JOURNAL OF BACTERIOLOGY, Jan. 1986, p. 161–166
0021-9193/86/010161-06$02.00/0

Vol. 165, No. 1

Nucleotide Sequence Corresponding to Five Chemotaxis Genes in *Escherichia coli*

NORIHIRO MUTOH AND MELVIN I. SIMON*

Division of Biology, California Institute of Technology, Pasadena, California 91125

Received 10 September 1985/Accepted 24 October 1985

The nucleotide sequence of DNA which contains five chemotaxis-related genes of *Escherichia coli*, *cheW*, *cheR*, *cheB*, *cheY*, and *cheZ*, and part of the *cheA* gene was determined. Molecular weights of the polypeptides encoded by these genes were calculated from translated amino acid sequences, and they were 18,100 for *cheW*, 32,700 for *cheR*, 37,500 for *cheB*, 14,100 for *cheY*, and 24,000 for *cheZ*. Nucleotide sequences which could act as ribosome-binding sites were found in the upstream region of each gene. After the termination codon of the *cheW* gene, a typical rho-independent transcription termination signal was observed. There are no other open reading frames long enough to encode polypeptides in this region except those which code for the two previously reported genes *tar* and *tap*.

Many of the genes that are required for bacterial chemotaxis have been identified, and their gene products have been characterized in both *Escherichia coli* and *Salmonella typhimurium* (1, 4, 10, 12, 17, 22). Two operons that are adjacent to each other on the bacterial genome encode 10 chemotaxis-related functions. The Mocha operon includes the *motA* and *motB* genes that are responsible for coupling flagella rotation to energy supplied by the electrochemical gradient across the cell membrane. Adjacent to these genes are the *cheA* and *cheW* genes that are required for chemotaxis (1, 12, 17). The second operon includes the *tar* and *tap* genes, which are responsible for the synthesis of transmembrane receptor proteins (2, 7), the *cheR* and *cheB* genes, which reversibly methylate the transmembrane proteins and are responsible for adaptation (5, 20, 23), and the *cheY* and *cheZ* genes, which are thought to play a central role in generating a signal that regulates bacterial flagella rotation (1, 3, 12, 15, 17). In the past 10 years, we have learned how to measure and analyze components of the chemotaxis system in a variety of sophisticated ways; however, we still do not understand the basis for the signal transduction process, i.e., how the binding of a specific attractant molecule to a receptor generates a signal that regulates flagellar rotation. One approach to this problem involves the manipulation of the levels of the chemotaxis gene products and the isolation and purification of the gene products to study their biochemical properties. This approach would be greatly facilitated by the availability of the nucleic acid sequence of the genes responsible for chemotaxis. The *tar* and *tap* gene sequences have been published (8). In this paper, we report the DNA sequence and the derived amino sequences for the *cheW*, *cheR*, *cheB*, *cheY*, and *cheZ* genes and for part of the *cheA* gene.

MATERIALS AND METHODS

Enzymes and chemicals. All restriction enzymes used in this study were purchased from Bethesda Research Laboratories, Inc., and New England BioLabs, Inc. T4 DNA ligase was obtained from Bethesda Research Laboratories. M13-mp7 replicative-form DNA was from Bethesda Research Laboratories, and M13-mp8 and M13-mp9 were purchased from P-L Biochemicals, Inc. Dideoxy nucleotide triphosphates and primers were purchased from Bethesda Research Laboratories. [α-^{32}P]dATP (3,000 Ci/mmol) was from Amersham Corp. DNA fragments were obtained by gel electrophoresis after digestion with restriction enzymes. Agarose gel (0.8%) was used to separate fragments more than 600-base-pairs long, and 5% polyacrylamide gels were used for fragments that were smaller than 600 base pairs. DNA sequencing was done by the procedures described by Heidecker et al. (7).

DNA sequencing. A 9.6-kilobase *Eco*RI fragment of pAK108 (2) was purified by gel electrophoresis. This fragment was cut with a variety of restriction enzymes, and the resulting fragments were cloned into M13-mp7, -mp8, or -mp9 phage. The hybrid phages were grown on *Escherichia coli* JM103 in L broth (1% tryptone [Difco Laboratories], 0.5% yeast extract, 0.5% sodium chloride). Phage DNA was purified by phenol extraction and ethanol precipitation. Sequencing was done with the purified phage DNA templates and the 26-base pair primer.

