Journal of Supramolecular Structure 13:147–163 (1980)
Membrane Transport and Neuroreceptors 81–97

Genetic Studies on Mechanisms of Protein Localization in Escherichia coli K-12

Michael N. Hall, Scott D. Emr, and Thomas J. Silhavy

Cancer Biology Program, NCI Frederick Cancer Research Center, Frederick, Maryland 21701

In the last few years, several laboratories have demonstrated that many proteins (both from eukaryotic and prokaryotic organisms) that are destined to be localized in noncytoplasmic locations initially are synthesized as a precursor with a 15–30 amino acid extension at the NH_2-terminal end of the molecule. This extra peptide has been termed the signal sequence, and it has been proposed that this signal plays a role in the localization of the extracytoplasmic protein. We are studying the process by which proteins are exported to the envelope region of Escherichia coli. Our work deals primarily with the outer membrane proteins, λ receptor, the product of the *lamB* gene, and the major outer membrane (porin) proteins 1a and 1b, products of the *ompF* and *ompC* genes.

Using techniques of gene fusion, we have demonstrated that information specifying the cellular location of the λ receptor is contained within the *lamB* gene. Furthermore, we have shown that this information is capable of directing even a normally cytoplasmic protein, β-galactosidase, to the outer membrane. Some of this information is contained within the signal sequence. Mutations that alter this sequence prevent export of the λ receptor protein. Again using techniques of gene fusion, we have shown that the signal sequence alone is not sufficient to cause export of β-galactosidase from the cytoplasm. Other information within the *lamB* gene is required.

Selection procedures have been developed to isolate mutations that exhibit a general alteration in the export process. Genetic analysis of these mutations has provided evidence for the involvement of the ribosome in the process of protein localization.

The structural genes for the porin proteins, 1a and 1b, are regulated at the transcriptional level by the *ompB* locus. This has permitted us to extend our studies on outer membrane protein localization to protein 1. With this genetic system, it should be possible to determine if E coli employs more than a single mechanism for the export of proteins to the outer membrane.

Key words: gene fusions, λ receptor, major outer membrane proteins, signal sequence mutations, ribosome

Michael N. Hall and Scott D. Emr are affiliated with the Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115.

Received April 14, 1980; accepted June 17, 1980.

0091-7419/80/1302-0147$03.20

Approximately 25% of the proteins synthesized by the gram-negative bacterium Escherichia coli are destined to be found in noncytoplasmic locations: the cytoplasmic membrane, the outer membrane, or the periplasmic space bounded by these two structures. The process of localizing noncytoplasmic proteins is selective and efficient, since proteins indigenous to one location are rarely, if ever, found in another. We wish to understand the genetic basis for this observed selectivity and the molecular mechanisms involved in the export of noncytoplasmic proteins to their respective cellular locations.

The set of genes determining the transport of maltose and maltodextrins in E coli provides a convenient system for the study of protein localization. These genes are clustered in one locus (*malB*) on the chromosome and comprise two operons transcribed in opposite directions [1] (Fig. 1). The products of two of the genes are well characterized: the *malE* gene encodes a periplasmic maltose-binding protein [2], and the *lamB* gene codes for an outer membrane protein. This latter protein serves as the cell surface receptor for bacteriophage λ [3], is essential for penetration of maltodextrins into the cell, and facilitates transport of maltose when the sugar is present in the medium at low concentrations [4]. Recent evidence indicates that the *malF* gene product is a protein located in the cytoplasmic membrane [5]. The location of the *malK* and *malG* gene products has not been determined [6].

The synthesis of these proteins is inducible by maltose, and their regulation is determined by the *malT* locus, which is located elsewhere on the chromosome. The *malT* gene codes for a positive control factor; inactivation of the *malT* gene by mutation results in a Mal$^-$ phenotype [7]. Since the regulation of the *malB* locus is well understood, and since its products are constituents of the three extracytoplasmic cellular compartments, the *malB* locus seemed ideally suited as a system for genetic analysis of protein export.

Randall et al [8] have shown that the λ receptor and the maltose-binding protein are preferentially synthesized on membrane-bound polysomes. In addition, evidence demonstrating that both proteins are synthesized initially in larger precursor form has been presented [8, 9]. Recent studies involving DNA sequencing of the early parts of the genes coding for these proteins and amino acid analysis of the protein precursors themselves have revealed that these precursors have a 25 and 26 amino acid signal sequence at the NH_2-terminal end of the molecule, respectively [10, 11].

To account for these results and the results obtained by genetic analysis of protein export, we have proposed a model depicting the various steps of the localization process (Fig. 2) [12]. This model is based, in part, on models depicting equivalent processes in eukaryotic cells [13–15]. Indeed, prokaryotic and eukaryotic organisms are strikingly similar with respect to protein localization.

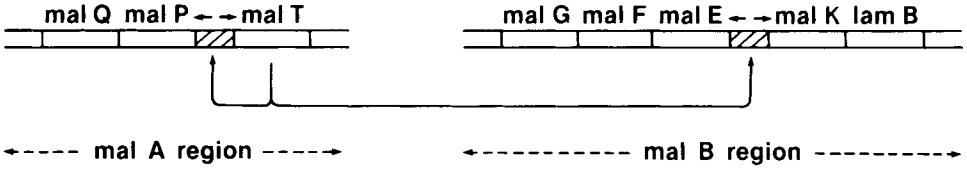

Fig. 1. The *malA* and *malB* regions of Escherichia coli. Genes *malP* and *malQ* are the structural genes encoding the enzymes maltodextrin phosphorylase and amylomaltase, respectively. All of the genes in the *malB* region code for proteins involved in maltose transport (see text). The product of the *malT* gene is a positive control factor that regulates transcription at the promoters for the three *mal* operons, as indicated. (Reproduced by permission of John Wiley & Sons, Inc. [12].)

