Scanning the *Corynebacterium glutamicum* R genome for high-efficiency secretion signal sequences

Keiro Watanabe,¹ Yoshiki Tsuchida,^{1,2} Naoko Okibe,¹
Haruhiko Teramoto,¹ Nobuaki Suzuki,¹ Masayuki Inui¹
and Hideaki Yukawa¹

Correspondence

Hideaki Yukawa
mmg-lab@rite.or.jp

¹Molecular Microbiology and Biotechnology Group, Research Institute of Innovative Technology for the Earth (RITE), 9-2, Kizugawadai, Kizugawa, Kyoto 619-0292, Japan

²Honda R&D Co., Ltd, 1-4-1 Chuo Wako, Saitama 351-0193, Japan

Systematic screening of secretion proteins using an approach based on the completely sequenced genome of *Corynebacterium glutamicum* R revealed 405 candidate signal peptides, 108 of which were able to heterologously secrete an active-form α -amylase derived from *Geobacillus stearothermophilus*. These comprised 90 general secretory (Sec)-type, 10 twin-arginine translocator (Tat)-type and eight Sec-type with presumptive lipobox peptides. Only Sec- and Tat-type signals directed high-efficiency secretion. In two assays, 11 of these signals resulted in 50- to 150-fold increased amounts of secreted α -amylase compared with the well-known corynebacterial secretory protein PS2. While the presence of an AXA motif at the cleavage sites was readily apparent, it was the presence of a glutamine residue adjacent to the cleavage site that may affect secretion efficiency.

Received 9 September 2008
Revised 11 November 2008
Accepted 25 November 2008

INTRODUCTION

Protein export from the synthesis compartment in the cytoplasm to other organelles or out of a cell is a common phenomenon that takes place in both prokaryotes and eukaryotes (Pohlschroder *et al.*, 1997; Riezman, 1997; Schatz & Dobberstein, 1996). Secretory proteins are synthesized as pre-proteins with proper N-terminal extensions (signal sequences) which control translocation of the secreted proteins. Signal sequences usually consist of 20–40 aa in the N-terminal regions of pre-proteins, and are removed from the secreted proteins by distinct signal peptidases upon penetration through the membrane (Dalbey *et al.*, 1997; Paetzel *et al.*, 1998). In general, a signal sequence requires three distinct domains for its function (Pugsley, 1993; Tjalsma *et al.*, 2000). The region adjacent to the N terminus, called the N-domain, contains several positively charged amino acids for attachment to the cell membrane (Akita *et al.*, 1990; Chen & Nagarajan, 1994; Gennity *et al.*, 1990). Next to the N-domain is the hydrophobic H-domain region, which forms an α -helical conformation in the membrane (Briggs *et al.*, 1986). Neighbouring the C terminus of the signal sequence is the C-domain, which harbours recognition

and cleavage sites for signal peptidases (Dalbey *et al.*, 1997; Paetzel *et al.*, 1998).

Two types of secretion-specific signal peptidases (SPases) (type I and type II) have been identified to date. Both are anchored on the plasma membrane and function to cleave signal sequences from their pre-proteins. In the general secretory (Sec) pathway, pre-protein is transported from the cytoplasm to the extracellular space in an unfolded form and a signal sequence is cut off by type I SPase (Dalbey & Von Heijne, 1992; Dalbey *et al.*, 1997; Paetzel *et al.*, 2002; Tjalsma *et al.*, 2000). Recently, a second general pathway, designated the twin-arginine translocator (Tat) pathway, was discovered. Signal sequences which utilize this pathway are thought to contain two conserved arginine residues ‘-RR-’ in the region near their N terminus. Tat pathway translocation is Sec-independent and employs a cytoplasmic chaperone in order to secrete proteins in a fully folded form (Berks *et al.*, 2000; Cline *et al.*, 1993; Mori & Cline, 2001; Yen *et al.*, 2002). Sequence analysis has shown that Tat signal sequences are also cleaved by type I SPases, because consensus sequences of cleavage sites of Tat signal sequences are similar to those of Sec signal sequences (Tjalsma *et al.*, 2000). The average length of N-domains of Tat signal sequences is twice that of typical Sec signal sequences, even though no clear differentiating features between the H-domains of the two types of signal sequences exist (Tjalsma *et al.*, 2000).

Abbreviations: Sec, general secretory; SPase, signal peptidase; Tat, twin-arginine translocator.

Two supplementary tables, listing oligonucleotide DNA primers used in this study and putative secretory proteins examined in this work, are available with the online version of this paper.

In addition to type I SPase, secretion of lipoprotein is dependent upon the action of lipoprotein-specific (type II) SPases (Sankaran & Wu, 1995). Lipoprotein precursors have a series of well-conserved amino acid residues, -Leu-Ala-Ala-Cys-, called the lipobox. Precursor lipoproteins are modified by lipoprotein diacylglyceryl transferase (LGT). This modification promotes cleavage of Lipo signal sequences by type II SPases and induces secretion through the cell membrane. After cleavage, prolipoprotein is acylated by *N*-acyltransferase (Inouye *et al.*, 1983; Qi *et al.*, 1995; Sankaran & Wu, 1995). Lipo signal sequences are shorter than their Sec counterparts and the cysteine in the lipobox must be situated between the 15th and the 35th amino acid (Sutcliffe & Harrington, 2002).

Corynebacterium glutamicum is a Gram-positive, non-sporulating and facultatively anaerobic bacterium (Nishimura *et al.*, 2008). *C. glutamicum* belongs to the group of actinomycetes that includes the mycobacteria. The cell wall of mycobacteria contains arabinogalactan anchoring to the wall peptidoglycan, which is built up of alternating α -1,4-linked *N*-acetylglucosamine and *N*-acetylmuramic acid. Mycolic acids, which possess very long chains (C60–C90) and oxygen function, are a typical feature. However, compared to mycobacteria, corynebacteria contain short-chain mycolates (C22–C36) (Collins *et al.*, 1982). Mycolic acids are covalently linked to the cell-wall arabinogalactan or esterifying trehalose and glycerol in the cell envelope. Components containing these mycolic acids may control the structure, function and permeability of the cell envelope (Brennan & Nikaido, 1995; Draper, 1998; Minnikin, 1982).