RESULTS

The plasmid pAK108 (2) carries an *Eco*RI fragment of 9.6 kilobases. This DNA includes the sequences corresponding to half of the *cheA* gene and the complete *cheW* gene and the chemotaxis operon which includes the two genes encoding the chemosensory transducers, *tar* and *tap*, and four other chemotaxis-related genes, *cheR*, *cheB*, *cheY*, and *cheZ*. Figure 1 shows the strategy used to determine the nucleotide sequence. Appropriate restriction fragments were cloned into the M13 vectors, and initial sequencing was done at random. Overlapping sequences were matched by scanning the sequences and compiling them with a computer program. All of the sequencing was done so that there were at least two separate determinations for each segment, and the ends of each sequence were overlapped by a sequence determined for a separate fragment. The numbering scheme chosen was essentially arbitrary, with number one corresponding to the middle of the *Eco*RI site that defined the start of the fragment carrying the *cheA* gene (Fig. 1A and 2). Nucleotide 1198 of this segment corresponded to the first nucleotide of the sequence that Krikos et al. (8) previously published for the *tar* gene. In the same way, the sequence of the second segment that encoded the *cheR*, *cheB*, *cheY*, and *cheZ* genes

* Corresponding author.

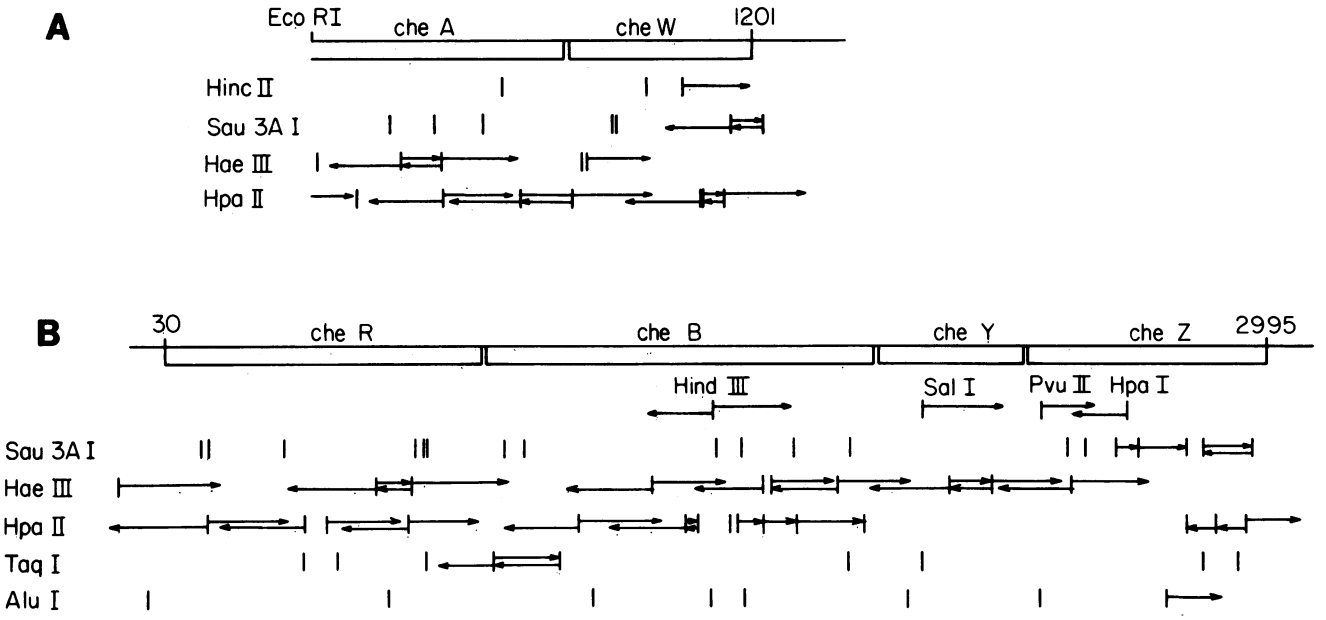

FIG. 1. Strategy for DNA sequencing. (A) Schematic drawing of the fragment extending from the *Eco*RI site in the middle of the *cheA* gene to the end of the *cheW* gene. The *cheW* gene ends at nucleotide 1201, with nucleotide number 1 corresponding to the G residue in the *Eco*RI site. The restriction fragments that were sequenced were derived by digestion with the enzymes shown on the right side. The arrows indicate the direction and the extent of sequencing. (B) Map of the region downstream from the *tap* gene. The numbering is arbitrary, with nucleotide number 30 corresponding to the A in the first codon of the *cheR* gene. The vertical lines indicate the position of specific restriction sites, and the horizontal arrows indicate the extent of sequencing of fragments obtained with specific restriction enzymes.

was numbered from 1 to 3,063 (Fig. 1B, 3, 4). Nucleotide 1 of this sequence corresponded to nucleotide 3450 of the *tap* sequence that was published previously (8). This data therefore links up the sequence previously published (8) and provides continuous sequence information covering seven chemotaxis genes and 7.7 kilobase pairs of DNA.

