Sequence of the flaA (cheC) Locus of Escherichia coli and Discovery of a New Gene

SCOT C. KUO AND DANIEL E. KOSHLAND, JR.

Department of Biochemistry, University of California, Berkeley, California 94720

Received 17 January 1986/Accepted 28 March 1986

The flaA (cheC) locus from Escherichia coli is important in controlling the rotational direction of flagella during chemotaxis. The locus was sequenced, and a site of transcriptional initiation was determined. Two reading frames, flaA and flaAI, span the locus. flaAI corresponds to certain flaA and cheC mutations, and has some unusual features in the predicted secondary structure. flaAI, however, has not been identified previously, but a flaAI deletion, which produced a truncated FlaA1 peptide in minicells, clearly identified the FlaA1 protein.

Peritrichous bacteria, such as Escherichia coli, swim in two modes, smooth swimming and tumbling, that alternate to produce a three-dimensional random walk (4). Chemotaxis, the migration in response to chemical gradients, is accomplished by altering the frequency and duration of each of these two modes while swimming, producing a biased random walk (4, 18). One gene, flaA (also known as cheC), plays a key role in mediating transitions between smooth swimming and tumbling (12, 28, 31).

Mutations in the flaA (cheC) locus exhibit a number of phenotypes. Some E. coli and Salmonella typhimurium mutants, including null mutants, lack flagella (17, 35, 36), while others exhibit paralyzed flagella (17, 34, 36), suggesting that flaA has roles in both flagellar structure and motility. There are also chemotaxis mutants (designated cheC) with strong biases toward smooth swimming or tumbling (3, 17, 25, 28, 36, 37), including a mutant with inverted response to attractants and repellents (12, 31). Although FlaA is not an integral membrane protein (7), the phenotype of flaA mutants is retained in spheroplasts which lack cytoplasmic material (29), suggesting that FlaA is membrane-associated. The cumulative evidence suggests that FlaA protein is part of a molecular switch at the motor controlling flagellar rotation during chemotaxis (12, 27, 28, 31).

Recently, the flaA locus from E. coli has been subcloned, and the effects of its overproduction on chemotaxis have been studied (7). Deletion mutants and two-dimensional gels of cheC mutants correlate flaA with a 37-kilodalton (kDa) protein. Overproduction of this 37-kDa protein affects the motility of cells by slowing the rate of swimming and reducing the frequency of tumbling.

We have sequenced the flaA locus from E. coli and determined that two open reading frames, designated flaA and flaAI, span the region. flaAI has not been identified previously, whereas flaAI corresponds to certain flaA and cheC mutations. Evidence is presented that flaAI is expressed in minicells.

MATERIALS AND METHODS

Bacterial strains. Only E. coli K-12 strains were used in this study. RP437 (F' thr-1 (Am)/leuB6 his-4 metF159(Am) edu-50 rpsL136 thi-1 ara-14 lacY1 mil-1 xyl-5 flaA31 tss-78), which is wild type for chemotaxis, was from J. S. Parkinson (26). Plasmid isolation and propagation was done in MM294 (hsdR17 (hsdR hsdM* endA1 thi) (23) with a deletion in recA (from B. Bochner, Genentech, San Francisco, Calif.). M13 phage isolation and propagation were done in JM103 (24).

Enzymes and nucleotides. Restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase, and the Klenow fragment of E. coli DNA polymerase were purchased from New England Biolabs and from Boehringer Mannheim Biochemicals. S1 nuclease and dideoxynucleotides were purchased from P-L Biochemicals.

[α-32P]dNTPs (400 Ci/mmol), [γ-32P]ATP, and deoxyadenosine 5'-α-[35S]dATP thiotriphosphate (αS) were purchased from Amersham Corp.

Sequencing and sequence analysis. Both Maxam-Gilbert (21) chemical cleavage and Sanger (32) deoxy chain termination were used to determine the sequence of both DNA strands from pCK210 (7). Chemical cleavage was used to determine the HindIII-BamHI sequence on both strands. The rest of the insert was sequenced by shotgun subcloning into M13mp8 (24) and by using [35S]dATP αS and dideoxynucleotides as described previously (5).

Sequences were analyzed by programs supplied by H. Martinez, University of California, San Francisco. Hydropathy was calculated by the algorithm of Kyte and Doolittle (15). Secondary structure was predicted by the algorithm of Chou and Fasman (6), using the updated data set (2).

