 kDa containing cpTatC and Hcf106, but not Tha4 has recently been obtained from digitonin-solubilized thylakoids (Cline and Mori 2001). Importantly, thylakoid ΔpH-dependent precursor proteins bind to the

cpTatC and Hcf106 within this complex. Although a direct interaction between the signal peptide and these components has not been demonstrated, binding requires both the twin-arginine motif and the hydrophobic core of the signal peptide (Cline and Mori 2001).

Translational fusion studies have shown that *E. coli* TatA, TatB and TatC are synthesized with an approximate ratio of 65:2.5:1 (Jack et al. 2001). Complexes containing TatA and TatB (Sargent et al. 2001), and TatA, TatB and TatC (Bolhuis et al. 2001) have been obtained from *E. coli* membrane fractions. In the latter case, the TatA content of the complex reduced during purification and reached a final TatA:TatB:TatC ratio of 1:1:1. In addition, it has been suggested that the TatA/B proportions compared to TatC in the translocon may vary according to the size and/or folding state of the protein to be translocated (Wu et al. 2000). Based on these observations, we suggest that the *tat* gene copy numbers may not necessarily reflect the ratios of the Tat components in a Tat translocon. Physiologically relevant stoichiometries may thus differ from those proposed on the basis of biochemical experiments performed to date.

The nature and distribution of mitochondrial Tat systems

Mitochondrially encoded TatC homologues have been identified in algae and photosynthetic eukaryotes as a result of the sequencing of their mitochondrial genomes. Homologues have not been identified in either yeast or animal mitochondrial or nuclear genomes, and no TatA homologues have been found in any eukaryote except higher plants. Because nuclear chromosomes of most of the organisms which exhibit mitochondrially encoded TatC homologues have not yet been sequenced, it cannot be concluded that TatA homologues are absent. It is possible that proteins exhibiting little or no sequence similarity with TatA may function with these TatC homologues, or that these proteins are yet to be identified. Biochemical analyses or extensive genome sequencing efforts will be required to resolve this interesting question.

Conclusions and perspectives

The analyses reported in this review extend the recent phylogenetic analyses of the TatA and TatC families reported by Wu et al. (2000). Our studies and those of Wu et al. allow identification of organisms that lack the Tat system as well as those that contain it. Furthermore, we have been able to specify the relative times of gene duplication events that gave rise to two or more paralogues of both TatA and TatC in the various organisms that have more than one such homologue. The surprising observation is the diversity of protein combinations found. Thus, the standard Tat systems consist of one TatC and two sequence-divergent TatA homologues (a TatA equivalent and a TatB equivalent). A few organisms have two or three TatC equivalents, but these may have one, two or three TatAs. Those that have three TatA paralogues all gained the third paralogue by duplication of the more conserved TatA equivalent rather than the less conserved TatB equivalent. The structural and functional explanations for these findings remain unknown, but they should minimally provide the scientific community with interesting food for thought. Further experimentation will be required to determine the functional consequences of these findings.

Acknowledgements This work was supported by NIH grants GM64368 and GM55434. We thank Mary Beth Hiller for her help in the preparation of this manuscript.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (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-L, Yamagishi A, Berenguer J, Wu L-F (2001) Export of *Thermus thermophilus* alkaline phosphatase via the twin-arginine translocation pathway in *Escherichia coli*. FEBS Lett 506:103–107

Berks BC (1996) A common export pathway for proteins binding complex redox cofactors? Mol Microbiol 22:393–404

Berks BC, Sargent F, De Leeuw E, Hinsley AP, Stanley NR, Jack RL, Buchanan G, Palmer T (2000a) A novel protein transport system involved in the biogenesis of bacterial electron transfer chains. Biochim Biophys Acta 1459:325–330

Berks BC, Sargent F, Palmer T (2000b) The Tat protein export pathway. Mol Microbiol 35:260–274

Bernhard M, Friedrich B, Siddiqui RA (2000) *Ralstonia eutropha* TF93 is blocked in Tat-mediated protein export. J Bacteriol 182:581–588

Blaudeck N, Sprenger GA, Freudl R, Wiegert T (2001) Specificity of signal peptide recognition in Tat-dependent bacterial protein translocation. J Bacteriol 183:604–610

Bogsch EG, Sargent F, Stanley NR, Berks BC, Robinson C, Palmer T (1998) An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria. J Biol Chem 273:18003–18006

Bolhuis A, Mathers JE, Thomas JD, Barrett CM, 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

Chanal A, Santini C-L, Wu L-F (1998) Potential receptor function of three homologous components, TatA, TatB and TatE, of the twin-arginine signal sequence-dependent metalloenzyme translocation pathway in *Escherichia coli*. Mol Microbiol 30:674–676

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

Cline K, Mori H (2001) Thylakoid ΔpH-dependent precursor proteins bind to a cpTatC-Hcf106 complex before Tha4-dependent transport. J Cell Biol 154:719–729

Dalbey RE, Robinson C (1999) Protein translocation into and across the bacterial plasma membrane and the plant thylakoid membrane. Trends Biochem Sci 24:17–22

Halbig D, Hou B, Freudl R, Sprenger GA, Klösgen RB (1999) Bacterial proteins carrying twin-R signal peptides are specifically targeted by the ΔpH-dependent transport machinery of the thylakoid membrane system. FEBS Lett 447:95–98