Figure 2 shows the last part of the sequence of the *cheA* gene and the complete sequence of the *cheW* gene. The *cheW* gene is the last gene in the Mocha operon, which comprises the *motA*, *motB*, *cheA*, and *cheW* genes. Downstream of this operon is the next operon in the chemotaxis series, which includes the *tar*, *tap*, *cheR*, *cheB*, *cheY*, and *cheZ* genes. In screening the sequence corresponding to *cheW*, we found a single open reading frame including nucleotides 701 to 1201. This open reading frame specified a polypeptide of 167 amino acid residues with a calculated molecular weight of 18,000. This value is only slightly larger than the molecular weight of the *cheW* gene product estimated from sodium dodecyl sulfate-polyacrylamide gel electrophoresis of the labeled *cheW* gene product (10, 18). The ATG start codon was selected because it gave the longest open reading frame and because 11 base pairs upstream from this ATG codon there is a sequence, AAGG, which could act as part of the ribosome-binding site for the translation of the *cheW* gene product. At the end of the *cheW* open reading frame, there is an inverted repeat sequence followed by a cluster of T residues. This is typical of a rho-independent transcription termination signal (14) and could represent the signal for the termination of the Mocha operon transcript. Preceding the open reading frame that represents the *cheW* sequence, there is another long, contiguous, uninterrupted open reading frame which extends to the end of the fragment. Figure 2 shows the translated product of this open reading frame; it corresponds to the C-terminal end of the *cheA* gene product.

Figure 3 shows the sequence starting at the end of the *tap* gene. The nucleotides TGA at positions 18, 19, and 20 correspond to the termination codon of the *tap* gene. There is a long open reading frame that starts with an ATG codon at nucleotide 30 and terminates with a TAA codon at nucleotide 890. This open reading frame is preceded by sequence AAGG, which could be part of the ribosome-binding site (16). The polypeptide encoded by this open reading frame has a calculated molecular weight of 32,700 and consists of 286 amino acids. The molecular weight of the *cheR* gene product has been reported to be 28,000 (10, 18) and is very similar to the molecular weight derived from the product encoded by this open reading frame. There is no other long, extensive open reading frame that we could discern that covers this region, and therefore, we have assigned this open reading frame to the *cheR* gene.

