The Nucleotide Sequence of the Gene for malF Protein, an Inner Membrane Component of the Maltose Transport System of Escherichia coli

REPEATED DNA SEQUENCES ARE FOUND IN THE malE-malF INTERCISTRONIC REGION*

(Received for publication, April 2, 1984)

Susan Froshauer† and Jon Beckwith‡
From the Department of Microbiology and Molecular Genetics, Harvard University Medical School, Boston, Massachusetts 02115

The malF gene product is an inner membrane component of the maltose transport system in Escherichia coli. Some gene fusions between malF and lacZ (encoding the normally cytoplasmic enzyme β-galactosidase) produce hybrid proteins which are membrane-bound while other fusions produce hybrid proteins which are cytoplasmic (Silhavy, T. J., Casadaban, M. J., Shuman, H. A., and Beckwith, J. R. (1976) Proc. Natl. Acad. Sci. U. S. A. 73, 3423–3427). To further analyze the localization properties of the different classes of fusion proteins and of the intact MalF protein, we have obtained the DNA sequence of 5 malF-lacZ fusions and the wild type malF gene. From the predicted amino acid sequence, MalF protein contains 514 amino acids and has a molecular weight of 56,947. Analysis of the hydrophobic character of MalF using the Kyte-Doolittle assignments (Kyte, J., and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105–132), indicates that the protein may have 2 or 3 amino-terminal membrane-spanning segments and 4 or 5 carboxy-terminal membrane-spanning segments separated by a region of 181 hydrophilic residues. Localization properties of the different fusion proteins correspond with degree of hydrophobicity.

By sequencing upstream from malF, the malE-malF intercistronic region was found to be 153 base pairs in length and to contain inverted repeats, homologous to intercistronic repeats of many other operons. Localization properties of the different fusion proteins correspond with degree of hydrophobicity.

By sequencing upstream from malF, the malE-malF intercistronic region was found to be 153 base pairs in length and to contain inverted repeats, homologous to intercistronic repeats of many other operons. Further analysis of this region may help in understanding the observed step-down in synthesis of the MalF protein.

* The research at the Frederick Cancer Research Facility was sponsored in part by the National Cancer Institute, Department of Health and Human Services, under Contract NO-CO125909 with Liton Biometrics Inc. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
† Sponsored by National Research Service Award in Genetics ST33GM070194.
‡ Sponsored by National Science Foundation Grant PCM8216464.
To whom reprint requests should be sent.
protein likely result from a reduced rate of synthesis (Shuman et al., 1980).

We have isolated gene fusions between the malF and lacZ (β-galactosidase) genes. When substantial portions of MalF protein are attached to P-galactosidase, the hybrid protein becomes membrane-bound (Silhavy et al., 1979a). Other fusions in which a much smaller amino-terminal segment of MalF is attached to β-galactosidase produce hybrid proteins which are cytoplasmic (Silhavy et al., 1979a).

In order to determine what components of the MalF protein can promote membrane localization, to study the step-down in gene expression from malE to malF, and to facilitate genetic studies on these localization and regulation problems, we have determined the DNA sequence of the malF gene, of 5 malF-lacZ fusions, and of the intercistronic region between malE and malF. We report that sequence in this paper.

### EXPERIMENTAL PROCEDURES

**Bacterial Strains, Phages, and Plasmids**—Bacterial strains, phages, and plasmids are listed in Table I. The β-galactosidase gene, of 5 malF-lacZ fusions, and of the intercistronic region between malE and malF. We report that sequence in this paper.

**Media and Growth Conditions**—Media were made according to Miller (1972). For liquid cultures either LB broth or minimal medium 63 (M63) supplemented with 5% LB broth was used. TYP agar was used for growth on rich plates and M63 agar, for minimal plates. Ampicillin (Amp) was used at 200 µg/ml, rifampicin (Rif) at 20 µg/ml, and 5-bromo-4-chloro-3-indolyl-β-D-galactoside (XG, Bachem, Inc.) at 40 µg/ml.