C. glutamicum has been used in industrial production of amino acids such as glutamate and lysine (Kinoshita, 1985; Malumbres *et al.*, 1995). The fermentation conditions for this bacterial species are well established for development of mass production methods (Liebl & Sinskey, 1990). *C. glutamicum* is also recognized as a favourable host for heterologous protein production, because extracellular proteolytic activity is not detected (Billman-Jacobe *et al.*, 1995). By using a transgenic approach, *C. glutamicum* has been made to secrete amylase, nuclease, protease, transglutaminase, subtilisin-like serine protease, epidermal growth factor, green fluorescence protein and protein glutaminase (Billman-Jacobe *et al.*, 1995; Date *et al.*, 2003, 2006; Kikuchi *et al.*, 2006, 2007; Liebl *et al.*, 1992; Meissner *et al.*, 2007; Salim *et al.*, 1997; Smith *et al.*, 1986). Despite signal sequences being essential for the secretion of target proteins, only eight, namely DNase, Rpf1, Rpf2, PS1, PS2 and three PS1-related secretion proteins, have been identified in *C. glutamicum* (Brand *et al.*, 2003; Hartmann *et al.*, 2004; Joliff *et al.*, 1992; Liebl & Sinskey, 1990; Peyret *et al.*, 1993). The most common *C. glutamicum* secretion protein is PS2, which is detected in culture media as one of two major proteins and has a function in the biogenesis of the surface layer in *C. glutamicum*. However, the biochemical function of PS2 protein is unknown

(Peyret *et al.*, 1993). Another major protein, PS1, is known as a corynebacterial mycolyltransferase, and its homologues are also found in mycobacteria. Loss-of-function analysis has shown that mycolyltransferases play an important role in the biogenesis of the cell envelope in *Corynebacterineae* (Kacem *et al.*, 2004). Secretion signals derived from PS2 are currently used in extracytoplasmic production in *C. glutamicum*. The main shortcoming of PS2-based secretion is that yields vary widely depending on the target protein (Date *et al.*, 2003, 2006). Consequently, applications of *C. glutamicum* in protein secretion are still comparatively limited and further information on protein secretion is needed.

C. glutamicum has long been considered to excrete only a limited number of proteins into its culture. However, extracellular protein identification using 2D SDS-PAGE has shown about 50 protein spots in a late-exponential-phase culture of *C. glutamicum* (Hermann *et al.*, 2001). Subsequently, the recent demonstration of 141 protein spots from the cell surface and extracellular space has raised the possibility that secretion signals exist in *C. glutamicum* that are not yet known (Hansmeier *et al.*, 2006a, b). In order to better understand the secretion mechanisms in *C. glutamicum*, we performed comprehensive analysis of its genome by using SignalP (Tjalsma *et al.*, 2000) to identify any new signal sequences. A total of 108 candidate signal sequences that could secrete heterologous α -amylase in *C. glutamicum* were identified, 106 of which were previously unknown.

METHODS

Bacterial strains and culture conditions. The bacterial strains used in this study are shown in Table 1. *Escherichia coli* was cultivated in Luria–Bertani (LB) medium (Sambrook *et al.*, 1989) at 37 °C with vigorous shaking. *C. glutamicum* R was grown at 33 °C in complex medium [urea, 2.0 g; (NH₄)SO₄, 7.0 g; K₂HPO₄, 0.5 g; KH₂PO₄, 0.5 g; MgSO₄·7H₂O, 0.5 g; FeSO₄·7H₂O, 6 mg; MnSO₄·H₂O, 4.2 mg; biotin, 200 µg; thiamin/HCl, 200 µg; yeast extract, 1.0 g; casamino acids, 1.0 g; glucose, 40.0 g; H₂O, 1.0 l]. Soluble starch (4%) (Becton Dickinson) was added to the growth medium to detect the activity of secreted α -amylase. Chloramphenicol (Wako Pure Chemical) was used at the following concentrations: for *E. coli*, 50 µg ml⁻¹; for *C. glutamicum*, 5 µg ml⁻¹.

Candidate signal sequences in *C. glutamicum* R. A total of 2990 ORFs have been predicted for the genome of *C. glutamicum* strain R (Yukawa *et al.*, 2007). Translated amino acid sequences based on ORFs were analysed using SignalP v. 3.0 (<http://www.cbs.dtu.dk/services/SignalP>). To predict secretory proteins, the first 60 residues of all annotated proteins of *C. glutamicum* R were analysed by SignalP. Tat-like signal sequences detected by SignalP were further analysed using the TatP Tat-signal-specific program (<http://www.cbs.dtu.dk/services/TatP/>).

DNA manipulations. *E. coli* plasmid DNA was isolated using a Qiaprep spin kit (Qiagen) according to the manufacturer's instructions. *E. coli* was transformed by the CaCl₂ method (Sambrook *et al.*, 1989). Purified plasmid DNA extracted from *E. coli* SCS110 was introduced into *C. glutamicum* by electroporation using Gene Pulser (Bio-Rad) (Kurusu *et al.*, 1990). High-throughput DNA manipula-

Table 1. Bacterial strains and plasmids in this study

Strain or plasmid	Relevant characteristics	Source or reference
<i>E. coli</i> strains		
JM109	<i>recA1 endA1 gyrA96 thi hsdR17 supE44 relA1 Δ(lac-proAB)/F [traD36 proAB⁺ lacI^q lacZΔM15]</i>	Takara
SCS110	<i>dam dcm endA1 supE44 hsdR17 thi leu rpsL1 lacY galK galT ara tonA thr Δ(lac-proAB)/F' [traD36 proAB⁺ lacI^q lacZΔM15]</i>	Toyobo
<i>C. glutamicum</i> strains		
R	Wild-type strain	Yukawa <i>et al.</i> (2007)
Δ <i>TatC</i>	R strain Δ <i>TatC</i> , deficient	This work
Plasmids		
pCRA429	Cm ^r ; pMB1/M13 <i>ori</i> ; α-amylase	AB437137
pCRC900	Cm ^r ; multiple cloning site (MCS) (with <i>EcoRV</i>); pMB1/M13 <i>ori</i> ; α-amylase without secretion signal; pBL1/coryneform bacterial <i>ori</i> ; shuttle vector	This work
pCRC911	Cm ^r ; signal-less α-amylase with <i>EcoRV</i>	This work
pCRC912	Cm ^r ; native signal reinsertion into pCRC900	This work
pCRC913	Cm ^r ; mutated signal reinsertion into pCRC900	This work
pCRC914Δ324	Cm ^r ; mutated 324 signal reinsertion into pCRC900	This work
pCRC915Δ800	Cm ^r ; mutated 800 signal reinsertion into pCRC900	This work
pCRC916Δ2791	Cm ^r ; mutated 2791 signal reinsertion into pCRC900	This work
pCRC917Δ949	Cm ^r ; mutated 949 signal reinsertion into pCRC900	This work
pUC4K	Km ^r ; <i>LacZ</i> /MCS cloning vector	Amersham
pHSG398	Cm ^r ; <i>LacZ</i> /MCS cloning vector	Takara

tion to construct secretion candidate plasmids was performed with a Biomek-FX laboratory automation workstation (Beckman Coulter) and a MultiScreen96 plasmid purification kit (Millipore). *C. glutamicum* genomic DNA was isolated using a GenomicPrep Cells and Tissue DNA Isolation kit (Amersham Bioscience). For PCR, KOD plus DNA polymerase (Toyobo) was used. Restriction endonucleases were purchased from Takara Bio.