The next long open reading frame begins at an ATG codon at nucleotide 893 and ends with a TAA codon at nucleotide 1942. This open reading frame specifies 349 amino acid residues encoding a polypeptide with a molecular weight of 37,500, which agrees with the 38,000 molecular weight reported for the *cheB* gene product. Interestingly, this open reading frame is preceded by a sequence, AAGGA, which could act as part of the ribosome-binding site (16). However, this sequence is also a portion of the coding region of the *cheR* gene. Thus, there is an apparent overlap between the sequences that encode the *cheR* gene and sequences that might act as the ribosome-binding site for the beginning of translation of the large open reading frame corresponding to the *cheB* gene. Further evidence to support the contention that this open reading frame corresponds to the *cheB* gene comes from the finding of an *Sst*II restriction endonuclease recognition site (CCGCGG, nucleotide 1070 to 1075) which was reported to exist in the *cheB* gene by Slocum and Parkinson (19). In previous labeling experiments in which

```
1    GA ATT CTG GCA AAA GCG GCC TCG CAA GGT TTG ACT GTC AGC GAA AAC ATG AGC GAC GAC GAA GTC GCG ATG CTG ATA TTT    80
        Ile Leu Ala Lys Ala Ala Ser Gln Gly Leu Thr Val Ser Glu Asn Met Ser Asp Asp Glu Val Ala Met Leu Ile Phe

81   GCA CCT GGC TTC TCC ACG GCA GAG CAG GTC ACC GAC GTC TCC GGG CGC GGC GTC GGC ATG GAC GTC GTT AAA CGT AAT ATC   161
     Ala Pro Gly Phe Ser Thr Ala Glu Gln Val Thr Asp Val Ser Gly Arg Gly Val Gly Met Asp Val Val Lys Arg Asn Ile

162  CAG AAG ATG GGC GGT CAT GTC GAA ATC CAG TCG AAG CAG GGT ACT GGC ACT ACG ATC CGC ATT TTA CTG CCG CTG ACG CTG   242
     Gln Lys Met Gly Gly His Val Glu Ile Gln Ser Lys Gln Gly Thr Gly Thr Thr Ile Arg Ile Leu Leu Pro Leu Thr Leu

243  GCC ATC CTC GAC GGC ATG TCC GTA CGC GTT GCG GAT GAA GTT TTC ATT CTG CCG CTG AAT GCT GTT ATG GAA TCA CTG CAA   323
     Ala Ile Leu Asp Gly Met Ser Val Arg Val Ala Asp Glu Val Phe Ile Leu Pro Leu Asn Ala Val Met Glu Ser Leu Gln

324  CCC CGT GAA GCC GAT CTC CAT CCA CTG GCC GGC GCG AGC GGG TGC TGG AAG TGC GGG GTG AAT ATC TGC CCA TCG TCG AAC   404
     Pro Arg Glu Ala Asp Leu His Pro Leu Ala Gly Ala Ser Gly Cys Trp Lys Cys Gly Val Asn Ile Cys Pro Ser Ser Asn

405  TGT GGA AAG GTG TTC AAC GTC GCG GGC GCG AAA ACC GAA GCC ACC CAG GGA ATT GTG GTG ATC TTA CAA ACT GGC GGT CGC   485