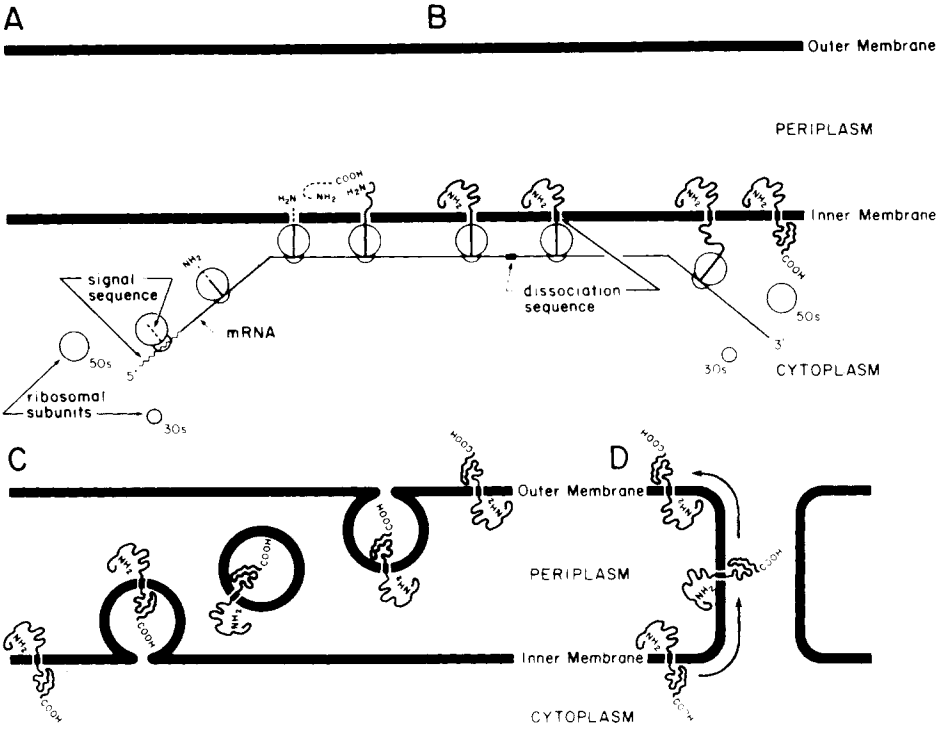

Fig. 2. Schematic illustration of the export of newly synthesized λ receptor to its normal outer membrane location. A. The signal sequence, represented by a jagged line at the 5′ end of the mRNA and by a dotted line in the nascent polypeptide chain, emerges from the ribosome and initiates attachment of the polysome to the cytoplasmic membrane. As translation proceeds, the nascent chain is transferred across the inner membrane in a vectorial fashion. Proteolytic processing of the signal sequence from the λ receptor precursor occurs outside of the cytoplasm. This section of the schematic is analogous to the signal hypothesis as proposed by Blobel and Dobberstein [13]. We propose that the *lamB* gene, which codes for the outer membrane protein λ receptor, has an additional information sequence relative to protein localization. We have termed the sequence "the dissociation sequence," represented here as a heavy dark line in both the mRNA and the λ receptor protein. B. Translation of this information sequence and subsequent emergence of the dissociation sequence region of the polypeptide chain from the ribosome triggers dissociation of the ribosome from the membrane. Subsequent translation of λ receptor mRNA completes the COOH-terminal end of the protein in the cytoplasm, leaving the protein tightly embedded in the inner membrane with its NH_2-terminus facing the periplasm and its COOH-terminus facing the cytoplasm. Two possible mechanisms by which the λ receptor, embedded in the inner membrane immediately after synthesis is complete, is subsequently translocated to its final outer membrane location are depicted. C. Vesicles blebbing off the inner membrane and subsequently fusing with the outer membrane. D. λ Receptor protein diffusing to the outer membrane through sites of inner membrane–outer membrane fusion. Both hypothetical mechanisms result in λ receptor incorporation into the outer membrane, with its NH_2-terminus facing the periplasm and its COOH-terminus exposed on the surface of the cell (Reproduced by permission of John Wiley & Sons, Inc. [12].)

INTRAGENIC INFORMATION SPECIFYING CELLULAR LOCATION

One of the initial goals in the study of protein export was to determine if information specifying cellular location was contained within the structural gene of a noncytoplasmic protein, as opposed to information in nontranslated regions of the mRNA, for example. Results summarized below demonstrate conclusively that such information is contained within the structural gene.

The Isolation and Characterization of *malB-lacZ* Fusions

Using techniques described previously [16, 17], we have constructed a variety of fusions between each of the *malB* genes and the structural gene for the cytoplasmic enzyme β-galactosidase (*lacZ*). All of these fusions are identical, in that each results in the production of a hybrid protein composed of an NH_2-terminal sequence from the *malB* gene product in question and a major functional portion of the COOH-terminal sequence of β-galactosidase. The amount of β-galactosidase in each of these hybrid proteins is essentially identical. This was shown by genetic mapping and, in some cases, DNA or protein sequencing [18]. Thus, different properties exhibited by different fusions cannot be attributed to differences in the amount of β-galactosidase sequences.

The *malE-lacZ* and *lamB-lacZ* fusions have been the most useful in elucidating mechanisms of protein export. As such, only these will be discussed here. However, it should be noted that *malG-lacZ*, *malF-lacZ*, and *malK-lacZ* fusions have been constructed also and have been useful for other purposes [5–7, 17, 19].

***lamB-lacZ* fusions.** Detailed descriptions of the isolation and genetic and biochemical characterization of some of these fusions have been described previously [12, 20, 21]. At present, four classes of *lamB-lacZ* fusions have been isolated. The relevant properties of these fusions are summarized in Table I. The large number of fusions belonging to class I have similar properties. The example shown in Table I, fusion 61-4, has been studied the most thoroughly [20, 21]. Recently, this hybrid protein has been purified, and the amino acid sequence of the NH_2-terminal portion of the molecule has been determined. Results show that only two amino acids coded for by *lamB* are present in the hybrid protein. Thus, it is not surprising that this protein remains in the cytoplasm [21].

The one class II fusion strain that exists is identical to the fusions of class I in all aspects, except that it contains slightly more *lamB* DNA by genetic mapping, and, accordingly, the hybrid protein produced is slightly larger. This protein has also been purified, and the amino acid sequence of the NH_2-terminal portion of the molecule has been determined. Results show that 39 amino acids corresponding to *lamB* are present in the hybrid protein. Thus, this protein contains the entire λ receptor signal sequence. Nevertheless, the hybrid protein remains in the cytoplasm [21]. More will be said about this below.

Fusion 42-1 is the most thoroughly studied of the many class III fusions. Genetic mapping suggests that fusion 42-1 contains a substantial portion of the *lamB* gene. A fraction (30–40%) of the hybrid protein produced by this fusion is located in the outer membrane [20].

Strains that carry the *lamB-lacZ* fusion 42-1 or any other class III fusion exhibit a very characteristic phenotype. These strains are sensitive to the presence of the inducer, maltose. The sensitivity to maltose is the result of the cells' inability to export properly the hybrid molecule. This Mal^S phenotype is described in more detail in a later section.

Fusions of class IV contain more *lamB* DNA than any other class of *lamB-lacZ* fusions and result in the production of the largest hybrid protein. In view of the results obtained with class III *lamB-lacZ* fusions, we expected that class IV fusions would be ex-

TABLE I. Relevant Properties of Various *lamB-lacZ* Fusion Strains and Hybrid Proteins

Fusion class	Example	Phenotype	Induced β-galactosidase activity (units)	Cellular hybrid protein localization (%)[a]				Approximate hybrid protein molecular weight	Approximate amount of *lamB* DNA present in hybrid gene[b]
				Inner membrane	Outer membrane	Periplasm	Cytoplasm		
I	61-4	Mal^+Lac^+	900–1,100	<10	<1	<1	90	115,000	<1/11
II	52-4	Mal^+Lac^+	2,000–3,000	<15	<1	<1	85	118,000	1/11
III	42-1	Mal^sLac^+	1,000–1,300	25	30	2	43	137,000	5/11
IV	42-12	Mal^+Lac^-	20–50	<10	<80	NT	<10[c]	141,000	7/11

[a]Techniques for cellular fractionation have been described [12, 22–24].