**Recombination of malFV49 onto pSF111**—The malFV49 ochre mutation was obtained on pSF111 by an in vivo recombination technique which can be used generally to transfer mutations onto plasmids. For this construction, pSF111 was transformed into EM9, a strain with an episome carrying the V49 mutation. A co-integrate of episome and plasmid forms as the result of recombination through regions of homology, in this case the malB locus. Subsequently, the episome-plasmid co-integrate was transferred to HS2012 by mating with the two strains for 1 h at 37°C (Miller, 1972) and then by growing in liquid M63 glucose containing rifampicin, to select against EM9, the donor, and ampicillin, to select for transfer of episome with integrated plasmids. Cells were grown for a few generations in this medium to allow episomes and plasmids to resolve and then dilutions were plated on M63 glucose plates containing XG, rifampicin, and ampicillin. Approximately 50% of the plasmids acquired the malFV49 mutation as a consequence of recombination with the V49-containing episome. These plasmids were easily detected by taking advantage of the polar effects of chain-terminating mutations such as V49. The original plasmid (pSF111) forms blue colonies on plates containing XG because of the production of the malF-lacZ hybrid protein, whereas plasmids with V49 (such as pSF113) fail to make the hybrid protein, and colonies are white on plates with XG. The presence of the V49 mutation on pSF113 was confirmed by testing for the ability of $\delta$supC% (an ochre suppressor-containing phage) to make colonies with pSF113 blue on plates with XG. We further established that our construction was correct by showing that $\delta$supC% (an amber suppressor-containing phage) failed to suppress the V49 mutation on pSF113. Finally, the structure of pSF113 was verified by restriction analysis.

**DNA Isolation**—Phage λ was grown and DNA was prepared by the methods of Maniatis et al. (1982). Large scale plasmid DNA preparations were obtained by modification of the method of Clewell (1972). Clear lysates were made with 4% Triton X-100 instead of Brij. Plasmid-containing cells were grown in either M63 glucose supplemented with 0.5% casamino acids or in LB broth with 0.4% glucose. Plasmid DNA to be used for sequencing was purified two times successively in cesium chloride gradients saturated with ethidium bromide. Rapid plasmid DNA preparations were obtained by the alkaline lysis method of Birnboim and Doly (1979).

**Restriction Enzyme Digestion of DNA**—Restriction enzymes were obtained from New England Biolabs, Inc., and digestion conditions were according to their recommendations.

**Cloning of Fusions**—Approximately 1 µg of vector and phase DNA were digested with EcoRI, mixed together, phenol-extracted, ethanol-precipitated, and resuspended in T4 DNA ligase buffer (New England Biolabs, Inc.). Ligation was carried out at 15°C overnight with 100 units of T4 DNA ligase (New England Biolabs, Inc.). Using the method of Dagert and Ehrlich (1979), competent MC1000 cells were prepared, transformed to ampicillin resistance with 1/4 of the ligation mixture, and spread on TYP Amp XG plates. Blue, Amp' transformants were purified and their plasmids were analyzed for the presence of the fusion-containing fragment.

**DNA Sequence Analysis**—DNA fragments were 3' end-labeled by filling in with the Klenow fragment of E. coli DNA polymerase I (New England Biolabs, Inc.) as described by Maniatis et al. (1982).
**Sequence of malF, an E. coli Maltose Transport Gene**

