JOURNAL OF BACTERIOLOGY, Aug. 1988, p. 3404–3414
0021-9193/88/083404-11$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 170, No. 8

Nucleotide Sequence of the *secA* Gene and *secA*(Ts) Mutations Preventing Protein Export in *Escherichia coli*

MICHAEL G. SCHMIDT, ELLEN E. ROLLO, JENNIFER GRODBERG, AND DONALD B. OLIVER*

Department of Microbiology, State University of New York at Stony Brook, Stony Brook, New York 11794-8621

Received 24 March 1988/Accepted 28 April 1988

The DNA sequence of the *secA* gene, essential for protein export in *Escherichia coli*, was determined and found to encode a hydrophilic protein of 901 amino acid residues with a predicted molecular weight of 101,902, consistent with its previously determined size and subcellular location. Sequence analysis of 9 *secA*(Ts) mutations conferring general protein export and *secA* regulatory defects revealed that these mutations were clustered in three specific regions within the first 170 amino acid residues of the SecA protein and were the result of single amino acid changes predicted to be severely disruptive of protein structure and function. The DNA sequence immediately upstream of *secA* was shown to encode a previously inferred gene, gene X. Sequence analysis of a conditionally lethal amber mutation, *am109*, previously inferred to be located proximally in the *secA* gene, revealed that it was located distally in gene X and was conditionally lethal due to its polar effect on *secA* expression. This and additional evidence are presented indicating that gene X and *secA* are cotranscribed.

Biochemical and genetic approaches have been used to study the molecular mechanisms responsible for protein localization in procaryotic and eucaryotic cells. Such studies reveal that protein translocation across the bacterial plasma membrane and the eucaryotic rough endoplasmic reticular membrane occur by apparently homologous and conserved mechanisms. Shared features include similar signal sequences and stop transfer sequences on presecretory protein precursors that are required for their correct localization and topology (46), frequent removal of amino-terminal signal peptides by signal peptidases of similar specificity (45), a dependence on an energy source such as ATP or electrochemical membrane potential for membrane translocation (6, 13, 14), and a requirement for some type of export machinery (29, 35, 44).

Biochemical analysis of an in vitro secretion system from mammalian endoplasmic reticulum has led to the identification of four components of the export machinery. Two components constitute a receptor-mediated system, consisting of a cytosolic signal recognition particle and its membrane receptor, which allow targeting of nascent presecretory precursors to sites in the reticular membrane (44). A third component is a recently identified integral membrane protein, the signal sequence receptor, which appears to interact with the signal peptide once it has been delivered to the membrane by the receptor mediated system (47). Finally, a membrane protein complex that contains the signal peptidase activity has been purified and found to consist of six different polypeptide chains (10). Biochemical analysis of an in vitro secretion system from *Escherichia coli* allowed the identification of three different components: two distinct signal peptidases for processing lipoprotein and general protein precursors (43, 49) and a cytoplasmic 12S particle required for the transfer of presecretory proteins to membrane vesicles (28).

Genetic selective approaches have been used in bacteria and yeasts to identify a set of genes whose products are required to promote secretion and which probably encode components of the export machinery. In *Saccharomyces cerevisiae*, distinct sets of genes have been identified which are involved in protein trafficking through each of the different secretory organelles (35). In *E. coli*, two prototypic genetic selections have been used to isolate mutants pleiotropically altered in protein export. One paradigm has been to select for extragenic suppressor mutations which restore the secretion of a previously export-defective presecretory protein. This selection has yielded mutants capable of suppressing defects in the signal peptide of exported proteins and has identified three genes: *prlA*/*secY*, *prlC*, and *prlD* (3, 9). The second paradigm has been to select for secretion-defective mutations which, by altering the localization of a presecretory precursor–β-galactosidase hybrid protein from the membrane to the cytoplasm, increase its enzymatic activity. This selection has yielded mutants defective in protein export and has identified three genes: *secA*, *secB*, and *secD* (11, 20, 30). In addition, a secretion-defective *secY* allele has been isolated by using localized mutagenesis techniques (39). Temperature-sensitive or cold-sensitive conditionally lethal alleles of *secA*, *secD*, and *secY* have been obtained which show temperature-dependent defects in the export of most periplasmic and outer membrane proteins (23, 48), indicating the essential nature of these genes and of the protein export process for cell viability. Conditionally lethal mutations of *secA*, *secD*, and *secY* also show a 10- to 20-fold increase in SecA levels when secretion blocks are imposed by temperature shift, indicating that *secA* expression is somehow regulated by protein export proficiency (11, 31, 32a).

Identification of these *sec* genes in *E. coli* opens the way to a molecular genetic and biochemical analysis to determine the structure and function of individual Sec proteins. The DNA sequence of the *secY* gene predicts a 49-kilodalton (kDa) highly hydrophobic protein, in agreement with the result that SecY was found to be an integral plasma membrane protein only solubilized in the presence of a strong detergent such as sodium dodecyl sulfate (SDS) (5, 16). Antiserum against a SecA-LacZ fusion protein was used to identify SecA as a polypeptide of approximately 92 kDa that is found either in the cytoplasm or loosely associated with the cytoplasmic membrane, depending on the ionic strength

* Corresponding author.

of the buffers used (31). In order to continue the analysis of the structure and function of the SecA protein in secretion, we have sequenced the *secA* gene and a number of different temperature-sensitive *secA* mutations. In addition, in order to begin to study how the regulation of *secA* is responsive to the protein export status of the cell, we have sequenced the region upstream of the *secA* gene.

MATERIALS AND METHODS

Bacterial strains. The wild-type *E. coli* strain MC4100 (F^- Δ*lacU169 araD136 relA rpsL thi*) and its derivative MM18, containing a *malE-lacZ* fusion, have been described (17). The following MC4100 derivatives containing *secA*(Ts) mutations were used: MM52 (30) and MQ5, MQ7, MQ9, MQ10, MQ11, MQ13, MQ15, and MQ16 (Quinlin, Oliver, and Beckwith, unpublished results). Strain DO265 is an MC4100 derivative containing a *supF*(Ts) mutation encoding a temperature-sensitive nonsense suppressor and a previously inferred amber mutation in the *secA* gene [*secA109*(Am) (31)]. Strain JM101 [Δ(*lac-pro*) *supE thi* F′ *traD36 proA*$^+$*B*$^+$ *lacI*^{q} *lacZ* M15) was used as a host for all work involving bacteriophage M13 (26). Strain MM294 (F^- *hsdR endA thi supE44*) was used to propagate all plasmids and was obtained from Mark Ptashne.

Media and reagents. Media used for growth of bacteria and bacteriophage have been described previously (27). Restriction enzymes were obtained from New England Biolabs, Bethesda Research Laboratories, and International Biotechnologies Incorporated, and were used as recommended by the manufacturers. DNA ligase and DNA polymerase I were obtained from New England Biolabs, and the Klenow fragment of DNA polymerase I and the "cyclone" M13 deletion kit were obtained from International Biotechnologies Inc. Radiochemicals were purchased as follows: [α-^{32}P]dATP (600 to 800 Ci/mmol) from ICN Radiochemicals or New England Nuclear, [^{35}S]methionine (~1,000 Ci/mmol) from Amersham, and [^{3}H]leucine (50 Ci/mmol) and [^{3}H]lysine (50 Ci/mmol) from ICN Radiochemicals. Isopropyl-β-D-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-GAL) were obtained from Sigma and Boehringer Mannheim Biochemicals, respectively. ATP, deoxynucleotide triphosphates, and dideoxynucleotide triphosphates were obtained from Pharmacia. Sequenase was purchased from United States Biochemical Corporation. Acrylamide, *N,N*′-methylenebisacrylamide, and *N,N,N*′*,N*′-tetramethylethylenediamine were obtained from Serva. Urea and cesium chloride were obtained from International Biotechnologies Inc. Formamide was obtained from Manufacturing Chemists Inc.