Sequencing. DNA sequencing was performed on an ABI Prime 3130xl and 3730xl genetic analyser using a Big Dye Terminator v3.1 cycle sequencing kit (both from Applied Biosystems) according to the manufacturer's instructions. DNA sequencing data were analysed using the Genetyx Win program (Genetyx).

Plasmid construction. The plasmids used in this work are listed in Table 1. The α-amylase AmyE from *Geobacillus stearothermophilus* NBRC 12550 was used as a marker to test the secretion ability of secretion signal candidates. To do this, a high-copy-number shuttle vector for *C. glutamicum* carrying a signal sequenceless α-amylase gene was constructed. A DNA fragment containing *amyE* was amplified by PCR with amyE-F and amyE-B primers (Supplementary Table S1). The fragment was digested with *Xho*I and *Sph*I, and ligated to the same site of pCRA429 (Table 1). The resultant plasmid was used as a PCR template, and a 6 kb DNA fragment was amplified with *EcoRV*-Amy-F and *EcoRV*-Amy-B primers (Supplementary Table S1). The fragment was then digested with *EcoRV*, circularized by self-ligation and transformed into *E. coli* cells. The extracted plasmid was designated pCRC900 (Fig. 1). As a control to check the α-amylase secretion, three plasmids, pCRC911, pCRC912 and pCRC913, were constructed. To obtain pCRC911, ATG was inserted into the *EcoRV* site of pCRC900 by PCR with Met-amy-F and Met-amy-B primers (Table 1, Supplementary Table S1). The coding regions of native or mutated AmyE signal sequences from the initiation methionine to the 34th amino acid residue were amplified by PCR with Bacillus-amy-F and Bacillus-amy-B or Bacillus-amy-F and Mut-amy-B primers, and ligated into the *EcoRV* site on pCRC900, in pCRC912 and pCRC913 (Table 1, Supplementary

Table S1). In pCRC913, three amino acid residues, -AKA-, from positions 32 to 34, were replaced with -FFF-.

To analyse the effect of the glutamine residue at position +1, the 39th, 27th and 22nd glutamine residues of CgR324, CgR800 and CgR2791 were replaced with an alanine residue, respectively. In contrast, the 31st leucine of CgR949 was altered to a glutamine

Fig. 1. Schematic view of plasmid used for secretion signal screening in *C. glutamicum*. The DNA sequence encoding a signal sequence was removed from *amyE* and replaced with a synthetic *EcoRV* recognition site. The signal sequence of AmyE was predicted by using the SignalP program. -GAT/ATC- indicates the synthetic *EcoRV* recognition site in pCRC900. Expression of α-amylase derived from *G. stearothermophilus* was induced by both *lac* and *tac* promoters.

residue. These mutageneses were carried out by PCR with 324-F and Mut-324-B, 800-F and Mut-800-B, 2791-F and Mut-2791-B or 949-F and Mut-949-B primer pairs, respectively. DNA fragments encoding mutated protein sequences were ligated to the *EcoRV* site of pCRC900. The resulting plasmids were designated pCRC914Δ324, pCRC915Δ800, pCRC916Δ2791 and pCRC917Δ949 (Table 1, Supplementary Table S1), respectively.

Screening of secretion signal sequences. DNA fragments encoding candidate signal sequences were amplified by PCR, purified by gel electrophoresis and extraction, digested with *EcoRV* and ligated into the same site of pCRC900 using Ligation high (Toyobo). After checking for PCR errors and direction of the inserted DNA, the plasmids were transformed into *E. coli* SCS110 and extracted plasmids were used to transform *C. glutamicum*. *C. glutamicum* cells possessing the plasmid were selected on complex solid medium plates containing 1.5% agar (Becton Dickinson) and chloramphenicol, and transferred to a new plate containing 4% soluble starch (Becton Dickinson). After 48 h, 3 ml iodine solution (I_2 , 0.1 g; KI, 2.0 g; H_2O , 300 ml) was added to the plates, which were incubated at room temperature for a few minutes. The α -amylase activity was visualized as a halo by iodine staining. Secreted α -amylase in liquid medium was detected with an EnzChek Ultra Amylase Assay kit (Invitrogen) according to the manufacturer's instructions. Dilution series of culture supernatants were prepared and fluorescent intensity derived from α -amylase activity was measured using a SpectraMax M2 (Molecular Devices).

Construction of a *tatC*-disrupted mutant of *C. glutamicum* R. *cgR1550*, the *tatC* orthologous gene for *C. glutamicum* R, was identified by homology search with the *tatC* sequence of *C. glutamicum* ATCC 13032. Two DNA fragments including 5'-terminal or 3'-terminal flanking regions of *cgR1550* were amplified by PCR from the *C. glutamicum* R genome with 1550-*EcoRI*-F and 1550-*KpnI*-B or 1550-*XbaI*-F and 1550-*Sall*-B primers, respectively (Supplementary Table S1). A DNA fragment containing the kanamycin-resistance gene was also amplified by PCR from the pUC4K plasmid with Km-F and Km-B primers (Supplementary Table S1). The three DNA fragments were individually digested with restriction endonucleases, *EcoRI* and *KpnI*, *XbaI* and *Sall* and *KpnI*, and *XbaI*, respectively, and successively ligated into the same site of pHSG398. The resulting plasmids were used to transform *E. coli* SCS110. Extracted plasmids were used to transform *C. glutamicum* R by electroporation, and the *tatC*-disrupted mutant ($\Delta tatC$) was screened on a solid medium plate containing kanamycin.