**TABLE I**

<table>
<thead>
<tr>
<th>Strain, phage, or plasmid</th>
<th>Genotype or phenotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC4106</td>
<td>F' araD139 ΔlacU169 thiA rpsL recA</td>
<td>Casadaban (1976)</td>
</tr>
<tr>
<td>MC1000</td>
<td>F' ΔlacX74 galE galK Δ(leu-ara)7897</td>
<td>Casadaban and Cohen (1980)</td>
</tr>
<tr>
<td>107011.3</td>
<td>F' trpJ hisG9 (Am) metE (o bo bg) rpsL</td>
<td>Laboratory collection</td>
</tr>
<tr>
<td>EM9</td>
<td>F' malE + malFV49 (Oc)/107011.3</td>
<td>H. Shuman</td>
</tr>
<tr>
<td>HS2012</td>
<td>MC4100 malFV49(Oc)/ropB</td>
<td>Silhavy et al. (1976)</td>
</tr>
<tr>
<td>MC4416</td>
<td>MC4100 (malF-lacZ)hyb11-1</td>
<td>Silhavy et al. (1979b)</td>
</tr>
<tr>
<td>MC4418</td>
<td>MC4100 (malF-lacZ)hyb14-4</td>
<td>Silhavy et al. (1979a)</td>
</tr>
<tr>
<td>MC4419</td>
<td>MC4100 (malF-lacZ)hyb32-1</td>
<td>Silhavy et al. (1979a)</td>
</tr>
<tr>
<td>MC4420</td>
<td>MC4100 (malF-lacZ)hyb41-2</td>
<td>Laboratory collection</td>
</tr>
<tr>
<td>MC4421</td>
<td>MC4100 (malF-lacZ)hyb53-1</td>
<td>Laboratory collection</td>
</tr>
<tr>
<td>Phage</td>
<td>MC4422 (malF-lacZ)hyb69-1</td>
<td>Laboratory collection</td>
</tr>
<tr>
<td>PFG1</td>
<td>Amp' LacZ+</td>
<td></td>
</tr>
<tr>
<td>PFG503</td>
<td>Amp' LacZ+ malE+ (malF-lacZ)hyb6-3</td>
<td>This study</td>
</tr>
<tr>
<td>PFG514</td>
<td>Amp' LacZ+ malE+ (malF-lacZ)hyb14-4</td>
<td>This study</td>
</tr>
<tr>
<td>PFG512</td>
<td>Amp' LacZ+ malE+ (malF-lacZ)hyb41-2</td>
<td>This study</td>
</tr>
<tr>
<td>PFG511</td>
<td>Amp' LacZ+ malE+ (malF-lacZ)hyb11-1</td>
<td>This study</td>
</tr>
<tr>
<td>PFG531</td>
<td>Amp' LacZ+ malE+ (malF-lacZ)hyb53-1</td>
<td>This study</td>
</tr>
<tr>
<td>PFG591</td>
<td>Amp' LacZ+ malE+ (malG-lacZ)hyb6-3</td>
<td>This study</td>
</tr>
<tr>
<td>PFG513</td>
<td>Amp' LacZ+ malE+ (malFV49(Oc)/lacZ)hyb11-1</td>
<td>This study</td>
</tr>
</tbody>
</table>

**RESULTS**

Cloning of the malF Gene—To determine the DNA sequence of malF, several malF-lacZ gene fusions were cloned from λ transducing phages into the plasmid vector pLG1 as shown in Fig. 1, b and c. The isolation and genetic analysis of all of these fusions and fusion-carrying phages, except for malF-lacZhyb32-1, have been described previously (Silhavy et al., 1979b). Basically, the fusions were isolated in vivo using Casadaban’s original approach (1976). The 41-2 fusion was isolated and characterized using these same techniques.

For cloning, phage DNA was cleaved with EcoRI to insert an EcoRI fragment containing the malEF promoter, malE, and all of the malF-lacZ fusion, except for the last 53 bp of lacZ, into pLG1 in an orientation that reconstructs the lacZ gene (Fig. 1, b and c). Five different malF-lacZ fusions (hybrids 6-3, 14-4, 41-2, 11-1, and 53-1) were cloned giving plasmids pSF630, pSF144, pSF142, pSF111, and pSF531 with 4.8-ki, 4.92-ki, 4.9-ki, 6.1-kb, and 6.4-kb EcoRI inserts, respectively. A malG-lacZ fusion (69-1) was also cloned in this manner to obtain the 5' end of malF on a plasmid. This plasmid (pSF501) contains a 6.7-ki EcoRI insert into plG1 which includes the malEF promoter, malE, malF, and approximately 300 bp of the malF-G intercistronic region and malG fused to lacZ.

**DNA Sequence Analysis of malF**—Using this set of plasmids, the nucleotide sequence of malF and some flanking sequences were obtained. Parallel polyacrylamide gel analysis of restriction digests of these fusion plasmids and the vector permitted identification of restriction fragments internal to malF as well as malF-lacZ fusion joint fragments. For example, the fusion joints were identified as the unique fragments found after comparison of the restriction patterns of the fusion-containing plasmids (data not shown). A series of digests using this approach facilitated isolation of relevant fragments for sequencing.

The sequencing strategy and a detailed restriction map generated from the nucleotide sequence of the 1744-bp DNA fragment from the 3' end of malE to the 3' end of malF. The solid box on the genetic map denotes the malE-malF intercistronic sequence. The arrows above and below the detailed restriction map indicate the sequencing experiments conducted on both strands and show the direction and extent of the nucleotide sequence determined from each fragment. Restriction sites are: A, AasII; B, BsrRI; D, DdeI; H, HincII; C, HinP1; M, MspI; R, Real; TaqI; B, BstEII.

