The EMBO Journal vol.9 no.10 pp.3209 - 3216, 1990

The *secD* locus of *E.coli* codes for two membrane proteins required for protein export

C.Gardel, K.Johnson, A.Jacq[1] and J.Beckwith

Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115, USA and [1]Institut de Microbiologie, Université Paris Sud, 91405 Orsay, France

Communicated by J.Beckwith

Cold-sensitive mutations in the *secD* locus of *Escherichia coli* result in severe defects in protein export at the non-permissive temperature of 23°C. DNA sequence of a cloned fragment that includes the *secD* locus reveals open reading frames for seven polypeptide chains. Both deletions and Tn*phoA* insertions in this clone have been used in maxicell and complementation studies to define the *secD* locus and its products. The *secD* mutations fall into two complementation groups, defining genes we have named *secD* and *secF*. These two genes comprise an operon, the first case of two genes involved in the export process being co-transcribed. The DNA sequence of the two genes along with alkaline phosphatase fusion analysis indicates that they code for integral proteins of the cytoplasmic membrane. We suggest that these two proteins may form a complex in the membrane which acts at late steps in the export process.

Key words: membrane proteins/protein secretion/*sec* genes/ Tn*phoA*

Introduction

The translocation of proteins through membranes during their synthesis requires a cellular machinery in both eukaryotic and prokaryotic cells (Pugsley, 1989). In both cell types, the precursor of the exported protein is recognized by cytosolic components of this machinery whose functions may include discrimination between cytosolic and secretory proteins, maintenance of the precursor in a 'secretion-competent' state, and delivery of protein precursors to the membrane. The membrane contains components of the secretory apparatus which promote the transfer of the protein through the membrane and the removal of the signal sequence.

In mammalian systems, transfer of proteins into the rough endoplasmic reticulum (RER) requires signal recognition particle, a cytosolic complex composed of six proteins and a 7S RNA molecule (Walter and Blobel, 1982). Several RER membrane proteins have also been implicated as components of the secretion machinery. These include the SRP receptor or docking protein (Gilmore *et al.*, 1982; Meyer *et al.*, 1982), integral membrane proteins thought to interact with signal sequences (Hartmann *et al.*, 1989; Krieg *et al.*, 1989), and a complex of proteins exhibiting signal peptidase activity (Evans *et al.*, 1986). Studies in yeast have revealed the involvement of heat-shock proteins (Chirico *et al.*, 1988; Deshaies *et al.*, 1988), a BiP-like protein (Vogel *et al.*, 1990) and three integral membrane proteins in the process (Deshaies and Schekman, 1989; Rothblatt *et al.*, 1989; C.Sterling and S.Schekman, personal communication).

In the prokaryote, *Escherichia coli*, several different genetic selections and screening procedures have yielded mutations in genes whose products are required for protein export. The *secB* gene codes for a cytoplasmic protein which is required for the export of a subset of cell envelope proteins (Kumamoto and Beckwith, 1985; Collier *et al.*, 1988; Kumamoto, 1989). The *secA* gene codes for a peripheral membrane protein (Oliver and Beckwith, 1982; Cabelli *et al.*, 1988) and the *secY/prlA* and *secE* genes code for integral proteins of the cytoplasmic membrane, all of which are essential components of the export mechanism (Emr *et al.*, 1981; Ito *et al.*, 1983; Schatz *et al.*, 1989). Genetic studies have also revealed the existence of the *secD* locus, mutations which cause pleiotropic defects in protein export (Gardel *et al.*, 1987). In addition, genes have been identified which code for two different membrane signal peptidases (*lepA* and *lspA*) involved in the processing of subsets of exported proteins (Silver and Wickner, 1983; Innis *et al.*, 1984; Yu *et al.*, 1984).

Other factors proposed to play a role in protein secretion in *E.coli* are trigger factor (Crooke and Wickner, 1987), 4.5S RNA (Poritz *et al.*, 1988) and the product of the *ffh* gene (Bernstein *et al.*, 1989; Romisch *et al.*, 1989). However, no genetic evidence exists for the function of trigger factor or the *ffs* and *ffh* products. We have argued elsewhere that because of the variety of genetic approaches used so far to detect *sec* genes, the collection of *sec* genes now known may represent all or nearly all of the genes essential for protein secretion (Schatz *et al.*, 1989).

In this paper, we report a detailed characterization of the *secD* locus, including DNA sequencing, Tn*phoA* insertion analysis and complementation of *secD* mutants. Our results show that there are at least two *sec* genes in this locus, *secD* and *secF*. Both of these genes are required for protein export and both appear to code for integral proteins of the cytoplasmic membrane.

Results

secD mutants

Mutations in the *secD* locus were originally isolated by a selection using strains in which β-galactosidase was fused to an exported protein (Gardel *et al.*, 1987). These *secD* mutants are cold-sensitive for growth and exhibit strong defects in protein export at low temperature (23°C). Subsequently, we used another approach to obtain *sec* mutants (Riggs *et al.*, 1988; Schatz *et al.*, 1989). This approach depended on the finding that the *secA* gene is regulated by the secretion needs of the cell. Lac^+ selections with a strain carrying a *secA–lacZ* fusion yielded a larger collection of *secD* and other *sec* mutants. All of the conditional lethal *secD* mutants obtained with this selection

were cold-sensitive. Several mutants were also isolated which were viable at all temperatures tested (non-conditional = nc mutants). So far, we have not obtained a temperature-sensitive mutation in this locus.

The mutants isolated more recently had not been examined for secretion defects. Therefore, we have tested certain of these mutants by pulse-labeling protein and looking for the appearance of precursors of a secreted protein. The mutants tested (10) all showed clear-cut export defects in the export of RBP at both 23°C (Figure 1A) and 37°C (Figure 1B). Some showed a stronger defect at 23°C.

Subcloning and restriction mapping of the secD locus

We previously described the cloning of the *secD* locus in a pBR322 derivative, pCG170 (Gardel *et al.*, 1987). We have further subcloned the chromosomal insert to obtain a 4.4 kb fragment which still complements the *secD1* mutation. The plasmid carrying this fragment is pCGSH1. Use of a wide variety or restriction enzymes gave the restriction map of the region shown in Figure 2A.

Sequence of the secD region

We have determined the DNA sequence of the entire 4.4 kb insert in pCGSH1 (Figure 3). Both DNA strands were sequenced. Analysis of this sequence reveals six potential open reading frames (ORFs) in one strand (ORFs 1−6), and a single ORF on the complementary strand, ORF7 (Figure 2A). ORFs 1−7 could code for proteins of 12.5, 12.7, 12, 65 or 67, 35, 12.5 and 19.5 kd, respectively. Analysis of the rare codon usage in the ORFs reveals that ORFs 1−6 contain a very low percentage of rare codons (1−2%) while ORF7 contains a percentage similar to that of non-coding DNA (11%) and, therefore, may not be translated. Several features of the proteins potentially coded for by these ORFs stand out. The sequence of ORF3 corresponds to a protein with a highly hydrophobic amino terminus which could be a signal sequence or a membrane spanning stretch. ORF4 and ORF5 both could code for proteins containing extensive hydrophobic regions (Figure 5), suggesting that these proteins may be localized to the cytoplasmic membrane. In addition to the extensive hydrophobic stretches, each of these proteins contains one particularly long hydrophilic sequence (420 amino acids in the case of ORF4 and 100 amino acids in the case of ORF5).