RESULTS

There are 108 *C. glutamicum* R protein coding sequences that secrete *G. stearothermophilus*-derived α -amylase

In order to identify candidate secretory proteins in *C. glutamicum* R, the first 60 aa from the N terminus of each of its 2990 annotated protein coding sequences was used to predict the presence of signal sequence cleavage sites using the software program SignalP. Of these, 405 proteins were predicted to possess signal sequences and their accompanying cleavage sites, on the basis of both Neural Network and Hidden Markov models (Bendtsen *et al.*, 2004). Determination of which of the 405 protein coding sequences actually secreted α -amylase necessitated construction of 405 chimeric proteins in which each of the

signal sequences was fused with an AmyE whose native signal sequence was lacking. This was performed by inserting each signal sequence into the *EcoRV* site of pCRC900 and separately transforming *C. glutamicum* R with each of the resultant plasmids.

As the insertions involved extra bases, it was imperative to ascertain the effect of the inserted sequence on the functional integrity of the signal sequences. A plasmid encoding native AmyE and three others in which the sequences encoding a methionine residue, the native signal sequence or a mutated signal sequence were inserted into *EcoRV* sites of pCRC900 were constructed (Fig. 2a). The resultant respective plasmids pCRA429, pCRC911, pCRC912 and pCRC913 (Table 1) were then used to

Fig. 2. Secreted α -amylase activity of *C. glutamicum* cells with modified α -amylase genes. (a) Schematic view of the constructed α -amylases. Native α -amylase from *G. stearothermophilus* (i), α -amylase with the signal sequence replaced with an *EcoRV* recognition site with an inserted methionine codon (ii), signal sequence of *amyE* (iii), and mutated signal sequence of *amyE* (iv). (b) Secreted α -amylase activity of *C. glutamicum* transformants. α -Amylase activity was detected by iodine staining. The distribution of α -amylase caused haloes around the cylindrical colonies in (i) and (iii). No halo was detected for (ii) and (iv).

transform *C. glutamicum* R. Transformants were incubated on a complex solid medium plate containing 4 % starch, and iodine solution was added to reveal clear haloes around pCRA429 and pCRC912 transformants only, which indicated active-form α -amylase secretion (Fig. 2b). The ability of pCRC912 to secrete active-form α -amylase implies that the presence of an artificially inserted *EcoRV* site in the *amyE* gene has no discernible effect on the functional integrity of the signal sequence. In contrast, removal of a signal sequence (pCRC911) or its mutation (pCRC913) resulted in colonies that did not show any clear halo, indicating that a proper signal sequence is required for AmyE secretion (Fig. 2b).

Sequences encoding each of the 405 predicted signal sequences including the 9 bp region at the end of each putative cleavage site of the signal sequence were thus separately fused onto AmyE, and the amylase activity of resultant transformants was assayed. A total of 108 samples showed clear haloes on iodine-stained plates (Supplementary Table S2), but 255 did not. The remaining 42 samples could not be successfully cloned into pCRC900 in *E. coli*. The putative function of the 108 coding sequences varied widely from membrane-related (26.8 %) and wall-related (13.8 %) proteins and secreted enzymes (15.7 %) to others (6.7 %) and unknown proteins (37.0 %). Furthermore, homology search experiments showed that 20 of 108 proteins in this study were identical to the proteins of 71 protein spots in a secretome analysis with 2D gel electrophoresis of *C. glutamicum* ATCC 13032 culture and cell surface proteins (Hansmeier *et al.*, 2006a). These data support the conclusion that the 108 protein coding sequences possess protein secretory functionality in *C. glutamicum*.

Of the signal sequences, 90.7 % are Sec type

Amino acid sequences of the 108 signal sequences allowed their classification into three types. Sec signal sequences, constituting the most common secretion pathway of bacteria, had an average of 36.3 residues per peptide, which is longer than that of the corresponding *Bacillus subtilis* signals, even though the proportions of their N-, H- and C-domains were similar to those of *B. subtilis* (data not shown) (Tjalsma *et al.*, 2000). Tat signal sequences have characteristic tandem arginine residues just before the H-region. However, there are cases known to have one arginine residue in the -RR- motif substituted with a lysine (Cristobal *et al.*, 1999; Stanley *et al.*, 2000). The Tat motif-specific TatP software was thus utilized to identify Tat signal sequences. Eighteen sequences identified from such a search were employed for further analysis by using a Tat pathway-deficient mutant. The *tatC* gene, which encodes one of the components of the Tat pathway, can be disrupted in *C. glutamicum* ATCC 13869, leading to non-lethal cessation of protein secretion using the Tat system (Kikuchi *et al.*, 2006). Only 10 of the 18 signals did not show extracellular α -amylase activity in a *tatC* disruption mutant (Δ *tatC*) of *C. glutamicum* R (Fig. 3). Finally, Sec signals with presumptive lipobox sequences, characterized by the lipobox consensus sequence, must have one cysteine residue between positions 15 and 35 from their N terminus. BLAST homology and PROSITE protein domain searches identified eight new Sec signals with a presumptive lipobox (Fig. 4). Taken together, 90 Sec signals (83.3 %), 10 Tat signals (9.2 %) and eight Sec signals with a presumptive lipobox (7.4 %) were newly identified in *C. glutamicum* R.

Gene product	Signal sequences	No. of a.a. residues	Secretion efficiency
CgR0079	MPSFKSARWRM <u>NRRL</u> FLGTSAAIIVGGVGGVQVVPYISSGEIQTSA / SST	48	++
CgR0120	MTSSFSRRQFLGLGLVAGTGA ⁺ ACTSDPGPAAS / APG	35	++++
CgR0124	MTTPTSPLLPLASDGCGCCAPSTPSATVSAPAVAA / ATD	35	+
CgR0900	MRRPVSRRAIFATSVLVAGVSIMSPSANA / AEA	29	++
CgR0949	MQINRRGFLKATAGLATIGAASMFMPKANA / LGA	30	++++
CgR1023	VRKGISRVLSVAVASSIGFGTVLTGTGIAAA / QDS	31	++++
CgR1448	MAQISRRHFLAAATVAGAGATLAACAGTGGSTSSS / SDS	35	+++
CgR2137	MPQLSRRQFLQTAVTAGLATFLGTPARA / EER	29	++++
CgR2627	MVNTLNSKTVNVPFRFAGVVAATAALFFGALVSLAPSALA / QEP	40	++++
CgR2926	MTQPAPMCSRRMFLGTATTTFAGAFLAACGTEPDQEVA / ATE	38	+++

Fig. 3. Protein coding sequences exhibiting Tat-type signal sequences in *C. glutamicum*. Underlined amino acids indicate arginine residues which are situated upstream of the H-domain. A solidus (/) indicates a putative cleavage site. The symbols ++++ (high), +++ (medium), ++ (low) and + (very low) represent the secretion efficiency of α -amylase estimated by measuring the size of haloes on a plate (see Supplementary Table S2).

Gene product	Signal sequences
CgR0040	MKNSKLLLIAAVSTASILLASC ₂₂
CgR0789	MRTSRVLGILAATLTVSLAAC ₂₂
CgR0865	VSKISTKLKALSAVLSVTTIVAGC ₂₄
CgR1522	MFKLSKPSKSMRVAVSTLAISTLALVGC ₂₈
CgR1819	MTLKSLAVTTAAALALS ₂₂
CgR2213	MSISRTVF ₂₃ GIAATAALSAA ₂₃ LVAC ₂₃
CgR2386	VRVFRGRRGAVAGSFLAVLAIGSLALTGC ₂₉
CgR2535	MADMKLLWTLPLPLVL ₂₁ LAGC ₂₁

Fig. 4. Protein coding sequences exhibiting Sec signals with presumptive lipobox sequences in *C. glutamicum*. Amino acid sequences of the eight newly identified Sec signals with a presumptive lipobox. Boxed regions indicate putative lipobox consensus sequences. Essential cysteines in a presumptive lipobox are shown at the C terminus of the signal sequences. The numbers in small type adjacent to the cysteine residues represent the distance in amino acid residues from the N terminus to the cysteines.

Eleven signal sequences showed 50–150 times the secretion efficiency of PS2

Since secretion efficiency is an important factor for protein production in bioindustry, the secretion efficiencies of the 108 signal sequences were analysed using α -amylase as a marker. Activities of α -amylases secreted using each of the 108 signal sequences were compared with that of α -amylase secreted using the PS2 signal sequence. To date, the PS2 signal sequence represents one of the most studied secretion signals for heterologous protein production in *C. glutamicum* (Date *et al.*, 2003,

2006). Based on extracellular amylase activity, determined by measuring the sizes of haloes around cell colonies on plates, the signal sequences were roughly classified into four classes: ‘high’ (greater than PS2), ‘medium’ (almost equal to PS2), ‘low’ and ‘very low’. In summary, 31 secretion signal sequences produced a halo larger than that of PS2, 26 of which were of Sec type and five of Tat type (Supplementary Table S2).

To evaluate the secretion efficiency more accurately, α -amylase activity of the supernatant of a 100 ml overnight culture in a flask was measured. A transformant expressing a chimeric protein of CgR949 and AmyE showed the highest α -amylase activity among the 31 signals. Its activity was over 150 times higher than that of PS2. Proteins homologous to CgR949 are found in many other bacteria, but their function remains annotated as unknown (data not shown). Furthermore, 11 samples consisting of nine Sec-type and two Tat-type signal sequences exhibited over 50 times higher activity than that of PS2 (Fig. 5).

The AXA motif and the +1 glutamine influence the characteristics of Sec- and Tat-type signal sequences

Of the signal sequences identified above, only Lipo-type signals were not found in the high-efficiency group of 31 signal sequences. Using SignalP, the 26 Sec- and five Tat-type signal sequences collectively revealed a bias towards alanine residues at positions -3 (~83 %) and -1 (~93 %), but no conserved amino acid residues were found at position -2 . This bias fell to ~62 % at position -3 and ~81 % at position -1 among the medium, low and very low efficiency signals (Table 2). This suggests that an -AXA- motif may be recognized at the cleavage site of Sec- and Tat-type signal sequences in *C. glutamicum*.

Fig. 5. Comparison of the amounts of secreted α -amylase of 32 *C. glutamicum* signal sequences. The α -amylase activities, shown as the mean of triplicates, were measured with an EnzChek Ultra Amylase Assay kit. Secretion types are indicated below the sample names.

Table 2. Amino acid frequencies in C-terminal signal sequences

Amino acid	Position*						
	-3	-2	-1	/	+1	+2	+3
High efficiency							
A	26	3	29		4	1	2
P	1		1			1	5
L	1	1			2		1
T	1	2			2	2	4
S	1	3	1		1	6	6
V	1	1			1		4
G		4				3	1
Y		1					
N		5				1	3
H		4					
Q		4			16	2	
R		1					3
E		2			4	10	1
D					1	4	
F							1
I						1	
Total	31	31	31		31	31	31
Medium, low and very low efficiency							
A	43	5	56		17		4
P	1		1		1	10	6
L	1	4	1			1	4
T	2	5	1		1	12	9
S	8	14	6		6	16	12
V	8	5			2	1	3
G	3	1	4		1	2	4
Y		1			1	1	1
N		5			1		6
H		2			1	1	1
Q	1	7			14		1
R					2		2
E	1	4			10	13	7
D		3			9	8	5
F		4			2	1	
I	1	3				2	4
K		3			1		
C		2				1	
M		1					
Total	69	69	69		69	69	69

*Position numbers indicate the same positions as in Fig. 6.

While the average length of signal sequences in the medium, low and very low efficiency categories was 35.5 amino acid residues, that of the high-efficiency category was 31.6. In addition, since the number of residues in the high-efficiency category varied from 21 to 48, it appears that the length of secretion signal may not affect secretion efficiency (Fig. 6). Subsequently, SignalP results predicted the hydrophobicities of 31 high-efficiency signals as being close to those of medium, low and very low category signals. No consensus sequences in the N- and H- domains were found from amino acid sequences of the 31 high-

efficiency samples (Fig. 6), even though a glutamine residue at position +1 was observed in ~51 % of these samples (Fig. 6 and Table 2, underlined), but in only 20.2 % of the medium, low and very low efficiency category samples. It follows therefore that although they are not located inside a signal sequence, glutamine residues flanking the cleavage site may indeed affect secretion efficiency in *C. glutamicum*.

A glutamine residue adjacent to the cleavage site is necessary for productive secretion of α -amylase in *C. glutamicum*

In order to determine the effect of the +1 glutamine (the glutamine residue neighbouring the predicted cleavage site) on secretion efficiency in *C. glutamicum*, point mutation analysis was carried out. Three of 31 high-efficiency signals, the 324, 800 and 2791 signals, containing glutamine residues at the +1 position, had alanine substituted for the +1 glutamine by PCR mutagenesis and were ligated into pCRC900 (Table 1). The results for strains with pCRC914 Δ 324, pCRC915 Δ 800 and pCRC916 Δ 2791 indicate that amylase activity was reduced in comparison with native signals (Table 1, Fig. 7). In contrast, the 949 signal, which was categorized as a high-efficiency signal, although the amino acid residue at the +1 position was leucine, was mutagenized from +1 leucine to +1 glutamine (Figs 6 and 7). The strain containing pCRC917 Δ 949 showed amylase activity three times higher than a strain with the native 949 signal (Table 1, Fig. 7). These results indicate that +1 glutamine residues have an influence on the accumulation of extracellular amylase in *C. glutamicum*.

DISCUSSION

Using simultaneous bioinformatic analysis and a high-throughput secretion assay, we determined 108 secretion signal sequences, including 106 new ones, in *C. glutamicum* R (Supplementary Table S2). The molecular masses of the 108 proteins which contain secretion signals ranged from 5 to 185 kDa and their types varied widely (Supplementary Table S2). These signal sequences may function to transport various proteins across the plasma membrane, suggesting the presence of a large number of unidentified secretory proteins in *C. glutamicum*. Recently, a study using 2D gel electrophoresis of *C. glutamicum* culture supernatants and cell surface proteins detected 141 secretory protein spots, of which 71 were identified by MALDI-TOF MS and tryptic peptide mass fingerprinting (Hansmeier *et al.*, 2006a). The number of secretory proteins identified in this work is less than the number of secretory protein spots identified by 2D gel electrophoresis. In addition, our bioinformatic results imply the existence of 405 secretory proteins in *C. glutamicum*. Therefore, this raises the possibility that unknown factors retarded the detection of secretion signals in our screening. On the other hand, 88 of 108 coding sequences in this

Gene product	Signal sequences	No. of a. a. residues
	-3-2-1/+1+2+3	
CgR0120	MTSSFSRRQFLGGLVLVAGTGAVAACSDPGPAAS /APG	35
CgR0201	VKIKSVFLSTALSASLLLGITPPVLGA / TIN	27
CgR0324	MVELSTRINPNERRFVFIAFLAMILVASGLATPAYA / QQE	38
CgR0337	MRFKSVAIALSTAVIMGGTASVANA / QQV	26
CgR0426	MQNITRKIAALAIAGTLILPATAHA / QSN	25
CgR0427	MRNIVSTVAATIIASSLLLPTTAVTASA / QSS	28
CgR0433	MKLLRRIAAPAIALGIAMSTIVTPSTAGA / AEV	29
CgR0800	MRRSLRHGFTALLTTWALLPTVAVA / QEP	26
CgR0949	MQINRRGFLKATAGLATIGAASMFMPKANA / LGA	30
CgR1023	VRKGISRVLSVAVASSIGFGTVLTGTGIAAA / QDS	31
CgR1053	MKTRHRTLACVAASVSLVSPGLAPTANA / QDR	29
CgR1068	LWLLPYLFHFHLSDLLRKPMKKLRFATIAAATVALTASLTPSASA / QDF	45
CgR1123	MEKKSFPARVIGIGVLGIAGMGILLWLAVTLSDPASPGA / QET	41
CgR1176	MSSASFTTKALSVLAALTAASAPLVAASPAHA / LAN	32
CgR1180	VTKTLPRLLTVAAALAIALPATPVASA / VTP	27
CgR1301	MKFKKIALVLAFLGLASCSSASGDPATNA / DGS	30
CgR1833	MRSFHTAAVAGLTALSVGSATVATA / EES	27
CgR1834	MRKFRNTAIALVSAAISLGGVTAATA / QEA	27
CgR2011	LSLRSTLTLVTASAVALSFTVPAQA / QSS	27
CgR2063	VRKGISRVLSVAVASSIGFGSVLSGTGIAAA / QDT	31
CgR2069	LNEVDRGFLKMFGRWVSVVASCVIALVLPVSHSGA / EEV	38
CgR2137	MPQLSRRQFLQTTAVTAGLATFLGTPARA / EER	29
CgR2146	MNFSVLHLRENVRRVSVTVAGAGALLISGPFFTAHTAEA / TET	40
CgR2165	VQGMKRKMVAAALATALLFGAAPAHA / SEL	26
CgR2218	LKVVEVKRNNLLVAPLTASLVFCNLAVANAVEVEA / ESP	36
CgR2236	MSVTNFRRLAGIAVIAACVAATPTAQA / QSS	28
CgR2495	MSRISARTLAIALAGATAASLAVVPAATANP / AGT	31
CgR2627	MVNTLSKTVNVPFARGVVAATALFFGALVSLAPSALA / QEP	40
CgR2697	VKIKKSASALSRMRIGIATITSTAMLGGLVAVPAHLLPTTAVAQA / QTV	48
CgR2766	MSVFTRAGEASRKLVVALATAAALMVVGQGTAAQ / ANR	37
CgR2791	MKQVFLLLFLAFLVLFPSNAHA / QSP	21

Fig. 6. List of high-efficiency secretion signal sequences in *C. glutamicum*. A glutamine residue that frequently appeared at the +1 position is underlined. Putative signal cleavage sites are represented by a solidus (/). Numbers above the first sequence indicate the positions of amino acids relative to the cleavage site.

study showed no homology to the secretory proteins identified by secretome analysis using 2D gel electrophoresis. The reasons may be as follows: separation of proteins on the 2D gel was carried out with a limited pH range, and our analysis was not reliant on the amount of secretory proteins and the effects of gene expression depending on cell growth because only the genome information was utilized.

Of the 108 secretion signals, 98 Sec-type and 10 Tat-type signals were included, implying that about 90.7 and 9.2% were Sec and Tat signals in *C. glutamicum*, respectively. In contrast, the proportion of Sec and Tat in all signal sequences of *B. subtilis* has been estimated to be about 55.7 and 4.7% through a protein database search (Tjalsma *et al.*, 2000). The higher proportion of Tat-type signals in *C. glutamicum* implies that Tat-dependent secretion might occur more frequently in *C. glutamicum* than in *B. subtilis*. The average length of the 10 new Tat-type signal sequences identified in this study was 35 amino acid residues, which is almost the same as that in *B. subtilis* (Tjalsma *et al.*, 2000). However, the -RRXFL- motif, which is characteristic of Tat-type signal sequences in the N-region (Berks *et al.*,

2000), was not observed in three of the 10 Tat-type signal sequences in this study (Fig. 3). The substitution of an arginine within a -RRXFL- motif triggers decreased amounts of protein secretion (Berks *et al.*, 2000; Cristobal *et al.*, 1999; Stanley *et al.*, 2000), but in *C. glutamicum*, neither CgR1023 nor CgR2627 has two tandem arginines or promotes strong secretion of α -amylase (Fig. 3). In contrast, CgR79, classified in the low secretion efficiency group, possessed the -RRXFL- motif (Fig. 3). Taken together, conservation of the -RRXFL- motif in Tat signal sequences is not an indispensable factor for high secretion competency in *C. glutamicum*. No other unique features of H- and C-domains of the 10 Tat-type signal sequences of *C. glutamicum* were found.

In addition, eight Sec signals with presumptive lipobox sequences were isolated. Their average length of 22.8 aa was shorter than that of Sec-type signal sequences. The essential cysteine residue in the putative lipobox was situated 21–29 residues from the N terminus (Fig. 4), accounting for the fact that the features of the eight new Sec signals with presumptive lipoboxes were identical to lipoproteins of other bacteria (Sutcliffe & Harrington, 2002).

Fig. 7. Effect of the +1 glutamine on extracellular amylase activity in *C. glutamicum*. After 24 h culture in flasks, α -amylase activities, shown as the mean of triplicates, were measured using an EnzChek Ultra Amylase Assay kit. The y axis indicates the relative intensity of amylase activity with respect to the native signal. Wild-type signals are represented by grey bars and mutated signals by white bars. Δ , Secretion signal constructs with a point mutation (Table 1). Letters indicate amino acid sequences of SPase recognition sites in the wild-type or the substituted sequences used in mutation analysis.

Thirty-one signal sequences gave higher protein secretion than that of PS2 (Fig. 5). These were classified into 26 Sec- and five Tat-type signals (Fig. 5). Although the 108 secretion signal sequences included only 10 Tat-type signals (9.2%), the proportion of Tat increased to 16.1% in 31 high-efficiency signal sequences, suggesting that Tat-dependent secretion may be the more prevalent secretion pathway in *C. glutamicum*. Any common features or motifs for the high-productivity signal sequences are not yet known. However, a glutamine residue often appeared just one amino acid downstream of a putative cleavage site (Table 2). Point mutation analyses of the glutamine (we designated it as the +1 glutamine) changed the amount of secretory amylase in *C. glutamicum* (Fig. 7). The role of the +1 glutamine in secretion is not yet clear, but the glutamine at this position, which might be adjacent to the cleavage site, may affect the state of pre-secretory proteins in the translocator complex.

ACKNOWLEDGEMENTS

We thank Dr C. Omumasaba for critical reading of the manuscript. This study was partly funded by the New Energy and Industrial Technology Development Organization (NEDO).

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.
- Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**, 783–795.
- Berks, B. C., Sargent, F. & Palmer, T. (2000). The Tat protein export pathway. *Mol Microbiol* **35**, 260–274.
- Billman-Jacobe, H., Wang, L., Kortt, A., Stewart, D. & Radford, A. (1995). Expression and secretion of heterologous proteases by *Corynebacterium glutamicum*. *Appl Environ Microbiol* **61**, 1610–1613.
- Brand, S., Niehaus, K., Pühler, A. & Kalinowski, J. (2003). Identification and functional analysis of six mycolyltransferase genes of *Corynebacterium glutamicum* ATCC 13032: the genes *cop1*, *cmt1*, and *cmt2* can replace each other in the synthesis of trehalose dicorynomycolate, a component of the mycolic acid layer of the cell envelope. *Arch Microbiol* **180**, 33–44.
- Brennan, P. J. & Nikaido, H. (1995). The envelope of mycobacteria. *Annu Rev Biochem* **64**, 29–63.
- 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.
- Chen, M. & Nagarajan, V. (1994). Effect of alteration of charged residues at the N termini of signal peptides on protein export in *Bacillus subtilis*. *J Bacteriol* **176**, 5796–5801.
- Cline, K., Henry, R., Li, C. & Yuan, J. (1993). Multiple pathways for protein transport into or across the thylakoid membrane. *EMBO J* **12**, 4105–4114.
- Collins, M. D., Goodfellow, M. & Minnikin, D. E. (1982). A survey of the structures of mycolic acids in *Corynebacterium* and related taxa. *J Gen Microbiol* **128**, 129–149.
- Cristobal, 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. & Von Heijne, G. (1992). Signal peptidases in prokaryotes and eukaryotes – a new protease family. *Trends Biochem Sci* **17**, 474–478.
- Dalbey, R. E., Lively, M. O., Bron, S. & van Dijk, J. M. (1997). The chemistry and enzymology of the type I signal peptidases. *Protein Sci* **6**, 1129–1138.
- Date, M., Yokoyama, K., Umezawa, Y., Matsui, H. & Kikuchi, Y. (2003). Production of native-type *Streptovercillium mobaraense* transglutaminase in *Corynebacterium glutamicum*. *Appl Environ Microbiol* **69**, 3011–3014.
- Date, M., Itaya, H., Matsui, H. & Kikuchi, Y. (2006). Secretion of human epidermal growth factor by *Corynebacterium glutamicum*. *Lett Appl Microbiol* **42**, 66–70.
- Draper, P. (1998). The outer parts of the mycobacterial envelope as permeability barriers. *Front Biosci* **3**, D1253–D1261.
- Gennity, J., Goilstein, J. & Inouye, M. (1990). Signal peptide mutants of *Escherichia coli*. *J Bioenerg Biomembr* **22**, 233–269.
- Hansmeier, N., Chao, T. C., Pühler, A., Tauch, A. & Kalinowski, J. (2006a). The cytosolic, cell surface and extracellular proteomes of the biotechnologically important soil bacterium *Corynebacterium efficiens* YS-314 in comparison to those of *Corynebacterium glutamicum* ATCC 13032. *Proteomics* **6**, 233–250.
- Hansmeier, N., Chao, T. C., Kalinowski, J., Pühler, A. & Tauch, A. (2006b). Mapping and comprehensive analysis of the extracellular and cell surface proteome of the human pathogen *Corynebacterium diphtheriae*. *Proteomics* **6**, 2465–2476.
- Hartmann, M., Barsch, A., Niehaus, K., Pühler, A., Tauch, A. & Kalinowski, J. (2004). The glycosylated cell surface protein Rpf2,

containing a resuscitation-promoting factor motif, is involved in intercellular communication of *Corynebacterium glutamicum*. *Arch Microbiol* **182**, 299–312.

Hermann, T., Pfefferle, W., Baumann, C., Busker, E., Schaffer, S., Bott, M., Sahm, H., Dusch, N., Kalinowski, J. & other authors (2001). Proteome analysis of *Corynebacterium glutamicum*. *Electrophoresis* **22**, 1712–1723.

Inouye, S., Franceschini, T., Sato, M., Itakura, K. & Inouye, M. (1983). Prolipoprotein signal peptidase of *Escherichia coli* requires a cysteine residue at the cleavage site. *EMBO J* **2**, 87–91.

Joliff, G., Mathieu, L., Hahn, V., Bayan, N., Duchiron, F., Renaud, M., Schechter, E. & Leblon, G. (1992). Cloning and nucleotide sequence of the *cspl* gene encoding PS1, one of the two major secreted proteins of *Corynebacterium glutamicum*: the deduced N-terminal region of PS1 is similar to the *Mycobacterium* antigen 85 complex. *Mol Microbiol* **6**, 2349–2362.

Kacem, R., De Sousa-D'Auria, C., Tropis, M., Chami, M., Gounon, P., Leblon, G., Houssin, C. & Daffé, M. (2004). Importance of mycolyltransferases on the physiology of *Corynebacterium glutamicum*. *Microbiology* **150**, 73–84.

Kikuchi, Y., Date, M., Itaya, H., Matsui, K. & Wu, L. F. (2006). Functional analysis of the twin-arginine translocation pathway in *Corynebacterium glutamicum* ATCC 13869. *Appl Environ Microbiol* **72**, 7183–7192.

Kikuchi, Y., Itaya, H., Date, M., Matsui, K. & Wu, L. F. (2007). Production of *Chryseobacterium proteolyticum* protein-glutaminase using the twin-arginine translocation pathway in *Corynebacterium glutamicum*. *Appl Microbiol Biotechnol* **78**, 67–74.

Kinoshita, S. (1985). Glutamic acid bacteria. In *Biology of Industrial Microorganisms*, pp. 115–146. Edited by A. L. Demain & N. A. Solomon. London: Cummings.

Kurusu, Y., Kainuma, M., Inui, M., Satoh, Y. & Yukawa, H. (1990). Electroporation-transformation system for coryneform bacteria by auxotrophic complementation. *Agric Biol Chem* **54**, 443–447.

Liebl, W. & Sinskey, A. J. (1990). *Coryneform expression and secretion system*. USA patent 4965197.

Liebl, W., Sinskey, A. J. & Schleifer, K. H. (1992). Expression, secretion, and processing of staphylococcal nuclease by *Corynebacterium glutamicum*. *J Bacteriol* **174**, 1854–1861.

Malumbres, M., Mateos, L. M. & Martin, J. F. (1995). Microorganisms for amino acid production: *Escherichia coli* and corynebacteria. In *Food Biotechnology Microorganisms*, pp. 423–469. Edited by Y. H. Hui & G. G. Kachourians. New York: VCH Publishers.

Meissner, D., Vollstedt, A., van Dijl, J. M. & Freudl, R. (2007). Comparative analysis of twin-arginine (Tat)-dependent protein secretion of a heterologous model protein (GFP) in three different Gram-positive bacteria. *Appl Microbiol Biotechnol* **76**, 633–642.

Minnikin, D. E. (1982). Lipids: complex lipids, their chemistry, biosynthesis and roles. In *The Biology of the Mycobacteria*, vol. 1, pp. 95–184. Edited by C. Ratledge & J. L. Stanford. London: Academic Press.

Mori, H. & Cline, K. (2001). Post-translational protein translocation into thylakoids by the Sec and Δ pH-dependent pathways. *Biochim Biophys Acta* **1541**, 80–90.

Nishimura, T., Teramoto, H., Vertes, A. A., Inui, M. & Yukawa, H. (2008). ArnR, a novel transcriptional regulator, represses expression

of the *narKGHJI* operon in *Corynebacterium glutamicum*. *J Bacteriol* **190**, 3264–3273.

Paetzel, M., Dalbey, R. E. & Strynadka, N. C. (1998). Crystal structure of a bacterial signal peptidase in complex with a β -lactam inhibitor. *Nature* **396**, 186–190.

Paetzel, M., Karla, A., Strynadka, N. C. & Dalbey, R. E. (2002). Signal peptidases. *Chem Rev* **102**, 4549–4580.

Peyret, J. L., Bayan, N., Joliff, G., Gulik-Krzywicki, T., Mathieu, L., Schechter, E. & Leblon, G. (1993). Characterization of the *cspl* gene encoding PS2, an ordered surface-layer protein in *Corynebacterium glutamicum*. *Mol Microbiol* **9**, 97–109.

Pohlschroder, 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.

Pugsley, A. P. (1993). The complete general secretory pathway in Gram-negative bacteria. *Microbiol Rev* **57**, 50–108.

Qi, H. Y., Sankaran, K., Gan, K. & Wu, H. C. (1995). Structure–function relationship of bacterial prolipoprotein diacylglycerol transferase: functionally significant conserved regions. *J Bacteriol* **177**, 6820–6824.

Riezman, H. (1997). The ins and outs of protein translocation. *Science* **278**, 1728–1729.

Salim, K., Haedens, V., Content, J., Leblon, G. & Huygen, K. (1997). Heterologous expression of the *Mycobacterium tuberculosis* gene encoding antigen 85A in *Corynebacterium glutamicum*. *Appl Environ Microbiol* **63**, 4392–4400.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sankaran, K. & Wu, H. C. (1995). Bacterial prolipoprotein signal peptidase. *Methods Enzymol* **248**, 169–180.

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

Smith, M. D., Flickinger, J. L., Lineberger, D. W. & Schmidt, B. (1986). Protoplast transformation in coryneform bacteria and introduction of an α -amylase gene from *Bacillus amyloliquefaciens* into *Brevibacterium lactofermentum*. *Appl Environ Microbiol* **51**, 634–639.

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.

Sutcliffe, I. C. & Harrington, D. J. (2002). Pattern searches for the identification of putative lipoprotein genes in Gram-positive bacterial genomes. *Microbiology* **148**, 2065–2077.

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.

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.

Yukawa, H., Omumasaba, C. A., Nonaka, H., Kos, P., Okai, N., Suzuki, N., Suda, M., Tsuge, Y., Watanabe, J. & other authors (2007). Comparative analysis of the *Corynebacterium glutamicum* group and complete genome sequence of strain R. *Microbiology* **153**, 1042–1058.

Edited by: C. W. Chen