     Cys Gly Lys Val Phe Asn Val Ala Gly Ala Lys Thr Glu Ala Thr Gln Gly Ile Val Val Ile Leu Gln Ser Gly Gly Arg

486  CGC TAC GCC TTG CTG GTG GAT CAA TTA ATT GGT CAA CAC CAG GTT GCG GTT AAA AAC CTT GAA AGT AAC TAT CGC AAA GTC   566
     Arg Tyr Ala Leu Leu Val Asp Gln Leu Ile Gly Gln His Gln Val Ala Val Lys Asn Leu Glu Ser Asn Tyr Arg Lys Val

567  CCC GGC ATT TCT GCT GCG ACC ATT CTT GGC GAC GGC AGC GTG GCA CTG ATT GTT GAT GTC TCC GCC TTG CAG GCG ATA AAC   647
     Pro Gly Ile Ser Ala Ala Thr Ile Leu Gly Asp Gly Ser Val Ala Leu Ile Val Asp Val Ser Ala Leu Gln Ala Ile Asn

648  CGC GAA CAA CGT ATG GCG AAC ACC GCC GCC TGA ATGAGTAAAA AGGTAACAAT ATG ACC GGT ATC ACG AAT GTA ACA AAG CTG    730
     Arg Glu Gln Arg Met Ala Asn Thr Ala Ala                          Met Thr Gly Met Thr Asn Val Thr Lys Leu

731  GCC AGC GAG CCG TCA GGC CAG GAA TTT CTG GTA TTT ACC CTT GGT GAT GAA GAG TAC GGT ATT GAT ATC CTG AAA GTC CAG   811
     Ala Ser Glu Pro Ser Gly Gln Glu Phe Leu Val Phe Thr Leu Gly Asp Glu Glu Tyr Gly Ile Asp Ile Leu Lys Val Gln

812  GAG ATC CGT GGC TAC GAT CAG GTA ACA CGC ATT GCG AAC ACG CCA GCG TTT ATC AAA GGC GTC ACG AAT CTG CGC GGC GTT   892
     Glu Ile Arg Gly Tyr Asp Gln Val Thr Arg Ile Ala Asn Thr Pro Ala Phe Ile Lys Gly Val Thr Asn Leu Arg Gly Val

893  ATT GTG CCG ATT GTT GAC TTA CGA ATT AAG TTC AGC CAG GTG GAT GTG GAC TAT AAC GAC AAC ACG GTA GTT ATC GTC CTG   973
     Ile Val Pro Ile Val Asp Leu Arg Ile Lys Phe Ser Gln Val Asp Val Asp Tyr Asn Asp Asn Thr Val Val Ile Val Leu

974  AAT CTC GGA CAG CGG GTG GTC GGC ATC GTG GTT GAC GGC GTC TCA GAC GTG CTT TCA TTG ACG GCC GAG CAA ATT CGT CCG  1054
     Asn Leu Gly Gln Arg Val Val Gly Ile Val Val Asp Gly Val Ser Asp Val Leu Ser Leu Thr Ala Glu Gln Ile Arg Pro

1055 GCA CCG GAA TTT GCC GTG ACG CTT TCA ACA GAA TAT CTC ACT GGA CTG GGC GCA CTG GGC GAC CGG ATG TTG ATT CTG GTG  1135
     Ala Pro Glu Phe Ala Val Thr Leu Ser Thr Glu Tyr Leu Thr Gly Leu Gly Ala Leu Gly Asp Arg Met Leu Ile Leu Val

1136 AAC ATC GAA AAA CTG CTG AAC AGC GAA GAG ATG GCG CTG TTA GAT AGC GCG GCG TCA GAA GTG GCG TAA TTCTCC CGATTTCCTC  1220
     Asn Ile Glu Lys Leu Leu Asn Ser Glu Glu Met Ala Leu Leu Asp Ser Ala Ala Ser Glu Val Ala

1221 AATTGAAATG AACCCGATGA TCTGCGCATC GGGTTTTTTA TTTCAATTTC GCGGCGGGTG GCATCAGCAA TAAAGTTTCC CCCCTCCTTG  1310

1311 CCGATAACGA GATCAACTTG TTTTCAGGAA GCTGCCTTAT GATTAACCGT  1360
```