[b]The *lamB* gene has been divided into 11 segments by deletion mapping. The value listed represents the fraction of segments that are present in the hybrid gene and that were not removed during construction [20, 25]. It should be noted that this is a genetic result, not a physical result, and accordingly these data are only an estimate. If deletion endpoints are nonrandom, it is possible that these estimates are substantially incorrect.

[c]Exact fractionation data with these strains are difficult to give because of the extremely low specific activity of the hybrid protein. These values represent estimates from gels of the various cellular fractions.

tremely Mal^S; because, the hybrid protein produced is 4,000 daltons larger than the class III hybrid proteins. However, this is not the case. Although the class IV fusion strains do exhibit some maltose sensitivity, this sensitivity is much less than that observed with class III fusion strains. Determining the cellular location of the class IV hybrid proteins has proved to be difficult. Several methods exist that are routinely used to separate inner and outer membranes [22, 23]. We have found that these different methods give quantitatively different results. Although more careful fractionation studies must be performed, we feel confident in stating that most of the class IV hybrid protein is localized to the outer membrane.

The results obtained with the *lamB-lacZ* fusions indicate that information specifying cellular location is contained within the *lamB* gene. Furthermore, this information is capable of directing even a large, normally cytoplasmic enzyme to the outer membrane.

***malE-lacZ* fusions.** The isolation and characterization of these fusions have been described in detail [26]. Five classes of *malE-lacZ* protein fusions have been characterized.

Although the maltose-binding protein is a protein normally exported to the periplasm, essentially no β-galactosidase activity is released from any of the fusion strains by procedures known to release periplasmic constituents. However, the relative amounts of the *malE* protein in the hybrid proteins of the different fusion size classes do determine where the hybrid protein is localized. In the case of class I fusions, the hybrid gene contains only a small portion of the *malE* gene. NH_2-terminal sequencing of one class I hybrid protein revealed that only the first 15 amino acid residues are derived from the maltose-binding protein precursor. As expected, this incomplete signal sequence is not sufficient to cause export from the cytoplasm [11].

Strains containing larger *malE-lacZ* fusions exhibit a Mal^S phenotype analogous to that exhibited by class III *lamB-lacZ* fusions. A portion of the hybrid protein produced by these *malE-lacZ* fusions is located in the cytoplasmic membrane. The Mal^S phenotype is the result of the cells' attempt to export the hybrid proteins to the periplasm.

The periplasmic maltose-binding protein and the outer membrane protein λ receptor appear to share common steps in their export pathways. Both *malE-lacZ* and *lamB-lacZ* fusions that contain the appropriate signal sequence region exhibit a characteristic Mal^S phenotype. In addition, evidence described elsewhere suggests that the export of the *malE* gene product may interfere directly with the export of the *lamB* gene product [26]. If a Mal^S *malE-lacZ* fusion strain is induced by the addition of maltose to the growth media, λ receptor precursor accumulates in the strain [26]. Conversely, if a Mal^S *lamB-lacZ* fusion strain is induced by maltose, precursor *malE* gene product can be detected in the cell [27]. Accordingly, it appears that the initial step, the signal sequence-mediated step, in the export of both λ receptor and maltose-binding protein employs the same localization machinery.

The model proposed in Figure 2 is consistent with the results obtained with the various fusions. First, the model offers an explanation of how the localization of the products of the *malE* and *lamB* genes can be similar in the initial steps, and yet, by diverging at a later stage, eventually deliver the products of the two genes to different cellular compartments. Furthermore, the model explains the following two differences observed between *malE-lacZ* and *lamB-lacZ* fusions: no *malE-lacZ* hybrids reach the periplasm, even though *lamB-lacZ* hybrids are incorporated into the outer membrane with high efficiency, and the Mal^S phenotype is proportional to the amount of *malE* DNA present in *malE-lacZ* fusions but not in *lamB-lacZ* fusions.

For a protein to be synthesized by cytoplasmic machinery and yet eventually be located in the periplasm, the molecule must cross a membrane. Since no *malE-lacZ* fusions produce a hybrid protein located in the periplasm (even though in one fusion nearly all of the *malE* gene is present in the hybrid gene), we believe that it is not possible for the export machinery to vectorially discharge into the periplasm molecules with a β-galactosidase moiety. In fact, we believe that the Mal^S phenotype is the result of a lethal jamming of the export machinery with β-galactosidase sequences.

The model does not require outer membrane proteins to pass completely through a membrane and, therefore, accounts for how it is possible to localize β-galactosidase to the outer membrane. Provided that the β-galactosidase is fused to the *lamB* gene at a point past the dissociation sequence (class IV fusions, see Fig. 2), the β-galactosidase portion of the hybrid would not enter the cytoplasmic membrane. The model states that only the NH_2-terminal portion of the hybrid molecule (that is, λ receptor sequences) has to pass through the membrane.

Mutations That Alter the Signal Sequence Region of the *lamB* Gene

In the previous section we described an unusual phenotype exhibited by class III *lamB-lacZ* fusion strains. Such strains are sensitive to the inducer, maltose (Mal^S). The Mal^S phenotype is the result of the inability of the cell to export large amounts of this hybrid protein efficiently [28]. This unusual phenotype has been exploited to isolate mutants that are defective in the export of the λ receptor. Genetic and biochemical characterizations of these mutants have provided conclusive proof that the signal sequence is necessary for protein export.

The Mal^S phenotype provides a selection for the isolation of mutations that prevent export of the hybrid protein. Since the Mal^S phenotype is a consequence of the defective export of the hybrid protein, selecting a maltose-resistant (Mal^r) phenotype should yield mutants in which export of the hybrid protein is blocked. To date, 47 Mal^r mutants of the Mal^S class III *lamB-lacZ* fusion strain, pop 3186, have been isolated [27]. Of these 47 mutants, 27 have been analyzed in detail. These mutants are not affected drastically in the level of expression of the *lamB-lacZ* hybrid gene, and they exhibit their mutant phenotypes both when the mutation is present in the *lamB-lacZ* fusion (Mal^r) and when recombined into an otherwise wild-type *lamB* gene. When these mutations are present in a wild-type *lamB* gene, the strain is Dex^- (unable to grow on maltodextrins) and λ^r. Genetic analysis of these mutations has revealed that they fall into three classes. Class I contains a single mutant strain, SE1050, which has a deletion mutation that fuses *malK* to *lamB*, which is in turn fused to *lacZ*. This unusual double fusion codes for a *malK-lamB-lacZ* tribrid protein that maintains both *malK* and *lacZ* activity and is localized in the inner membrane of the cell. Class II contains a group of 12 deletion mutations which, unlike class I, are deletions internal to the *lamB* gene. These deletions all map extremely in the *lamB* gene and cause the *lamB-lacZ* hybrid protein to accumulate in the cell cytoplasm. Class III contains 14-point mutations, all mapping in a very small region extremely early in the *lamB* gene. These mutations lead to the accumulation of the *lamB-lacZ* hybrid protein in the cell cytoplasm, and, when present in a wild-type *lamB* gene, they also cause precursor λ receptor to accumulate in the cell cytoplasm.