Preparative techniques. Restriction enzymes were used as recommended by the supplier. DNA manipulations were generally carried out as described by Maniatis et al. (19). Minicells were purified and labeled as described previously (22), except that l-[35S]methionine, at 50 μCi/ml in Difco methionine assay medium, was used to label minicells at an optical density (650 nm) of 1. Protein gels were done by the methods of Laemmli (16) and Ames (1). Molecular weight calibrations for protein gels were performed with the Pharmacia calibration kit.

S1 nuclease mapping. RNA minilysates were prepared from mid-log-phase cells grown at 30°C by a modification of the lysozyme-lysis procedure of Davis et al. (8) in which the diethyl ether and RNase A treatments were omitted. RNA content was estimated spectrophotometrically, with the bulk of the minilysate nucleic acid assumed to be RNA. By standard methodology (19), the HindIII-BamHI probe from pCK210 was 5' labeled with T4 polynucleotide kinase, and strands were separated on a native 5% polyacrylamide gel. Strands were identified by Maxam-Gilbert sequencing.
The sequence of the flaA (cheC) locus. The sequence of the flaA (cheC) locus was determined by using both Maxam-Gilbert and Sanger dideoxy sequencing on both strands, as described in Materials and Methods and in the legend to Fig. 2. Two open reading frames span the insert, and the predicted amino acid sequences are shown. The site of transcriptional initiation, labeled as +1, was determined by S1 nuclease digestion as shown in Fig. 4. A portion of λ sequence from the original λ fla36Δ7 clone is also indicated.
FIG. 2. Sequencing strategy. A summary of sequence information and restriction mapping information is presented at the top of the figure. The thin line represents pBR322 sequence. The first 590 nucleotides of sequence was determined by chemical cleavage of both strands of DNA. The tail of an arrow represents the 32P-labeled end of a fragment, either 5' labeled or 3' labeled, as indicated in the figure. The length of the arrow shows the extent of sequence determination with that fragment. The rest of the DNA sequence was determined by dideoxy sequencing in both directions with M13mp8 subclones. Arrows in this region of the sequence indicate the direction and extent of sequence determination. Abbreviations: A, HainIII; B, BaxHI; E5, EcoRV; Hpi, HpaI; M, MluI; N, NraI; P1, PvuI; P2, PvuII; S, SmaI; X, XmaII.

Sequence of the flaA locus. Previously, Clegg and Koshland subcloned the Escherichia coli flaA (cheC) locus from a λ clone into pBR322, constructing pCK210 (7). The DNA sequence of this flaA subclone is shown in Fig. 1, and the sequencing strategy is shown in Fig. 2. Two open reading frames, flaAI and flaAII, span the region and encode polypeptides with predicted molecular masses of 17 and 38 kDa, respectively. Mutants in flaAI have yet to be identified, but there already exists evidence that flaAII corresponds to flaA and cheC. Previous deletion studies by Clegg and Koshland correlate FlaA with a 37-kDa protein that is truncated by a deletion spanning the PvuI sites of pCK210 (7). Since the flaAII gene should encode a 38-kDa polypeptide, and since it spans a PvuI site, flaAII corresponds to the gene identified by certain flaA and cheC mutants. Immediately following flaAI on pCK210 is a portion of λ sequence from the original lambda subclone, λflaAI (28), which confirms previous restriction mapping of λflaAI deletions (7).

The predicted polypeptides were analyzed for hydrophathy (Fig. 3). FlaAI had a strongly hydrophobic amino-terminal region and an overall hydrophathy that was higher than average (0.03 versus –0.26 for globular proteins), suggesting that it is membrane associated. Whether or not FlaAI inserts into the inner membrane has yet to be determined. FlaAII, on the other hand, exhibited an average overall hydrophathy. It had two strongly hydrophobic regions and only one strongly hydrophobic region.