DNA sequencing and data analysis. Bacteriophage lambda DO20 was the primary source of the *secA* gene (32). Lambda JFL40 DNA was the source of the 0.46-kilobase (kb) *Pvu*II-*Eco*RI fragment containing sequences upstream of *secA* (24). M13mp8, M13mp9, M13mp18, and M13mp19 DNAs were purchased from New England Biolabs. Two M13 junction clones were constructed in order to connect DNA sequence derived from the 0.46-kb *Pvu*II-*Eco*RI, 0.8-kb *Eco*RI, and 2.5-kb *Eco*RI restriction enzyme fragments. A *Pvu*II-*Hin*III fragment spanning the first junction was subcloned from pLG552 (18) into M13mp18, forming M13 MS300. A *Hin*III fragment spanning the second junction was subcloned from lambda DO20 into M13mp9, forming M13 LL8. The nucleotide sequences of all the cloned fragments were determined by the dideoxy chain termination method of Sanger et al. (34) as adapted to M13 (36). Ambiguities in the DNA sequence were resolved by the use of dITP in the sequencing reactions with the Sequenase kit according to the manufacturer's instructions. Sequencing gels of 6% acrylamide, 7 M urea, and TBE (25) were run at 30 mA for 1.3 to 3 h. Gels were fixed in 10% glacial acetic acid for 15 min, rinsed in water for 15 min, dried under vacuum on a Hoeffer gel drier, and subjected to autoradiography. Sequence data obtained were compiled and analyzed with a gel reader and the Sequence Analysis System computer program for the IBM PC, purchased from International Biotechnologies, Inc.

Amino-terminal protein sequence analysis. *E. coli* MM18, which overproduces SecA protein 10- to 20-fold when grown in the presence of maltose, was used to prepare SecA for amino-terminal sequence analysis. MM18 was grown in M63 minimal medium containing 0.4% glycerol at 30°C until early logarithmic phase, when maltose was added to 2%, and growth proceeded for an additional 3 h. The culture was split into two 4-ml portions, and each was double labeled with 204 μCi of [^{35}S]methionine and 200 μCi of [^{3}H]leucine or 204 μCi of [^{35}S]methionine and 200 μCi of [^{3}H]lysine for 30 min. Cultures were chilled on ice and sedimented at 12,000 × *g* for 10 min, and cell pellets were recovered. Protein solubilization, SecA protein immunoprecipitation, and SDS-polyacrylamide gel electrophoresis (PAGE) were done as described previously (23). After PAGE, the gel was dried immediately without fixing and subjected to autoradiography. The band containing the SecA protein was excised from the dried gel, and the protein was eluted from the gel matrix via an ISCO electrophoretic sample concentrator as described by Semler et al. (38). The sample was applied to the spinning cup of the Beckman 890C protein sequencer directly in the electrophoretic elution buffer, and the sample was dried with the Beckman sample application subroutine. Apomyoglobin (2 mg) was applied to the sample cup in a separate application to serve as a carrier protein and as an internal standard. Sequence analysis was done essentially as described (2) except that a 0.1 M Quadrol buffer program (Beckman no. 030176) was used for all samples and phenylisothiocyanate was omitted during the first sequencing cycle (residue 0).

Construction of *secA*-CAT gene fusion. The 0.8-kb and 2.5-kb *Eco*RI fragments comprising the *secA* gene were cloned into the *Eco*RI site of pBR325, resulting in plasmid psecA, in which *secA* is in the same direction as the chloramphenicol acetyltransferase (CAT) gene. A *secA*-CAT gene fusion was constructed by partially cleaving psecA with *Nco*I at two out of three sites and religating the plasmid. The resulting plasmid, pAJ1, contained the correct *secA*-CAT gene fusion as determined by restriction enzyme analysis and DNA sequence analysis across the fusion joint.

Genetic mapping and cloning of *secA* mutations. In order to map the *secA*(Ts) and *secA*(Am) mutations, a set of cosmids (all pBR322 derivatives) were constructed which carried various restriction enzyme fragments of the *secA* gene. Bacteriophage lambda 616 was grown on each of the cosmid-containing strains to produce a lysate, and 0.02 ml of each lysate was spotted onto TYE plates or lactose tetrazolium plates that had been spread with 0.1 ml of an overnight culture of a given mutant. The plates were incubated overnight at 42°C, and the production of wild-type Ts$^+$ or Am$^+$ recombinants within a given spot was scored. In certain cases the sensitivity of this mapping technique was increased as follows. A given cosmid was transduced into a given *secA* mutant by spotting the appropriate lysate on a TYE plate containing ampicillin (100 μg/ml) that had been prespread with the *secA* mutant. After overnight growth at 30°C, the

ampicillin-resistant transductants were purified and tested at 42°C for the ability to form Ts^+ or Am^+ recombinants. Chromosomal DNA was prepared for each *secA* mutant, digested with the appropriate restriction enzymes, and cloned into M13mp18 or M13mp19 vectors. After transfection and plating, phage plaques were transferred to nitrocellulose filters and probed with a *secA*-specific probe to the desired region essentially as described by Maniatis et al. (25). Hybridizing plaques were repurified and retested, and phage stocks were prepared for DNA sequence analysis.

Analysis of protein secretion defects. Bacteria were grown in M63 minimal medium containing 0.4% glycerol and 0.4% maltose at the appropriate temperature to 2×10^8 to 4×10^8 cells per ml. Cultures (0.5 ml) were labeled with 10 μCi of L-[^{35}S]methionine for 2 min, followed by the addition of an equal volume of ice-cold 10% trichloroacetic acid to halt radioisotope incorporation. Cells were sedimented in a Microfuge for 1 min at 4°C, washed with 0.5 ml of cold acetone, and resedimented. The cell pellet was suspended in 0.05 ml of ABB buffer (1% SDS, 1 mM EDTA, 10 mM Tris, pH 8) and heated at 100°C for 2 min. Immunoprecipitation, SDS-PAGE, and autoradiography were done as described previously (23) except that autoradiography was carried out on Kodak X-Omat AR film that had been flashed according to the manufacturer's specifications. Protein levels were determined by densitometry of the autoradiograms with a Hoeffer GS-300 scanning densitometer.

RESULTS

Sequence of the *secA* gene. We wanted to determine the DNA sequence of the *secA* gene and of the region upstream of *secA* by using the dideoxy chain termination method (34) as applied to bacteriophage M13 (36). Two *Eco*RI fragments of 0.8 kb and 2.5 kb, which together encode the *secA* gene, and an adjacent *Pvu*II-*Eco*RI fragment of 0.46 kb, comprising sequences upstream of *secA*, were cloned into M13mp18 and M13mp19 vectors. Starting with these primary M13 clones, two methods were used to generate secondary M13 clones containing sets of overlapping deletions. First, deletions were constructed by the judicious use of one or two restriction enzymes which together cleaved both within the insert and polylinker regions, thereby deleting the intervening DNA sequences after blunting (when necessary) and ligation of the two DNA ends. Second, the method of Dale et al. (8) was used to construct a nested set of deletions. Briefly, this method involves making deletions on single-stranded M13 DNA by restriction enzyme cleavage of the DNA at a short duplex created by oligonucleotide annealing, 3′ to 5′ exonuclease treatment of the DNA with T4 DNA polymerase, homopolymer tailing of the shortened molecule with terminal transferase, and noncovalent recircularization with an oligonucleotide complementary to both the 5′ nondeleted and 3′ tailed regions. These two methods generated a set of secondary M13 clones sufficient to determine the complete sequence of these three restriction fragments. In addition, two junction clones (MS300 and LL8, Fig. 1) were constructed as described in Materials and Methods in order to allow the nucleotide sequence of the three fragments to be connected. The sequence information collected from each M13 clone used in this study is shown in Fig. 1. It is apparent that the DNA sequences of both strands of the entire region have been determined.