The DNA sequence of 15 of these mutations has been determined [29] (Fig. 3). These mutations and similar mutations isolated in the *malE* system [11] represent the first genetic proof of the functional role of the signal sequence in protein export or secretion.

The amino acid alterations caused by these mutations denote components of the signal sequence that are critical for transmembrane protein transfer.

A comparison between known prokaryotic and eukaryotic signal sequences indicates that they have several similar characteristics [32]. The seven prokaryotic signal sequences available at present (Fig. 4) can be broken down into two segments: a short NH_2-terminal hydrophilic, basic segment followed by a segment of predominantly hydrophobic amino acids that extends up to the site of processing. There are from one to three basic amino acids (arg and/or lys) in the basic portion of the signal sequence. These positively charged residues may play a role in the attachment of the polysomes to the negatively charged inner surface of the cytoplasmic membrane [32]. The hydrophobic segment directly follows the last basic residue of the hydrophilic segment. Charged amino acids, basic or acidic, are completely absent from this portion of the signal sequence. It has been suggested that this hydrophobic region loops into the membrane lipid bilayer or becomes associated with specific membrane protein(s), thereby initiating the transmembrane transfer of the exported polypeptides [12–15, 32, 33] (Fig. 2).

It is interesting that, in both the λ receptor and MBP systems, multiple copies of the same mutational events were found in independent isolates. Based on the DNA sequence of the λ receptor, 17 of the 18 amino acid residues in the hydrophobic portion of the signal sequence can be changed to charged amino acids by single base changes in the DNA. Despite this fact, of 14 independently isolated point mutations, only 4 separate amino acids were changed. This suggests that changes in only a specific subset of the signal sequence amino acids affects export drastically. A more extensive mutant search is necessary to confirm such a contention.

The mutations described here for the λ receptor signal sequence and those described for the maltose-binding protein [11] argue strongly that the hydrophobic segment of the signal sequence plays a critical role in the export process. The 4 different single amino acid changes in the λ receptor signal sequence and 4 of the 5 amino acid changes in the maltose-binding protein signal sequence are changes from hydrophobic (leu, met, val, ala) or weakly hydrophilic (thr) amino acids to very hydrophilic charged amino acids (arg, lys, glu, asp). Consequently, the hydrophobicity of the signal sequence must be functionally essential.

Other Information Within the *lamB* Gene Specifying Cellular Location

The signal sequence mutants described above demonstrate that an intact signal sequence is necessary for protein export. We can now ask if the signal sequence alone is sufficient to cause export. Results discussed in this section provide strong evidence that this is not the case.

The most compelling evidence to support the contention that a signal sequence alone is not sufficient to cause export comes from the amino acid sequence analysis of the class II *lamB-lacZ* fusion, 52-4 [21]. Until residue 39, the NH_2-terminal sequence of the 52-4 protein is that of the *lamB* protein precursor. Residue 41 of the hybrid protein corresponds to residue 20 of β-galactosidase. Therefore, the 52-4 protein possesses not only a complete signal sequence of the *lamB* protein precursor but also the first 15 amino acids of the mature *lamB* protein. Despite this, the 52-4 protein is located in the cytoplasm since (1) it does not sediment with membranes upon centrifugation of extracts and is not released with periplasmic proteins when the cells are submitted to an osmotic shock; 2) its signal sequence is not cleaved to any significant extent; and 3) its synthesis in large amounts does not lead to cell death as is the case with *malE-lacZ* or larger *lamB-lacZ* hy-

brid proteins. This result, plus the fact that classes III and IV *lamB-lacZ* fusions are exported, demonstrates that, for β-galactosidase at least, a signal sequence alone is not sufficient to cause export. Other information within the *lamB* gene must be required.

In addition to the information within *lamB* that specifies export, it seems likely that additional information exists within the structural gene that determines specifically an outer membrane location as opposed to a periplasmic location. As is summarized in Figure 2, we believe that the export of the periplasmic maltose-binding protein and the outer membrane protein λ receptor share a common initial step, diverging at a later stage to result in localization in different cellular compartments. We suggest that this divergence is mediated by the presence of a "dissociation sequence" in the *lamB* gene and the corresponding absence of such a sequence in the *malE* gene. Reasons for proposing this sequence were described earlier. However, at present, we have no direct evidence to support either the existence or the function of such a sequence.

OTHER CELLULAR COMPONENTS INVOLVED IN THE EXPORT PROCESS

An important goal of a genetic analysis of protein export is to identify cellular components (eg, the ribosome or the membrane) that are involved in the export process. Mutations that alter the localization of a protein (or a group of proteins) but are not in the structural gene for the protein itself should prove to be the key to such genetic studies. For example, if ribosome–membrane interactions are important, a ribosomal alteration may prevent attachment of the polysome to the membrane and, hence, prevent protein export. Another possibility is that mutations may be isolated that alter a pore or channel through which proteins pass during the export process or that block the proteolytic precursor processing activity. In any event, other genetic loci are likely to be involved in the rather complex process of protein localization. Characterization of unlinked mutations could provide us with insight into the nature of these other components and, possibly, the mechanism of the export process.

It is also possible that unlinked mutations that affect the localization of either the λ receptor or the maltose-binding protein would affect the process of protein localization in general and, as such, would be lethal. So far, all the characterized Mal^r mutants of the various Mal^s fusions are the result of linked mutations. Accordingly, other genetic techniques may be required to uncover new genetic loci involved in the export process. One approach involves seeking mutants that are export defective. Another approach is to devise selections for mutants in which an internalized protein is exported. The mutations we have isolated in which the precursor of the *lamB* gene product is found in the cytoplasm provide such a selection.

When the *lamB*S69 mutation, a signal sequence point mutation, is present in an otherwise wild-type *lamB* gene, reversion of the mutation can be detected by selecting for return of the wild-type Dex^+ phenotype. The mutation responsible for reversion must restore some degree of export of λ receptor to the outer membrane. Many such revertants appear to be "true revertants," in that the reversion mutation was found to map at the site of the original mutation. Such true revertants produce apparently normal amounts of λ receptor protein and, as expected, such revertants exhibit wild-type sensitivity to phage λ.