Secondary structure predictions were performed by using the algorithm of Chou and Fasman. Two notable regions of FlaAII are predicted to α-helical and are included in Fig. 3. These regions are very amphiphilic, which strengthens the Chou-Fasman prediction of α-helicity for these two regions (11, 33). The amphiphilicity of the second helical region, Asp-248 through Leu-276, is particularly striking since most of the hydrophobic residues are arranged in a strip along one side of the helix and there are strong possibilities of salt bridges between ionized residues on adjacent turns of the helix. If this region is a single, continuous α-helix, the sequence Asp-248 through Leu-276 is predicted to produce a helix 4.2 nm long, possibly spanning the length of the protein. Assuming a molecular volume of 0.00123 mm3/dalton (20) and a spherical shape, FlaAII would have a diameter of 4.5 nm.

flaAI expression. Previous mapping of flaA mutants produced by Mu phage insertions identified flaA as the promoter-proximal gene of the operon (35). Our sequence suggests that the previously unknown flaAI is actually the first gene of the operon.

If flaAI is the first gene of its operon, transcription should start immediately upstream of the open reading frame. Since previous physical mapping suggested that the BamHI-HindIII fragment of pCK210 contains the promoter (7), this fragment was used to map the 5′ end of the transcript by protection from S1 nuclease digestion. Transcripts were isolated from bacteria transformed with pCK210. The protected fragment comigrated with a sequencing ladder fragment of 80 nucleotides in length (Fig. 4). Assuming the protected fragment is 80 nucleotides long, transcription starts at an adenine 31 nucleotides upstream of flaAI which is identified as +1 in Fig. 1. The region 10 nucleotides upstream of this adenine also provides the best match to the consensus sequence for RNA polymerase-binding sites (30). The −35 region showed a poorer match to the consensus sequence.

Since transcription is consistent with flaAI expression, we attempted to identify the 17-kDa FlaAI polypeptide. Because of the high background labeling in bands of this molecular mass range, Clegg and Koshland (7) could not clearly observe a polypeptide from flaAI in minicells transformed with pCK210. To eliminate such ambiguity, a truncated version of FlaAI was constructed by a 96-base-pair in-frame deletion fusing Ile-16 to Arg-49 (Fig. 5). Protein expression from both the wild-type flaAI gene and the flaAI deletion was examined in minicells produced by aberrant division of minB mutants (22). Expressed proteins were labeled with 35S methionine and resolved on polyacrylamide gels (Fig. 5). Both plasmids express the FlaAI polypeptide at 37 kDa, as reported previously (7). Wild-type flaAI expressed a polypeptide with an apparent molecular mass of 18.3 kDa rather than the predicted 17 kDa. Such discrepancies between predicted and apparent molecular masses are not uncommon and have been observed for other proteins (e.g., reference 10). The 18.3-kDa band was absent in minicells containing the flaAI deletion. Instead, a 15-kDa...
polypeptide was observed, clearly identifying the FlaAI polypeptide. The difference of 3.3kDa in apparent molecular weights is consistent with a loss of 32 amino acids (96-base-pair deletion).

**DISCUSSION**

Sequence determination of the flaA (cheC) locus revealed two open reading frames which span the locus. The second reading frame, flaAII, was identified previously by physical mapping as the coding sequence for the 37-kDa FlaA protein. A previously unknown gene, designated flaAI, encodes a 17-kDa polypeptide and precedes flaAII. The discovery of flaAI has also been made by J. Malakooti and P. Matsumura (personal communication).

Since flaAI was missed in previous genetic studies, it was important to show that flaAI was actually transcribed and expressed. S1 nuclease digestion identified the site of transcriptional initiation that is consistent with complete transcription of flaAI. Furthermore, the −10 region matches very well with the bacterial consensus promoter sequence (30), but the −35 region is not as clear a match. Protein gels of minicells containing flaAI plasmids showed a polypeptide with an apparent molecular weight of 18.3kDa that became 3.3kDa smaller when a 96-base-pair in-frame deletion was constructed in flaAI. This result unambiguously identifies the FlalAI polypeptide.

In light of the discovery of a new gene in the flaA locus, the designation of cheC mutations and their interpretations must be reexamined. Both smooth swimming and tumbling phenotypes in *E. coli* (3, 25, 28) and *S. typhimurium* (17, 36, 37), as well as an inverted response phenotype (11, 31), have been observed for cheC. Since these genetic studies were conducted before the existence of flaAI was known, there is some doubt now whether all of these phenotypes are due to lesions in a single gene. To date, there is no evidence that a gene corresponding to *E. coli* flaAI exists in *S. typhimurium*.