The DNA sequence upstream of and encoding the *secA* gene is shown in Fig. 2. The first 14 nucleotides of the sequence correspond to the end of the *E. coli* cell division gene *envA*, which terminates in an ochre codon (4). Starting at nucleotide position 44 and extending through nucleotide position 96 was a region capable of forming a double stem-loop structure followed by a run of T's (arrows in Fig. 2), which is characteristic of a rho-independent transcriptional terminator. In fact, this sequence functions as a strong transcriptional terminator in vivo (4). This transcriptional terminator should effectively uncouple the transcription of the upstream cell division genes *ftsQ*, *ftsA*, *ftsZ*, and *envA* from similarly oriented downstream genes. Starting at nucleotide position 317 and extending through nucleotide position 757 was an open reading frame, commencing with ATG and terminating with TAA, which would encode a polypeptide of 147 amino acid residues with a predicted molecular weight of 16,033. Alternatively, if GTG is used as an initiation codon, as has been seen in *E. coli* (15), then this open reading frame could be extended at its 5′ end to begin at either nucleotide position 173 or 248 (boxed nucleotides in Fig. 2). These two extended open reading frames would encode polypeptides of 193 or 170 amino acid residues with predicted molecular weights of 21,705 and 18,846, respectively. We know that one of these three polypeptides was in fact made, since we have sequenced an amber mutation in this open reading frame (G to A transition at nucleotide position 711; see below). The existence of this region was first noted by Oliver and Beckwith (32), who found that there was sufficient distance between *envA* and *secA* to encode a protein of 20,000 to 30,000 daltons. Sullivan and Donachie (41) later named this region gene X. They used in vitro transcription and runoff assays to map a promoter within this region approximately 295 nucleotides upstream from the *Eco*RI site at nucleotide position 463. Based on −35 and −10 promoter consensus elements and spacing constraints (33), there were two potential promoters of average strength located between nucleotides 195 and 249 (underlined nucleotides in Fig. 2). The 5′ promoter would allow the synthesis of either the 16,033-dalton or 18,846-dalton form of the gene X polypeptide, whereas the 3′ promoter would only allow the synthesis of the shorter polypeptide.

At 64 nucleotides past the end of gene X there was a very long open reading frame, starting at nucleotide position 822 and extending through nucleotide position 3524, which encoded the SecA polypeptide with a predicted 901 amino acid residues, corresponding to a molecular weight of 101,902. There did not appear to be a separate promoter for *secA* at the end of gene X or within the gene X-*secA* intergenic region. In fact, it appeared that *secA* was cotranscribed with gene X (see below). There was a GAG sequence 8 nucleotides upstream of *secA* (underlined nucleotides in Fig. 2) which could serve as a ribosome-binding site (40). The predicted size of the SecA protein (101,902 daltons) is larger than that originally reported (92,000 daltons) (32). However, this estimate was based on comigration of SecA protein with

FIG. 1. *secA* sequencing strategy. The DNA sequences collected from the M13 clones used in this study are depicted at the top of the figure by arrows drawn to scale, with the clone designation given above the arrow. A map of common restriction enzyme sites used in these studies is presented in the middle part of the figure. The positions of gene X and *secA* on this map are indicated at the bottom of the figure. In this figure the start site for gene X is assumed to be nucleotide position 317.

MS 110
MS 115
MS 111
MS 116
MS 137
MS 140
LL 14
JGS 3
BF 1
LL 4
JGS 7
MS 122
BF 2
SL 1
JG 621
JG 1
JG 134
MS 117
LL 8
JG 612
JG 144
MS 136
MS 300
MS 118
BF 3
MS 134
JG 631
MS 275
MS 125
MS 157
MS 119
MS 133
JG 641
MS 121
MS 135
MS 250
LL 13
MS 131
JG 651
MS 132
MS 200
MS 120
MS 130

Pvu II
Pst I
Hpa
Eco R I
BssH II
Hpa I
Nco I
Hind III
Eco R I
Hind III
Mlu I
BssH II
Sal I
Hind III
Bgl II
Nru I
Kpn I
Pst I
SnaB I
Pst I
Nco I
Mlu I
EcoR I

0.187
0.418
0.463
0.152
0.375
0.448
0.555
0.80
0.269
0.500
0.528
0.748
0.821
1.064
1.184
1.678
2.153
2.247
2.384
2.548

Gene X
SecA

```
                    End EnvA
CAGC TGT ACT GGC ATA ACG ACA TTT ATA CTG TCG TAT AAA ATT CGA CTG GCA AAT CTG GCA    61

 CTC TCT CCG GCC AGG TGA ACC AGT CGT TTT TTT TTG AAT TTT ATA AGA GCT ATA AAA AAC   121

 GGT GCG AAC GCT GTT TTC TTA AGC ACT TTT CCG CAC AAC TTA TCT TCA TTC GTG CTG TGG   181
                         -35                          -10     -35
 ACT GCA GGC TTT AAT GAT AAG ATT TGT GCG CTA AAT ACG TTT GAA TAT GAT CGG GAT GGC   241
   -10
 AAT AAC GTG AGT GGA ATA CTG ACG CGC TGG CGA CAG TTT GGT AAA CGC TAC TTC TGG CCG   301

 CAT CTC TTA TTA GGG ATG GTT GCG GCG AGT TTA GGT TTG CCT GCG CTC AGC AAC GCC GCC   361
                     Met Val Ala Ala Ser Leu Gly Leu Pro Ala Leu Ser Asn Ala Ala

 GAA CCA AAC GCG CCC GCA AAA GCG ACA ACC CGC AAC CAC GAG CCT TCA GCC AAA GTT AAC   421
 Glu Pro Asn Ala Pro Ala Lys Ala Thr Thr Arg Asn His Glu Pro Ser Ala Lys Val Asn

 TTT GGT CAA TTG GCC TTG CTG GAA GCG AAC ACA CGC CGC CCG AAT TCG AAC TAT TCC GTT   481
 Phe Gly Gln Leu Ala Leu Leu Glu Ala Asn Thr Arg Arg Pro Asn Ser Asn Tyr Ser Val

 GAT TAC TGG CAT CAA CAT GCC ATT CGC ACG GTA ATC CGT CAT CTT TCT TTC GCA ATG GCA   541
 Asp Tyr Trp His Gln His Ala Ile Arg Thr Val Ile Arg His Leu Ser Phe Ala Met Ala

 CCG CAA ACA CTG CCC GTT GCT GAA GAA TCT TTG CCT CTT CAG GCG CAA CAT CTT GCA TTA   601
 Pro Gln Thr Leu Pro Val Ala Glu Glu Ser Leu Pro Leu Gln Ala Gln His Leu Ala Leu

 CTG GAT ACG CTC AGC GCG CTG CTG ACC CAG GAA GGC ACG CCG TCT GAA AAG GGT TAT CGC   661
 Leu Asp Thr Leu Ser Ala Leu Leu Thr Gln Glu Gly Thr Pro Ser Glu Lys Gly Tyr Arg

 ATT GAT TAT GCG CAT TTT ACC CCA CAA GCA AAA TTC AGC ACG CCC GTC TGG ATA AGC CAG   721
 Ile Asp Tyr Ala His Phe Thr Pro Gln Ala Lys Phe Ser Thr Pro Val Trp Ile Ser Gln

 GCG CAA GGC ATC CGT GCT GGC CCT CAA CGC CTC ACC TAA CAA CAA TAA ACC TTT ACT TCA   781
 Ala Gln Gly Ile Arg Ala Gly Pro Gln Arg Leu Thr ---
                                         RBS
 TTT TAT TAA CTC CGC AAC GCG GGG CGT TTG AGA TTT TATTATG CTA ATC AAA TTG TTA ACT   842
                                                  Met Leu Ile Lys Leu Leu Thr

 AAA GTT TTC GGT AGT CGT AAC GAT CGC ACC CTG CGC CGG ATG CGC AAA GTG GTC AAC ATC   902
 Lys Val Phe Gly Ser Arg Asn Asp Arg Thr Leu Arg Arg Met Arg Lys Val Val Asn Ile

 ATC AAT GCC ATG GAA CCG GAG ATG GAA AAA CTC TCC GAC GAA GAA CTG AAA GGG AAA ACC   962
 Ile Asn Ala Met Glu Pro Glu Met Glu Lys Leu Ser Asp Glu Glu Leu Lys Gly Lys Thr

 GCA GAG TTT CGT GCA CGT CTG GAA AAA GGC GAA GTG CTG GAA AAT CTG ATC CCG GAA GCT  1022
 Ala Glu Phe Arg Ala Arg Leu Glu Lys Gly Glu Val Leu Glu Asn Leu Ile Pro Glu Ala

 TTC GCC GTG GTA CGT GAG GCA AGT AAG CGC GTC TTT GGT ATG CGT CAC TTC GAC GTT CAG  1082
 Phe Ala Val Val Arg Glu Ala Ser Lys Arg Val Phe Gly Met Arg His Phe Asp Val Gln

 TTA CTC GGC GGT ATG GTT CTT AAC GAA CGC TGC ATC GCC GAA ATG CGT ACC GGT GAA GGA  1142
 Leu Leu Gly Gly Met Val Leu Asn Glu Arg Cys Ile Ala Glu Met Arg Thr Gly Glu Gly

 AAA ACC CTG ACC GCA ACG CTG CCT GCT TAC CTG AAC GCA CTA ACC GGT AAA GGC GTG CAC  1202
 Lys Thr Leu Thr Ala Thr Leu Pro Ala Tyr Leu Asn Ala Leu Thr Gly Lys Gly Val His

 GTA GTT ACC GTC AAC GAC TAC CTG GCG CAA CGT GAC GCC GAA AAC AAC CGT CCG CTG TTT  1262
 Val Val Thr Val Asn Asp Tyr Leu Ala Gln Arg Asp Ala Glu Asn Asn Arg Pro Leu Phe

 GAA TTC CTT GGC CTG ACT GTC GGT ATC AAC CTG CCG GGC ATG CCA GCA CCG GCA AAG CGC  1322
 Glu Phe Leu Gly Leu Thr Val Gly Ile Asn Leu Pro Gly Met Pro Ala Pro Ala Lys Arg

 GAA GCT TAC GCA GCT GAC ATC ACT TAC GGT ACG AAC AAC GAA TAC GGC TTT GAC TAC CTG  1382
 Glu Ala Tyr Ala Ala Asp Ile Thr Tyr Gly Thr Asn Asn Glu Tyr Gly Phe Asp Tyr Leu

 CGC GAC AAC ATG GCG TTC AGC CCT GAA GAA CGT GTA CAG CGT AAA CTG CAC TAT GCG CTG  1442
 Arg Asp Asn Met Ala Phe Ser Pro Glu Glu Arg Val Gln Arg Lys Leu His Tyr Ala Leu

 GTG GAC GAA GTG GAC TCC ATC CTG ATC GAT GAA GCG CGT ACA CCG CTG ATC ATT TCC GGC  1502
 Val Asp Glu Val Asp Ser Ile Leu Ile Asp Glu Ala Arg Thr Pro Leu Ile Ile Ser Gly
```