For 5' end labeling, fragments were treated with calf intestine alkaline phosphatase (Boehringer Mannheim Biochemicals), phenol-extracted extensively, and labeled using T4 polynucleotide kinase (Boehringer Mannheim Biochemicals or New England Biolabs, Inc.) according to the conditions of Maxam and Gilbert (1980). Labeled DNA fragments were purified, and chemical sequencing reactions were performed and analyzed on sequencing gels using the methods of Maxam and Gilbert (1980). All 32P-labeled nucleotide triphosphates were purchased from Amersham.

![Fig. 2. Restriction map and sequencing strategy for the 1744 bp DNA fragment from the 3' end of malE to the 3' end of malF. The solid box on the genetic map denotes the malE-malF intercistronic sequence. The arrows above and below the detailed restriction map indicate the sequencing experiments conducted on both strands and show the direction and extent of the nucleotide sequence determined from each fragment. Restriction sites are: A, AasII; B, BsrRI; D, DdeI; H, HincII; C, HinP1; M, MspI; R, Real; TaqI; B, BstEII.](image-url)
Fig. 3. Nucleotide sequence of the noncoding strand of malF and flanking sequences. The numbering is relative to the proposed initiation codon for malF which is designated +1. Less frequent restriction sites are shown. The lines above the sequence starting at positions −11 and +154 indicate possible Shine-Dalgarno sequences for malF and malG, respectively. The V48(Gcc) mutation is indicated at position +13. Locations of the 5 malF-lacZ fusion joints are marked by arrows at positions +43, +139, +154, +1286, and +1501. Amino acid sequences of the carboxy terminus of the protein are printed below the DNA sequence.

frame for the beginning of the MalF protein, we sequenced the malF-lacZjoint on pSF630 which, based on genetic data, was predicted to contain only a small amount of malF DNA. We sequenced from the PvuII site at base pair 102 of lacZ (Kalinins et al., 1983) (amino acid 34 of β-galactosidase) through the fusion joint (amino acid 22 of β-galactosidase) and into malF sequences (see position +43, Fig. 3). The reading frame established from the known lacZ sequence reveals two possible initiation codons (both AUGs) at positions +1 and −15 (Fig. 3). These potential start sites are found in the region downstream from the end of malF (position −154, see below). A possible Shine and Dalgarno ribosome recognition sequence is found 7 nucleotides upstream from only one of the AUGs, at positions −7 to −11 (GGAAG) (Shine and Dalgarno, 1974). We suggest, therefore, that the AUG at position +1 encodes the beginning of the MalF protein. Based on this, the 6-3 fusion protein contains only the first 14 amino acids of MalF, a result consistent with the
For this program, specific hydropathy values have been as-
signed to each amino acid, based on a composite analysis of
the location of amino acid side chains in a number of struc-
turally well-characterized proteins as well as consideration of
the water vapor transfer free energies. Using a moving seg-
ment approach, the program determines the average hy-
dropathy of a protein over a specified range of amino acids,
termed the window.

As the result of an analysis of a large number of soluble
proteins, the value \(-0.4\) has been designated the midpoint on
the Kyte and Doolittle hydropathy scale. The average hy-
dropathy of MalF protein was determined using this scale and
determined to be \(0.29\). Bacteriorhodopsin (Khorana
et al., 1979), a protein which crosses the membrane seven times, has
an average hydropathy of \(0.70\). On the other hand, we deter-
mined that two proteins thought to span the membrane only
once or twice, the chemotaxis proteins, Tar (Russo and Kosh-
land, 1983) and Tsr (Boyd et al., 1983), have mean hydropa-
thies of \(0.09\) and \(0.07\), respectively.

The results of the SOAP program analysis of MalF protein
using a window of 11 are plotted in Fig. 4. Two or three
hydrophobic regions which extend from around amino acid
17 through 93 are evident at the amino terminus and are
indicated by Segments 1–3 in Fig. 4. Segment 1 has an average
hydropathy of \(1.56\); Segment 2, \(1.27\); and Segment 3, \(1.5\).
The results also indicate that the internal portion of MalF
protein, shown by the dashed line in Fig. 4, is not very
hydrophobic. This segment which extends 181 amino acids
from residue 94 to 274 has an average hydropathy of only
\(0.7\). From amino acid 274 to the carboxy terminus of the
protein, there are 4 or 5 hydrophobic segments with average
hydropathy values of \(2.0\) (Segment 4), \(1.76\) (Segment 5),
\(0.96\) (Segment 6), \(1.0\) (Segment 7), and \(1.7\) (Segment 8).
The **malE-malF Intercistronic Region**—We determined the DNA sequence of the intercistronic region between **malE** and **malF** by sequencing 5' to **malF**. The structure of this region was of interest in terms of the regulation of MalF protein synthesis because of the 10- to 100-fold step down in synthesis of MalF with respect to maltose binding protein that has been observed (Shuman et al., 1980).