In the only work correlating FlaA protein with cheC, flaA locus clones were generated by homologous recombination (7) of Afla36 (13) with two *E. coli* flaA (cheC) mutants. The mutant loci were further subcloned into pBR322 and examined for protein expression on two-dimensional gels. The smooth swimming cheC497 allele from AW674 (3) is clearly associated with flaAII because the mutation alters the pl of its 37-kDa polypeptide. However, assigning the tumbling allele of cheC from AW675 (14) to flaAII is less clear cut, as no change in pl for the 37-kDa protein was observed in this strain compared with its parent. Until further studies are done on AW675, the possibility of a dominant tumbling mutation outside flaAII still remains.

Chou-Fasman analysis of the FlaAII sequence produces predictions of α-helicity at Val-47 through Arg-75 and Asp-248 through Leu-276. The plausibility of the prediction for the second region is greatly enhanced by both amphipathicity and possibilities of salt bridges. Recent genetic studies suggest a particular function for these two α-helical regions. Yamaguchi et al. have isolated several *S. typhimurium* flaQ (cheC) mutants and used deletions to map the sites of mutations (36). Of 22 deletion-defined segments,
FIG. 4. S1-Nuclease mapping the start of transcription. Probe is the 307 nucleotide BamHI-HindIII fragment of the promoter region of the flaA locus labeled at the 5' end of the BamHI site. RNA was isolated from RP437 (wild-type chemotaxis) transformed with pCK210. Lanes 1 and 2, Sequencing ladder generated by Maxam-Gilbert chemical cleavage (A+G and C+T, respectively) of the probe region labeled 3' at the BamHI site with the Klenow fragment of DNA polymerase. Note that the Maxam-Gilbert fragments are one nucleotide out of register from enzymatically produced fragments. Lane 3, S1 nuclease digestion of probe alone (no RNA added). Lane 4, Full-length probe (no S1 nuclease digestion). Lanes 5 through 7, RNA-protected fragments after increasing S1 nuclease digestion (P-L Biochemicals units) at 37°C for 30 min and 120, 200, and 580 U, respectively.

all 15 of the clockwise (tumbling bias) mutations map to two segments, 5 and 18. Assuming equal segment lengths (see reference 36) and a one-to-one correspondence between the E. coli flaAII and S. typhimurium flaQ genes, these two segments correspond exactly to portions of the two α-helical regions predicted in flaAII. Of the five counterclockwise (smooth-swimming) cheC mutants, only one mutation maps to a segment corresponding to the first α-helical region of flaAII. Clearly, the sequences of these S. typhimurium mutants will aid the exploration of this intriguing correlation between these predicted α-helical regions and clockwise rotation.

ACKNOWLEDGMENTS

We thank P. Matsumura for his helpful correspondence via computer mail, and J. S. Parkinson for bacterial strains. This work was supported by Public Health Service grants from the National Institutes of Health (D.E.K.) and a predoctoral fellowship from the National Science Foundation (S.C.K.).

LITERATURE CITED

1. Ames, G. F. L. 1974. Resolution of bacterial proteins by polyacrylamide gel electrophoresis on slabs. J. Biol. Chem. 249:634-644.
2. Argo, P., M. Haneil, and R. M. Garavito. 1978. The Chou-Fasman secondary structure prediction method with an ex-
1012 KUO AND KOSHLAND