The sequences of ORF1, ORF2 and ORF6 would correspond to proteins without long hydrophobic stretches, and thus would probably make cytoplasmic proteins, if expressed. Translation of ORF4 could initiate with either an AUG or a GUG codon, the latter giving rise to a protein 11 amino acids longer than the former. Both the GUG and AUG codons are preceded by potential ribosome binding sites. Translation of ORF5 would most likely imitate at a GUG codon which is preceded by a ribosome binding site.

Maxicell analysis

The proteins coded for by pCGSH1 were analyzed by maxicell analysis. To determine which of the ORFs code for the proteins seen in these experiments, derivatives of pCGSH1 containing transposon insertions and the deletion plasmid pCGEE were also examined. The location of these insertions and the deletions are shown in Figure 2B and C; the insertion mutations are described in a subsequent section. Three proteins of mol. wts 12, 12.5 and 65 kd were identified as products of pCGSH1 in maxicells (Figure 4, lanes A and F). Only the 12 kd protein was expressed by the deletion plasmid pCGEE, which contains ORF1, ORF2 and ORF3 (lane G). This suggests either that only one of these three ORFs is expressed, or that their gene products co-migrate in our gel system. The transposon insertion, 5.3, which is in ORF6, eliminated production of the 12.5 kd protein (lane D). Two insertions in ORF4, 6.2 and 9.2, eliminated production of the 65 kd protein (lanes C and E). We detected no effect of the insertion in ORF5 on the synthesis of any protein (lane B). Despite repeated attempts, we could not identify the proteins which would be encoded by ORF5 and ORF7, nor could we determine which of three ORFs, ORF1, ORF2 or ORF3, encoded the 12 kd protein(s).

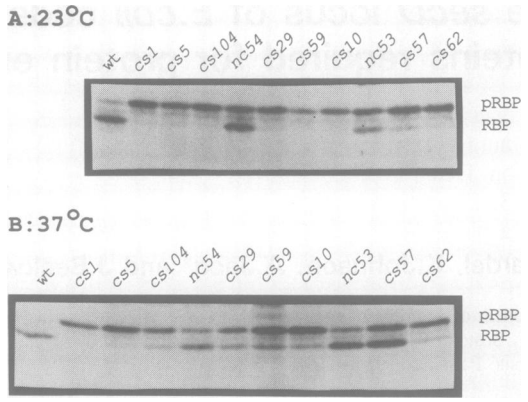

Fig. 1. Accumulation of preRBP in the *secD* mutants. Cultures were pulse-labeled for 30 s either (**A**) 3 h after a shift from 37°C to 23°C or (**B**) at 37°C. Labeled cells were then immunoprecipitaed with anti-RBP. nc = non-conditional.

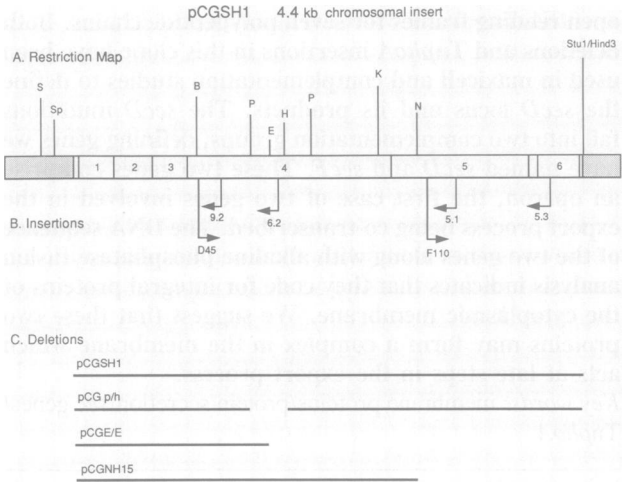

Fig. 2. (**A**) Restriction map of the 4.4 kb chromosomal insert of pCGSh1. The open bar represents the insert; the numbers within the open bar signify open reading frames. The shaded region represents the flanking pBR322 sequences. B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; N, *Nsi*I; P, *Pvu*II; S, *Sal*I; Sp, *Sph*I. (**B**) Tn*phoA* transposon insertions within the insert. Arrows denote the orientation of insertions. (**C**) Deletion plasmids with different amounts of the chromosomal insert. Lines represent DNA from the insert present in the plasmids.

Complementation of secDcs mutations by pCGSH1 derivatives

To determine which of the ORFs was *secD*, we tested the ability of derivatives of pCGSH1 carrying deletions and transposon insertions to complement the cold sensitivity of *secD* mutations. Initial studies were done with the *secDcs1* mutation. Deletion plasmid pCGEE, which carries ORFs 1, 2 and 3, failed to complement this mutation. On the other hand, a plasmid carrying a transposon insertion in ORF6 did complement *secDcs1*. Finally, insertions in either ORF4 or ORF5 in pCGSH1 resulted in an inability to complement the mutation. These data suggested either that *secDcs1* required both the ORF4 and ORF5 gene products for complementation or that ORF4 and ORF5 comprised an operon and *secDcs1* mapped to ORF5. According to this latter explanation the failure of insertions in ORF4 to complement would be due to polarity effects on ORF5.

To distinguish between these possibilities, we carried out complementation studies with the seven *secDcs* mutants described above, along with *secDcs1*. Transposon insertions in ORF4 (pD45) and ORF5 (pF110) were used (Table I). We found that the insertion in ORF4 (pD45) eliminated complementation of all *secD* mutations. The insertion in ORF5 (pF110) complemented some of the *secD* mutations (class I), indicating that this class comprised mutations in ORF4. With a second class of mutations (class II), this same plasmid failed to complement, but also resulted in transformants that grew poorly at the permissive temperature. As we discuss in a subsequent section, pF110 produces a fusion protein in which the ORF5 gene product is fused to alkaline phosphatase. Since it seemed possible that the sickly growth phenotype was due to the fusion protein, we also tested a plasmid in which ORF5 and ORF6 had been deleted (pCGNH15, Figure 2B). However, this plasmid showed the same complementation behavior as pF110 (Table I), including the effects on growth rate with certain mutants. The results suggest that the class II mutants carry mutations in ORF5, and that expressing ORF4 at high levels in the absence of ORF5 is deleterious to the cell.

The existence of two complementation groups raised the possibility that ORF4 and ORF5 comprised an operon of secretion genes, and suggested that the *secD* mutations included those in both genes. The failure of insertions in ORF4 to complement any *secD* mutation could be due to polar effects of insertion in ORF4 on ORF5.

Complementation with a non-polar deletion within ORF4

To obtain further evidence for an operon of secretion genes, we have generated a mutation in ORF4 which should not be polar on downstream genes. Using restriction enzymes, we constructed an in-frame deletion in this gene. This was made possible by the existence of *Pvu*II and *Hpa*I sites in appropriate sequences. Cutting with the two restriction enzymes and blunt end religation should yield an in-frame deletion which removes 132 amino acids from the protein. A construct which exhibited the correct restriction map was sequenced and shown to have generated the in-frame deletion. The plasmid carrying this deletion was named pCGp/h (Figure 2C).