The results show that the intercistronic space is 153 bp long as indicated in Fig. 3 at positions −153 to −1. The location of the 3' end of **malE** at −154 was established by comparing the nucleotide sequence to preliminary carboxy-terminal amino acid sequencing data provided by Audree Fowler. We found no potential ribosome-binding sites or open reading frames capable of encoding a small peptide in this region. Further analysis of the intercistronic sequence reveals an extensive region of dyad symmetry that when transcribed could form a long, stable stem and loop structure in which the gene at position +1 is based on the finding of the end of the **malE** gene and on the existence of an in-frame AUG codon preceded by a Shine-Dalgarno sequence (GGAG) at the appropriate position. This initiation codon is followed by a 514-amino acid open reading frame which is confirmed in part by sequence analysis of a **malF ochre** (chain-terminating) mutation and of several gene fusions between **malF** and lacZ. As a result of this analysis, the reading frame is established from residue 1 of the presumed amino acid sequence to 51 and then from residue 428 to the end at residue 514. The first chain-terminating codon in this open reading frame is found at positions +1543 and +1552. These presumably define the end of the gene. It is conceivable that the AUG starting at position +1560 is the initiation codon for **malG** since a potential Shine-Dalgarno sequence can be found starting 8 bp upstream from this site.

The choice of the start site allows the prediction of a molecular weight for the hybrid protein produced by gene fusion 6-3 of 115,598 which is similar to that estimated from sodium dodecyl sulfate-polyacrylamide gel analysis. However, the definitive determination of the start point of this gene awaits amino acid sequence analysis of the **MalF** protein.

**DISCUSSION**

The DNA sequence in Fig. 3 from positions +1 to +1540 is that of the **malF** gene. The suggested translation start site for the gene at position +1 is based on the finding of the end of the **malE** gene and on the existence of an in-frame AUG codon preceded by a Shine-Dalgarno sequence (GGAG) at the appropriate position. This initiation codon is followed by a 514-amino acid open reading frame which is confirmed in part by sequence analysis of a **malF ochre** (chain-terminating) mutation and of several gene fusions between **malF** and lacZ.

**Mu DNA in the **MalF-lacZ** Fusion Joints**—Further analysis of the nucleotide sequence of four of the **malF-lacZ** fusion joints reveals the presence of DNA from the right end (the S-end) of bacteriophage Mu (Kalman and Kamp, 1979). The fusion joint from 41-2 contains no Mu DNA. The sequences of these fusion joints are illustrated in Fig. 5. Mu DNA is indicated in *boldface* type. The 14-4 fusion joint occurs within the coding for amino acid 19 of **β-galactosidase**, the remaining joints occur at or within the coding for amino acid 22 of **β-galactosidase**. The 6-3 fusion has 29 bp from Mu, the 11-1 fusion, 10 bp, and the 53-1 fusion, 13 bp. The 14-4 fusion has at least 30 bp of Mu, 1 bp of which could not be determined because of a compression which commonly occurs in the sequence at these joints. In the 14-4 fusion, at least 5 bp of Mu are inverted and in a different orientation from the remaining Mu DNA (see arrows above the sequence in Fig. 6b). In several cases (see Fig. 6, a, c, and d), Mu and **malF** DNA are homologous at the fusion joints and whether the joint was formed from Mu or **malF** is not clear. Mu DNA has been detected previously at the junctions of gene fusions generated in vitro and is thought to result from the deletion event that produces the fusion. Inverted Mu sequences also have been observed.

An unexpected finding is the presence of the tetranucleotide CGCT in the wild-type **malF** sequence at the position of the joint in four of the fusions. The joints occur either between the C and T, as is the case for fusions 6-3, 14-4, and 53-1, or after the T, the case for fusion 11-1. A similar result has been observed in a **thra-lacZ** fusion. These findings may help elucidate the mechanism of Mu insertion, perhaps by providing some evidence for insertional specificity and a requirement for homology between Mu ends and the target site. Analysis of a large number of Mu insertions in **malE** and lamB supports this possibility (Raibaud et al., 1979; Silhavy et al., 1979b).