A second class of revertants has also been obtained. Unlike true revertants, these revertants localize to the outer membrane barely detectable levels of a protein that corresponds to the λ receptor. Genetic mapping studies reveal that these reversion mutations

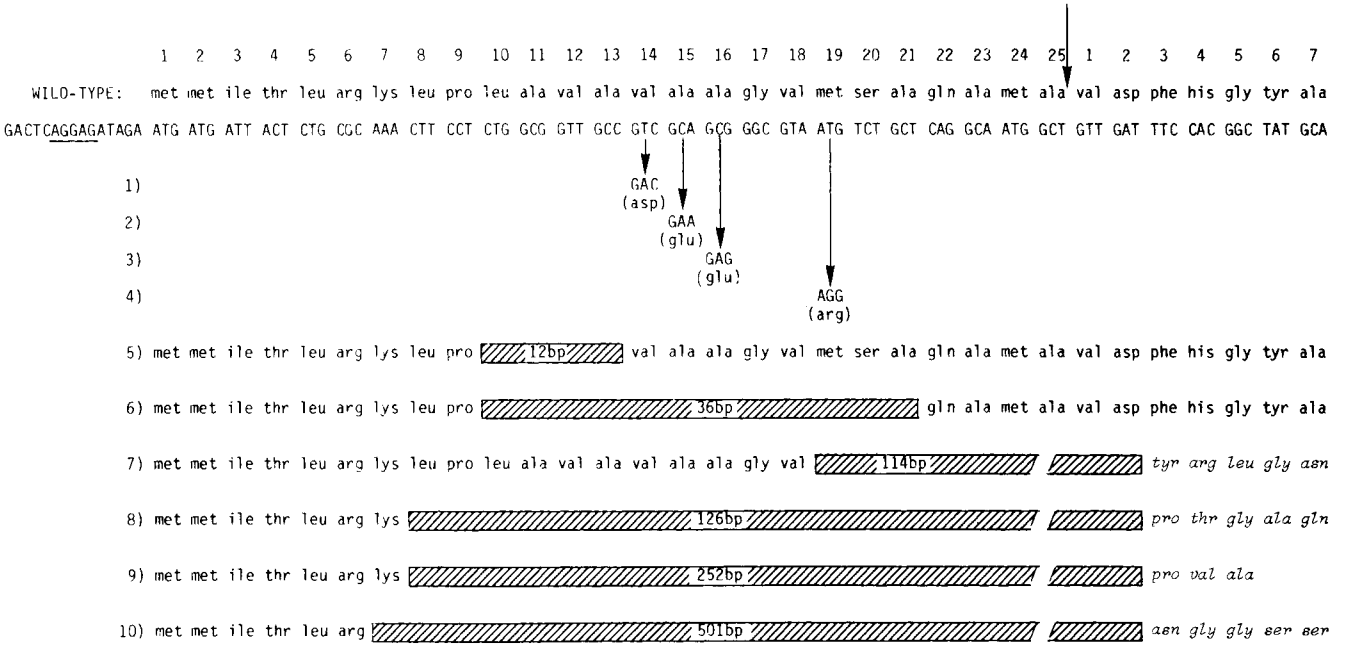

Fig. 3. Sequences shown were determined using the chain termination procedure [30]. Sequence reactions were run on thin 8% polyacrylamide gels [31]. Gels were autoradiographed (Kodak No-screen X-ray film) for 12–24 h at 20°C before reading off the above sequences. The signal sequence of wildtype λ receptor (top) and that of 15 *lamB* signal sequence mutations are shown. These mutations correspond to those present in the following E coli strains: 1) SE2071; 2) SE2070, SE2079, and SE2105; 3) SE2099; 4)SE2069, SE2088, and SE2100; 5) SE2078; 6) SE2060 and SE2084; 7) SE2087; 8) SE2068; 9) SE2089; and 10) SE2106. These strains are all described in detail elsewhere [26]. The amino acids to the right of deletions 7–10 above have not yet been confirmed by wild-type DNA sequencing. For this reason, they are shown in italics. In 8, 9, and 10 above, the deletion endpoints are within codons. The amino acid directly following each of these deletions is, therefore, the result of a fused codon that may encode an amino acid not normally found at this position in the wild-type sequence (reproduced by permission of *Nature* [29]).

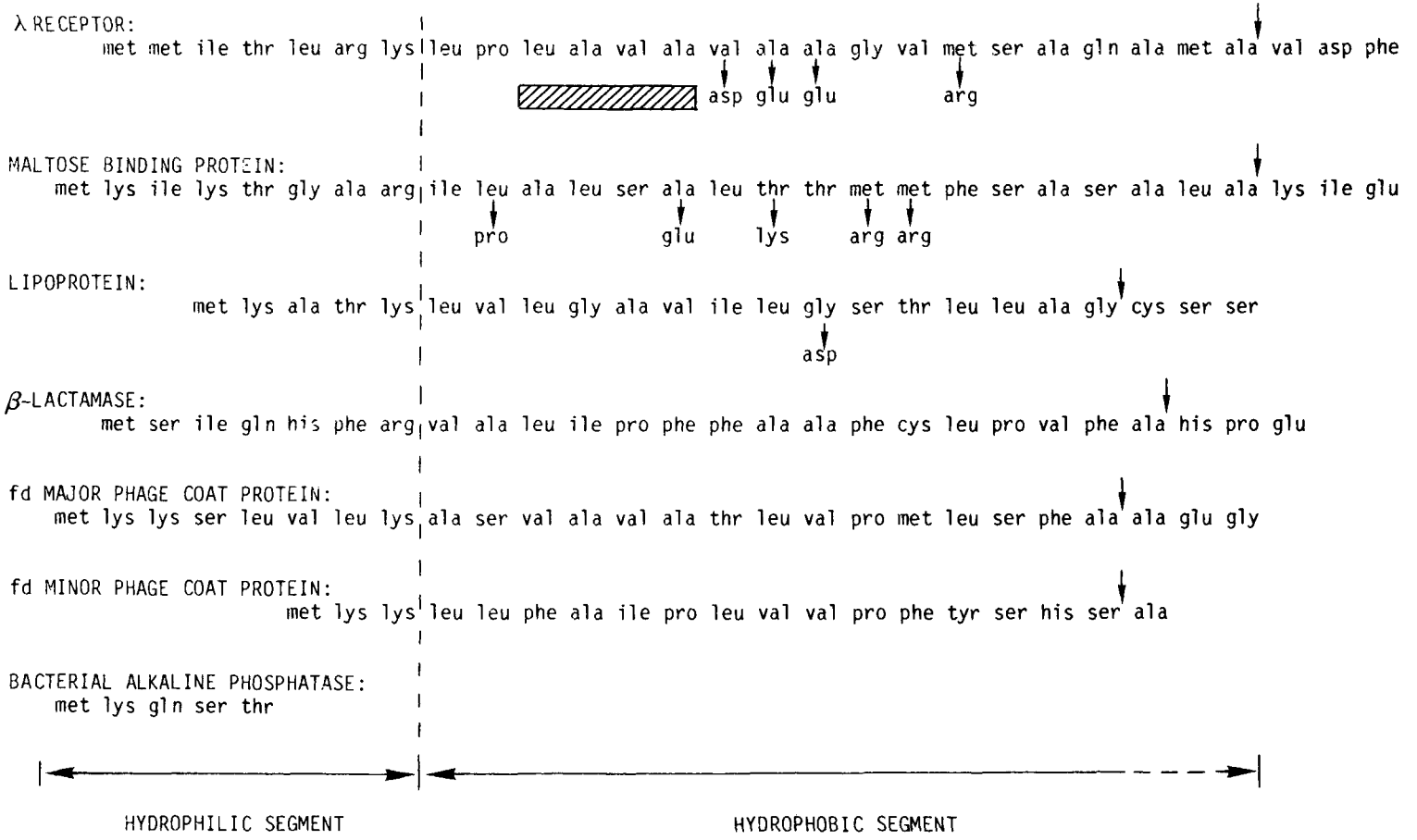

Fig. 4. The amino acid sequences of seven known prokaryotic signal sequences are shown. These include λ receptor [9, 10], maltose-binding protein [11], lipoprotein [34], β-lactamase [35], fd major phage coat protein [36], fd minor phage coat protein [37], and alkaline phosphatase [38]. Mutational alterations in the λ receptor signal sequence [29], the maltose-binding protein signal sequence [11], and the lipoprotein signal sequence [39] are indicated below their corresponding wild-type sequences. The site of processing for these signal sequences is indicated above each sequence by an arrow (reproduced by permission of *Nature* [29]).