The in-frame deletion should not be polar on the expression of downstream genes. Thus, if the failure of insertions in ORF4 to complement mutations which appeared to be in ORF5 was due to polar effects, this in-frame deletion should obviate the problem. When pCGp/h was tested for complementation of the collection of *secD* mutations, it failed to complement those mutations thought to lie in ORF4 (class I). In contrast, those mutations which were complemented by pCGp/h were ones that were not complemented by the insertion in ORF5. These results suggest that the *secD* locus is made up of at least two *sec* genes which together comprise an operon. We propose that ORF4 be called *secD* and ORF5 be called *secF*.

One of the mutants presumed to be in ORF4, *secDcs104*, when transformed with the plasmid pCGp/h, showed an enhanced growth defect similar to that seen with transformation of class II mutants with pF110. The mutant *secDcs1* behaved in the same fashion. The properties of *secDcs1* suggest either that it is a double mutation in *secD* or *secF* or that for some other reason it requires both ORF4 and ORF5 for complementation.

TnphoA analysis of the secD plasmid

Tn*phoA*, a derivative of the transposon Tn*5*, carries the gene for alkaline phosphatase (*phoA*) such that upon transposition the alkaline phosphatase can be fused to the product of the gene in which it is inserted (Manoil and Beckwith, 1985; Manoil *et al.*, 1990). The hybrid protein produced will exhibit alkaline phosphatase enzymatic activity if, and only if, the other gene product contains a protein export signal. For this reason, Tn*phoA* can be used to ascertain which genes code for cell envelope proteins as well as to determine the location of genes on a plasmid. Insertions of Tn*phoA* into either plasmid pCG170 or pCGSH1 were obtained and inserts which did and did not show alkaline phosphatase activity were examined. The sites of the Tn*phoA* insertions were determined by restriction mapping and DNA sequencing (Figure 2B).

We found active Tn*phoA* fusions in *secD* and *secF*. Two such fusions were used in the complementation analysis, pD45 and pF110. The existence of these active fusions indicates that the genes for these proteins are, in fact, expressed, and, further, that their gene products are components of the cell envelope as suggested by an inspection of the deduced amino acid sequences. Two insertions which exhibited high alkaline phosphatase activity, one in *secD* and one in *secF*, were determined by DNA sequencing to lie within the large hydrophilic domains of the encoded polypeptides. These results suggest that the hydrophilic domains are localized to the periplasmic space (Manoil and Beckwith, 1986). The orientation of the Tn*phoA* insertions confirmed the direction of transcription and translation predicted by the orientation of the ORFs.

secD and secF code for proteins with sequence similarity

The *secD* and *secF* gene products have highly similar hydropathy profiles (Figure 5), especially in their carboxy termini. This finding raised the possibility that the two proteins might share sequences in common. In fact, the hydrophobic carboxy-terminal portions of the two proteins show substantial similarity over 176 amino acids (Figure 6). The two proteins share 26% identical amino acids and 48% similar amino acids in this region. It could be argued that

the similarity between the two proteins results from the bias towards hydrophobic amino acids in these domains rather than from a common origin of these sequences. To test this possibility, the sequences of the appropriate region of SecD was randomized and the randomized sequences aligned with the corresponding region of SecF. If the bias towards hydrophobic amino acids explains the high degree of similarity between SecD and SecF, then the scores of the alignments of the randomized sequences should be close to the score of the true alignment. This is not the case. The average aligned score of the SecF randomized sequences with SecD is 12 standard deviations below that of the alignment of the actual sequences.

Similarity searches with the SecD and SecF proteins

The predicted amino acid sequence of the SecD protein was compared with other known peptides entered in the NIH and Swiss Protein data libraries and no significant similarities involving the entire SecD sequence were found (Lipman and Pearson, 1985). However, among the most similar proteins were the DedA protein (Nonet *et al.*, 1987) of *E.coli* (22% identity in 142 amino acids), the yeast ATPase protein 3 [Macino and Tzagoloff (1980) with 27% identity in 60 amino acids], the acetylcholine receptor [Baldwin *et al.* (1988) with 21% identity in 101 amino acids] and the tetracycline resistance protein [Sanchez-Pescador *et al.* (1988) with 29% identity in 63 amino acids]. The DedA protein has no known function and was identified as an ORF located distal to *hisJ* at 50 min on the *E.coli* chromosome. The similarities with the ATPase 3 and the acetylcholine receptor are matches involving putative transmembrane domains and hence may reflect a bias towards hydrophobic amino acids. The similarity with tetracycline resistance protein involves the long hydrophilic domain of the SecD protein. The significance of these similarities remains unclear.