FIG. 2. Nucleotide sequence of the *secA* region. The predicted amino acid sequences of gene X and *secA* are given below the nucleotide sequence. Boxed nucleotides indicate positions of potential translational start and stop signals. Underlined nucleotides locate potential promoter elements (−35 and −10) and ribosome-binding sites (RBS). The arrows indicate the position of a hyphenated dyad symmetry element identified as a transcriptional terminator. ORF, Open reading frame.

```
CCG GCA GAA GAC AGC TCG GAA ATG TAT AAA CGC GTG AAT AAA ATT ATT CCG CAC CTG ATC 1562
Pro Ala Glu Asp Ser Ser Glu Met Tyr Lys Arg Val Asn Lys Ile Ile Pro His Leu Ile

CGT CAG GAA AAA GAA GAC TCC GAA ACC TTC CAG GGC GAA GGC CAC TTC TCG GTG GAC GAA 1622
Arg Gln Glu Lys Glu Asp Ser Glu Thr Phe Gln Gly Glu Gly His Phe Ser Val Asp Glu

AAA TCT CGC CAG GTG AAC CTG ACC GAA CGT GGT CTG GTG CTG ATT GAA GAA CTG CTG GTG 1682
Lys Ser Arg Gln Val Asn Leu Thr Glu Arg Gly Leu Val Leu Ile Glu Glu Leu Leu Val

AAA GAG GGC ATC ATG GAT GAA GGG GAG TCT CTG TAC TCT CCG GCC AAC ATC ATG CTG ATG 1742
Lys Glu Gly Ile Met Asp Glu Gly Glu Ser Leu Tyr Ser Pro Ala Asn Ile Met Leu Met

CAC CAC GTA ACG GCG GCG CTG CGC GCT CAT GCG CTG TTT ACC CGT GAC GTC GAC TAC ATC 1802
His His Val Thr Ala Ala Leu Arg Ala His Ala Leu Phe Thr Arg Asp Val Asp Tyr Ile

GTT AAA GAT GGT GAA GTT ATC ATC GTT GAC GAA CAC ACC GGT CGT ACC ATG CAG GGC CGT 1862
Val Lys Asp Gly Glu Val Ile Ile Val Asp Glu His Thr Gly Arg Thr Met Gln Gly Arg

CGC TGG TCC GAT GGT CTG CAC CAG GCT GTG GAA GCG AAA GAA GGT GTG CAG ATC CAG AAC 1922
Arg Trp Ser Asp Gly Leu His Gln Ala Val Glu Ala Lys Glu Gly Val Gln Ile Gln Asn

GAA AAC CAA ACG CTG GCT TCG ATC ACC TTC CAG AAC TAC TTC CGT CTG TAT GAA AAA CTG 1982
Glu Asn Gln Thr Leu Ala Ser Ile Thr Phe Gln Asn Tyr Phe Arg Leu Tyr Glu Lys Leu

GCG GGG ATG ACC GGT ACT GCT GAT ACC GAA GCT TTC GAA TTT AGC TCA ATC TAC AAG CTG 2042
Ala Gly Met Thr Gly Thr Ala Asp Thr Glu Ala Phe Glu Phe Ser Ser Ile Tyr Lys Leu

GAT ACC GTC GTT GTT CCG ACC AAC CGT CCA ATG ATT CGT AAA GAT CTG CCG GAC CTG GTC 2102
Asp Thr Val Val Val Pro Thr Asn Arg Pro Met Ile Arg Lys Asp Leu Pro Asp Leu Val

TAC ATG ACT GAA GCG GAA AAA ATT CAG GCG ATC ATT GAA GAT ATC AAA GAA CGT ACT GCG 2162
Tyr Met Thr Glu Ala Glu Lys Ile Gln Ala Ile Ile Glu Asp Ile Lys Glu Arg Thr Ala

AAA GGC CAG CCG GTG CTG GTG GGT ACT ATC TCC ATC GAA AAA TCG GAG CTG GTG TCA AAC 2222
Lys Gly Gln Pro Val Leu Val Gly Thr Ile Ser Ile Glu Lys Ser Glu Leu Val Ser Asn

GAA CTG ACC AAA GCC GGT ATT AAG CAC AAC GTC CTG AAC GCC AAA TTC CAC GCC AAC GAA 2282
Glu Leu Thr Lys Ala Gly Ile Lys His Asn Val Leu Asn Ala Lys Phe His Ala Asn Glu

GCG GCG ATT GTT GCT CAG GCA GGT TAT CCG GCT GCG GTG ACT ATC GCG ACC AAT ATG GCG 2342
Ala Ala Ile Val Ala Gln Ala Gly Tyr Pro Ala Ala Val Thr Ile Ala Thr Asn Met Ala

GGT CGT GGT ACA GAT ATT GTG CTC GGT GGT AGC TGG CAG GCA GAA GTT GCC GCG CTG GAA 2402
Gly Arg Gly Thr Asp Ile Val Leu Gly Gly Ser Trp Gln Ala Glu Val Ala Ala Leu Glu

AAT CCG ACC GCA GAG CAA ATT GAA AAA ATT AAA GCC GAC TGG CAG GTA CGT CAC GAT GCG 2462
Asn Pro Thr Ala Glu Gln Ile Glu Lys Ile Lys Ala Asp Trp Gln Val Arg His Asp Ala

GTA CTG GAA GCA GGT GGC CTG CAT ATC ATC GGT ACC GAG CGT CAC GAA TCC CGT CGT ATC 2522
Val Leu Glu Ala Gly Gly Leu His Ile Ile Gly Thr Glu Arg His Glu Ser Arg Arg Ile

GAT AAC CAG TTG CGC GGT CGT TCT GGT CGT CAG GGG GAT GCT GGT TCT TCC CGT TTC TAC 2582
Asp Asn Gln Leu Arg Gly Arg Ser Gly Arg Gln Gly Asp Ala Gly Ser Ser Arg Phe Tyr

CTG TCG ATG GAA GAT GCG CTG ATG CGT ATT TTT GCT TCC GAC CGA GTA TCC GGC ATG ATG 2642
Leu Ser Met Glu Asp Ala Leu Met Arg Ile Phe Ala Ser Asp Arg Val Ser Gly Met Met

CGT AAA CTG GGT ATG AAG CCA GGC GAA GCC ATT GAA CAC CCG TGG GTG ACT AAA GCG ATT 2702
Arg Lys Leu Gly Met Lys Pro Gly Glu Ala Ile Glu His Pro Trp Val Thr Lys Ala Ile

GCC AAC GCC CAG CGT AAA GTT GAA AGC CGT AAC TTC GAC ATT CGT AAG CAA CTG CTG GAA 2762
Ala Asn Ala Gln Arg Lys Val Glu Ser Arg Asn Phe Asp Ile Arg Lys Gln Leu Leu Glu

TAT GAT GAC GTG GCT AAC GAT CAG CGT CGC GCC ATT TAC TCC CAG CGT AAC GAA CTG TTG 2822
Tyr Asp Asp Val Ala Asn Asp Gln Arg Arg Ala Ile Tyr Ser Gln Arg Asn Glu Leu Leu

GAT GTC AGC GAT GTG AGC GAA ACC ATT AAC AGC ATT CGT GAA GAT GTG TTC AAA GCG ACC 2882
Asp Val Ser Asp Val Ser Glu Thr Ile Asn Ser Ile Arg Glu Asp Val Phe Lys Ala Thr

ATT GAT GCC TAC ATT CCA CCA CAG TCG CTG GAA GAA ATG TGG GAT ATT CCG GGG CTG CAG 2942
Ile Asp Ala Tyr Ile Pro Pro Gln Ser Leu Glu Glu Met Trp Asp Ile Pro Gly Leu Gln
```

FIG. 2—*Continued.*

```
GAA CGT CTG AAG AAC GAT TTC GAC CTC GAT TTG CCA ATT GCC GAG TGG CTG GAT AAA GAA 3002
Glu Arg Leu Lys Asn Asp Phe Asp Leu Asp Leu Pro Ile Ala Glu Trp Leu Asp Lys Glu

CCA GAA CTG CAT GAA GAG ACG CTG CGT GAC GGC ATT CTG GCG CAG TCC ATC GAA GTG TAT 3062
Pro Glu Leu His Glu Glu Thr Leu Arg Asp Gly Ile Leu Ala Gln Ser Ile Glu Val Tyr

CAG CGT AAA GAA GAA GTG GTT GGT GCT GAG ATG ATG CGT CAC TTC GAG AAA GGC GTC ATG 3122
Gln Arg Lys Glu Glu Val Val Gly Ala Glu Met Met Arg His Phe Glu Lys Gly Val Met

CTG CAA ACG CTT GAC TCC CTG TGG AAA GAG CAC CTG GCA GCG ATG GAC TAT CTG CGT CAG 3182
Leu Gln Thr Leu Asp Ser Leu Trp Lys Glu His Leu Ala Ala Met Asp Tyr Leu Arg Gln

GGT ATC CAC CTG CGT GGC TAC GCA CAG AAA GAT CCG AAG CAG GAA TAC AAA CGT GAA TCG 3242
Gly Ile His Leu Arg Gly Tyr Ala Gln Lys Asp Pro Lys Gln Glu Tyr Lys Arg Glu Ser

TTC TCC ATG TTT GCA GCG ATG CTG GAG TCG TTG AAA TAT GAA GTT ATC AGT ACG CTG AGC 3302
Phe Ser Met Phe Ala Ala Met Leu Glu Ser Leu Lys Tyr Glu Val Ile Ser Thr Leu Ser

AAA GTT CAG GTA CGT ATG CCT GAA GAG GTT GAG GAG CTG GAA CAA CAG CGT CGT ATG GAA 3362
Lys Val Gln Val Arg Met Pro Glu Glu Val Glu Glu Leu Glu Gln Gln Arg Arg Met Glu

GCC GAG CGT TTA GCG CAA ATG CAG CAG CTT AGC CAT CAG GAT GAC GAC TCT GCA GCC GCA 3422
Ala Glu Arg Leu Ala Gln Met Gln Gln Leu Ser His Gln Asp Asp Asp Ser Ala Ala Ala

GCT GCA CTG GCG GCG CAA ACC GGA GAG CGC AAA GTA GGA CGT AAC GAT CCT TGC CCG TGC 3482
Ala Ala Leu Ala Ala Gln Thr Gly Glu Arg Lys Val Gly Arg Asn Asp Pro Cys Pro Cys

GGT TCT GGT AAA AAA TAC AAG CAG TGC CAT GGC CGC CTG CAA TAA AAG CTA ACT GTT GAA 3542
Gly Ser Gly Lys Lys Tyr Lys Gln Cys His Gly Arg Leu Gln ---
                                                           Begin ORF——►
                                              RBS
GTA AAA GGC GCA GGA TTC TGC GCC TTT TTT ATA GGT TTA AGA CAA TGA AAA AGC TGC AAA 3602

TTG CGG TAG GTA TTA TTC GCA ACG AGA ACA ATG AAA TCT TTA TAA CGC GTC GCG CAG CAG 3662

ATG CGC ACA TGG CGA ATA AAC TGG AGT TTC CCG GCG GTA AAA TTG AAA TGG GTG AAA CGC 3722

CGG AAC AGG CGG TGG TGC GTG AAC TTC AGG AAG AAG TCG GGA TTA CCC CCC AAC ATT TTT 3782

CGC TAT TTG AAA AAC TGG AAT ATG AAT TC 3811
```

FIG. 2—*Continued.*

rabbit muscle glycogen phosphorylase b on SDS-PAGE gels. Phosphorylase b has been reported to be 92,500 daltons (37), 97,400 daltons (42), and 100,000 daltons (7). The 97,400-dalton estimate is based on the amino acid sequence of the protein, whereas the 100,000-dalton estimate is based on mobility in SDS-PAGE gels with known molecular weight markers. Therefore, the predicted size of 101,902 daltons for SecA is in good agreement with its comigration with phosphorylase b.

We wanted to confirm that the beginning and end of the predicted amino acid sequence of SecA in fact matched that of the bona fide protein. To corroborate the beginning of the *secA* gene, we performed amino-terminal sequence analysis of SecA protein immunopurified from cells that were doubly labeled with either [^{35}S]methionine and [^{3}H]leucine or [^{35}S] methionine and [^{3}H]lysine. These results are shown in Fig. 3. The positions of methionine, leucine, and lysine predicted from our sequence precisely matched those experimentally determined for SecA protein for the first 25 amino acid residues, indicating a correct assignment for the start of *secA*. To corroborate the end of the *secA* gene, we used gene fusion techniques to show that the *secA* open reading frame extended beyond the *Nco*I site at nucleotide position 3509. To this end, we made a fusion between the *secA* gene and the CAT gene by using this *Nco*I site in *secA* and the *Nco*I site within the CAT gene (see Materials and Methods for details). The plasmid which contained this gene fusion, pAJ1, should produce a hybrid protein in which the last 3 amino acid residues of SecA are replaced with 22 amino acid residues from an unused reading frame of CAT (1). It can be seen in Fig. 4 that SecA immunoprecipitated from cells carrying this plasmid not only produced the wild-type form of SecA, but also overproduced the slightly higher molecular weight fusion protein (SecA^{+} in Fig. 4). In addition, the altered *secA* gene contained on pAJ1, in contrast to its wild-type counterpart on psecA, was only able to partially complement the temperature sensitivity of the *secA*(Ts) mutant MM52. Therefore, we conclude that the *Nco*I site at nucleotide position 3509 is within the *secA* coding sequences and that this gene must encode a polypeptide of at least 101,500 daltons.

At 62 nucleotides past the end of the *secA* gene there was an open reading frame, starting at nucleotide position 3587 and continuing to the end of our sequence, which would encode a polypeptide of greater than 75 amino acid residues. There was an indication that this open reading frame was in fact translated. First, the sequence AGGT 8 nucleotides upstream of the ATG initiation codon could serve as a potential ribosome-binding site. Second, the codon usage and amino acid composition of this predicted polypeptide are in accordance with those of a typical *E. coli* protein (data not shown).

The predicted sequence of the SecA protein contained 29% charged amino acid residues distributed throughout the length of the protein and rich in glutamic acid (10.2% of all amino acid residues). The hydropathy profile (21) of SecA protein is characteristic of a soluble protein, lacking stretches of sufficient length and hydrophobicity to span a membrane (data not shown). This is in keeping with the observed biochemical localization of SecA protein, which

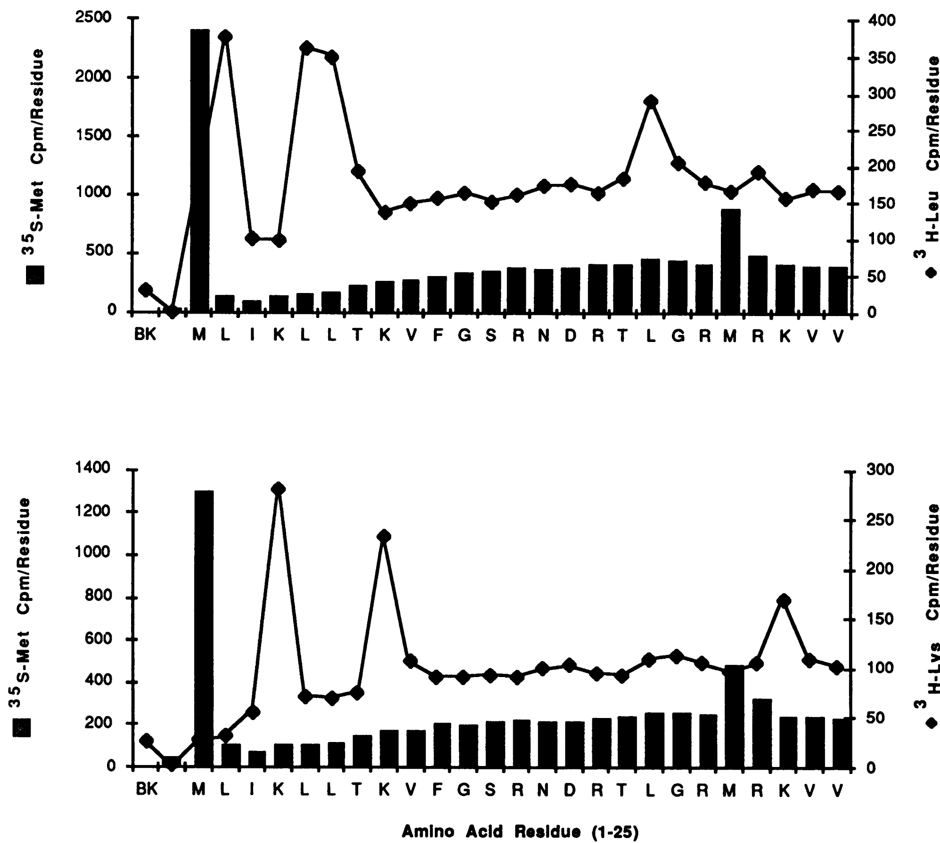

FIG. 3. Determination of the amino-terminal sequence of SecA protein. The amount of [^{35}S]methionine and [^{3}H]leucine or [^{3}H]lysine per residue is given on the ordinate versus the amino acid residue from successive Edman degradation on the abscissa. The amino-terminal amino acid sequence predicted from the *secA* DNA sequence is given below the abscissa. BK, Background sample.

indicated that it is largely soluble but can interact weakly with membranes (31). Comparison of the predicted amino acid sequence of SecA protein with protein sequences contained in the Protein Identification Resource (release 6.0, 1985) compiled by the National Biomedical Research Foundation by using the programs developed by Lipman and Pearson (22) revealed no significant similarities.

Analysis of *secA*(Ts) mutants. Having defined the sequence of the *secA* gene, we wanted to determine the nucleotide

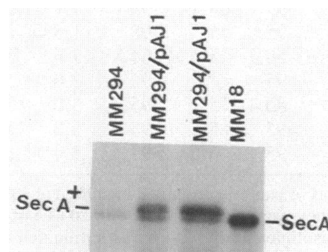

FIG. 4. Analysis of a *secA*-CAT gene fusion. Strains were grown in M63 medium supplemented with 0.4% glucose and tetracycline (2.5 μg/ml) for plasmid-containing strains at 37°C when they were labeled with [^{35}S]methionine (20 μCi/ml) for 1 min. Radioisotope incorporation was terminated by the addition of ice-cold trichloroacetic acid, and samples were prepared for SecA immunoprecipitation and PAGE analysis as described in Materials and Methods. MM18 was grown in maltose to derepress SecA. The positions of SecA and the carboxy-terminal extended form of SecA (SecA^{+}) are indicated.

alterations present in a collection of nine *secA*(Ts) mutants. These mutants were selected as partially defective in the export of a hybrid protein between maltose-binding protein and β-galactosidase, thereby increasing enzymatic activity (30). The mutants were characterized as carrying conditionally lethal temperature-sensitive mutations and showing temperature-dependent export blocks. In order to determine the location of the temperature-sensitive mutations, we constructed a mapping series of cosmids containing defined restriction fragments from the sequenced region and prepared lambda lysates from strains containing individual cosmids. This permitted rapid mapping of the *secA*(Ts) mutations by simple spot tests of these lysates on a given mutant and subsequent scoring for the production of Ts^{+} recombinants arising at the restrictive temperature (see Materials and Methods for details). As shown in Table 1, the nine *secA*(Ts) mutants fell into three classes, based on their map position within the *secA* gene. Using this information, we cloned the appropriate restriction enzyme fragment from a given mutant into M13 and determined the complete DNA sequence of the map interval. This allowed the unique assignment of each of the nine temperature-sensitive mutations given in Table 1.

There were several notable features in this collection of *secA*(Ts) mutations. First, all of the mutations mapped within the sequences encoding the first 170 amino acid residues of the 901-amino-acid-residue SecA protein. Second, despite the fact that there were at least two mutations of independent origin in each of the three classes, mutations in

TABLE 1. Temperature-sensitive mutations in *secA*[a]

Class (nt)	Strain(s)	Nucleotide	Mutation	Amino acid affected	Change
I (910–1018)	MM52, MQ11, MQ16	949	T→C	43	Leu→Pro
II (1018–1263)	MQ7	1198	T→A	126	Val→Glu
	MQ9	1185–1187	ACC→del	122	Thr→del
III (1263–1532)	MQ10, MQ13	1327	C→A	169	Ala→Asp
	MQ5, MQ15	1329	T→G	170	Tyr→Asp

[a] The *secA*(Ts) mutations were grouped into three classes based on genetic mapping results. The nucleotide (nt) positions of these 3 map intervals are given beside the class designations. del, Deletion mutation.

each class were no more than 4 amino acid residues apart, making it appear as if three sites were being defined. Whether these two features are the result of the specific constraints of the genetic selection that was used to obtain these mutants or reflect some specific biochemical property of the SecA protein or both is unclear and must await further studies. Third, the amino acid changes caused by the mutations were all disruptive of SecA protein structure. These alterations either disrupted predicted secondary structures (12) within the protein (Leu-43 to Pro, Thr-122 deleted, Val-126 to Glu) or exchanged amino acid residues that are very different in their hydrophilic-hydrophobic nature (Val-126 to Glu, Ala-169 to Asp, Tyr-170 to Asp).

In order to correlate the mutational alterations in the *secA* gene given in Table 1 with a biological phenotype, we measured the strength of the export blocks and the derepression of SecA levels in strains that carried these *secA*(Ts) mutations. The export of the outer membrane protein OmpA and the periplasmic maltose-binding protein (MalE) and the level of synthesis of SecA protein were measured at the permissive (30°C) and nonpermissive (41°C) temperatures by pulse-labeling cells with [^{35}S]methionine and immunoprecipitating and analyzing these proteins on PAGE gels by quantitative autoradiography. These results are presented in Table 2. Although all of the *secA*(Ts) mutants examined showed temperature-dependent export defects and elevated levels of SecA synthesis, there were quantitative differences in these properties among the different mutants. Whereas most of the *secA*(Ts) mutants examined were export proficient at 30°C (<5% precursor) and export defective at 41°C (>95% precursor), MQ7 showed a strong export defect at 30°C for both secretory proteins examined (~65% precursor). In contrast, protein export in the MQ10 and MQ13 mutant pair was not completely blocked after 2 h of growth at 41°C (60 to 80% precursors). All of the *secA*(Ts) mutants examined showed 1.3- to 3-fold-elevated levels of SecA synthesis at 30°C and 4- to 9-fold-higher levels at 41°C relative to the *secA*$^+$ wild-type strain. Although there was a basic correlation between SecA synthesis levels and protein export proficiency, this correlation was not precise. For example, SecA synthesis levels were two- to threefold higher under conditions in which protein export showed a major defect (~65% precursor for MQ7 at 30°C) or only a minor one (<5% precursor for MQ13 at 30°C). This indicates that there is some impression in the mechanism by which the cell derepresses *secA* in response to protein export defects.