are not linked to the *lamB* gene. At present, three different second-site mutations have been identified. Only one has been genetically characterized in detail. This mutation is 98% cotransducible by bacteriophage P1, with the mutation conferring resistance to the antibiotic spectinomycin. The latter mutation, *rpsE*, maps at approximately 72 minutes on the current E coli linkage map in a region of the chromosome commonly referred to as the "ribosomal cluster." In this tightly linked cluster there are approximately 30 genes coding for proteins that are a part of both the large and the small ribosomal subunits. In light of these mapping results, we suspect that these reversion mutations (presently termed *prl* for protein localization) lie within a gene coding for a ribosomal protein. The possibility that *prl* is a mutation causing ambiguity in the translation process does not seem likely, since the phenotype exhibited by *prl* mutants is quite different from that of known ribosomal ambiguity mutants. We favor the idea that *prl* is a mutation altering some aspect of the localization process. This contention is strengthened by the observation that *prl* also phenotypically suppresses several of the *malE* signal sequence mutations (P. Bassford, unpublished observations]. This result provides genetic evidence that export of the λ receptor protein occurs in a cotranslation manner. Membrane-bound ribosome and vectorial transfer are central features of the model shown in Figure 2.

DO MULTIPLE EXPORT MECHANISMS EXIST?

The export of proteins is an extremely selective and complex process. As discussed above, the bacterial cell has to have different mechanisms for the localization of proteins to different noncytoplasmic compartments. Furthermore, bacteria may even employ different mechanisms to localize different proteins to the same compartment. For example, it has been reported that the *lamB* protein is translocated at the septal region of dividing cells [40], whereas protein 1, another outer membrane protein, and corresponding S typhimurium proteins are incorporated at the adhesion sites located throughout the cell surface [41]. This suggests that E coli sorts outer membrane proteins for export by different mechanisms, thereby creating another level of complexity in protein localization. This complexity in export makes E coli similar to higher organisms, where many proteins destined for different cellular locations are initially translocated through the endoplasmic reticulum. This diversity can be analyzed with the gene-fusion technique, as described above, by isolating fusions between protein 1 and β-galactosidase.

E coli protein 1 is a major outer membrane protein. In K-12 strains, protein 1 can be separated electrophoretically into two components, 1a and 1b. Functionally, these proteins aggregate to form aqueous channels or pores in the outer membrane, allowing diffusion of certain metabolites. As such, these proteins are often referred to as porins. The porin proteins also act as receptors for various phages. Mutations in at least three widely spaced chromosomal loci are known to affect the production of proteins 1a and 1b. The loci *ompF*, mapping at 21 minutes, and *ompC* (48 minutes) are defined by mutations that result in the loss of proteins 1a and 1b, respectively. A mutation at another locus *ompB* (74 minutes), results in the loss of expression of 1a or 1b, or both [for review, see 32]. Conflicting models have been proposed to account for the existence of three genetic loci but only two known functions. Bassford et al [42] have reported that 1a and 1b have similar cyanogen bromide fragments and therefore suggest that *ompB* is the structural gene for a protein precursor that is modified to produce two species. Other investigators [43, 44] have presented biochemical evidence indicating that the two proteins are different. They

suggest that *ompF* and *ompC* are structural genes for the respective proteins. Genetic evidence also supports the contention that *ompF* is the structural gene for protein 1a [45] and *ompC* is the structural gene for protein 1b [46, 47].

REGULATION OF THE OUTER MEMBRANE PROTEINS 1a AND 1b

The expression of proteins 1a and 1b is regulated. The relative and absolute amounts of 1a and 1b fluctuate, depending on the growth medium [42, 48]. For example, in tryptic soy broth (TSB), protein 1b is preferentially synthesized, while in nutrient broth (NB) the inverse is the case. Protein 1b also is expressed preferentially in media of high osmotic strength [49, 50].

Before a study of the export of protein 1 can be undertaken, the location of the structural genes and the nature of the regulation of protein 1 have to be determined. In order to do this, we have again employed the technique of gene fusion. It is possible to construct two different types of fusions [16]. The *lamB-lacZ* and *malE-lacZ* fusions discussed above are "protein fusions" – ie, fusions that result in the production of a hybrid protein caused by the fusion of two structural genes. A second type of fusion, "operon fusion," places the *lac* operon under the regulatory control of a different promoter. Operon fusions do not result in the production of a hybrid protein. The genetic structure of operon fusions is such that the transcriptional regulatory properties of the promoter in question are reflected in the levels of β-galactosidase produced. We have constructed a series of *ompC-lac* [46] and *ompF-lac* operon fusions (manuscript in preparation). Analysis of these fusions has established a significant role for transcriptional regulation in the expression of the *ompC* and *ompF* gene products.

The levels of β-galactosidase expression in the *ompF-lac* and *ompC-lac* fusion strains [46] coincide with the observed regulatory phenomena of the major outer membrane proteins 1a and 1b (see Table II). First, depending on the growth medium, either protein 1a or 1b is expressed in large amounts (Fig. 5). By growing the fusion strains in two different media that elicit these two extremes in protein 1 production, we have observed a similar disparity in β-galactosidase activities. The activity of the *ompC-lac* fusion strains cultured in TSB is 3 to 4 times that of what it is when they are grown in NB. The reciprocal is the case for the *ompF-lac* fusion strains grown in the same two media. Furthermore, in media of high osmotic strength, the β-galactosidase activity of the *ompC-lac* fusion is significantly greater than that of the *ompF-lac* fusion strains. In media of low osmotic strength, expression of the *ompF-lac* fusion is favored. Second, production of protein 1 can be prevented by a mutation at *ompB*. When an *ompB* mutation that fails to express either one or both of the two major outer membrane proteins is introduced into the fusion strains, expression of the corresponding fusion is greatly reduced. When an *ompB* mutation that fails to express only one of the two proteins is introduced, the media-determined fluctuation of the expressed protein (or fusion) is altered.