3. Armstrong, J. B., and J. Adler. 1969. Complementation of nonchemotactic mutants of Escherichia coli. Genetics 61:61-66.
4. Berg, H. C., and D. A. Brown. 1972. Chemotaxis in Escherichia coli analysed by three-dimensional tracking. Nature (London) 239:500-504.
5. Biggin, M. D., T. J. Gibson, and G. F. Hong. 1983. Buffer gradient gels and 35S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA 80:3963-3965.
6. Chou, P. Y., and G. D. Fasman. 1973. Structural and functional role of leucine residues in proteins. J. Mol. Biol. 74:263-281.
7. 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.
8. Davis, R. W., M. Thomas, J. Cameron, T. P. St. John, S. Scherer, and R. A. Padgett. 1980. Rapid DNA isolations for enzymatic and hybridization analysis. Methods Enzymol. 65:404-411.
9. Gilman, M. Z., and M. J. Chamberlin. 1983. Developmental and genetic regulation of Bacillus subtilis genes transcribed by α-pro RNA polymerase. Cell 35:285-293.
10. Higgins, C. F., P. D. Haag, K. Nikaido, F. Ardeshr, G. Garcia, and G. F.-L. Ames. 1982. The histidine transport operon of Salmonella typhimurium: complete nucleotide sequence, four protein components. Nature (London) 298:723-727.
11. Kaiser, E. T., and F. J. Kendy. 1984. Amphillic secondary structure: design of peptide hormones. Science 223:249-255.
12. 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.
13. Komeda, T., K. Shimada, and T. Iino. 1977. Isolation of specialized lambda transducing bacteriophages for flagellar genes (fla) of Escherichia coli. K-12. J. Virol. 22:654-661.
14. Kondoh, H. 1982. Tumbling chemotaxis mutants of Escherichia coli: possible gene dependent effect of methionine starvation. J. Bacteriol. 142:527-534.
15. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157:105-132.
16. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680-685.
17. Macnab, R. M. 1984. Bacterial motility and the bacterial flagellar motor. Annu. Rev. Biophys. Bioeng. 13:51-83.
18. Macnab, R. M., and D. E. Koshland, Jr. 1972. Gradient sensing mechanism in bacterial chemotaxis. Proc. Natl. Acad. Sci. USA 69:2509-2512.
19. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. Matthews, B. W. 1968. Solvent content of protein crystals. J. Mol. Biol. 33:491-497.
21. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific cleavages. Methods Enzymol. 65:499-560.
22. Meagher, R. B., R. B. Tait, M. Betlach, and H. W. Boyer. 1977. Protein expression in Escherichia coli minicells by recombinant plasmids. Cell 10:521-536.
23. Meselson, M., and R. Yuan. 1968. DNA restriction endonuclease from Escherichia coli. Nature (London) 217:1110-1114.
24. Messing, J., R. Crea, and P. Seeburg. 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. 9:309-321.
25. Parkinson, J. S. 1976. cheA, cheB, and cheC genes of Escherichia coli and their role in chemotaxis. J. Bacteriol. 126:758-770.
26. Parkinson, J. S. 1978. Complementation analysis and deletion mapping of Escherichia coli mutants defective in chemotaxis. J. Bacteriol. 135:45–53.
27. Parkinson, J. S., and S. R. Parker. 1979. Interaction of the cheC and cheZ genes is required for chemotactic behavior in Escherichia coli. Proc. Natl. Acad. Sci. USA 76:2390-2394.
28. Parkinson, J. S., S. R. Parker, P. B. Talbert, and S. E. Houts. 1983. Interactions between chemotaxis genes and flagellar genes in Escherichia coli. J. Bacteriol. 155:265-274.
29. Ravid, S., and M. Eisenbach. 1984. Direction of flagellar rotation in bacterial cell envelopes. J. Bacteriol. 158:222-230.
30. Rosenberg, M., and D. Court. 1979. Regulatory sequences involved in the promoter and termination of RNA transcription. Ann. Rev. Genet. 13:319-353.
31. Rubik, B. A., and D. E. Koshland, Jr. 1978. Potentiation, desensitization and inversion of response in bacterial sensing of chemical stimuli. Proc. Natl. Acad. Sci. USA 75:2820-2824.
32. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-termination inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
33. Schiffer, M., and A. B. Edmundson. 1967. Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. Biophys. J. 7:121-135.
34. Silverman, M., and M. Simon. 1973. Genetic analysis of flagellar mutant in Escherichia coli. J. Bacteriol. 113:105-113.
35. Silverman, M., and M. Simon. 1973. Genetic analysis of bacteriophage Mu-induced flagellar mutant in Escherichia coli. J. Bacteriol. 116:114-122.
36. Yamaguchi, S., H. Fujita, A. Ishihara, S.-I. Aizawa, and R. M. Macnab. 1986. Subdivision of multifunctional flagellar genes of Salmonella typhimurium into regions responsible for assembly, rotation, and switching. J. Bacteriol. 166:187-193.
37. Warrick, H. M., B. L. Taylor, and D. E. Koshland, Jr. 1977. Chemotactic mechanism of Salmonella typhimurium: preliminary mapping and characterization of mutants. J. Bacteriol. 130:223-231.