Heikkilä MP, Honisch U, Wunsch P, Zumft WG (2001) Role of the Tat transport system in nitrous oxide reductase translocation and cytochrome, cd_1 biosynthesis in *Pseudomonas stutzeri*. J Bacteriol 183, 1663–1671

Jack RL, Sargent F, Berks BC, Sawers G, Palmer T (2001) Constitutive expression of *Escherichia coli tat* genes indicates an important role for the twin arginine translocase during aerobic and anaerobic growth. J Bacteriol 183:1801–1804

Jongbloed JDH, Martin U, Antelmann H, Hecker M, Tjalsma H, Venema G, Bron S, van Dijl JM, Müller J (2000) TatC is a specificity determinant for protein secretion via the twin-arginine translocation pathway. J Biol Chem 275:413500–41357

Keegstra K, Cline K (1999) Protein import and routine systems of chloroplasts. Plant Cell 11:557–570

Mori H, Cline K (1998) A signal peptide that directs non-Sec transport in bacteria also directs efficient and exclusive transport on the thylakoid ΔpH pathway. J Biol Chem 273:11405–11408

Mori H, Summer EJ, Cline K (2001) Chloroplast TatC plays a direct role in thylakoid ΔpH-dependent protein transport. FEBS Lett 501:65–68

Oresnik IJ, Ladner C, Turner RJ (2001) Identification of a twin-arginine leader-binding protein. Mol Microbiol 40:323–331

Saier MH Jr (2000a) A functional/phylogenetic classification system for transmembrane solute transporters. Microbiol Mol Biol Rev 64:354–411

Saier MH Jr (2000b) Vectorial metabolism and the evolution of transport systems. J Bacteriol 182:5029–5035

Saier MH Jr (2001) Evolution of transport proteins. In: Setlow JK (ed) Genetic engineering. Principles and methods, vol 23. Kluwer Academic/Plenum, New York, pp 1–10

Saier MH Jr, Tseng T-T (1999) Evolutionary origins of transmembrane transport systems. In: Broome-Smith JK, Baumberg S, Stirling CJ, Ward FB (eds) Transport of molecules across microbial membranes, Symposium 58, Society for General Microbiology, Cambridge University Press, Cambridge, UK, pp 252–274

Sambasivarao D, Turner RJ, Simala-Grant JL, Shaw G, Hu J, Weiner JH (2000) Multiple roles for twin arginine leader sequence of dimethyl sulfoxide reductase of *Escherichia coli*. J Biol Chem 275:22526–22531

Santini C-L, Ize B, Chanal A, Muller M, Giordano G, Wu L-F (1998) A novel Sec-independent periplasmic protein translocation pathway in *Escherichia coli*. EMBO J 17:101–112

Santini C-L, Bernadac A, Zhang M, Chanal A, Ize B, Blanco C, Wu L-F (2001) Translocation of jellyfish green fluorescent protein via the Tat system of *Escherichia coli* and change of its periplasmic localization in response to osmotic up-shock. J Biol Chem 276:8159–8164

Sargent F, Bogsch EG, Stanley NR, Wexler M, Robinson C, Berks BC, Palmer T (1998) Overlapping functions of components of a bacterial Sec-independent protein export pathway. EMBO J 17:3540–3650

Sargent F, Gohlke U, de Leeuw E, Stanley NR, Palmer T, Saibil HR, Berks BC (2001) Purified components of the *Escherichia coli* Tat protein transport system form a double-layered ring structure. Eur J Biochem 268:3361–3367

Settles AM, Martienssen R (1998) Old and new pathways of protein export in chloroplasts and bacteria. Trends Cell Biol 8: 494–501

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

Stanley NR, Palmer T, Berks BC (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

Stanley NR, Findlay K, Berks BC, Palmer T (2001) *Escherichia coli* strains blocked in Tat-dependent protein export exhibit pleiotropic defects in the cell envelope. J Bacteriol 183:139–144

Summer EJ, Mori H, Settles AM, Cline K (2000) The thylakoid ΔpH-dependent pathway machinery facilitates RR-independent N-tail protein integration. J Biol Chem 275:23483–23490

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Voordouw G (2000) A universal system for the transport of redox proteins: early roots and latest developments. Biophys Chem 86:131–140

Weiner JH, Bilous PT, Shaw GM, Lubitz SP, Frost L, Thomas GH, Cole JA, Turner RJ (1998) A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins. Cell 93:93–101

Wexler M, Bogsch EG, Klösgen RB, Palmer T, Robinson C, Berks BC (1998) Targeting signals for a bacterial Sec-independent export system direct plant thylakoid import by the ΔpH pathway. FEBS Lett 431:339–342

Wexler M, Sargent F, Jack RL, Stanley NR, Bogsch EG, Robinson C, Berks BC, Palmer T (2000) TatD is a cytoplasmic protein with DNase activity. No requirement for TatD family proteins in Sec-independent protein export. J Biol Chem 275:16717–16722

Wu L-F, Ize B, Chanal A, Quentin Y, Fichant G (2000) Bacterial twin-arginine signal peptide-dependent protein translocation pathway: evolution and mechanism. J Mol Microbiol Biotechnol 2:179–189

Yahr TL, Wickner WT (2001) Functional reconstitution of bacterial Tat translocation in vitro. EMBO J 20:2472–2479