These results indicate that an *ompB* gene product is a diffusable positive regulatory element that is required for transcriptional expression of both *ompC* and *ompF*. The transcriptional activity at both *ompC* and *ompF*, as monitored by β-galactosidase activity in the fusion strains, is virtually turned off by a mutation at *ompB*, located many minutes away on the chromosome. A role for *ompB* as a positive regulatory element also accounts for why a mutation at either one of two genetic loci, *ompB* and *ompC* or *ompF*, results in the absence of protein 1a or 1b. If an *ompB* gene product were exclusively a modifying

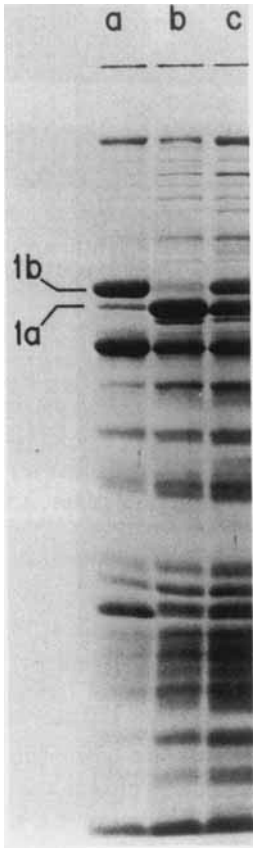

Fig. 5. Media-determined fluctuation in expression of major outer membrane proteins 1a and 1b in a wild-type strain. Outer membrane fractions were prepared after growth in the indicated media and electrophoresed on an urea-sodium dodecyl sulfate-polyacrylamide gel. The acrylamide concentration was 12%. a) Strain grown in tryptic soy broth (TSB). b) Strain grown in nutrient broth (NB). c) Strain grown in tryptone yeast (TY) extract broth. Major outer membrane proteins 1a and 1b are labeled.

or processing enzyme acting at a post-transcriptional level, β-galactosidase activity produced by the operon fusion strains would be unaffected. The results also indicate that the media-determined fluctuation of proteins 1a and 1b is controlled at the transcriptional level. The regulation of this fluctuation is also mediated by the *ompB* locus.

A regulatory role for *ompB* and the pronounced similarities between the regulation of β-galactosidase in the fusion strains and proteins 1a and 1b in a wild-type strain suggest that *ompF* and *ompC* are the structural genes for proteins 1a and 1b, respectively. Further evidence for this is provided by the different types of resistance to bacteriophage hy-2 (this phage uses protein 1b as a receptor) exhibited by *ompB* and *ompC* mutants. The *ompC* mutants are completely resistant to the phage, a predictable phenotype for a strain lacking a receptor. The *ompB* mutants are still slightly sensitive to the phage, indicating perhaps that the strain is not lacking a receptor but rather only has reduced amounts. This

TABLE II. β-Galactosidase Activities of *ompF-lac* and *ompC-lac* Operon Fusion Strains*

Fusion	*ompB* allele	β-Galactosidase activity TSB	NB	Ratio (NB/TSB)
ompF-lac	ompB$^+$	165	715	4.3
(1a)	1a$^+$1b$^-$	344	887	2.6
	1a$^-$1b$^+$	27	353	13.1
	1a$^-$1b$^-$	5	9	1.8
Fusion	*ompB* allele	β-Galactosidase activity TSB	NB	Ratio (TSB/NB)
ompC-lac	ompB$^+$	594	168	3.5
(1b)	1a$^+$1b$^-$	72	20	3.6
	1a$^-$1b$^+$	1,246	1,618	0.77
	1a$^-$1b$^-$	27	25	1.1

*β-Galactosidase activities of *ompF-lac* and *ompC-lac* operon fusion strains containing the indicated wild-type and mutant *ompB* alleles, grown in tryptic soy broth (TSB) and nutrient broth (NB). All values are the average of various independent fusion strains and two *ompB* alleles with each mutant phenotype. The fusion strains were grown to mid-logarithmic phase in the indicated media. β-Galactosidase activity was assayed as described [51].

is consistent with the model that *ompC* is the structural gene for the receptor and *ompB* is a regulatory element, in the absence of which only basal levels of phage receptor are expressed.

Confident that *ompF* and *ompC* are the structural genes for proteins 1a and 1b, and equipped with an understanding of how to regulate the expression of these two proteins, we have extended the study described above on the localization of λ receptor to protein 1. Protein fusions of protein 1a and β-galactosidase have been isolated and partially characterized. These OmpF-LacZ hybrid proteins fall into two size classes. The smaller, represented by only one fusion, is 116K daltons, the size of β-galactosidase. The second size class is approximately 3,000 daltons larger. The protein fusions exhibit the same regulatory properties described for the operon fusions.

In order to determine if the localization of the β-galactosidase activity in the cell was altered by gene fusion to *ompF*, we prepared membrane fractions from each of the fusion strains. We find that essentially all of the β-galactosidase activity (90%) produced by strains containing the smaller hybrid protein is located in the soluble fraction. Fractionation studies performed with strains containing the larger hybrid protein reveal that approximately 25% of the β-galactosidase activity is membrane-bound. We do not yet know if any of this hybrid protein reaches the outer membrane.

Most of the β-galactosidase activity produced by both classes of *ompF-lacZ* fusions remains in the cytoplasm. In this regard they appear to be similar to the two smallest classes of *lamB-lacZ* fusions. We have, however, detected what may be an important difference. Certain *ompB* alleles (1a$^+$1b$^-$) substantially increase expression of protein 1a. This increase in expression is reflected by an increase in the β-galactosidase activity produced

by strains containing the larger *ompF-lacZ* protein fusion or an *ompF-lac* operon fusion. When this *ompB* allele is introduced into strains that produce the smaller OmpF-LacZ hybrid protein, growth of the strain is impaired and viability is decreased. This phenotype is similar to the Mal^S phenotype exhibited by class III *lamB-lacZ* fusion strains. It is intriguing that only the smaller *ompF-lacZ* protein fusion confers this phenotype, and it may reflect a fundamental difference in the mechanisms of localization of the two outer membrane proteins, λ receptor and protein 1a. Further characterization of these fusions should reveal whether multiple export mechanisms are employed by E coli for outer membrane proteins.

PROJECTIONS

Although it is clear that enormous strides have been taken in recent years toward an understanding of the process of protein localization, it is also clear that much remains unknown. Despite sequencing data obtained from a plethora of exported proteins, and despite the existence of a number of mutations, we still do not understand how the signal sequence functions. The nature of the additional intragenic export information also is unknown, as is information other than that specifying export, which must be present to direct a protein to the correct extracytoplasmic location. Available evidence indicates that proteins destined for several different cellular locations may be exported in a similar manner. Such export pathways must diverge at some point.

What about cellular components? Are there membrane pores or channels through which proteins pass during the export process? At present, we simply do not know. Finally, are all proteins localized via a common mechanism? Even if we confine ourselves to outer membrane proteins in E coli, this does not seem likely. Many such proteins contain an NH_2-terminal signal sequence that is removed during export [32]; many other outer membrane proteins are not processed [52]. Certain newly synthesized outer membrane proteins appear in the membrane at the septal region of dividing cells [40]; others appear at sites located throughout the cell surface [41]. Answers to many of these questions will come from the isolation and characterization of mutants altered in each step of the complex process of protein localization.

