JOURNAL OF BACTERIOLOGY, Oct. 2000, p. 5592–5595 Vol. 182, No. 19
0021-9193/00/$04.00+0

Revised Translation Start Site for *secM* Defines an Atypical Signal Peptide That Regulates *Escherichia coli secA* Expression

SHAMEEMA SARKER,[1] KENNETH E. RUDD,[2] AND DONALD OLIVER[1]*

Department of Molecular Biology and Biochemistry, Wesleyan University, Middletown, Connecticut 06459,[1] *and Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, Miami, Florida 33101*[2]

Received 30 November 1999/Accepted 6 July 2000

The secretion-responsive regulation of *Escherichia coli secA* occurs by coupling its translation to the translation and secretion of an upstream regulator, *secM* (formerly geneX). We revise the translational start site for *secM*, defining a new signal peptide sequence with an extended amino-terminal region. Mutational studies indicate that certain atypical amino acyl residues within this extended region are critical for proper *secA* regulation.

The eubacterial protein secretion machinery consists of a number of soluble and membrane-associated components (5). One critical element is SecA ATPase, which acts as a molecular motor to promote protein secretion at translocation sites that consist of SecYE, the SecA receptor, and SecG and SecDFyajC proteins, which regulate SecA membrane cycling (6, 7, 9, 15, 24). SecA appears to directly recognize the preprotein and guide its entry into the translocon by utilizing its membrane insertion and retraction activity (1, 8, 10, 12). Since SecA appears to initiate the first committed step in the protein translocation cycle, its level and activity are likely to be carefully regulated.

secA has been shown previously to be regulated at the translational level by the protein secretion-proficient state of the *Escherichia coli* cell, with derepression occurring when protein secretion becomes rate limiting (17, 20). Repression occurs by an autogenous mechanism whereby SecA binds to its translational operator site on geneX *secA* mRNA to block or dislodge ribosomes that initiate at the *secA* ribosome-binding site (21, 22). *secA* translation initiation requires a translational coupling mechanism that employs the upstream gene, geneX (11). Recently we have shown that the basis for the observed secretion-responsive regulation of *secA* relies on the secretability of geneX preprotein, since signal sequence mutations in geneX rendered *secA* expression constitutive and *prlA* signal sequence suppressor alleles restored *secA* regulation in this context (16). These results demonstrated that geneX is an important regulator of *secA*, and accordingly we suggest that it should be renamed *secM* (for "secretion monitor").

In the original sequencing of the *envA secM secA* region, it was suggested that *secM* begins with an AUG codon to encode a 147-amino-acid-residue protein (2). Although it was subsequently pointed out that there are also two potential upstream GUG start sites for *secM* (23), the AUG initiator has been generally accepted as the *secM* start site. However, we show here that this is not the case.

No homologs of *secM* were found in GenBank, but two homologs were identified in the unfinished *Salmonella enterica* serovar typhi and *Yersina pestis* genome sequences kindly made available by the Sanger Centre (the *S. typhi* and *Y. pestis secM* sequence data were produced by the S. typhi and Y. pestis Sequencing Groups at the Sanger Centre and can be obtained from ftp://ftp.sanger.ac.uk/pub/pathogens), and one homolog was found in the *Klebsiella pneumoniae* genome sequence produced by the Genome Sequencing Center at Washington University, St. Louis, Mo. (personal communication). Examination of the homologous sequences strongly suggested that the GUG codon 69 nucleotides upstream of the presumed AUG initiator is the correct translation start for several reasons. First, the protein sequence similarity extends upstream of the current AUG start and stops at the proposed GUG start, which is present in all four organisms (Fig. 1A). Second, the proposed GUG start sites are preceded by Shine-Dalgarno sequences that are able to promote translation initiation, whereas the downstream AUG codons are not (Fig. 1B and data not shown). Third, although it was demonstrated previously that *E. coli secM* possesses a signal sequence (18), the AUG-based signal peptide lacks a basic amino-terminal region (N region) and contains a charged residue (Glu) within its hydrophobic core region (H region); such features impair signal peptide function. In addition, the previously proposed cleavage site, Ala-Lys-Ala, is not conserved in the homologs. By contrast, the newly proposed *secM* signal peptides contain several basic amino acyl residues within their atypically long N region as well as a more typical H region. Furthermore, a new cleavage site nine residues before the old site is predicted by the SignalP program for all four *secM* sequences (14). Since the newly proposed *secM* signal peptide appeared to be more plausible, we performed a genetic test to verify this prediction.