Analysis of a gene X amber mutant. An amber mutation (*am109*) has been assigned to the *secA* gene based on complementation and fine structure genetic mapping studies (31). Strains that carry *am109* and a temperature-sensitive suppressor tRNA [*supF*(Ts)] are temperature sensitive for growth, show protein export defects, and have undetectable SecA levels at the nonpermissive temperature (31). In mapping *am109* along with the other *secA*(Ts) mutations, we found that it mapped between nucleotide positions 615 and 838, making it unlikely that the mutation was within the *secA* gene (which starts at nucleotide position 822). An appropriate restriction enzyme fragment containing this entire region was cloned from the *am109*-containing mutant into M13 (see Materials and Methods for details), and the DNA sequence of this region was determined. We found that the *am109* mutation lay within the gene X open reading frame (G to A transition at nucleotide position 711; Fig. 2), resulting in the premature termination of gene X protein, now missing the last 16 amino acid residues. This result implies that gene X is polar on *secA* expression. Consistent with this result, we found that a strain containing the *am109* mutation, a *supF*(Ts) suppressor, and a plasmid encoding only an intact *secA* gene (psecA) was no longer temperature sensitive for growth.

DISCUSSION

In order to begin to study the structure of the SecA protein, its role in secretion, and how it is regulated by cellular protein export rates, we sequenced the *secA* gene and a region upstream of this gene. Our sequence predicts that SecA is a large protein of 901 amino acid residues with

TABLE 2. Protein export proficiency and SecA levels in *secA*(Ts) mutants[a]

Strain	% Precursor[b]				SecA level[c]			
	OmpA		MalE		A		B	
	30°C	41°C	30°C	41°C	30°C	41°C	30°C	41°C
MC4100	<5	<5	<5	<5	1.0	1.0	1.0	1.0
MM52	<5	>95	<5	>95	1.4	3.7	1.3	4.7
MQ5	<5	>95	11.2	>95	1.4	6.2	1.3	7.8
MQ7	66.5	>95	63.4	>95	1.9	8.9	2.3	9.0
MQ9	<5	>95	<5	>95	1.5	8.9	1.6	9.6
MQ10	<5	58.7	<5	64.8	1.6	7.6	2.1	7.2
MQ11	<5	>95	<5	<95	2.0	5.5	1.6	4.8
MQ13	<5	80.4	<5	>95	3.0	5.5	2.2	6.7
MQ15	<5	>95	<5	>95	3.8	8.1	3.1	8.8
MQ16	<5	>95	<5	>95	3.4	5.5	3.1	5.3

[a] Growing cultures of bacteria remained unshifted at 30°C or were shifted to 41°C for 2 h when they were pulse-labeled with 20 μCi of [^{35}S]methionine per ml for 2 min. Radioisotope incorporation was terminated by the addition of an equal volume of ice-cold 10% trichloroacetic acid, and the samples were prepared for immunoprecipitation and PAGE analysis as described in Materials and Methods.

[b] Determined by using the formula % precursor = (precursor/[precursor + mature form]) × 100.

[c] The SecA levels given in A are the ratio of SecA to total OmpA found in a given mutant divided by the comparable ratio found in MC4100. The SecA levels given in B are the ratio of SecA to RNA polymerase subunits RpoB and RpoC for a given mutant divided by the comparable ratio found in MC4100. By normalizing SecA levels to OmpA or RpoB plus RpoC levels determined within the same immunoprecipitation reaction, quantitation of SecA levels was made more accurate.

a predicted molecular weight of 101,902, compatible with its observed comigration with phosphorylase b. Furthermore, the predicted sequence was correlated with the amino-terminal sequence of SecA protein and found to be correct. The predicted sequence of SecA contains 29% charged amino acid residues distributed throughout the length of the protein, is rich in glutamic acid (10.2% of all amino acid residues), and is free of any predicted membrane-spanning segments. These observations are entirely consistent with the finding that most of the cellular SecA protein was found in the cytoplasm or in loose association with the cytoplasmic membrane (31). Comparison of the predicted amino acid sequence of SecA protein with protein sequences contained in a data base compiled by the National Biomedical Research Foundation revealed no significant similarities. However, we will be interested to compare our sequence with the sequences of other *sec* genes of *E. coli* and *S. cerevisiae* as well as mammalian signal recognition particle when these sequences become available.

We sequenced nine *secA*(Ts) mutations derived from a single genetic selection that all showed temperature-dependent protein export defects. We found that all of these mutants contained single point mutations and that all of the alterations were tightly clustered at three sites within the first 170 amino acid residues of the SecA protein. The amino acid alterations found in these mutants were severe in nature and would either disrupt predicted protein secondary structure or cause a strong alteration in the hydrophobic-hydrophilic property of a given amino acid residue. It is clear that all of these *secA*(Ts) mutants showed strong protein export defects at the nonpermissive temperature as well as elevated levels of SecA protein synthesis, although individual mutants differed somewhat in the severity of these defects. It should be noted that the genetic selection used to isolate these mutations was constrained, since it required a hybrid protein between maltose-binding protein and β-galactosidase to be synthesized, show a secretion defect prior to membrane association, and result in temperature-sensitive mutants (30). These specific requirements or less apparent ones may have severely limited the type of *secA* mutations isolated. Since mutations in only three regions in the first 20% of the *secA* gene were obtained, additional *secA* mutants are needed to dissect the function of this large protein.

Part of the rationale behind our sequence analysis was to begin to determine which DNA sequences around the *secA* gene are responsible for the coregulation of *secA* with cellular protein export rates. In this regard, we determined the DNA sequence of 800 nucleotides upstream of the *secA* gene. We found that there was a sequence that apparently functioned as a rho-independent transcriptional terminator at the end of the *envA* gene (4), which should effectively uncouple the transcription of the upstream cell division genes *ftsQ*, *ftsA*, *ftsZ*, and *envA* from that of the downstream genes. Downstream of this terminator is a predicted gene, gene X, which precedes *secA*. Several possible promoters and translation start sites for gene X were contained in our sequence. That gene X is indeed translated was indicated by the fact that the *am109* mutation previously assigned to the proximal portion of *secA* was instead located within the distal portion of gene X. The lethality of this mutation in a suppressor tRNA-deficient host is solely due to insufficient SecA protein levels, since a *trans* copy of the *secA* gene lacking functional gene X sequences was sufficient to restore cell viability. This result implies that the production of a gene X protein lacking its carboxy-terminal 16 amino acid residues is compatible with cell growth and therefore that it may be a nonessential gene. The failure to synthesize detectable SecA protein in the nonsuppressed gene X *am109*-containing strain indicates that gene X and *secA* are cotranscribed. Two additional lines of evidence support this contention. First, we found no indication in our sequence for a promoter for *secA* at the end of gene X or in the intergenic region preceding *secA*. Second, plasmids that contained the *secA* gene and as much as 358 nucleotides upstream of *secA* produced low levels of SecA protein and showed no *secA* regulation in response to protein export defects, whereas the opposite result was seen when such vectors contained an additional 800 nucleotides of DNA upstream of *secA* (M. Faecke, M. G. Schmidt, and D. B. Oliver, unpublished observations). Whether this polarity is due to rho-mediated transcriptional termination, some type of translation coupling, or another novel mechanism remains to be determined. Additional work will be required to determine whether *secA* derepression in response to protein secretion defects occurs at the transcriptional or translational level or both and involves coordinate or uncoupled regulation of gene X and *secA*.

ACKNOWLEDGMENTS

We thank the following people for help with various stages of the work: Laura Liss, Barri Fessler, and Shanli Tsui for constructing M13 clones, Holger Rohl and Torsten Krude for sequencing *secA*(Ts) mutations, Rob Cabelli for reading sequencing gels and preparation of radiolabeled SecA protein for sequence analysis, and Carl Anderson for sequence analysis of SecA protein. We thank J. Lutkenhaus for sending his manuscript prior to publication.

This work was supported by Public Health Service grant 1RO1GM32958 from the National Institute of General Medical Sciences and grant NP-573 from the American Cancer Society to D.B.O. M.G.S. and J.G. were supported by training grant 5T32CA09176 from the National Cancer Institute.

ADDENDUM IN PROOF

We recently found that the open reading frame distal to *secA* encodes an essential gene product as defined by conditionally lethal temperature-sensitive mutations that map between nucleotides 3530 and 3705.