ACKNOWLEDGMENTS

We would like to express our gratitude to Jon Beckwith. He, more than any other person, is responsible for applying the techniques of gene fusion to the study of protein localization. Without his expertise and foresight, none of the work described here would have been possible. We also would like to thank our many collaborators; in particular Maxime Schwartz, who taught us the ins and outs of *malB* genetics and biochemistry; Maurice Hofnung and Joe Hedgpeth for their help with DNA sequencing; and Audree Fowler for her work on protein sequencing. We also would like to thank Phil Bassford and Howard Shuman for sharing their results with us and for many hours of fruitful discussion.

This work was supported in part by the National Cancer Institute under contract NO1-CO-75380 with Litton Bionetics, Inc.

REFERENCES

1. Hofnung M: Genetics 76:169, 1974.
2. Kellermann O, Szmelcman S: Eur J Biochem 47:139, 1974.
3. Randall-Hazelbauer L, Schwartz M: J Bacteriol 116:1436, 1973.
4. Szmelcman S, Schwartz M, Silhavy TJ, Boos W: Eur J Biochem 65:13, 1976.

5. Shuman HA, Silhavy TJ, Beckwith JR: J Biol Chem 255:168, 1980.
6. Silhavy TJ, Brickman E, Bassford PJ, Casadaban MJ, Shuman HA, Schwartz V, Guarente L, Schwartz M, Beckwith JR: Mol Gen Genet 174:249, 1979.
7. Debarbouille M, Shuman HA, Silhavy TJ, Schwartz M: J Mol Biol 124:359, 1978.
8. Randall LL, Josefsson L-G, Hardy SJS: Eur J Biochem 92:411, 1978.
9. Marchal G, Perrin D, Hedgpeth J, Hofnung M: Proc Natl Acad Sci USA 77:1491, 1980.
10. Hedgpeth J, Clement JM, Marchal C, Perrin D, Hofnung M: Proc Natl Acad Sci 77:2621, 1980.
11. Bedouelle H, Bassford PJ, Fowler AV, Zabin I, Beckwith J, Hofnung M: Nature 285:78, 1980.
12. Silhavy TJ, Bassford PJ, Beckwith JR: In Inouye M (ed): "Bacterial Outer Membranes." New York: John Wiley & Sons, 1979, pp 203–254.
13. Blobel G, Dobberstein B: J Cell Biol 67:835, 1975.
14. Sabatini DD, Kreibich G: Martonosi A (eds): "The Enzymes of Biological Membranes." New York: Plenum Press, 1976, pp 531–579.
15. Lodish HF, Rothman NE: Sci Am 240:48, 1980.
16. Casadaban M: J Mol Biol 104:541, 1976.
17. Silhavy T, Casadaban MJ, Shuman HA, Beckwith JR: Proc Natl Acad Sci USA, 73:3423, 1976.
18. Brickman E, Silhavy TJ, Bassford PJ, Shuman HA, Beckwith JR: J Bacteriol 139:13, 1978.
19. Shuman HA, Beckwith JR: J Bacteriol 137:365, 1979.
20. Silhavy T, Shuman H, Beckwith J, Schwartz M: Proc Natl Acad Sci USA 74:814, 1977.
21. Moreno F, Fowler AV, Hall M, Silhavy TJ, Zabin I, Schwartz M: Nature 286:356, 1980.
22. Diedrich DL, Summers AO, Schnaitman CA: J Bacteriol 131:598, 1977.
23. Osborn MJ, Gander JE, Parisi E, Carson J: J Biol Chem 247:3962, 1972.
24. Neu HC, Heppel LA: J Biol Chem 240:3685, 1965.
25. Raibaud O, Roa M, Braun-Breton C, Schwartz M: Mol Gen Genet 174:241, 1979.
26. Bassford PJ, Silhavy TJ, Beckwith JR: J Bacteriol 139:19, 1979.
27. Emr SD, Silhavy TJ: J Mol Biol 141:63, 1980.
28. Emr SD, Schwartz M, Silhavy TJ: Proc Natl Acad Sci USA 75:5802, 1978.
29. Emr SD, Hedgpeth J, Clement JM, Silhavy TJ, Hofnung M: Nature 285:82, 1980.
30. Sanger F, Nicklen S, Coulson AR: Proc Natl Acad Sci USA 74:5463, 1977.
31. Sanger F, Coulson R: FEBS Lett 87:107, 1978.
32. DiRienzo JM, Nakamara K, Inouye M: Annu Rev Biochem 47:481, 1978.
33. Chan SN, Patzelt C, Dugericd NR, Quinn P, Labrecque A, Noyes B, Kein P, Heinrikson RL, Steiner DF: In Russell RT (ed): "From Gene to Protein: Information Transfer in Normal and Abnormal Cells." New York, Academic Press (in press).
34. Inouye S, Wang S, Sekizowa J, Halegona S, Inouye M: Proc Natl Acad Sci USA 74:1004, 1972.
35. Sutcliffe JG: Proc Natl Acad Sci USA 75:3737, 1978.
36. Fugimoto K, Sugisaki H, Okomoto T, Takanami MJ: J Mol Biol 111:487, 1977.
37. Schaller H, Beck E, Takanami M: In Denhardt D, Dressler D, Ray D (eds): "The Single-Stranded DNA Phages." Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1978, pp 139–153.
38. Sarthy A, Fowler AV, Zabin I, Beckwith J: J Bacteriol 139:932, 1979.
39. Lin JJC, Kanazawa H, Wu HC: Proc Natl Acad Sci USA 75:4891, 1978.
40. Ryter A, Shuman H, Schwartz M: J Bacteriol 122:295, 1975.
41. Smit J, Nikaido H: J Bacteriol 135:687, 1978.
42. Bassford PJ, Diedrich DL, Schnaitman CA, Reeves P: J Bacteriol 131:608, 1977.
43. Ichihara S, Mizushima S: J Biochem 83:1095, 1978.
44. Verhoef C, Lugtenberg B, van Boxtel R, deGraaff P, Verheij H: Mol Gen Genet 169:137, 1979.
45. Sato T, Yura T: J Bacteriol 139:468, 1979.
46. Hall MN, Silhavy TJ: J Bacteriol 140:342, 1979.
47. vanAlphen L, Lugtenberg B, van Boxtel R, Hack AM, Verhoef C, Havekes L: Mol Gen Genet 169:147, 1979.
48. Lugtenberg B, Peters R, Bernheimer H, Berendsen W: Mol Gen Genet 147:251, 1976.
49. vanAlphen W, Lugtenberg B: J Bacteriol 131:623, 1977.
50. Kawagi H, Mizuno T, Mizushima S: J Bacteriol 140:843, 1979.
51. Achtman M, Manning PA, Edelbluth C, Herrlich P: Proc Natl Acad Sci USA 76:4837, 1979.
52. Miller JH: "Experiments in Molecular Genetics." Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1972.