**Mu DNA in the **MalF-lacZ** Fusion Joints**—Further analysis of the nucleotide sequence of four of the **malF-lacZ** fusion joints reveals the presence of DNA from the right end (the S-end) of bacteriophage Mu (Kalman and Kamp, 1979). The fusion joint from 41-2 contains no Mu DNA. The sequences of these fusion joints are illustrated in Fig. 5. Mu DNA is indicated in *boldface* type. The 14-4 fusion joint occurs within the coding for amino acid 19 of **β-galactosidase**, the remaining joints occur at or within the coding for amino acid 22 of **β-galactosidase**. The 6-3 fusion has 29 bp from Mu, the 11-1 fusion, 10 bp, and the 53-1 fusion, 13 bp. The 14-4 fusion has at least 30 bp of Mu, 1 bp of which could not be determined because of a compression which commonly occurs in the sequence at these joints. In the 14-4 fusion, at least 5 bp of Mu are inverted and in a different orientation from the remaining Mu DNA (see arrows above the sequence in Fig. 6b). In several cases (see Fig. 6, a, c, and d), Mu and **malF** DNA are homologous at the fusion joints and whether the joint was formed from Mu or **malF** is not clear. Mu DNA has been detected previously at the junctions of gene fusions generated in vitro and is thought to result from the deletion event that produces the fusion. Inverted Mu sequences also have been observed.

An unexpected finding is the presence of the tetranucleotide CGCT in the wild-type **malF** sequence at the position of the joint in four of the fusions. The joints occur either between the C and T, as is the case for fusions 6-3, 14-4, and 53-1, or after the T, the case for fusion 11-1. A similar result has been observed in a **thra-lacZ** fusion. These findings may help elucidate the mechanism of Mu insertion, perhaps by providing some evidence for insertional specificity and a requirement for homology between Mu ends and the target site. Analysis of a large number of Mu insertions in **malE** and lamB supports this possibility (Raibaud et al., 1979; Silhavy et al., 1979b).

**DISCUSSION**

The DNA sequence in Fig. 3 from positions +1 to +1540 is that of the **malF** gene. The suggested translation start site for the gene at position +1 is based on the finding of the end of the **malE** gene and on the existence of an in-frame AUG codon preceded by a Shine-Dalgarno sequence (GGAG) at the appropriate position. This initiation codon is followed by a 514-amino acid open reading frame which is confirmed in part by sequence analysis of a **malF ochre** (chain-terminating) mutation and of several gene fusions between **malF** and lacZ. As a result of this analysis, the reading frame is established from residue 1 of the presumed amino acid sequence to 51 and then from residue 428 to the end at residue 514. The first chain-terminating codon in this open reading frame is found at positions +1543 and +1552. These presumably define the end of the gene. It is conceivable that the AUG starting at position +1560 is the initiation codon for **malG** since a potential Shine-Dalgarno sequence can be found starting 8 bp upstream from this site.

The choice of the start site allows the prediction of a molecular weight for the hybrid protein produced by gene fusion 6-3 of 115,598 which is similar to that estimated from sodium dodecyl sulfate-polyacrylamide gel analysis. However, the definitive determination of the start point of this gene awaits amino acid sequence analysis of the **MalF** protein.

The molecular weight of the **malF** protein is predicted to be 56,947 from the sequence analysis. The molecular weight estimated from sodium dodecyl sulfate-polyacrylamide gels was 40,000 (Shuman et al., 1980). This difference is not unexpected in the case of highly hydrophobic membrane proteins, as a similar large discrepancy was noted between the estimated and true molecular weights of the lacY gene product, the lactose permease (Büchel et al., 1980). On the other hand, the **malF** gene product seen in cell extracts may be the result of a processing event.

We have done very short labelings (30 s) of the **MalF** protein and failed to see any indication of a precursor form. Further, conditions in which **E. coli** can be made to accumulate precursors of proteins with signal sequences have failed to reveal a similar precursor for **MalF** protein. These results and an inspection of the amino-terminal sequence suggest that **MalF**, like many other inner membrane proteins, may not have an amino-terminal cleavable signal sequence.