When the SecF protein was compared with the Swiss Protein Database, no proteins were found which showed a high degree of similarity to its entire sequence. However, among the top five most similar proteins were the human multiple drug resistance (MDR) gene [Chen *et al.* (1986a)

```
GATCTTTGATGAGTG   CCGTATCCTGCTGACTGGGATTACGCAAAACGCTCCATGGAGATGTCTCTGCGTTGGGCGAACGTAGCCGT    90
                                                                M  G  E  R  S  R
                                                               ORF1
GAGCGTTTTGACAGTC CGGAAACAAAAATGCGCTGTTTGGTATCATCCAGGGCAGCGTTTACGAAGATTTACGTGATATTTCTGTTAAA   180
E  R  F  D  S  L  G  N  K  N  A  L  F  G  I  I  Q  G  S  V  Y  E  D  L  R  D  I  S  V  K

GGTCTGGTAGATATCGGTTTTGATCGGTACGCTGTCGGCGGTCTGGCTGTGGGTGAGCCGAAAGCAGATATGCACCGCATTCTGGAGCAT   270
G  L  V  D  I  G  F  D  R  Y  A  V  G  G  L  A  V  G  E  P  K  A  D  M  H  R  I  L  E  H

GTATGCCCGCAAATTCCGGCAGACAAACCGCGTTACCTGATGGGCGTTGGTAAACCAGAAGACCTGGTTGAAGGCGTACGTCGTGGTATC   360
V  C  P  Q  I  P  A  D  K  P  R  Y  L  M  G  V  G  K  P  E  D  L  V  E  G  V  R  R  G  I

GATATGTTTGACTCGTAATGCCAACCCCCCAACGCCCGAATGGTCATTTGTTCGTGACCGATGGCGTGGTGAAAATCCGCAATGCGAAGT   450
D  M  F  D  S  *M  P  T  P  Q  R  P  N  G  H  L  F  V  T  D  G  V  V  K  I  R  N  A  K  Y
                 ORF2
ATAAGAGCGATACTGGCCCACTCGATCCTGAGTGTGATTGCTACACCTGTCGCAATTATTCACGCGCTTACTTGCATCATCTTGACCGTT   540
  K  S  D  T  G  P  L  D  P  E  C  D  C  Y  T  C  R  N  Y  S  R  A  Y  L  H  H  L  D  R  C

GCAACGAAATATTAGGCGCGCGACTCAACACCATTCATAACCTTCGTTACTACCAGCGTTTGATGGCGGGTTTACGCAAGGCTATTGAAG   630
  N  E  I  L  G  A  R  L  N  T  I  H  N  L  R  Y  Y  Q  R  L  M  A  G  L  R  K  A  I  E  E

AGGGTAAATTAGAGAGCTTCGTAACTGATTTTTACCAGCGTCAGGGGCGAGAAGTACCACCTTTGAACGTTGATTAATATTAATAATGAG   720
  G  K  L  E  S  F  V  T  D  F  Y  Q  R  Q  G  R  E  V  P  P  L  N  V  D  *

GGAAATTTAATGAGCTTTTTTATTTCTGATGCGGTAGCGGCAACGGGTGCACCGGCGCAAGGTAGCCCGATGTCTTTGATTTTGATGCTG   810
        M  S  F  F  I  S  D  A  V  A  A  T  G  A  P  A  Q  G  S  P  M  S  L  I  L  M  L
       ORF3
GTGGTATTCGGTCTGATTTTCTATTTCATGATCCTGCGTCCACAGCAGAAGCGCACCAAAGAACACAAAAAGCTGATGGACTCCATTGCC   900
V  V  F  G  L  I  F  Y  F  M  I  L  R  P  Q  Q  K  R  T  K  E  H  K  K  L  M  D  S  I  A

AAAGGTGATGAAGTTCTGACGAACGGTGGCCTGGTTGGTCGCGTAACCAAAGTAGCGGAAAACGGCTACATTCGTATCGCGCTGAATGAC   990
K  G  D  E  V  L  T  N  G  G  L  V  G  R  V  T  K  V  A  E  N  G  Y  I  R  I  A  L  N  D

ACCACTGAAGTAGTTATTAAACGTGACTTCGTAGCTGCCGTCCTGCCGAAAGGCACCATGAAGGCGCTGTAATTAAAATTTTTCCCTAAG  1080
T  T  E  V  V  I  K  R  D  F  V  A  A  V  L  P  K  G  T  M  K  A  L  *

GGAATTGCCGTGTTAAACCGTTATCCTTTGTGGAAGTACGTCATGCTGATCGTGGTGATTGTCATCGGTCTGCTGTATGCGCTTCCCAAC  1170
        M  L  N  R  Y  P  L  W  K  Y  V  M  L  I  V  V  I  V  I  G  L  L  Y  A  L  P  N
       secD
CTGTTTGGTGAGGATCCGGCTGTTCAGATCACTGGTGCGCGCGGAGTCGCCGCCAGTGAGCAAACGCTGATCCAGGTCCAGAAAACGTTA  1260
L  F  G  E  D  P  A  V  Q  I  T  G  A  R  G  V  A  A  S  E  Q  T  L  I  Q  V  Q  K  T  L

CAAGAAGAAAAAATAACTGCTAAGTCTGTGGCACTGGAAGAGGGCGCTATTCTTGCGCGCTCCGACTCCACTGACACCCAGTTGCGCGCT  1350
Q  E  E  K  I  T  A  K  S  V  A  L  E  E  G  A  I  L  A  R  S  D  S  T  D  T  Q  L  R  A

CGTGAAGCATTAATGGGCGTTATGGGTGACAAATACGTCGTGGCGCTTAACCTTGCCCCGGCAACGCCGCGCTGGCTGGCAGCTATTCAC  1440
R  E  A  L  M  G  V  M  G  D  K  Y  V  V  A  L  N  L  A  P  A  T  P  R  W  L  A  A  I  H

GCTGAGCCGATGAAGCTCGGCCTTGACCTGCGTGGCGGCGTTCACTTCCTGATGGAAGTGGATATGGACACCGCGCTTGGCAAACTCCAG  1530
A  E  P  M  K  L  G  L  D  L  R  G  G  V  H  F  L  M  E  V  D  M  D  T  A  L  G  K  L  Q

GAACAAAATATCGATAGCCTAGCCAGTGACCTGCGCGAAAAGGGCATCCCGTATACCACTGTTCGTAAAGAAAACAACTACGGCCTGAGC  1620
E  Q  N  I  D  S  L  A  S  D  L  R  E  K  G  I  P  Y  T  T  V  R  K  E  N  N  Y  G  L  S

ATCACTTTCCGCGATGCTAAAGCTCGTGATGAAGCCATTGCGTATCTGAGCAAGCGCCATCCGGACCTGGTGATTAGCAGCCAGGGCAGC  1710
I  T  F  R  D  A  K  A  R  D  E  A  I  A  Y  L  S  K  R  H  P  D  L  V  I  S  S  Q  G  S

AACCAGCTGCGTGCGGTAATGAGCGATGCTCGTCTGAGTGAAGCGCGTGAATATGCGGTGCAGCAGAACATTAATATCCTGCGTAACCGT  1800
N  Q  L  R  A  V  M  S  D  A  R  L  S  E  A  R  E  Y  A  V  Q  Q  N  I  N  I  L  R  N  R

GTAAACCAACTTGGCGTGGCGGAGCCGGTGGTTCAGCGTCAGGGTGCTGACCGTATCGTTGTTGAACTGCCAGGTATTCAGGACACTGCG  1890
V  N  Q  L  G  V  A  E  P  V  V  Q  R  Q  G  A  D  R  I  V  V  E  L  P  G  I  Q  D  T  A

CGTGCGAAAGAGATTCTGGGTGCGACGGCAACGCTGGAATTCCGTCTGGTAAACACCAACGTTGACCAGGCCGCTGCGGCATCCGGTCGC  1980
R  A  K  E  I  L  G  A  T  A  T  L  E  F  R  L  V  N  T  N  V  D  Q  A  A  A  A  S  G  R

GTACCGGGCGACTCTGAAGTGAAACAGACCCGCGAAGGTCAGCCAGTTGTGCTGTACAAACGCGTAATTCTGACCGGTGACCATATCACC  2070
V  P  G  D  S  E  V  K  Q  T  R  E  G  Q  P  V  V  L  Y  K  R  V  I  L  T  G  D  H  I  T

GACTCCACTTCCAGCCAGGACGAATACAACCAACCGCAGGTTAACATCTCGCTCGATAGCGCTGGTGGTAACATCATGTCTAACTTCACT  2160
D  S  T  S  S  Q  D  E  Y  N  Q  P  Q  V  N  I  S  L  D  S  A  G  G  N  I  M  S  N  F  T

AAGGACAACATCGGCAAACCGATGGCAACCCTGTTTGTGGAGTACAAAGACAGCGGTAAGAAAGATGCAAATGGTCGTGCGGTTCTGGTG  2250
K  D  N  I  G  K  P  M  A  T  L  F  V  E  Y  K  D  S  G  K  K  D  A  N  G  R  A  V  L  V
```