Since we have been previously unsuccessful in raising antibody to SecM for its detection and quantification, we constructed a *secM-phoA* fusion for measurement of SecM expression levels. PCR methods were utilized to amplify a 1.89-kb *secM-phoA* fragment from pCB9 (18) by using a forward primer upstream of *secM* and a reverse primer at the end of *phoA* that also contained a *Bam*HI recognition sequence. This fragment was cut with *Bst*BI and *Bam*HI to generate a 1.6-kb fragment that was cloned into the analogous sites of pPhIF (16), thereby replacing the *secM secA-lacZ* region of pPhIF with the *secM-phoA* fusion to generate pSS1. Two mutations, GTG to GTA (Val) (*secM1*) and ATG to TGC (Cys) (*secM2*), were constructed to test the importance of the potential GUG and AUG initiation codons, respectively (Fig. 2). These mutations were chosen to abolish the potential of the respective start codons to initiate translation while preserving a chemically similar amino acid residue if the codon was noninitiating. The mutations were made in pPhIF using the Quik Change proce-

* Corresponding author. Mailing address: Department of Molecular Biology and Biochemistry, Wesleyan University, Middletown, CT 06459. Phone: (860) 685-3556. Fax: (860) 685-2141. E-mail: doliver@wesleyan.edu.

A

```
SecMec      MSGILTRWRQFGKRYFWPHLLLGMVAASLGLPA-LSNA/AEPNAPAKA//
SecMst      MSGILTRWRQLGRRYFWPHLLLGMVAASFGLPA-LSNA/AETNTPART
SecMkp      MSGLLTRWRQFGRRYFWPHLLLGMVAASLGLPA-LSNGHETA/APAKA//
SecMyp      MIGILNRWRQFGRRYFWPHLLLGMVAASLGVPLNLSGVPDHA/ALANT
```

B

```
                AGGA
secMec      ATGATCGGGATGGCAATAAC-GTGAGTGGA
secMst      CATATCAGGTGGCAATGAC--GTGAGTGGA
secMkp      AAAGAAAGGTAATGAAC----GTGAGTGGA
secMyp      CTCCGGCGGATTTAAACTC-GTGATCGGT
```

FIG. 1. (A) Predicted amino acid sequence of the 5′ portion of *secM* of *E. coli* (ec), *S. typhi* (st), *K. pneumoniae* (kp), and *Y. pestis* (yp) utilizing the proposed GUG initiator. The translation start site from the previously proposed AUG initiator is in bold. The predicted new and old signal peptide processing sites are indicated by the slash and double slash, respectively. (B) DNA sequences upstream of the proposed GTG initiator (in bold) of *secM*. The most conserved portion of the Shine-Dalgarno sequence is aligned above the corresponding region upstream of *secM*.

dure as described by the manufacturer (Stratagene) and confirmed by DNA sequence analysis. This resulted in pPhIF derivatives pSS2 and pSS3 that contained *secM1* and *secM2*, respectively. In order to generate *secM-phoA* fusions containing the *secM1* and *secM2* alleles, pSS2 and pSS3 were cut with *Bst*BI and *Bam*HI and the vector fragment was isolated and ligated to the 1.6-kb *Bst*BI-*Bam*HI DNA fragment containing the *secM-phoA* fusion to generate pSS4 and pSS5, respectively. These latter two plasmids are allelic with pSS1. CC118 containing each of these plasmids was assayed for alkaline phosphatase activity (Fig. 3). The results suggest strongly that the GUG codon and not the downstream AUG codon is the correct translation start site of *secM*, since the *secM1* mutant essentially lacked alkaline phosphatase activity, while the *secM2* mutant had alkaline phosphatase activity that was similar to that of the wild-type strain containing the *secM-phoA* fusion.