LITERATURE CITED

1. **Alton, N., and D. Vapnek.** 1979. Nucleotide sequence analysis of the chloramphenicol resistance transposon Tn9. Nature (London) **282:**864–869.
2. **Anderson, C. W., and J. B. Lewis.** 1980. Amino terminal sequence of adenovirus type 2 proteins: hexon, fiber component IX, and early protein 1B-15K. Virology **104:**27–41.
3. **Bankaitis, V., and P. Bassford.** 1985. Proper interaction between at least two components is required for efficient export of proteins to the *Escherichia coli* cell envelope. J. Bacteriol. **161:** 169–178.
4. **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.
5. **Cerrettii, D., D. Dean, G. Davis, D. Bedwell, and M. Nomura.** 1983. The *spc* ribosomal protein operon of *Escherichia coli*: sequence and cotranscription of the ribosomal protein genes and a protein export gene. Nucleic Acids Res. **11:**2599–2616.
6. **Chen, L., and P. Tai.** 1985. ATP is essential for protein translocation into *Escherichia coli* membrane vesicles. Proc. Natl. Acad. Sci. USA **82:**4384–4388.
7. **Cohen, P., T. Duewar, and E. Fischer.** 1971. Phosphorylase from dogfish skeletal muscle. Purification and a comparison of its physical properties to those of rabbit muscle phosphorylase. Biochemistry **10:**2683–2694.
8. **Dale, R. M. K., B. A. McClure, and J. B. Houchins.** 1985. A

rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. Plasmid **13**:31–40.
9. **Emr, S., S. Hanley-Way, and T. Silhavy.** 1981. Suppressor mutations that restore export of a protein with a defective signal sequence. Cell **23**:79–88.
10. **Evans, E., R. Gilmore, and G. Blobel.** 1986. Purification of microsomal signal peptidase complex. Proc. Natl. Acad. Sci. USA **83**:581–585.
11. **Gardel, C., S. Benson, J. Hunt, S. Michaelis, and J. Beckwith.** 1987. *secD*, a new gene involved in protein export in *Escherichia coli*. J. Bacteriol. **169**:1286–1290.
12. **Garnier, J., D. J. Ogsthorpe, and B. Robson.** 1978. Analysis of the accuracy and implications of simple methods for predicting secondary structure of globular proteins. J. Mol. Biol. **120**:97–120.
13. **Geller, B., N. Movva, and W. Wickner.** 1986. Both ATP and the electrochemical potential are required for optimal assembly of pro-OmpA into *Escherichia coli* inner membrane vesicles. Proc. Natl. Acad. Sci. USA **83**:4219–4222.
14. **Hansen, W., P. Garcia, and P. Walter.** 1986. In vitro protein translocation across the yeast endoplasmic reticulum: ATP-dependent post-translational translocation of the prepro-α-factor. Cell **45**:397–406.
15. **Inouye, H., W. Barns, and J. Beckwith.** 1982. Signal sequence of alkaline phosphatase of *Escherichia coli*. J. Bacteriol. **149**:434–439.
16. **Ito, K.** 1984. Identification of the SecY (*prlA*) gene product involved in protein export in *Escherichia coli*. Mol. Gen. Genet. **197**:204–208.
17. **Ito, K., P. Bassford, and J. Beckwith.** 1981. Protein localization in *E. coli*: is there a common step in the secretion of periplasmic and outer membrane proteins? Cell **24**:707–717.
18. **Jones, C., and I. Holland.** 1984. Inactivation of essential cell division genes, *ftsA*, *ftsZ*, suppresses mutations at *sfiB*, a locus mediating division inhibition during the SOS response in *E. coli*. EMBO J. **3**:1181–1186.
19. **Kumamoto, C., and J. Beckwith.** 1983. Mutations in a new gene, *secB*, cause defective protein localization in *Escherichia coli*. J. Bacteriol. **154**:253–260.
20. **Kumamoto, C., and J. Beckwith.** 1985. Evidence for specificity at an early step in protein export in *Escherichia coli*. J. Bacteriol. **163**:267–274.
21. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
22. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
23. **Liss, L., and D. Oliver.** 1986. Effects of *secA* mutations on the synthesis and secretion of proteins in *Escherichia coli*: evidence for a major export system for cell envelope proteins. J. Biol. Chem. **261**:2299–2303.
24. **Lutkenhaus, J., H. Wolf-Watz, and W. Donachie.** 1980. Organization of genes in the *ftsA-envA* region of the *Escherichia coli* genetic map and identification of a new *fts* locus (*ftsZ*). J. Bacteriol. **142**:615–620.
25. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Messing, J., R. Crea, and P. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
27. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
28. **Müller, M., and G. Blobel.** 1984. Protein export in *Escherichia coli* requires a soluble activity. Proc. Natl. Acad. Sci. USA **81**:7737–7741.
29. **Oliver, D.** 1985. Protein secretion in *Escherichia coli*. Annu. Rev. Microbiol. **39**:615–648.
30. **Oliver, D., and J. Beckwith.** 1981. *E. coli* mutant pleiotropically defective in the export of secreted proteins. Cell **25**:765–772.
31. **Oliver, D., and J. Beckwith.** 1982. Regulation of a membrane component required for protein secretion in *Escherichia coli*. Cell **30**:311–319.
32. **Oliver, D., and J. Beckwith.** 1982. Identification of a new gene (*secA*) and gene product involved in the secretion of envelope proteins. J. Bacteriol. **150**:686–691.
32a. **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.
33. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-termination inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
35. **Schekman, R.** 1985. Protein localization and membrane traffic in yeast. Annu. Rev. Cell Biol. **1**:115–143.
36. **Schreier, P., and R. Cortese.** 1979. A fast and simple method for sequencing DNA cloned in the single-stranded bacteriophage M13. J. Mol. Biol. **129**:169–172.
37. **Seery, V., E. Fischer, and D. Teller.** 1970. Subunit structure of glycogen phosphorylase. Biochemistry **9**:3591–3598.
38. **Semler, B. L., C. W. Anderson, N. Kitamura, P. G. Rothberg, W. L. Wishart, and E. Wimmer.** 1981. Poliovirus replication proteins: RNA sequence encoding P3-1b and the sites of proteolytic processing. Proc. Natl. Acad. Sci. USA **78**:3464–3468.
39. **Shiba, K., K. Ito, T. Yura, and D. Cerretti.** 1984. A defined mutation in the protein export gene within the *spc* ribosomal protein operon of *Escherichia coli*: isolation and characterization of a new temperature-sensitive *secY* mutant. EMBO J. **3**:631–635.
40. **Shine, J., and L. Dalgarno.** 1974. The 3′ terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
41. **Sullivan, N., and W. Donachie.** 1984. Transcriptional organization within an *Escherichia coli* cell division gene cluster: direction of transcription of the cell separation gene *envA*. J. Bacteriol. **160**:724–732.
42. **Titani, K., A. Koide, J. Hermann, L. Ericsson, S. Kumar, R. Wade, K. Walsh, H. Neurath, and E. Fischer.** 1977. Complete amino acid sequence of rabbit muscle glycogen phosphorylase. Proc. Natl. Acad. Sci. USA **74**:4762–4766.
43. **Tokunaga, M., J. Loranger, and H. Wu.** 1984. Prolipoprotein modification and processing enzymes in *Escherichia coli*. J. Biol. Chem. **259**:3825–3830.
44. **Walter, P., and V. Lingappa.** 1986. Mechanism of protein translocation across the endoplasmic reticulum membrane. Annu. Rev. Cell Biol. **2**:499–516.
45. **Watts, C., W. Wickner, and R. Zimmermann.** 1983. M13 procoat and pre-immunoglobulin share processing specificity but use different membrane receptor mechanisms. Proc. Natl. Acad. Sci. USA **80**:2809–2813.
46. **Wickner, W., and H. Lodish.** 1985. Multiple mechanisms of protein insertion into and across membranes. Science **230**:400–407.
47. **Wiedmann, M., T. Kurzchalia, E. Hartmann, and T. Rapoport.** 1987. A signal sequence receptor in the endoplasmic reticulum membrane. Nature (London) **328**:830–833.
48. **Wolfe, P., M. Rice, and W. Wickner.** 1985. Effects of two *sec* genes on protein assembly into the plasma membrane of *Escherichia coli*. J. Biol. Chem. **260**:1836–1841.
49. **Wolfe, P., W. Wickner, and J. Goodman.** 1983. Sequence of the leader peptidase gene of *Escherichia coli* and the orientation of leader peptidase in the bacterial envelope. J. Biol. Chem. **258**:12073–12080.