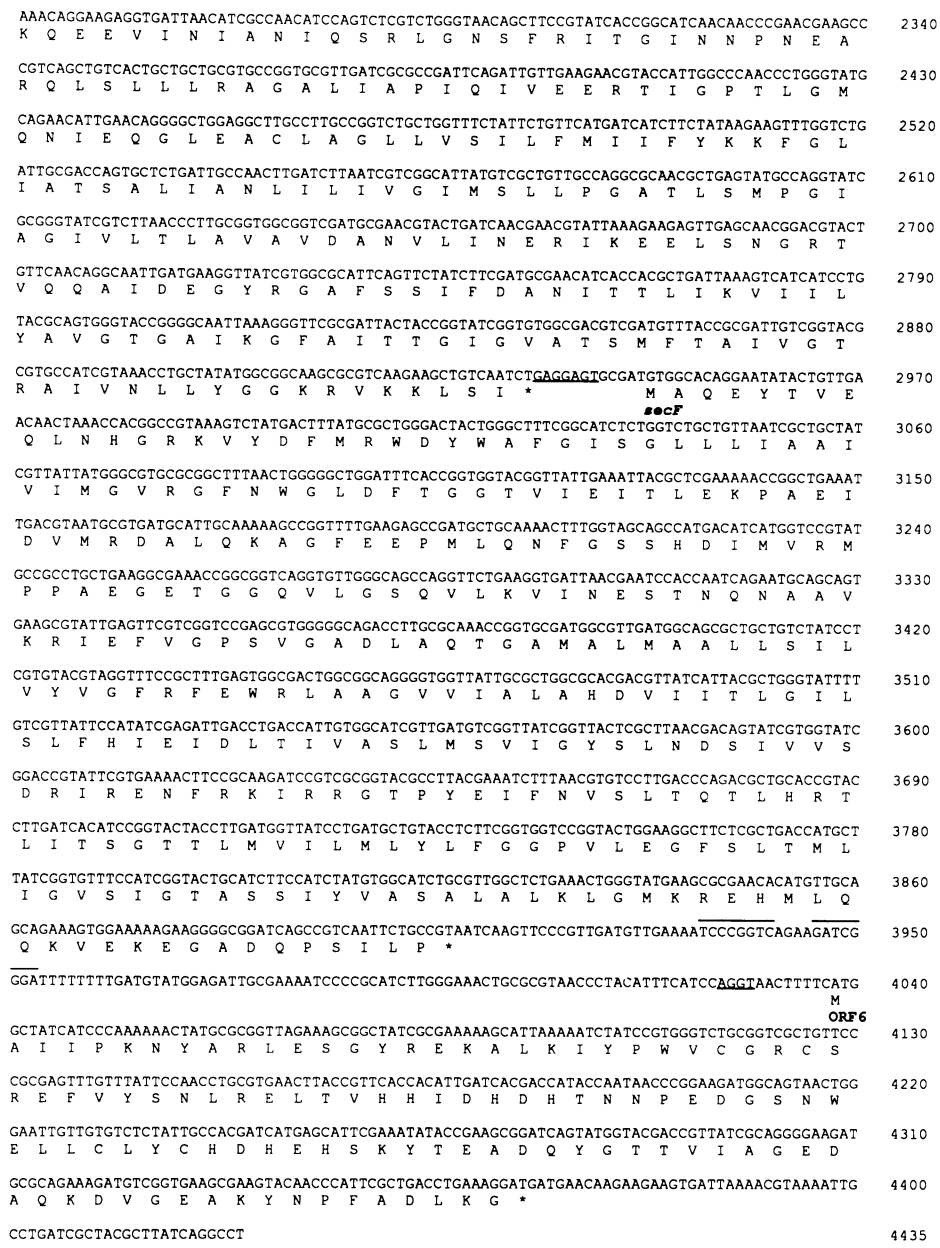

Fig. 3. DNA sequence of the 4.4 kb region encoding the *secD* operon. The deduced amino acid sequences of the open reading frames are presented below the DNA sequence. Termination codons are indicated by an asterisk, potential ribosome binding sites are underlined, and a potential transcriptional terminator for the *secD* operon is overlined.

with 27% identity in 56 amino acids], the *oppB* gene of the *E.coli* oligopeptide permease operon [Hiles *et al.* (1987) with 27% identity in 60 amino acids] and the *arsB* gene of the *E.coli* arsenate resistance operon [Chen *et al.* (1986b) with 24% identity in 156 amino acids). These proteins each contain sequences similar to the hydrophobic carboxy-terminus of SecF, but yet show little sequence similarity to one another or to SecD. Therefore, the significance of these similarities is unclear. However, we note that both *oppB* and *arsB* code for membrane proteins of transport systems which include ATP binding proteins, while MDR exhibits sequence similarity to others of these membrane proteins.

Discussion

The sequencing and complementation studies reported here demonstrate that the *secD* locus is composed of two genes involved in protein export. Tn*phoA* insertions in and a deletion of the downstream gene, *secF*, complemented a subset (class I) of export-defective mutants. An in-frame deletion in the upstream gene, *secD*, complemented all but one (*secD1*) of the remaining mutations (class II). Partial complementation of *secD1* was seen with the 5.1 insert in *secF* and full complementation with a *secD*$^+$ plasmid. This mutant may be composed of two mutations, one each in *secD* and *secF*, or in another gene in this region or may be a partially dominant mutation.

The finding that an in-frame deletion in *secD* will complement *secF* mutants, but Tn*phoA* insertions in the same gene will not, strongly suggests that the two genes comprise an operon. We have not yet established whether the operon extends beyond these two genes. Tn*phoA* insertions in *secD* or *secF* do not eliminate the synthesis of the 12.5 kd protein coded for by the gene downstream of *secF* indicating that

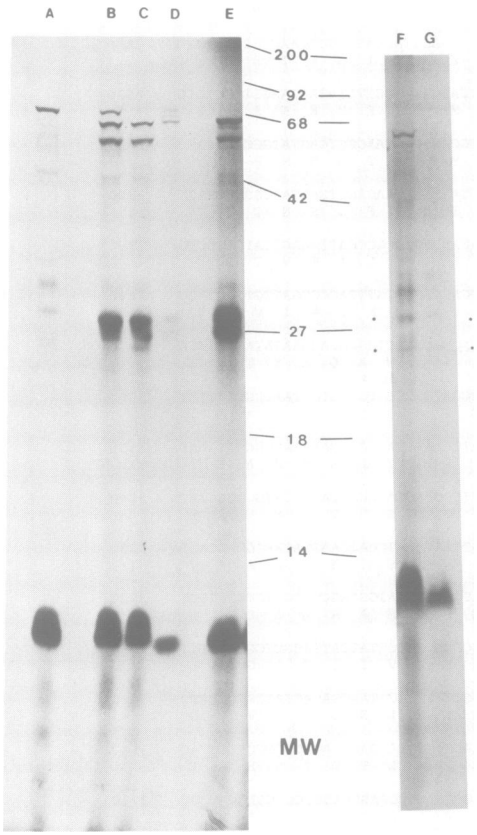

Fig. 4. Maxicell analysis of the deletion and insertion plasmids. The procedure is discussed in Materials and methods. Lanes A and F, pCGSH1; lane B, pCGSH1-5.1; lane C, pCGSH1-6.2; lane D, pCGSH1-5.3; lane E, pCGSH1-9.2; lane G, pCGE/E. Mol. wts of standard proteins are shown. Proteins found on the parent pBR322 are marked with black dots. The Tn*phoA* transposon encodes the Tn*5* proteins with mol. wts of 58, 54 and 25 kd.

this gene is not in the operon. However, it is still possible that the three ORFs upstream of *secD* are part of the operon. This possibility is currently under investigation.