From the detailed examination of several transmembrane proteins, it appears that a stretch of approximately 20–23 amino acids which are uncharged and enriched for hydrophobic amino acids will act as a membrane-spanning segment. Kyte and Doolittle (1982), using their hydrophobicity assignments, propose that any such segment will have an average hydrophy of greater than +1.22. In the case of **MalF**, a
Sequence of malF, an E. coli Maltose Transport Gene

Kyte-Doolittle hydrophathy analysis indicates several such stretches in this protein. Two or three of these occur close to the amino terminus of the protein. These are followed by a long sequence of approximately 181 amino acids which are not notably hydrophobic. Then, a more hydrophobic central region which is not hydrophobic may either be extramembraneous or may be intramembraneous, but either interacting with hydrophilic faces of other proteins (e.g. one of the other maltose transport proteins, such as MalG3) or forming part of an aqueous channel involved in maltose transport.

This work was initiated to study the mechanism of incorporation of MalF protein into the membrane. The sequence analysis of three gene fusions between malF and lacZ (Fig. 6) shown here provides preliminary information on this question. The latter two fusions code for a hybrid β-galactosidase which contains at least 5 or 6 of the hydrophobic stretches; these proteins are incorporated into the inner membrane (see Fig. 4). In contrast, the shorter fusion, 6-3, codes for a protein with none of these stretches and the protein is found, not surprisingly, in the cytoplasm.

We expect that a further analysis of such gene fusions, as well as of deletion mutants obtained in vivo and in vitro will indicate which portions of the protein can promote membrane insertion. Some indication of the structure and orientation of MalF transmembrane segments may also be obtained.

In the case of certain pairs of membrane proteins, especially those involved in transport, extensive amino acid sequence homology has been observed suggesting possible evolutionary and functional relationships (Gilson et al., 1982a). Using a computer program developed by Anthony Beckwith, Craig Werner, and Richard Ebright (Harvard University Medical School), we have compared the amino acid sequence of MalF to the membrane protein products of the hisM, hisP, and hisQ genes of Salmonella typhimurium (Higgins et al., 1982b).

These genes code for membrane components of a binding protein transport system (for histidine) similar to the maltose system of E. coli. However, no significant sequence homologies were found. We have also compared MalF to other E. coli inner membrane proteins including, Enzyme IIβ (the mltA gene product) (Lee and Saier, 1983), the lactose permease (Büchel et al., 1980), and several ATPase subunits (Nielsen et al., 1981). Once again, there were no significant homologies. Finally, we considered the possibility that MalF protein may have specific sequences in common with the other proteins involved in maltose transport or maltose chemotaxis. As a result, we compared the amino acid sequence of MalF with that of the MalK protein (Gilson et al., 1982b), with LamB (Clément and Hofnung, 1981), with maltose binding protein (Duplay et al., 1984), and with Tar (Russo and Koshland, 1983), the maltose chemoreceptor. No significant sequence homology was detected even though in all cases the stringency used can detect even small homologies, for example, that between the arabinose and galactose binding proteins (Argos et al., 1981).

By sequencing upstream of malF, we determined that the 153-bp malE-malF intercistronic space contains palindromic sequences which are strikingly homologous to sequences found in several other regions of the E. coli and S. typhimurium chromosomes (Fig. 5) (Anderson and Roth, 1978; Higgins et al., 1982a). Whether these palindromes function in the step-down in MalF protein synthesis or whether they are vestiges of the evolution of the chromosome, as has been suggested (Anderson and Roth, 1978; Higgins et al., 1982a), is not known. Of interest is the finding that in the histidine transport and biosynthetic operons of S. typhimurium, a similar regulation in the levels of synthesis of the promotor-distal gene product has been observed when these sequences

---

9 C. A. Lee, personal communication.
occur in the intercistronic spaces (Higgins et al., 1982a). Using the malF-lacZ gene fusions, we can isolate and characterize mutants affecting malF synthesis to help elucidate a possible regulatory role for this intercistronic sequence.

Acknowledgments—We are extremely grateful to Michael Berman for his expert advice on DNA sequencing techniques and for allowing the initial sequencing experiments to be conducted in his laboratory at the Frederick Cancer Research Facility, Frederick, MD 21701. S. F. especially thanks Anthony Beckwith for excellent assistance with computer analyses and word processing.

REFERENCES

Miller, J. H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY