JOURNAL OF BACTERIOLOGY, Apr. 1986, p. 244–252
0021-9193/86/040244-09$02.00/0

Vol. 166, No.1

Nucleotide Sequence of the *Escherichia coli motB* Gene and Site-Limited Incorporation of Its Product into the Cytoplasmic Membrane

JOAN STADER,[1]† PHILIP MATSUMURA,[1]* DOMINICK VACANTE,[1]‡ GARY E. DEAN,[2]§ AND ROBERT M. MACNAB[2]

Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois 60680,[1] and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511[2]

Received 8 November 1985/Accepted 24 January 1986

The *motB* gene product of *Escherichia coli* is an integral membrane protein required for rotation of the flagellar motor. We have determined the nucleotide sequence of the *motB* region and find that it contains an open reading frame of 924 nucleotides which we ascribe to the *motB* gene. The predicted amino acid sequence of the gene product is 308 residues long and indicates an amphipathic protein with one major hydrophobic region, about 22 residues long, near the N terminus. There is no consensus signal sequence. We postulate that the protein has a short N-terminal region in the cytoplasm, an anchoring region in the membrane consisting of two spanning segments, and a large cytoplasmic C-terminal domain. By placing *motB* under control of the tryptophan operon promoter of *Serratia marcescens*, we have succeeded in overproducing the MotB protein. Under these conditions, the majority of MotB was found in the cytoplasm, indicating that the membrane has a limited capacity to incorporate the protein. We conclude that insertion of MotB into the membrane requires the presence of other more hydrophobic components, possibly including the MotA protein or other components of the flagellar motor. The results further reinforce the concept that the total flagellar motor consists of more than just the basal body.

Escherichia coli cells swim by rotating their flagella (4). Paralyzed strains possessing apparently intact, but nonrotating, flagella are termed Mot⁻ mutants (2). Rotation requires the flagellar motor, an intact cell membrane, and a proton motive force. No cytoplasmic component appears to be part of the rotation mechanism, since cell envelopes, devoid of cytoplasm, can sustain flagellar rotation if an artificial proton motive force is applied (36). For a review of bacterial motility and the flagellar motor, see reference 27.

At least five proteins are essential for motor rotation (41, 47), but none of these have been found within the flagellar basal body (1, 20), even though the basal body is considered a major part of the motor. Two of these proteins, MotA and MotB, are integral to the cell membrane (6, 37, 41). They are not necessary for assembly of the flagellum and do not copurify with it (20). They can be synthesized after flagellar assembly and used to activate the motor; paralyzed *motA motB* mutants acquire the ability to rotate their flagella after MotA and MotB synthesis under the direction of lambda-*E. coli* hybrid bacteriophage (41). Recently, it was shown (6) that, at least in the case of MotB, this acquisition of motility proceeds by quantum increases in flagellar rotation rate, presumably as a result of successive incorporation of subunits of MotB protein. The fact that *motA* and *motB* null mutants are nonetheless flagellated makes these genes different from all other flagellum-associated genes and will have to be taken into account in considering how their products may assemble and function.

Other flagellar and motility proteins in addition to MotA and MotB are known or believed to be associated with the cell membrane, among them the *flaBI*, *flaBII*, and *flaBIII* gene products (D. H. Bartlett, Ph.D. thesis, University of Illinois, 1985). Also, the *flaBII* and *flaAII* gene products, which are important for motor rotation and switching and can be responsible for general chemotaxis defects (3, 10, 12, 14, 23, 47; J. Malakooti and P. Matsumura, unpublished data), have been shown to behave as membrane components in cell envelope assays of motor function (36). Therefore, it is possible that MotA and MotB are part of a multicomponent membrane structure which interacts with the chemosensory system.

Recently, we reported the *motA* gene sequence (15). We now report the sequence of the *motB* gene and also the subcloning of this gene downstream of the *Serratia marcescens* tryptophan operon (*trp*) promoter (21), with consequent overproduction of the MotB protein. Under overproduction conditions, MotB appears predominantly in the cytoplasm, suggesting that its incorporation into the membrane is site limited.

MATERIALS AND METHODS

Bacterial strains, bacteriophage, and plasmid vectors. Bacterial strains and plasmids used (Table 1) include C600, used to harbor recombinant plasmids; JC411, used for colicin E1 preparation; JM103, used for infection with bacteriophage M13; MS5037 and MS5038, used in complementation analysis; and X1488, a minicell-producing strain used in the labeling of plasmid-encoded proteins. Strains of bacteriophage M13 used for subcloning and sequencing were mp8 and mp9. Plasmid vectors used were pBR322, which carries resistance to ampicillin and tetracycline, and pMK2004,

* Corresponding author.
† Present address: Department of Molecular Biology, Princeton University, Princeton, NJ 08544.
‡ Present address: National Institute of Neurological and Communicative Disorders and Stroke, Bethesda, MD 20503.
§ Present address: Department of Microbiology and Molecular Genetics, University of Cincinnati College of Medicine, Cincinnati, OH 45267.

TABLE 1. Bacterial strains and plasmids

Strain	Genotype	Reference
C600	F⁻ *thi-1 thr-1 leuB6 lacY1 fhuA21 supE44*	28
JC411	F⁺ ColE1⁺ *arg his leu met thy*	8
JM103	Δ(*lac pro*) *thi rpsL supE endA sbcB15 hsdR4* F′ *traD36 proAB lacI^qZΔM15*	33
MS5037	*lac-1 gal-1 gal-2 ara-2 xyl thr leu thi fhuA* Str^r λ*t3 motA*	43
MS5038	*lac-1 gal-1 gal-2 ara-2 xyl thr leu thi fhuA* Str^r λ*t3 motB*	43
X1488	F⁻ Str^r *hst minA minB purE pdxC his ile met ade ura hsdR hsdM⁺*	32
pBR322	Cloning vector	28
pLC1-28	ColE1-*E. coli* hybrid containing *motB*	8, 15, 30
pMK2004	Cloning vector	22
pGD2	Subclone containing *motA* and *motB*	15
pJS15	Subclone containing *motA* and *motB*	This paper
M13mp8	Subcloning and sequencing vector	33
M13mp9	Subcloning and sequencing vector	33
M13mp8(14)	M13mp8 clone containing portion of *motB*	This paper
M13mp8(8)	M13mp8 clone containing portion of *motB*	This paper
pPM6	*motB* overexpression plasmid	This paper
pDV4	*cheA* overexpression plasmid	Vacante and Matsumura, unpublished work
pDV41	*cheY-motB* overexpression plasmid	This paper
pRL22	*cheY* overexpression plasmid	29

which carries resistance to ampicillin, tetracycline, and kanamycin.

For the sequencing aspects of this study, we also used bacterial strains and plasmids described in our report of the *motA* sequence (15).

Media, buffers, and chemicals. Luria broth and Luria agar were used for procedures which required rich media. Motility was assayed on plates containing 0.4% (wt/vol) agar. Antibiotics, when needed, were used at the following concentrations (micrograms per milliliter): ampicillin, 150; kanamycin, 50; and tetracycline, 30. Colicin E1 was prepared from strain JC411 and used as described by Spudich et al. (45). M9 salts medium (13) was used for growth of minicell cultures and was supplemented as required. Antibiotics, amino acids, vitamins, and β-3-indoleacrylic acid were purchased from Sigma Chemical Co., St. Louis, Mo.

Restriction endonucleases, exonucleases, and nucleotides were purchased from commercial sources and used according to the assay conditions of the manufacturers. DNA polymerase I, [^{32}P]dATP, and [^{32}P]dCTP were purchased from Amersham Corp., Arlington Heights, Ill. Polynucleotide kinase was purchased from PL Biochemicals, Milwaukee, Wis. T4 ligase was prepared by the procedure of Pirrotta and Bickle (35). [^{35}S]deoxycytidine 5′-[α-thio]triphosphate and [^{35}S]methionine were purchased from New England Nuclear Corp., Boston, Mass.

DNA isolation. Plasmid DNA was isolated as described by Clewell (11), except Triton X-100 was used for cleared lysis and 1 g of CsCl and 0.5 mg of ethidium bromide were used per ml of lysate. Unique restriction fragments for sequencing and cloning were isolated from pJS15 (Fig. 1) or pGD2 (15) by the method of Maxam and Gilbert (31), with 15 mM NaCl–1.5 mM sodium citrate (pH 7.5) for elution from gels.

Genetic and molecular genetic techniques. Cloning, transformation, and minicell procedures have been described previously (28, 30). Genetic complementation was observed as swarming on motility agar containing any required antibiotics.

DNA sequencing. Isolated DNA restriction fragments were 3′ end labeled with DNA polymerase I or 5′ end labeled with polynucleotide kinase and sequenced by the method of Maxam and Gilbert (31). In some cases, fragments were cloned into M13 bacteriophage and sequenced by the Sanger method (39) as described in the pamphlet entitled "M13 Cloning/DNA Sequencing System" (New England BioLabs, Inc., Beverly, Mass.).

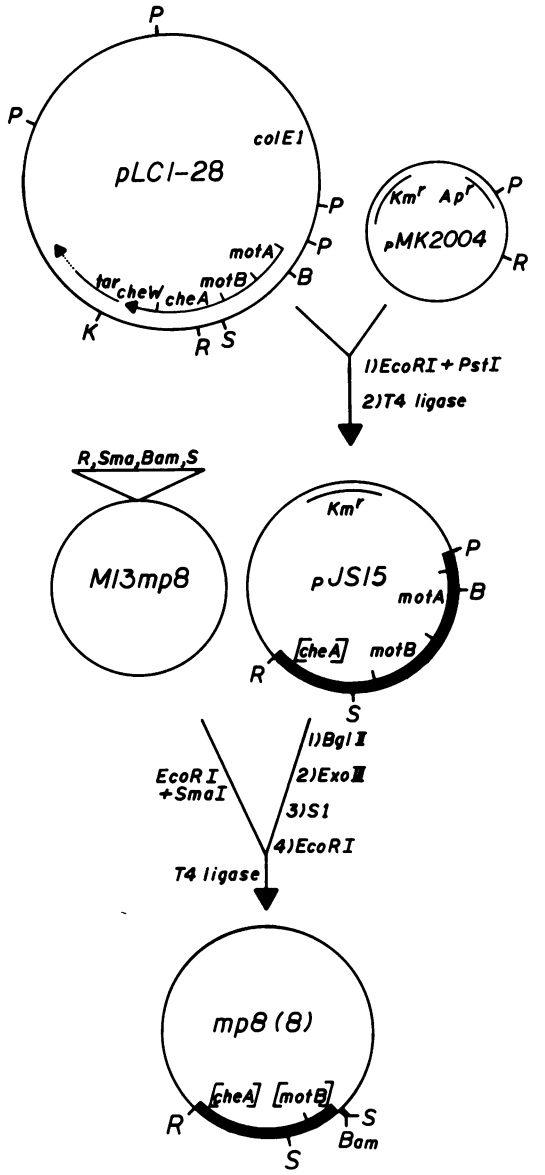

FIG. 1. Cloning of the *motB* gene for DNA sequencing. *E. coli* (ColE1) plasmid pLC1-28, containing the entire Mocha operon, was used as the initial source. Details of cloning, selection, and screening procedures are given in Materials and Methods. Heavy lines indicate donor DNA in the recombinant plasmids. Restriction enzyme abbreviations are as follows: B, *Bgl*II; Bam, *Bam*HI; K, *Kpn*I; P, *Pst*I; R, *Eco*RI; S, *Sal*I. Partial genes are enclosed in square brackets.

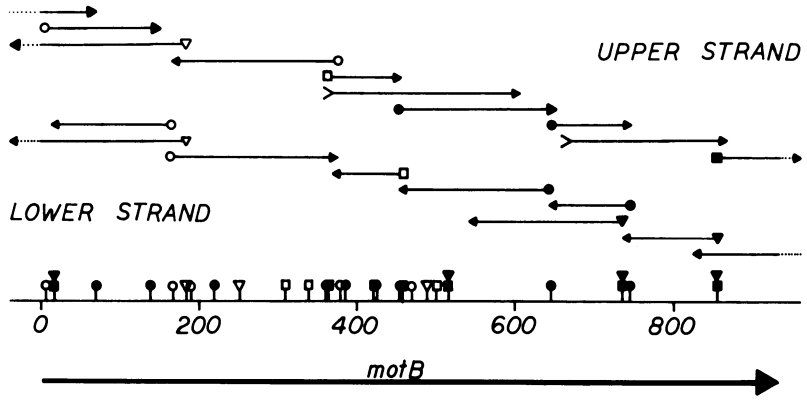

FIG. 2. Strategy for *motB* DNA sequencing, applied to plasmids pJS15 and pGD2. Restriction fragments used to obtain the upper (nontemplate) and lower (template) strands are shown. The symbol at the tail of each arrow denotes the restriction enzyme used prior to labeling (when the Maxam and Gilbert method was employed) or cloning (when the Sanger method was employed). The two upper-strand arrows that do not contain a symbol represent sequences of M13 clones generated by *Exo*III-S1 deletion (see text). Restriction enzyme symbols are as follows: ●, *Sau*3A; ■, *Hha*I; ○, *Hin*fI; ▼, *Hin*pI; □, *Taq*I; ▽, *Hpa*II.

***Exo*III-S1-generated deletions.** *Exo*III-S1-generated deletions were performed by modified versions of the methods of Roberts and Lauer (38) and Guo and Wu (19). DNA from plasmid pJS15 (10 μg) was digested with 10 U of *Bgl*II in a total volume of 50 μl of buffer (10 mM Tris (pH 7.4), 10 mM $MgSO_4$, 1 mM dithiothreitol) at 37°C for 2 h. After digestion, the NaCl was added to 66 mM, and Tris (pH 7.4) was added to 50 mM (total volume, 65 μl). *Exo*III (1 μl [66 U]) was then added, and the mixture was incubated at 22°C for 2 h. After incubation, 65 μl of 2× S1 buffer (100 mM sodium acetate (pH 4.0), 300 mM NaCl, 2 mM $ZnSO_4$) and 1 μl (45 U) of S1 were added. The mixture was left at 20°C for 2 h. Protein was removed by extraction with phenol-chloroform, and the DNA was precipitated with ethanol.

Overexpression of *motB* and cell fractionation. Expression of *motB* was induced by adding β-3-indoleacrylic acid (21). A C600(pPM6) culture was grown to mid-log phase in 1 liter of Luria broth, and 5 ml of β-3-indoleacrylic acid in ethanol was added to a final concentration of 100 μg ml^{-1}. Incubation, with shaking, was continued at 37°C. Maximum expression occurred 3 h after induction, at which point cells were harvested by centrifugation at 6,000 × *g* for 5 min. The pellets were frozen at −70°C. The frozen pellets were thawed and suspended in 50 ml of 50 mM Tris (pH 7.9), and the cells were ruptured by sonication (five 30-s bursts). Whole cells were removed by centrifugation at 6,000 × *g* for 10 min, and the supernatant was centrifuged in a fixed-angle Beckman 60 Ti rotor at 40,000 rpm (113,000 × *g*) for 1 h at 4°C. Membranes were recovered in the pellet; cytoplasm and periplasm were recovered in the supernatant.

Sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis. Analytical and preparative 12.5% acrylamide electrophoretic gels were prepared and used as described by Laemmli (26).

Preparation and amino acid analysis of CheY-MotB fusion protein. The *cheY-motB* fusion gene, on pDV41 in strain C600, was induced with β-3-indoleacrylic acid as described for CheY overproduction (29), and the cells were pelleted and subjected to SDS-gel electrophoresis in 7-mm-thick preparative gels. The position of the fusion protein was determined by Coomassie blue staining of vertical strips cut from the margins of the gel, and a 5-mm-wide horizontal strip at that position was placed in dialysis tubing, submerged in 10 mM NH_4HCO_3 in 0.01% SDS (pH 7.5), and electroeluted overnight at 30 mA. The gel slice was removed, and the eluted protein was extensively dialyzed against 10 mM NH_4HCO_3 (pH 7.5) and lyophilized. Amino acid analysis was carried out by ion-exchange chromatography with a Beckman model 121M analyzer.

RESULTS

Cloning of *motB*. Screening of the Clarke-Carbon *E. coli*(ColE1) hybrid plasmid library (9) for clones containing *motA* and *motB* has been described previously (15, 30); one of these plasmids, pLC1-28 (Fig. 1), carries the entire Mocha operon, consisting of *motA*, *motB*, *cheA*, and *cheW*.

The *motA* and *motB* genes were subcloned by inserting a *Pst*I-*Eco*RI fragment from the donor plasmid pLC1-28 into pMK2004 digested with the same two restriction endonucleases (Fig. 1). Transformants were selected for kanamycin resistance and then screened for ampicillin sensitivity and the ability to complement *motA* and *motB* mutants. pJS15 is one such recombinant plasmid that restored motility in paralyzed *E. coli* MS5037 (*motA*) and MS5038 (*motB*). From the *Pst*I site, the insert in pJS15 contains (i) the last 109 nucleotides of *flaI*, (ii) termination signals for the *flbB flaI* operon (D. H. Bartlett, B. Frantz, and P. Matsumura, manuscript in preparation; cf. reference 15), (iii) the promoter region of the Mocha operon, (iv) *motA* and *motB*, and (v) part of *cheA* (43).

As part of the strategy for sequencing *motB*, deletions into the gene were made by *Exo*III-S1 digestion from the *Bgl*II site within *motA*. This treatment was followed by *Eco*RI digestion and insertion of the fragments, containing various extents of *motB*, into *Sma*I- and *Eco*RI-digested M13mp8. Two useful clones were recovered: M13mp8(14) and M13mp8(8), in which *motB* was deleted to nucleotides 370 and 670, respectively. The structure of M13mp8(8) is given in Fig. 1.

DNA sequence of *motB*. Using the sequencing strategy shown (Fig. 2), we obtained the nucleotide sequence of the region containing *motB* (Fig. 3). The longest reading frame downstream of *motA* extends from nucleotide 1 to 924 and is preceded by a potential ribosome-binding site (Shine-Dalgarno sequence; 40), which lies within the *motA* gene. The translation product encoded in this region would be expected to have a molecular weight of 34,147. In our gel system, the protein migrated with an M_r of 42,000; other studies have reported an M_r of 39,000 (30, 41). The next-largest coding region is in the same reading frame and

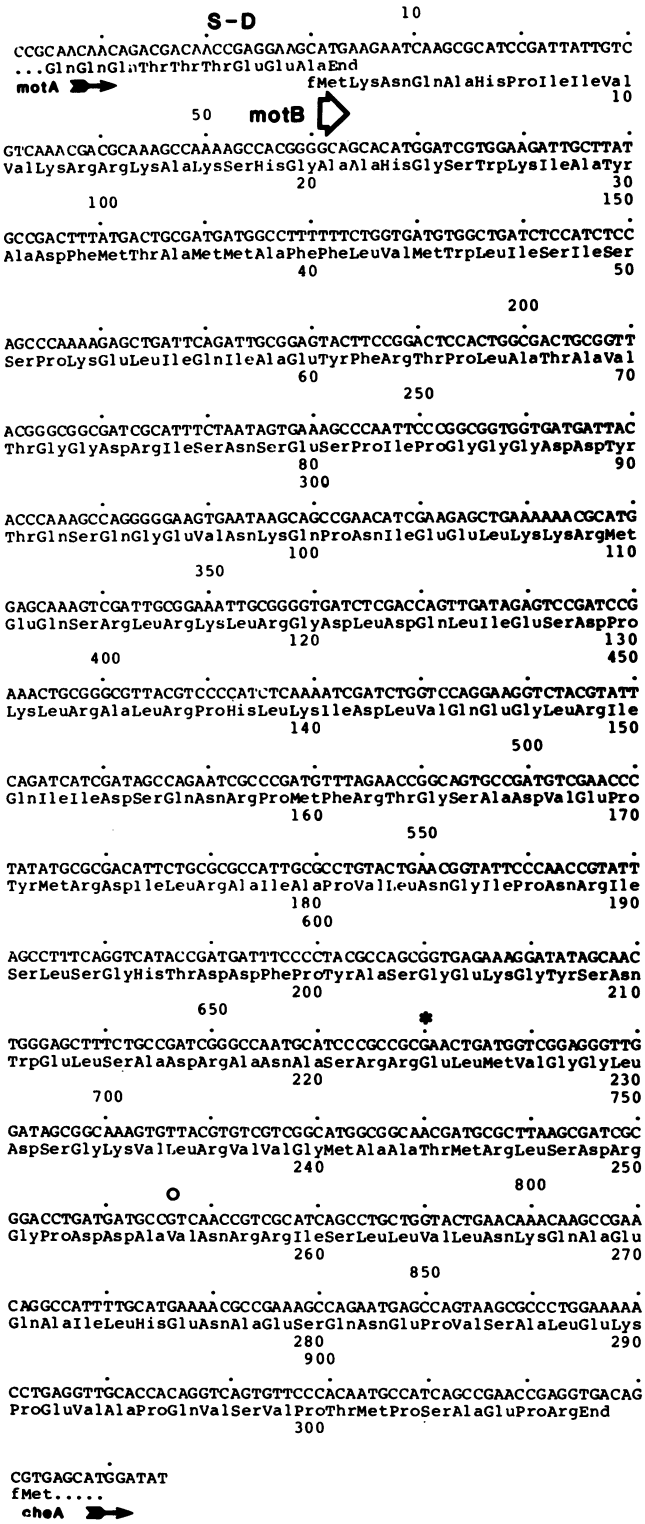

FIG. 3. Nucleotide sequence of the *motB* gene and amino acid sequence of the gene product. The sequence is numbered commencing at the putative initiation codon of the *motB* gene. The 3′ end of *motA* and the 5′ end of *cheA*, i.e., the first and third genes of the Mocha operon, respectively, are also shown. The start of the MotB portion of the CheY-MotB fusion protein is Glu224 (*); the C-terminal MotB residue of the truncation protein mentioned in the text is Val 256 (O).

extends from nucleotide 100 to 924. It would encode a polypeptide of only 30,384 daltons, its putative start codon is not preceded by a recognizable Shine-Dalgarno sequence, and it would also leave a substantial intergenic region between *motA* and *motB*. For these reasons, we favor the earlier start site.

The termination codon of *motA* and the predicted initiation codon of *motB* form a 4-base overlap, ATGA. There is no overlap between *motB* and *cheA*, the next gene in the operon. *cheA* starts, after an intergenic region of four nucleotides, at nucleotide 932 on our numbering system in Fig. 3 (E. Kofoid and J. S. Parkinson, personal communication).

Analysis of coding probability. A diagnostic program called Testcode (17) was used to determine the probability that the putative *motB* gene (nucleotides 1 through 924) was an authentic coding sequence; the program predicted that it was, at a confidence level of 88%. The open reading frame from nucleotide 100 through 924 was also predicted to be an authentic coding sequence (confidence level of 88%).

Analysis of codon usage. Grosjean and Fiers (18) have compared codon usage between highly and weakly expressed genes in *E. coli* and have found definite differences in codon preferences. These preferences are thought to be

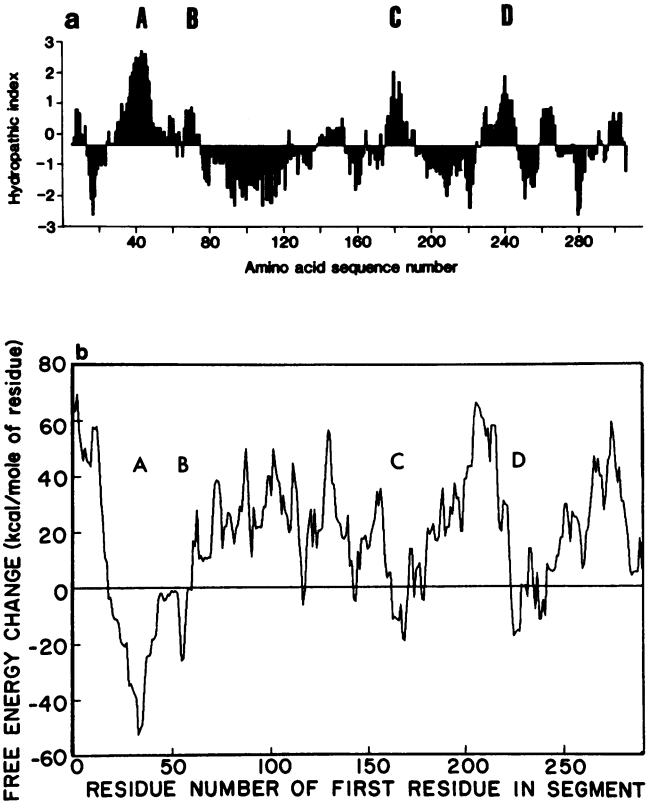

FIG. 4. Hydropathy profile of the MotB protein as determined by two models. (a) Kyte-Doolittle model (25). Hydropathy values are averages for a moving segment of nine amino acids, centered around the numbered residue. (b) Engelman-Steitz model (16). Free energy changes are for the transfer, from water to a nonpolar environment, of a moving α-helical segment of 18 residues, beginning at the numbered residue. In panel a, positive values represent hydrophobicity, and negative values represent hydrophilicity; in panel b, the reverse is true. Hydrophobic regions A through D are discussed in the text.

based on the energy involved in codon-anticodon pairing and the composition of the tRNA population. Furthermore, there are "modulator" codons that are rarely, if ever, used by highly expressed genes. We scrutinized *motB* for codon preferences and found that they were consistent with those of genes expressed at low levels.

We also found differences in codon usage between *motA* and *motB* which were consistent with the observation that *motB* expression is lower than that of *motA* (15, 37). Relative to *motA*, *motB* contains more codons found in weakly expressed genes (34% of total versus 26% of total), fewer codons found in strongly expressed genes (14 versus 19%), and more modulator codons (6 versus 4%).

Predicted amino acid sequence and hydropathy analysis of MotB. The predicted amino acid sequence of MotB is given in Fig. 3. It is not as hydrophobic as might have been expected for an integral membrane protein (Leu+Ile+Val+Met+Phe+Pro = 34 mol%; Asp+Glu+Arg+Lys = 26 mol%). The corresponding figures for MotA are 40 and 18 mol%, respectively. MotB is predicted to be 13 residues larger than MotA but to contain 14 fewer hydrophobic residues, 25 more charged residues, and 2 more amphipathic residues.

There are several noteworthy features of the MotB sequence. (i) The N terminus is very basic. Among the first 27 residues, there are 5 Lys, 2 Arg, and 3 His and no acidic residues. (ii) Hydrophobic residues are fairly uniformly distributed throughout the sequence, except for one concentrated 17-residue region between residues 33 and 49, in which there are 12 hydrophobic residues (Leu, Ile, Val, Met, and Phe) and no charged residues. (iii) Although the sequence as a whole is not especially proline rich (20 residues, 6 mol%), 6 of these residues occur within the last 25 residues at the C terminus.

Kyte and Doolittle (25) have developed an algorithm for analysis of the hydropathic properties of a protein. We subjected the MotB amino acid sequence to this analysis (Fig. 4a). There is one region of the sequence (region A, residues 28 to 49) that is strongly hydrophobic and of sufficient length (21 amino acids) to traverse the membrane. There are two other, less pronounced, hydrophobic regions: C (residues 175 to 187) and D (residues 225 to 245). Analysis of the MotB sequence by an alternative procedure developed by Engelman and Steitz (16), which calculates the free energy of transfer of an α-helical peptide from water to a hydrophobic environment, was consistent with the above description (Fig. 4b) but indicated an additional hydrophobic region B (residues 55 to 72), which was less evident in the Kyte-Doolittle analysis.

Construction of a plasmid to overexpress *motB*. Figure 5 shows the construction of a plasmid designed to overexpress *motB*. Plasmid pDV4 (D. McNally, D. Vacante, and P. Matsumura, manuscript in preparation) was derived from pRL22, a plasmid that overexpresses *cheY* (29). pDV4 contains the same DNA as pRL22 from the ampicillin resistance marker clockwise through the first *Sal*I site. Specifically, it contains the first 120 bases of *cheY*. Digestion of pLC1-28 (Fig. 1) with *Bgl*II and *Kpn*I and insertion of the *motB*-containing fragment into pDV4 digested with *Bam*HI and *Kpn*I produced plasmid pPM6. The ligation of the *Bgl*II and *Bam*HI sites resulted in the loss of both sites and also, since the *Bgl*II site in *motA* is at nucleotide 324 of the coding sequence (15), generated an out-of-phase *cheY-motA* gene fusion. Therefore, translation of the transcript that includes *motB* first initiates at the start codon of *cheY* and terminates shortly after the *cheY-motA* junction. It then reinitiates at

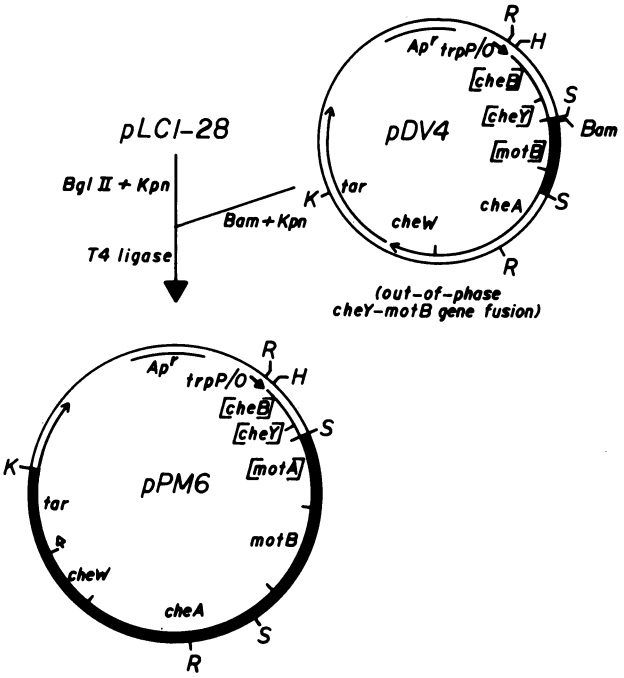

FIG. 5. Construction of plasmid pPM6, used to overexpress *motB*. Refer to the legend for Fig. 1 for symbols, restriction enzyme abbreviations, and a diagram of pLC1-28. See the text for an explanation.

the start codon of *motB*. We calculate the untranslated region between the end of the *cheY-motA* fusion and the beginning of *motB* to be 498 bases. Plasmid pPM6 was able to complement *E. coli* MS5038 (*motB*), confirming that *motB* was intact and indicating that the untranslated intergenic region had no noticeable polar effects on its expression. The absence of a polar effect from untranslated mRNA was also noted in the overproduction of CheY in pRL22 (29).

Protein synthesis from plasmid pPM6. The proteins synthesized in minicells when strain X1488 was transformed with pPM6 are shown in Fig. 6 (lane a). MotB can be seen at an M_r of 42,000 with the same mobility was the band produced by pJS15 (lane b). In addition to *motB*, pPM6 was found to direct expression of the *cheA*, *cheW*, and *tar* (MCPII) genes. Although MotB is greatly overproduced in induced exponential-phase cultures (see below), the amount of radiolabeled protein produced in minicell preparations does not reflect this.

The natural promoter for *motA* and *motB* is, like those of all other motility-related operons, under control of the *flbB flaI* operon which is itself under control of the cyclic AMP-cyclic AMP receptor protein complex (24, 42). Consistent with this fact and with the construction of our plasmids, MotA and MotB synthesis from pJS15 was repressed by 50 mM glucose—the phenomenon of catabolite repression—whereas MotB synthesis from pPM6, under *trp* promoter control, was not (data not shown).

Overexpression of *motB* and localization of the MotB protein. *E. coli* C600 cells transformed with pPM6 overexpressed *motB* when the *trp* inducer β-3-indoleacrylic acid was added to the culture (21). Figure 7 shows cytoplasmic and membrane fractions of induced C600 cells containing pPM6 compared with the same cells without plasmid. The cytoplasmic fraction of the plasmid-containing culture shows an intense band at an M_r of 42,000 (lane b), which was detected

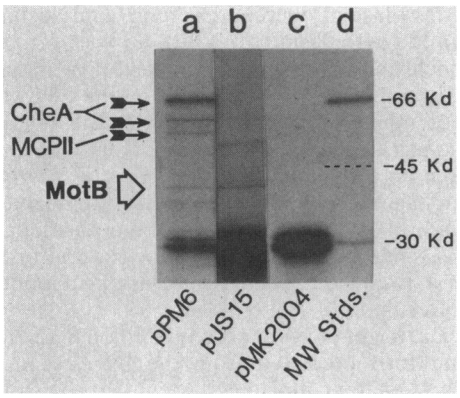

FIG. 6. Autoradiograph of SDS-polyacrylamide gel containing polypeptides produced in minicell strain X1488 containing the following plasmids: pPM6 (lane a), pJS15 (lane b), and pMK2004 (lane c). Lane d contains molecular weight standards (66 Kd, bovine serum albumin; 45 Kd, chicken ovalbumin, with radiolabel detectable at longer exposures, at the position shown by the dotted line; and 30 Kd, carbonic anhydrase). Lane b is from a longer film exposure of the same gel, done to render the MotB band visible at the same density as in lane a; the register of the bands shown here represents that seen on a single film.

only after induction and increased over time thereafter (data not shown). It was not seen in the cytoplasmic fraction of the culture without plasmid (lane c). The membrane fraction of C600 cells containing pPM6 shows, in contrast, only a faint band at the same position (lane d), this band again being absent in cells lacking pPM6 (lane e). (pPM6-mediated synthesis and insertion of small amounts of MotB into the cell membrane was confirmed in radiolabeled minicell preparations, which are not shown.) These results indicate that *E. coli* cells have a limited capacity to incorporate MotB protein into the membrane.

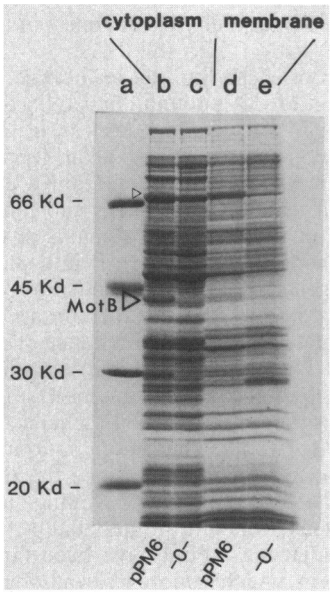

FIG. 7. SDS-polyacrylamide gel electrophoretograms demonstrating the overproduction and localization of MotB protein. Each lane contains fractions from 5×10^7 cells. Lane a, Molecular weight standards as for Fig. 6, plus 20 Kd trypsin inhibitor; lane b, cytoplasmic fraction of C600 cells containing plasmid pPM6; lane c, cytoplasmic fraction of C600 alone; lane d, membrane fraction of C600 cells containing pPM6; lane e, membrane fraction of C600 alone. The position of MotB (M_r = 42,000) is indicated by the large arrowhead; CheA (higher-molecular-weight form, M_r = 76,000) is indicated by a small arrowhead; CheW is not within the M_r range of this gel, but has been detected under other conditions (not shown); MCPII was being synthesized under the control of its natural promoter and was not detected.

Amino acid compositional analysis of a CheY-MotB fusion protein. We wished to obtain amino acid compositional information to reinforce our conclusions, based on DNA sequence, concerning the MotB protein. Although we had no convenient method of obtaining the MotB protein itself in high purity for such an analysis, we had constructed and sequenced a fusion gene that coded for a protein consisting of the first 41 amino acids of CheY, 4 amino acids coded for by the M13 polylinker, and the last 85 amino acids of MotB (Vacante, Stader, and Matsumura, unpublished data). The fusion gene was under *S. marcescens trp* control on plasmid pDV41 (pDV4 [Fig. 5] with four bases added to the *Bam*HI site). When induced with β-3-indoleacrylic acid, synthesis of the fusion protein was such that it represented 20% of cytoplasmic protein and could be purified by electroelution from 7-mm preparative SDS-polyacrylamide gels. The amino acid composition of this protein was in good agreement with that predicted from the gene sequence (Table 2).

DISCUSSION

Identification and sequencing of the *motB* gene. We have identified the DNA sequence given in Fig. 3 (bases 1 to 927) as being the *motB* gene, based on the following criteria. (i) Plasmids pJS15 and pGD2 used to obtain this sequence both complemented *motB* mutants. (ii) The sequence immediately upstream has previously been identified as the *motA* gene (15), and the sequence immediately downstream has been identified as the *cheA* gene (Kofoid and Parkinson, personal communication); this is the physical order of these genes, the first three of the Mocha operon, as deduced by detailed mapping studies (44). (iii) The sequence is an open reading frame of about the right size to code for the MotB protein (but see discussion below regarding the anomalous electrophoretic mobility of MotB).

TABLE 2. Amino acid composition of CheY-MotB fusion protein

Amino acid	No. of residues	
	Predicted	Observed[a]
Ala	12	11.3
Arg	8	9.0
Asx	16	15.8
Cys	0	0.0
Glx	17	17.9
Gly	8	10.7
His	1	1.2
Ile	4	3.9
Leu	14	14.1
Lys	6	6.2
Met	6	3.9
Phe	3	3.0
Pro	8	7.8
Ser	9	8.3
Thr	3	3.2
Trp	0	ND
Tyr	0	0.0
Val	15	13.8

[a] Normalized to the predicted total of 130. ND, Not determined.

Amino acid compositional data from a CheY-MotB fusion protein support the correctness of the sequence data.

Arrangement of MotB in the lipid bilayer. The *motB* gene sequence enables us to predict the primary sequence of the gene product and then to consider possible higher-order organization of the MotB protein.

MotB has previously been shown to be in the cytoplasmic membrane and to remain there in the presence of salt or EDTA (37), suggesting a hydrophobic interaction between MotB and the lipid bilayer. The hydropathic profile of MotB (Fig. 4), however, reveals a polypeptide that is not particularly hydrophobic except for one major region (region A) and a few minor ones (regions B, C, and D). Region A is as hydrophobic as other sequences that have been shown to traverse the lipid bilayer (25), but regions B, C, and D contain charged amino acids (2, 1, and 3, respectively) and are probably no more hydrophobic than interior regions of soluble proteins. Since (i) there is no consensus signal sequence at the N terminus, (ii) 10 of 27 of the N-terminal amino acids are charged, and (iii) it is known from in vitro translation studies that the MotB protein is not cleaved posttranslationally (7), it seems unlikely that the N terminus of MotB is translocated to the outside of the membrane. It is possible, however, that the highly hydrophobic region A is an "insertion signal" similar to the C-terminal end of cytochrome *b* (5, 34) and would form the first of two segments spanning the membrane. A nearby region (indicated as B in Fig. 4), from residue 55 to 72, is moderately hydrophobic and might constitute the return spanning segment. In this case, all hydrophilic regions except for the short region between residues 50 and 55 would remain on the cytoplasmic side of the membrane. This predicted organization of MotB in the membrane (Fig. 8) is not likely to be energetically as favorable as that predicted (Fig. 6 of reference 15) for MotA, which is a considerably more hydrophobic protein. We shall return to this question when the localization of overproduced MotB is discussed.

Overproduction of MotB protein. We have placed *motB* under the control of the high-level promoter of the *S. marcescens trp* operon and have shown that *E. coli* cells carrying this plasmid overexpress *motB* and accumulate its product in the cytoplasm.

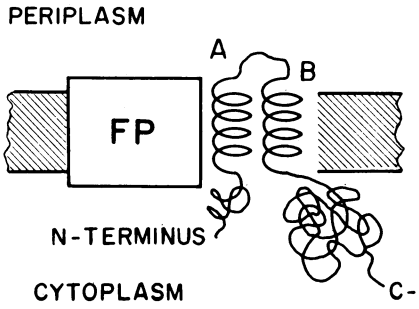

FIG. 8. Postulated organization of MotB in the cell membrane, based on the assumption that the protein is not translocated to the exterior by a signal sequence (see Results). The first 26 or so amino acids at the N terminus contain a high density of positive charge and form a small cytoplasmic domain. The strongly hydrophobic region A (Fig. 4) spans the membrane and, after an abrupt turn, possibly involving Pro 52, the moderately hydrophobic region B (Fig. 4) returns across the membrane. The remaining 75% of the sequence towards the C terminus constitutes a large cytoplasmic domain. The membrane-spanning hairpin may be only marginally stable and require association with other flagellar motor components, indicated schematically by FP (flagellar proteins) in the adjacent block in the membrane.

Our conclusion that the overproduced protein with an M_r of 42,000 is authentic MotB is based on the following. (i) In samples of minicells containing pPM6, the overproduced protein comigrated in electrophoretic gels with MotB produced from its own promoter on pJS15. (ii) *motB* mutants were complemented when transformed with the overproducing plasmid, pPM6. (iii) Synthesis of the 42-kilodalton protein on plasmids containing either the natural or a foreign promoter showed the expected patterns of catabolite repression (see Results).

Molecular weight of MotB protein. Although the molecular weight of MotB predicted from the DNA sequence was 34,147, the apparent molecular weight on SDS-polyacrylamide gels was in the range of 39,000 to 42,000. Interestingly, the CheY-MotB fusion protein, which contained only the C-terminal 85 amino acids of MotB, also displayed a substantial discrepancy (M_r = 20,000; predicted molecular weight, 14,109; Stader and Matsumura, unpublished data). In contrast, a truncated protein consisting of the first 256 residues of MotB and 5 additional residues runs close to normal on gels (M_r = 30,000; predicted molecular weight, 28,924; M. L. Wilson and R. M. Macnab, unpublished data). Apparently, the C-terminal end of MotB is necessary and sufficient to cause aberrantly low electrophoretic mobility. This region of the sequence is remarkable in one regard: there are 6 prolines within the final 25 residues. It may be that this difference either prevents the protein from binding the usual amount of SDS or causes it to adopt a conformation with a different frictional geometry. In another biopolymer, duplex DNA, the introduction of bends is known to cause abnormally low mobility in polyacrylamide gels (46).

Incorporation of MotB into the membrane. When present at normal levels, MotB is found entirely in the cytoplasmic membrane (37). In a previous study (7), in which synthesis was under *lac* control, MotB was again found in the membrane; however, the level of production was not high. We have found that when the protein is substantially overproduced (as for Fig. 7), most of it remains in the cytoplasm. The fact that incorporation of MotB into the membrane is site limited suggests that other components are necessary for its stable incorporation into the membrane. As was noted above, the predicted amino acid sequence of MotB indicates a protein that would by itself interact only moderately strongly with a lipid bilayer. Placing *motB* under the control of a high-level promoter not only changes the absolute amount of MotB, but also changes its ratio to other components of the motor. Specifically, since the plasmid did not synthesize MotA, the normally prevailing stoichiometry of these two proteins, which are functionally related and may be physically adjacent, would have been far from normal. The MotA protein, which is more hydrophobic than MotB, is not site limited with respect to incorporation into the membrane, even at high levels of overproduction (M. L. Wilson and R. M. Macnab, unpublished data). One possibility is that, in the normal *E. coli* cell, MotA-MotB complexes form that have a sufficiently high mean hydrophobicity to stably incorporate into the membrane.

It is not clear, however, what the natural stoichiometry is between MotA and MotB. When expressed from a lambda-*E. coli* hybrid phage, but using the natural Mocha promoter, MotA appears in a substantial excess over MotB (about 20:1) (15). When we published the *motA* gene sequence, we noted the region of overlap between the last four nucleotides of

motA and the first four nucleotides of motB and suggested that this arrangement could be important in controlling the relative stoichiometry between the two gene products (15). It is interesting that there is no such overlap between motB and cheA; there is also no reason, based on what is known of the functions of the gene products, to suspect that the stoichiometry of MotB relative to CheA is especially important.

Finally, we may recall that the only part of the flagellar motor of E. coli that is currently recognizable in morphological terms is the basal body. This structure does not contain the MotA and MotB proteins or the Fla proteins that are involved in energization and switching (1, 20, 47). It is therefore almost certainly not the "total motor" (27). Localization and morphological identification of the missing components such as MotB are important next steps to an understanding of bacterial motility.

ACKNOWLEDGMENTS

We thank E. Kofoid and J. S. Parkinson for showing us unpublished DNA sequence information concerning the motB-cheA junction region, M. Silverman for sending us strains MS5037 and MS5038, and B. Frantz for technical assistance.

This work has been supported by Public Health Service grants AI18985 (to P.M.) and AI12202 (to R.M.M.).

LITERATURE CITED

1. **Aizawa, S.-I., G. E. Dean, C. J. Jones, R. M. Macnab, and S. Yamaguchi.** 1985. Purification and characterization of the flagellar hook-basal body complex of *Salmonella typhimurium*. J. Bacteriol. **161**:836–849.
2. **Armstrong, J. B., and J. Adler.** 1967. Genetics of motility in *Escherichia coli*: complementation of paralyzed mutants. Genetics **56**:363–373.
3. **Bartlett, D. H., and P. Matsumura.** 1984. Identification of *Escherichia coli* region III flagellar gene products and description of two new flagellar genes. J. Bacteriol. **160**:577–585.
4. **Berg, H. C., and R. A. Anderson.** 1973. Bacteria swim by rotating their flagellar filaments. Nature (London) **245**:380–382.
5. **Blobel, G.** 1980. Intracellular protein topogenesis. Proc. Natl. Acad. Sci. USA **77**:1496–1500.
6. **Block, S. M., and H. C. Berg.** 1984. Successive incorporation of force-generating units in the bacterial rotary motor. Nature (London) **309**:470–472.
7. **Boyd, A., G. Mandel, and M. I. Simon.** 1982. Integral membrane proteins required for bacterial motility and chemotaxis, p. 123–137. *In* W. B. Amos and J. G. Duckett (ed.), Prokaryotic and eukaryotic flagella. Cambridge University Press, Cambridge.
8. **Clarke, L., and J. Carbon.** 1975. Biochemical construction and selection of hybrid plasmids containing segments of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA **72**:4361–4365.
9. **Clarke, L., and J. Carbon.** 1976. A colony bank containing synthetic col E1 hybrid plasmid representative of the entire *E. coli* genome. Cell **9**:91–99.
10. **Clegg, D. O., and D. E. Koshland, Jr.** 1985. Identification of a bacterial sensing protein and effects of its elevated expression. J. Bacteriol. **162**:398–405.
11. **Clewell, D. B.** 1972. Nature of col E_1 plasmid replication in *Escherichia coli* in the presence of chloramphenicol. J. Bacteriol. **110**:667–676.
12. **Collins, A. L. T., and B. A. D. Stocker.** 1976. *Salmonella typhimurium* mutants generally defective in chemotaxis. J. Bacteriol. **128**:754–765.
13. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. A manual for genetic engineering: advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
14. **Dean, G. E., S.-I. Aizawa, and R. M. Macnab.** 1983. *flaAII* (*motC*, *cheV*) of *Salmonella typhimurium* is a structural gene involved in energization and switching of the flagellar motor. J. Bacteriol. **154**:84–91.
15. **Dean, G. E., R. M. Macnab, J. Stader, P. Matsumura, and C. Burks.** 1984. Gene sequence and predicted amino acid sequence of the *motA* protein, a membrane-associated protein required for flagellar rotation in *Escherichia coli*. J. Bacteriol. **159**:991–999.
16. **Engelman, D. M., and T. A. Steitz.** 1984. On the folding and insertion of globular membrane proteins, p. 87–114. *In* D. B. Wetlaufer (ed.), The protein-folding problem. Westview Press, Boulder, Colo.
17. **Fickett, J. W.** 1982. Recognition of protein coding regions in DNA sequences. Nucleic Acid Res. **10**:5303–5318.
18. **Grosjean, H., and W. Fiers.** 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene **18**:199–209.
19. **Guo, L.-H., and R. Wu.** 1982. New rapid methods for DNA sequencing based on exonuclease III digestion followed by repair synthesis. Nucleic Acids Res. **10**:2065–2084.
20. **Hilmen, M., and M. Simon.** 1976. Motility and the structure of bacterial flagella, p. 35–45. *In* R. Goldman, T. Pollard, and J. Rosenbaum (ed.), Cell motility. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Joachimiak, A., R. L. Kelley, R. P. Gunsalus, C. Yanofsky, and P. B. Sigler.** 1983. Purification and characterization of *trp* aporepressor. Proc. Natl. Acad. Sci. USA **80**:668–672.
22. **Kahn, M., R. Kolter, C. Thomas, D. Figurski, R. Meyer, E. Remaut, and D. R. Helinski.** 1979. Plasmid cloning vehicles derived from plasmids ColE1, F, R6K, and RK2. Methods Enzymol. **68**:268–280.
23. **Khan, S., R. M. Macnab, A. L. DeFranco, and D. E. Koshland, Jr.** 1978. Inversion of a behavioral response in bacterial chemotaxis: explanation at the molecular level. Proc. Natl. Acad. Sci. USA **75**:4150–4154.
24. **Komeda, Y.** 1982. Fusions of flagellar operons to lactose genes on a Mu *lac* bacteriophage. J. Bacteriol. **150**:16–26.
25. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
26. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
27. **Macnab, R. M., and S.-I. Aizawa.** 1984. Bacterial motility and the bacterial flagellar motor. Annu. Rev. Biophys. Bioeng. **13**:51–83.
28. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Matsumura, P., J. J. Rydel, R. Linzmeier, and D. Vacante.** 1984. Overexpression and sequence of the *Escherichia coli cheY* gene and biochemical activities of the CheY protein. J. Bacteriol. **160**:36–41.
30. **Matsumura, P., M. Silverman, and M. Simon.** 1977. Synthesis of *mot* and *che* gene products of *Escherichia coli* programmed by hybrid ColE1 plasmids in minicells. J. Bacteriol. **132**:996–1002.
31. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
32. **Meagher, R. B., R. C. Tait, M. Betlach, and H. W. Boyer.** 1977. Protein expression in *E. coli* minicells by recombinant plasmids. Cell **10**:521–536.
33. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
34. **Okada, Y., A. B. Frey, T. M. Guenthner, F. Oesch, D. D. Sabatini, and G. Kreibich.** 1982. Studies on the biosynthesis of microsomal membrane proteins. Site of synthesis and mode of insertion of cytochrome b_5, cytochrome b_5 reductase, cytochrome P-450 reductase and epoxide hydrolase. Eur. J. Biochem. **122**:393–402.
35. **Pirrotta, V., and T. A. Bickle.** 1980. General purification schemes for restriction endonucleases. Methods Enzymol. **65**: 89–95.
36. **Ravid, S., and M. Eisenbach.** 1984. Minimal requirements for rotation of bacterial flagella. J. Bacteriol. **158**:1208–1210.

37. **Ridgway, H. F., M. Silverman, and M. I. Simon.** 1977. Localization of proteins controlling motility and chemotaxis in *Escherichia coli*. J. Bacteriol. **132**:657–665.
38. **Roberts, T. M., and G. D. Lauer.** 1979. Maximizing gene expression on a plasmid using recombination in vitro. Methods Enzymol. **68**:473–482.
39. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
40. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
41. **Silverman, M., P. Matsumura, and M. Simon.** 1976. The identification of the *mot* gene product with *Escherichia coli*-lambda hybrids. Proc. Natl. Acad. Sci. USA **73**:3126–3130.
42. **Silverman, M., and M. Simon.** 1974. Characterization of *Escherichia coli* flagellar mutants that are insensitive to catabolite repression. J. Bacteriol. **120**:1196–1203.
43. **Silverman, M., and M. Simon.** 1976. Operon controlling motility and chemotaxis in *E. coli*. Nature (London) **264**:577–580.
44. **Silverman, M., and M. Simon.** 1977. Identification of polypeptides necessary for chemotaxis in *Escherichia coli*. J. Bacteriol. **130**:1317–1325.
45. **Spudich, J. A., V. Horn, and C. Yanofsky.** 1970. On the production of deletions in the chromosome of *Escherichia coli*. J. Mol. Biol. **53**:49–67.
46. **Wu, H.-M., and D. M. Crothers.** 1984. The locus of sequence-directed and protein-induced DNA bending. Nature (London) **308**:509–513.
47. **Yamaguchi, S., H. Fujita, A. Ishihara, S.-I. Aizawa, and R. M. Macnab.** 1986. Subdivision of flagellar genes of *Salmonella typhimurium* into regions responsible for assembly, rotation and switching. J. Bacteriol. **166**:187–193.