Interestingly, nearly all the *sec* genes detected so far are in operons. The *secY* gene is part of an operon of ribosomal proteins (Shultz *et al.*, 1982). The *secE* gene is in an operon with a gene involved in transcription termination (Schatz *et al.*, 1989). The *secA* gene is preceded in the same operon by a gene whose function is not known and followed by the *mutT* gene (M.Schmidt and D.Oliver, unpublished results). In addition, the *lepA* gene is preceded in its operon by a gene coding for a non-essential protein of unknown function with an apparent GTP binding site (Ahnn *et al.*, 1986). The operon described in this paper is the first in which two genes involved in the export process are grouped in a single operon.

In maxicells, we only observed a subset of the proteins potentially coded for by plasmid pCGSH1. Included among the missing proteins is the *secF* gene product, a 35 kd protein. We know that this gene is expressed *in vivo* because of the complementation studies and the active *phoA* fusion to *secF* which produces a hybrid protein of the appropriate mol. wt when visualized on SDS gels after pulse-labeling and precipitation with anti-alkaline phosphatase (unpublished data). The failure to see the SecF protein in maxicells may be related to some peculiarity of expression of the operon in maxicells or it may be due to an altered mobility of the

Table I. Complementation of *secD* mutants

'*secD*' allele	Plasmids						
	pBR322	pCGSH1	pD45	pF110	pCGNH15	pCGp/h	Class
cs29	−	+	−	+	+	−	I
cs57	−	+	−	+	+	−	I
cs59	−	+	−	+	+	−	I
cs104	−	+	−	+	+	ST	I
cs5	−	+	−	ST	ST	+	II
cs10	−	+	−	ST	ST	+	II
cs62	−	+	−	ST	ST	+	II
cs1	−	+	−	ST	ST	ST	III

Plasmids were transformed into *recA* derivatives of PR478 containing the mutations indicated. Complementation was tested by the ability of the plasmids to correct the cold-sensitive growth defect conferred by the mutations (at 23°C). A + indicates that the plasmids complements the growth defect; a − indicates that it did not; ST indicates that the transformants displayed reduced viability at 37°C, compared with the same mutant transformed with pBR322, and did not grow at 23°C.

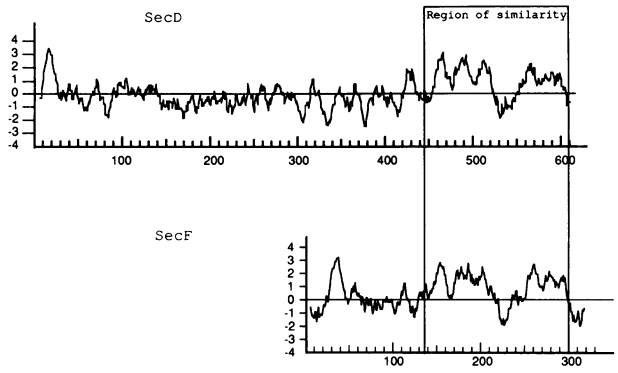

Fig. 5. Hydropathy plot of the SecD and SecF proteins. A Kyte–Doolittle hydropathy plot of the SecD and SecF proteins was performed using a scanning window of 11 amino acids. The region in which the two proteins show sequence similarity is boxed.

```
                130        140        150        160        170        180
SecF      ...VKRIEFVGPSVGADLAQTGAMA-LMAALLSILVYVGFRFEWRLAAGVVIALAHD
                 .::..: .  . :  : :...:.:::. . :. .. : :   ::. .
SecD      ...IVEERTIGPTLGMQNIEQGLEACLAGLLVSILFMIIFYKKFGLIA--TSALIAN
               440        450        460        470        480

                190        200        210        220        230
SecF      VIITLGILSLFH-IEIDLTIVASLMSVIGYSLNDSIVVSDRIRENFRKIRRGTPYEIFN
          .:...::.::.  ....  .:... ... ..........::.:.... :  :. ....
SecD      LILIVGIMSLLPGATLSMPGIAGIVLTLAVAVDANVLINERIKEELSNGR--TVQQAID
           490        500        510        520        530        540

               240        250        260        270        280        290
SecF      VSLTQTLHRTLITSGTTLMVILMLYLFGGPVLEGFSLTMLIGVSIGTASSIYVASA-LA
           . . .... .  . :::. ...:.:.:  ...::.:  :::...  ..:  ..: ..
SecD      EGYRGAFSSIFDANITTLIKVIILYAVGTGAIKGFAITTGIGVATSMFTAIVGTRAIVN
               550        560        570        580        590        600

                300        310        320
SecF      LKLGMKREHMLQQKVEKEGADQPSILP
          :  : :: ..:
SecD      LLYGGKRVKKLSI
               610
```

Fig. 6. Alignment of the SecD and SecF proteins. The SecD and SecF proteins were aligned using the FASTP program for sequence comparison. The resulting alignment was optimized using the ALIGN program from BIONET. The two proteins share 26% identical and 48% similar amino acids in this region.

protein on gels. Since SecF is predicted to be a membrane protein, altered mobility on gels would not be surprising. Other examples of substantially different mobilities from those expected have been reported with membrane proteins (Froshauer and Beckwith, 1984; Akiyama and Ito, 1985).

The amino acid sequences of the SecD and SecF proteins deduced from the DNA sequences reveal substantial

hydrophobic stretches in both proteins. These sequences are typical of proteins integral to the cytoplasmic membrane. Preliminary topological analysis of these two proteins, using Tn*phoA* analysis, is consistent with a membrane location, and suggests that they have large periplasmic domains of unknown function. This is in contrast to the SecY and SecE proteins, which appear to have most of their sequence integrated into the membrane (Akiyama and Ito, 1987; Schatz *et al.*, 1989).

There is extensive sequence similarity between the SecD and SecF proteins. This similarity lies in the carboxy-terminal hydrophobic region of the proteins. One possible role for these similar sequences could be in the formation of a complex either between the two proteins or to another protein with which they both interact. Further, the SecF protein shows sequence similarity with the hydrophobic regions of certain of the class of proteins related to and including the MDR protein of eukaryotic cells. Some of these proteins are involved in export of small molecules (Chen *et al.*, 1986a and b; Gros *et al.*, 1986) and others in the export of small peptides or larger polypeptide chains (Gerlach *et al.*, 1986; McGrath and Varshavsky, 1989). Although the sequence similarity is not extensive, it does raise the possibility of a common mechanism for many of these processes.

Extensive genetic selection and screening procedures have been used to detect mutations in genes involved in protein export in *E.coli*. We have argued elsewhere that most or all of the genes required for the process may have been detected (Schatz *et al.*, 1989). The genes that have now been characterized are *secA*, *secB*, *secD*, *secE*, *secF* and *secY/prlA*. Four of the six proteins coded for by these genes are membrane proteins. No evidence for an SRP-like complex of six proteins and an RNA emerges from the genetic studies nor from the *in vitro* studies. The soluble apparatus for protein translocation appears to be much simpler than that of mammalian cells. However, the finding of four membrane proteins indicates that a complex apparatus does exist in the membrane to promote the transfer of proteins through the lipid bilayer. Whether some or all of these proteins comprise a pore in the membrane or are simply there to facilitate passage of proteins through the membrane is not yet clear.

Two findings raise the possibility that the SecD and SecF proteins act at later steps of the export process than SecE and SecY. First, the apparent large periplasmic domains in the two former proteins suggest that part of the function of these proteins may be at steps subsequent to the initial interaction with the membrane. These periplasmic domains could be involved in final steps of export or in promoting interaction of exported proteins with the signal peptidases. Second, an extensive search for mutations (*prl* mutations) that suppress the defect of signal sequence mutations yielded mutations in only the *secA*, *secE* and *prlA/secY* genes (Stader *et al.*, 1989). The properties of these mutations suggest that the products of these *sec* genes interact with signal sequences. The failure to find suppressors in the *secD* and *secF* genes, which represent a sizeable target for mutations, could mean that their gene products act at steps after recognition of the signal sequence has taken place.

Genetic studies with the yeast *Saccharomyces cerevisiae* have implicated at least five genes in the process of secretion into the RER. Depletion of the three 70 kd heat-shock proteins leads to a defect in protein secretion (Deshaies *et al.*, 1988). One or all of these proteins may be involved in maintaining proteins in a secretion-competent state. This activity would be analogous to the proposed function for SecB in *E.coli*. A second gene codes for a homolog of the BiP protein which, in mammalian cells, is located in the lumen of the RER, and binds to unfolded or misfolded proteins (Vogel *et al.*, 1990). The remaining three genes appear to code for membrane proteins (Deshaies and Schekman, 1989; Rothblatt *et al.*, 1989). One of these shares significant sequence similarity with the *E.coli secY* gene (C.Sterling and R.Schekman, personal communication). No evidence exists so far for an SRP-like complex in this yeast. Therefore, it may be that the yeast and bacterial systems share more features than either does with the mammalian system.

Materials and methods

Plasmid constructions

pCGSH1 (SecD$^+$) was constructed by digesting pCG170 with *Hin*dIII and *Stu*I to yield a 2 kb deletion, according to the methods prescribed by New England Biolabs. The unique *Hin*dIII site within pBR322 and the unique *Stu*I within the chromosomal insert were ligated together after filling in the *Hin*dIII site with Klenow fragment and deoxynucleotide triphosphates.

pCGE/E was constructed by digesting pCG170 with *Eco*RI, phenol extracting, and ligating. pCGE/E contains 1100 bases of the chromosomal insert.

pCGp/h (having an internal deletion within *secD*) was constructed in the following manner. Within the pCGSH1 was a unique *Hpa* site gene (in *secD*) and three *Pvu*II sites (two within *secD* and one within the vector). pCGSH1 was cut completely with *Hpa*I and the linearzied plasmid was partially digested with *Pvu*II (1/5 enzyme) for 5 min at 23°C. After heat inactivation of the enzymes, the DNA was ligated and transformed into a *rec* deficient MC1000 strain. Colonies were screened for deletion plasmids and pCGp/h, with an internal in-frame deletion within *secD* was obtained. To verify the deletion, the fusion joint was sequenced.

TnphoA analysis

Tn*phoA* insertions were isolated in pCG170 and pCGSH1 as described (Manoil and Beckwith, 1986). Briefly, pCG170 or pCGSH1 was transformed into CC118 (PhoA$^-$ RecA$^-$) and the strains were grown to 2×10^9 cells/ml. The cultures were mixed with λTn*phoA* at an MOI of 0.1 and allowed to adsorb for 15 min at 37°C. After outgrowth in LB (in 10 separate tubes) at 34°C for 3 h, 0.1 ml aliquots were plated onto kanamycin (kan) (300 μg/ml), ampicillin (amp) (100 μg/ml), XP (60 μg/ml), TYE medium and incubated for 2 days at 34°C. The colonies on each plate were pooled and plasmids were harvested and transformed into Rec$^-$ cells on Kan Amp XP TYE medium. Colonies were purified and tested for Amp^r and Kan^r, and grown for harvesting plasmids. Restriction analysis identified those insertions within the complementing chromosomal insert. Plasmids were transformed into CG29 (MC1000 *secD1 recA*) and tested for complementation of *secD1* at 23°C. Insertion plasmids failing to complement fully the *secD1* mutation included pCGSH1-5.1, pCGSH1-6.2, pCGSH1-9.2, pD45 and pF110. Insertion plasmids pCG170-10.2 and pCGSH1-5.3 fully complemented the *secD1* mutation.

Maxicell analysis

Maxicells (Sancar *et al.*, 1979) were produced from CG51 (MC1000 *recA*) in the following manner. Ten ml of cells, grown to a Klett of 40 in M63 supplemented with 0.4% maltose, 2.5 μg/ml thiamine, 1.0% casamino acids and 50 μg/ml ampicillin, were UV irradiated for 50 s (25 J/m^2) followed by shaking at 37°C for 1 h. After this aeration, D-cycloserine (200 μg/ml) was added to kill any growing cells, and the culture was further incubated overnight. After thorough washing and starvation in minimal medium, these maxicells were labeled with 10 μCi/ml [^{35}S]methionine for 10 min. Cells were collected, lysed in sample buffer and heated at 85°C for 5 min before loading onto polyacrylamide gels [10–15%; Laemmli (1970)].

Sequencing

Cloning inserts into M13. Phage M13tg130 (Kieny *et al.*, 1983) was exclusively used for single stranded sequencing of one strand of the *secD* complementing insert (~4.4 kb). Two constructions were required to achieve

this. The KS phage had the 2.86 kb *Kpn*I–*Sph*I fragment of pCG170 cloned into the *Sph*I and *Kpn*I sites of the vector. The ES phage was composed of the 2.3 kb *Eco*RI–*Stu*I fragment cloned into the *Eco*RI and *Eco*RV sites of the M13 vector. No manipulations with the ends other than cutting and ligating were required for the above constructions.

M13 deletion constructions. In order to obtain a set of nested deletions required for single stranded sequencing into the inserts of the two phage constructs, the strategy employed was to isolate a single cut within the insert and then ligate this end to a unique site within the vector. Partial digests of replicative forms were made in the presence of 250 μg/ml ethidium bromide using the frequent cutting enzymes *Alu*I, *Bst*NI, *Dde*I, *Hgi*AI, *Hin*fI and *Rsa*I enzymes (Shortle *et al.*, 1981). These preparations were further digested with a site in the vector close to the primer and ligated and then transformed into strain 71-18. Purified phage were grown and lysed in a 1% SDS, 0.5 M EDTA and 50% glycerol solution. The phage DNA was screened on agarose sizing gels to observe deletions. In this manner, assorted deletion phages were obtained for single stranded sequencing.

Single stranded sequencing. In general, dideoxy sequencing was performed with Sequenase Version I in the manner prescribed by USB, using single stranded DNA (made from 5 h grown phage stocks) prepared as recommended by Amersham (Sanger *et al.*, 1977). Samples were loaded onto 5–8% urea gels that were run at 2000 V. Gels were washed in 10% MeOH and 10% acetic acid, dried and put on X-ray. For areas of sequence difficult to read because of compressions, dITP nucleotides were substituted for dGTP in the sequencing reactions.

Plasmid sequencing. Regions of the pCGSH1 insert which could not be sequenced using the deletions in M13 described above were analyzed by double stranded sequencing using synthetic oligonucleotide primers. The procedure used for plasmid preparation has been described by Zagursky *et al.* (1980). The plasmids were sequenced as described in the USB sequenase manual.

Acknowledgements

We thank Theresia Luna for excellent technical assistance and Ann McIntosh for excellent assistance in the preparation of this manuscript. This work was supported by grants from the American Cancer Society and the National Institutes of Health to J.B., a postdoctoral fellowship from the American Cancer Society of Massachusetts to A.J. and a predoctoral fellowship to K.J. from the National Science Foundation.

References

Ahnn,J., March,P.C., Takiff,H.E. and Inouye,M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8849–8853.
Akiyama,Y. and Ito,K. (1985) *EMBO J.*, **4**, 3351–3356.
Akiyama,Y. and Ito,K. (1987) *EMBO J.*, **6**, 3465–3470.
Baldwin,T.J., Yoshihara,C.M., Blackmer,K., Kintner,C.R. and Burden,S.J. (1988) *J. Cell Biol.*, **106**, 469–478.
Bernstein,H.D., Poritz,M.A., Strub,K., Hoben,P.J., Brenner,S. and Walter,P. (1989) *Nature*, **340**, 482–486.
Cabelli,R.J., Chen,L., Phang,C.T. and Oliver,D.B. (1988) *Cell*, **55**, 683–692.
Chen,C., Chin,J.E., Ueda,K., Clark,D.P., Pastan,I., Gottesmann,M.M. and Roninson,I.B. (1986a) *Cell*, **47**, 381–389.
Chen,C.M., Misra,T.K., Silver,S. and Rosen,B.P. (1986b) *J. Biol. Chem.*, **261**, 15030–15038.
Chirico,W.J., Waters,M.G. and Blobel,G. (1988) *Nature*, **332**, 805–810.
Collier,D.N., Bankaitis,V.A., Weiss,J.B. and Bassford,P.J., Jr (1988) *Cell*, **53**, 273–283.
Crooke,E. and Wickner,W. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5216–5220.
Deshaies,R.J. and Schekman,R. (1989) *J. Cell Biol.*, **109**, 2653–2664.
Deshaies,R.J., Koch,B.D., Werner-Washburne,M., Craig,E.A. and Schekman,R. (1988) *Nature*, **332**, 800–805.
Emr,S.D., Hanley-Way,S. and Silhavy,T. (1981) *Cell*, **23**, 79–88.
Evans,E.A., Gilmore,R. and Blobel,G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 581–585.
Froshauer,S. and Beckwith,J. (1984) *J. Biol. Chem.*, **259**, 1896–1903.
Gardel,C., Benson,S., Hunt,J., Michaelis,S. and Beckwith,J. (1987) *J. Bacteriol.*, **169**, 1286–1290.
Gerlach,J.H., Endicot,J.A., Juranka,P.F., Henderson,G., Sarangi,F., Deuchars,K.L. and Ling,V. (1986) *Nature*, **324**, 485–489.
Gilmore,R., Walter,P. and Blobel,G. (1982) *J. Cell Biol.*, **46**, 470–477.
Gros,P., Croop,J. and Housman,D. (1986) *Cell*, **47**, 371–380.
Hartmann,E., Wiedmann,M. and Rapoport,T.A. (1989) *EMBO J.*, **8**, 2225–2229.
Hiles,I.D., Gallagher,M.P., Jamieson,D.J. and Higgins,C.F. (1987) *J. Mol. Biol.*, **195**, 125–142.
Innis,M.A., Tokunaga,M., Williams,M.E., Loranger,J.M., Chang,S.-Y., Chang,S. and Wu,H.C. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3708–3712.
Ito,K., Wittekind,W., Nomura,M., Miura,A., Shiba,K., Yura,T., Miura,A. and Nashimoto,H. (1983) *Cell*, **32**, 789–797.
Kieny,M.P., Lathe,R. and Lecocq,J.P. (1983) *Gene*, **26**, 91–99.
Krieg,U.C., Johnson,A.E. and Walter,P. (1989) *J. Cell Biol.*, **109**, 2033–2043.
Kumamoto,C.A. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5320–5324.
Kumamoto,C.A. and Beckwith,J. (1985) *J. Bacteriol.*, **163**, 267–274.
Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 456–460.
Macino,G. and Tzagoloff,A. (1980) *Cell*, **20**, 507–517.
Manoil,C. and Beckwith,J. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8129–8133.
Manoil,C. and Beckwith,J. (1986) *Science*, **233**, 1403–1408.
Manoil,C., Mekalanos,J.J. and Beckwith,J. (1990) *J. Bacteriol.*, **172**, 515–518.
Meyer,D.I., Krause,E. and Dobberstein,B. (1982) *Nature*, **297**, 503–508.
McGrath,J.P. and Varshavsky,A. (1989) *Nature*, **340**, 400–404.
Nonet,M.L., Marvel,C.C. and Tolan,D.R. (1987) *J. Biol. Chem.*, **262**, 12209–12217.
Oliver,D. and Beckwith,J. (1982) *Cell*, **30**, 311–319.
Poritz,M.A., Strub,K. and Walter,P. (1988) *Cell*, **55**, 4–6.
Pugsley,A.P. (1989) *Protein Targeting*. Academic Press, San Diego.
Riggs,P., Derman,A. and Beckwith,J. (1988) *Genetics*, **118**, 571–579.
Romisch,K., Webb,J., Herz,J., Prehn,S., Frank,R., Vingron,M. and Dobberstein,D. (1989) *Nature*, **340**, 478–482.
Rothblatt,J.A., Deshaies,R.J., Sanders,S.L., Daum,G. and Schekman,R. (1989) *J. Cell Biol.*, **109**, 2641–2652.
Sancar,A., Hack,A.M. and Rupp,W.D. (1979) *J. Bacteriol.*, **137**, 692–693.
Sanchez-Pescador,R., Brown,J.T., Roberts,M. and Urdea,M.S. (1988) *Nucleic Acids Res.*, **16**, 1216–1217.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schatz,P.J., Riggs,P.D., Jacq,A., Fath,M.J. and Beckwith,J. (1989) *Genes Dev.*, **3**, 1035–1044.
Shortle,D., Dimaio,D. and Nathans,D. (1981) *Annu. Rev. Genet.*, **15**, 265–294.
Shultz,J., Silhavy,T.J., Berman,M.L., Fill,N. and Emr.S.D. (1982) *Cell*, **31**, 227–235.
Silver,P. and Wickner,W. (1983) *J. Bacteriol.*, **154**, 569–572.
Stader,J., Gansheroff,L.J. and Silhavy,T.J. (1989) *Genes Dev.*, **3**, 1045–1052.
Vogel,J., Misra,L. and Rose,M. (1990) *J. Cell Biol.*, **110**, 1885–1895.
Walter,P. and Blobel,G. (1982) *Nature*, **99**, 691–698.
Yu,F., Yamada,H., Daishima,K. and Mizushima,S. (1984) *FEBS Lett.*, **173**, 264–268.
Zagursky,R.J., Baumeister,K., Lomax,N. and Berman,M.L. (1980) *Gene Anal. Techn.*, **255**, 7973–7977.

Received on May 3, 1990; revised on June 21, 1990