We next examined the effects of the *secM1* and *secM2* mutations on *secA* regulation using pPhIF that contains *secM* and a *secA-lacZ* translational fusion (11) or derivatives containing the *secM1* and *secM2* alleles. Secretion-proficient and secretion-defective conditions for examining *secA* regulation were created by utilizing a wild-type strain (CG155) or its isogenic *secD1*(Cs) derivative (CG29) as hosts. The *secM1* alteration essentially abolished *secA* expression under both secretion-proficient and secretion-defective conditions, while *secA* expression and its regulation were normal with the *secM2* mutation (Fig. 4A). These results are also consistent with the GUG start of *secM* as well as the documented translational coupling between *secM* and *secA* (11).

The newly proposed *secM* signal peptide (Fig. 1) is unusually long for a gram-negative organism due to its extended N region. Interestingly, this region can be modeled as an amphipathic helix (Fig. 5). Since the magnitude of the positive charge in the N region has been shown to be important for the affinity of SecA for signal peptides (1), the *secM* signal peptide may have a unique interaction with SecA that is related to its regulatory function. The Sec-independent twin arginine (TAT) signal peptides also possess a long N region, accounting for their larger size (4). Although the proposed *secM* signal peptides do not contain a twin arginine motif per se, we note that they do possess Arg-Arg or Lys-Arg pairs (Fig. 1). Furthermore, a comparison of the positional preferences of TAT- and Sec-dependent signal peptides from gram-negative bacteria for particular amino acid residues identified a striking preference for proline in the −6 position relative to the cleavage site for the former but not latter system (4). *E. coli* and *Salmonella* serovar typhi *secM* signal peptides do contain a proline at this position. As noted previously, a Pro-to-Arg substitution at this position causes derepression of *secA* expression (16).

In light of these intriguing observations and the fact that the TAT and Sec pathways may be in competition with one another (4), we investigated the effect of *tat* gene deletions on

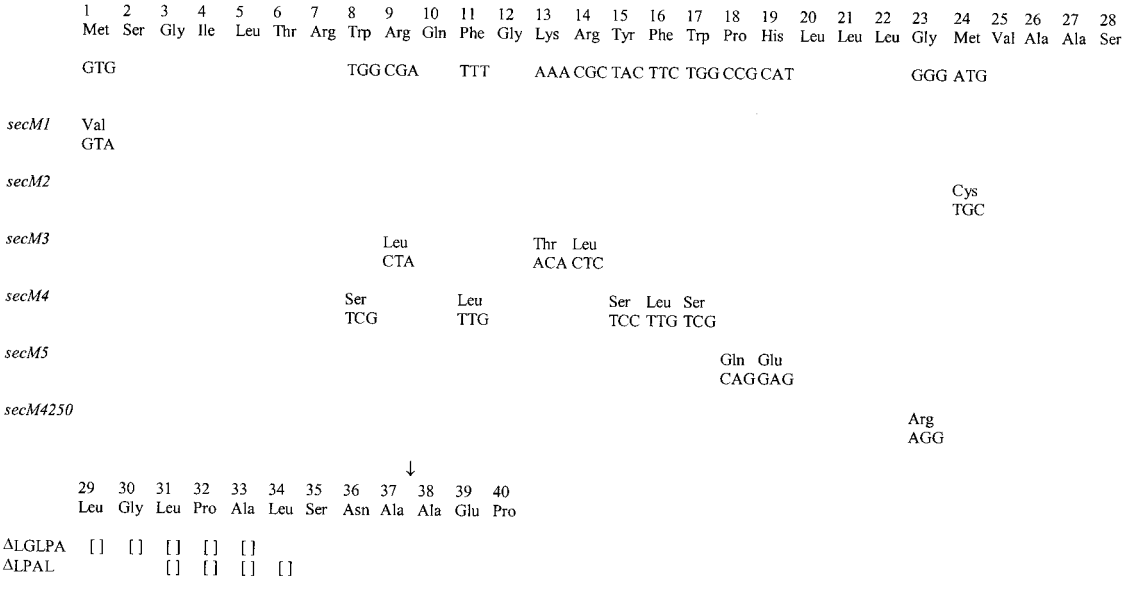

FIG. 2. *secM* signal sequence alleles constructed in this study. The codon and amino acid substitutions for each allele are depicted. Empty brackets denote deletions. The presumed signal peptide processing site is indicated by an arrow.

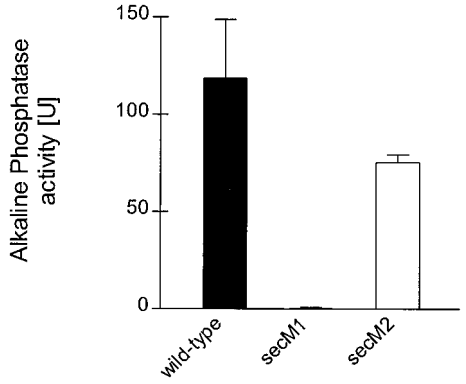

FIG. 3. CC118 [MC1000 *phoA20 rpsE rpoB argE*(Am) *recA1*] containing pSS1 (wild type) or the indicated allelic derivative was grown in Luria broth containing 100 μg of ampicillin per ml at 37°C to mid-logarithmic phase. Alkaline phosphatase assays were performed in duplicate for each of two duplicate cultures as described previously (18). The average result is given, with the error bar indicating the standard deviation.

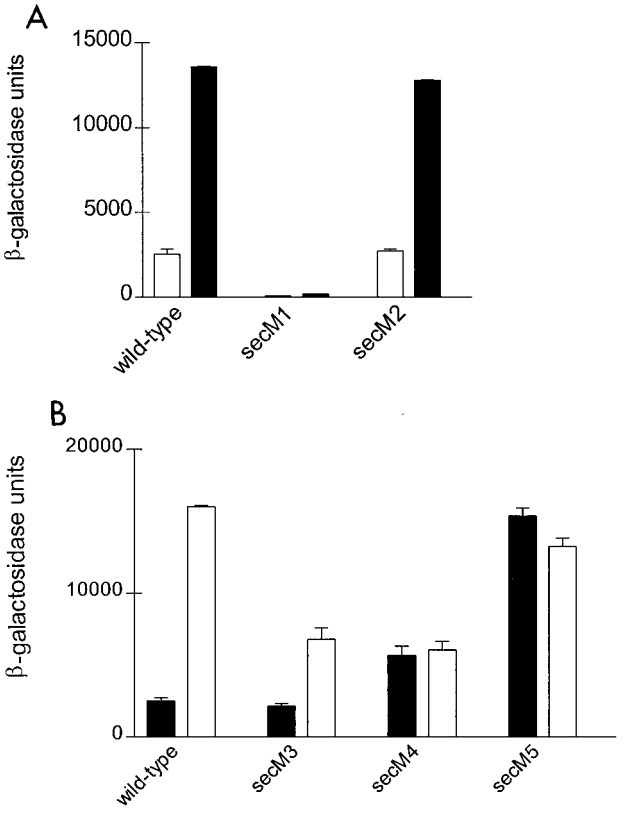

FIG. 4. CG155 (MC1000 *recA*) (□) and CG29 [MC1000 *recA1 secD1*(Cs) *phoR srl*::Tn*10*] (■) containing pPhIF (wild type) or the indicated allelic derivative was grown in Luria broth containing 100 μg of ampicillin per ml at 39°C to mid-logarithmic phase when the culture was shifted to 23°C for 4 h. β-Galactosidase assays were performed in duplicate for each of two duplicate cultures as described previously (13). The average result is given with the error bar indicating the standard deviation.

secA regulation by utilizing Δ*tatA*, Δ*tatE*, Δ*tatAE*, and Δ*tatC* mutants (4) containing pPhIF. *secA* expression was normal in all of these strains as indicated by β-galactosidase activities that were similar to wild type (data not shown). We next examined the importance of the positively charged and aromatic amino acid residues within the N region of the *E. coli secM* signal peptide on *secA* regulation. For this purpose the *secM3* and *secM4* alleles were made employing the Quik Change procedure on pPhIF to generate pSS6 and pSS7, respectively (Fig. 2). Examination of *secA* regulation showed that the *secM3* alteration caused a modest decrease in *secA* expression during a protein secretion block (Fig. 4B), suggesting that the highly positively charged N region and the Lys Arg motif were not essential for some level of *secA* regulation. By contrast, the *secM4* alteration resulted in an increased basal level of *secA* expression and an inability to derepress *secA* expression further. Collectively these results suggest that the *secM* signal peptide is recognized by the Sec pathway and that certain aromatic amino acyl residues within its N region may be important for proper *secA* regulation.

Reassignment of the translation start site of *secM* is consistent with and illuminates a number of previous observations. First, the mutations that we assigned previously to the *secM* signal sequence are still within this structure (ΔLGLPA and ΔLPAL) (Fig. 2), although they are located in the distal rather than the proximal end of the H region (16). Second, in an effort to overexpress *secM*, we previously engineered a better Shine-Dalgarno sequence immediately preceding the presumed AUG start site of *secM* (*secM5*) (Fig. 2). Based on our results here, this mutation causes polar and charged amino acid residues (Gln and Glu) to be substituted into the H region of the *secM* signal peptide, and it should lead to derepression of *secA* expression similarly to other *secM* signal sequence defects (16). Indeed, this prediction is precisely what was observed (Fig. 4B). Third, Cook and Kumamoto (3) reported recently that *secA* overexpression can compensate for certain types of *secB* defects. The *secM4250* allele (referred to as *secA4250*) that overproduces SecA protein 12-fold was suggested to affect *secM* translation, since it mapped three nucleotides upstream of the previously proposed *secM* translation start site (Fig. 2). However, consistent with its phenotype, this mutation is predicted to cause a severe disruption in the function of the reassigned *secM* signal peptide, resulting in a Gly-to-Arg substitution within the middle of the H region. Of note, all of the *secM* signal sequence mutations weaken the SignalP prediction relative to the wild type, but in most cases the various scored parameters do not drop below threshold values. This may indicate that *secM*-mediated regulation of *secA* is particularly sensitive to signal sequence variations. Finally, Riggs et al. (19) utilized a genetic selection based on up-regulation of a *secA-lacZ* fusion to select for *sec* mutants. Surprisingly, the vast

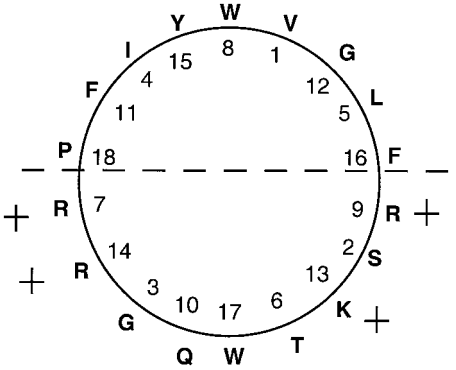

FIG. 5. Helical wheel diagram of the first 18 amino acyl residues of the *E. coli secM* signal peptide. A plus sign indicates the positively charged amino acid residues, and the dashed line depicts the boundary between the hydrophilic and hydrophobic faces of the α helix.

majority of Lac$^+$ mutants mapped to the *secA-lacZ* fusion. This result is readily understood since loss-of-function mutations can be obtained within the *secM* signal sequence as well as within novel *sec* genes that would have this phenotype. Additional studies will be required to investigate the unique features of the *secM* signal sequence and their relevance to promoting proper *secA* regulation.

We thank the Genome Sequencing Center, Washington University, St. Louis, Mo., for communication of the *K. pneumoniae* sequence data prior to publication. We thank Gunnar von Heijne for the kind gift of the *tat* mutant strains and Koreaki Ito for the suggestion of the *secM* nomenclature.

This work was supported by grants GM42033 and GM58560 from the National Institutes of Health to D.O. and K.R., respectively.

REFERENCES

1. **Akita, M., S. Sasaki, S. Matsuyama, and S. Mizushima.** 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.
2. **Beall, B., and J. Lutkenhaus.** 1987. Sequence analysis, transcriptional organization, and insertional mutagenesis of the *envA* gene of *Escherichia coli*. J. Bacteriol. **169:**5408–5415.
3. **Cook, H. A., and C. A. Kumamoto.** 1999. Overproduction of SecA suppresses the export defect caused by a mutation in the gene encoding the *Escherichia coli* export chaperone SecB. J. Bacteriol. **181:**3010–3017.
4. **Cristobal, S., J.-W. de Gier, H. Nielsen, and G. von Heijne.** 1999. Competition between Sec- and TAT-dependent protein translocation in *Escherichia coli*. EMBO J. **18:**2982–2990.
5. **Driessen, A. J. M., P. Fekkes, and J. van der Wolk.** 1998. The Sec system. Curr. Opin. Microbiol. **1:**216–222.
6. **Duong, F., and W. Wickner.** 1997. Distinct catalytic roles of the SecYE, SecG, and SecDFyajC subunits of preprotein translocase holoenzyme. EMBO J. **16:**2756–2768.
7. **Duong, F., and W. Wickner.** 1997. The SecDFyajC domain of preprotein translocase controls preprotein movement by regulating SecA membrane cycling. EMBO J. **16:**4871–4879.
8. **Economou, A., and W. Wickner.** 1994. SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. Cell **78:**835–843.
9. **Hartl, F.-U., S. Lecker, E. Schiebel, J. P. Hendrick, and W. Wickner.** 1990. The binding cascade of SecB to SecA to SecY/E mediates preprotein targeting to the E. coli plasma membrane. Cell **63:**269–279.
10. **Lill, R., W. Dowhan, and W. Wickner.** 1990. The ATPase activity of SecA is regulated by acidic phospholipids, SecY, and the leader and mature domains of precursor proteins. Cell **60:**271–280.
11. **McNicholas, P., R. Salavati, and D. Oliver.** 1997. Dual regulation of *Escherichia coli secA* translation by distinct upstream elements. J. Mol. Biol. **265:**128–141.
12. **Miller, A., L. Wang, and D. Kendall.** 1998. Synthetic signal peptides specifically recognize SecA and stimulate ATPase activity in the absence of preprotein. J. Biol. Chem. **273:**11409–11412.
13. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
14. **Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne.** 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10:**1–6.
15. **Nishiyama, K.-I., T. Suzuki, and H. Tokuda.** 1996. Inversion of the membrane topology of SecG coupled with SecA-dependent preprotein translocation. Cell **85:**71–81.
16. **Oliver, D., J. Norman, and S. Sarker.** 1998. Regulation of *Escherichia coli secA* by cellular protein secretion proficiency requires an intact gene X signal sequence and an active translocon. J. Bacteriol. **180:**5240–5242.
17. **Oliver, D. B., and J. Beckwith.** 1982. Regulation of a membrane component required for protein secretion in Escherichia coli. Cell **30:**311–319.
18. **Rajapandi, T., K. M. Dolan, and D. B. Oliver.** 1991. The first gene in the *Escherichia coli secA* operon, gene X, encodes a nonessential secretory protein. J. Bacteriol. **173:**7092–7097.
19. **Riggs, P. D., A. I. Derman, and J. Beckwith.** 1988. A mutation affecting the regulation of a *secA-lacZ* fusion defines a new *sec* gene. Genetics **118:**571–579.
20. **Rollo, E. E., and D. B. Oliver.** 1988. Regulation of the *Escherichia coli secA* gene by protein secretion defects: analysis of *secA*, *secB*, *secD*, and *secY* mutants. J. Bacteriol. **170:**3281–3282.
21. **Salavati, R., and D. Oliver.** 1995. Competition between ribosome and SecA binding promotes *Escherichia coli secA* translational regulation. RNA **1:**745–753.
22. **Schmidt, M. G., and D. B. Oliver.** 1989. SecA protein autogenously represses its own translation during normal protein secretion in *Escherichia coli*. J. Bacteriol. **171:**643–649.
23. **Schmidt, M. G., E. E. Rollo, J. Grodberg, and D. B. Oliver.** 1988. Nucleotide sequence of the *secA* gene and *secA*(Ts) mutations preventing protein export in *Escherichia coli*. J. Bacteriol. **170:**3404–3414.
24. **van der Wolk, J. P. W., J. G. de Wit, and A. J. M. Driessen.** 1997. The catalytic cycle of the *Escherichia coli* SecA ATPase comprises two distinct preprotein translocation events. EMBO J. **16:**7297–7304.