FIG. 2. Nucleotide sequence of the *cheA* fragment and the *cheW* gene. The sequences are shown, with the reading frame indicated by the spacing of the codons. The corresponding amino acids are listed below each codon. The underlined sequences represent the cluster of thymidylate residues that may mark the termination of the *mocha* transcription unit.

the *cheR* and *cheB* gene products were synthesized and labeled with [^{35}S]methionine, the *cheR* gene product always appeared to be synthesized at lower levels than the *cheB* gene product (18). One reason for this observation may come from the finding that the open reading frame corresponding to the *cheB* gene product included approximately 20 methionine residues, while the open reading frame for the *cheR* gene product only included 6 methionine residues. Thus, the specific activity of the two gene products could vary by more than a factor of three even if they were synthesized at exactly the same rates.

Figure 4 shows the sequences corresponding to the *cheY* and *cheZ* genes. The product of the *cheY* gene was reported to be a small polypeptide of molecular weight 8,000 (10, 18). The open reading frame corresponding to this polypeptide was found to start at nucleotide 1957 and extend to nucleotide 2343. This reading frame encoded a polypeptide comprising 129 amino acid residues with a molecular weight of 14,100. The calculated molecular weight was slightly higher than that estimated from the sodium dodecyl sulfate-polyacrylamide gel electrophoresis data. The putative ribosome-binding sequence was AGGAG, corresponding to nucleotides 1946 to 1950. Slocum and Parkinson (19) indicated the presence of an *Sal*I restriction endonuclease recognition site in the *cheY* gene. A *Sal*I recognition site was found at nucleotides 2074 to 2079 in this opening reading frame. The *cheB* gene ended at nucleotide 1943 and, thus, there was a space of 15 base pairs between the translated regions of the two genes. This space included a sequence that could serve as a ribosome-binding site. Matsumura et al. (9) reported the nucleotide sequence of the *cheY* gene. The sequence shown here is identical to the one that they reported. After the *cheY* gene, there was another long open reading frame, which we assigned to the *cheZ* gene. The *cheY* gene ended with a termination nucleotide TGA, corresponding to nucleotide 2347. There was then a space of 11 nucleotides before the next ATG, which started the long open reading frame extending from nucleotide 2359 to 2999. This could encode a polypeptide corresponding to 214 amino acid residues with a molecular weight of 24,000. This molecular weight corresponded well to that of the *cheZ* gene product estimated by polyacrylamide gel electrophoresis. There was a *Pvu*II restriction endonuclease recognition site 38 to 43 nucleotides downstream of the ATG start codon, and this again was consistent with the report of Slocum and Parkinson (19) that a *Pvu*II recognition site is found in the early part of the *cheZ* gene. A ribosome-binding site could correspond to the sequence AGGA which is 11 base pairs from the start of the *cheZ* gene. There would be a 1-base overlap between the termination codon of the open reading frame corresponding to *cheY* and the putative ribosome-binding site for the *cheZ* gene product. We explored the sequence further down-

```
     che R
1    TGGTATCCTG AAGTGATTGA GAAGCGGCT ATG ACT TCA TCT CTG CCC TGT GGG CAA ACG TCT TTA TTG TTA CAG ATG ACC GAG CGC    86
                                     Met Thr Ser Ser Leu Pro Cys Gly Gln Thr Ser Leu Leu Leu Gln Met Thr Glu Arg

87   CTG GCG CTT TCC GAC GCG CAT TTT CGG CGG ATA AGT CAA TTG ATC TAT CAA CGA GCC GGG ATC GTT CTG GCT GAC CAT AAA   167
     Leu Ala Leu Ser Asp Ala His Phe Arg Arg Ile Ser Gln Leu Ile Tyr Gln Arg Ala Gly Ile Val Leu Ala Asp His Lys

168  CGC GAC ATG GTT TAC AAC CGA CTG GTT CGT CGT TTG CGT TCG CTG GGA CTG ACG GAT TTC GGT CAT TAT CTG AAC TTG CTG   248
     Arg Asp Met Val Tyr Asn Arg Leu Val Arg Arg Leu Arg Ser Leu Gly Leu Thr Asp Phe Gly His Tyr Leu Asn Leu Leu

249  GAA TCT AAT CAG CAC AGC GGT GAG TGG CAG GCG TTT ATC AAT TCG CTG ACC ACG AAT CTG ACG GCA TTT TTC CGT GAG GCA   329
     Glu Ser Asn Gln His Ser Gly Glu Trp Gln Ala Phe Ile Asn Ser Leu Thr Thr Asn Leu Thr Ala Phe Phe Arg Glu Ala

330  CAT CAT TTC CCT CTG CTC GCG GAT CAC GCA CGT CGC GGT TCT GGC GAG TAT CGC GTA TGG AGC GCG GCG GCT TCG ACC GGC   410
     His His Phe Pro Leu Leu Ala Asp His Ala Arg Arg Gly Ser Gly Glu Tyr Arg Val Trp Ser Ala Ala Ala Ser Thr Gly

411  GAA GAG CCG TAC AGC ATT GCG ATG ACG CTG GCT GAC ACA TTG GGC ACC GCG CCC GGA CGC TGG AAA GTG TTT GCC AGT GAT   491
     Glu Glu Pro Tyr Ser Ile Ala Met Thr Leu Ala Asp Thr Leu Gly Thr Ala Pro Gly Arg Trp Lys Val Phe Ala Ser Asp

492  ATC GAC ACC GAA GTG CTG GAA AAA GCC AGA AGC GGT ATC TAT CGC CAT GAA GAG TTG AAA AAC CTG ACG CCG CAG CAA CTG   572
     Ile Asp Thr Glu Val Leu Glu Lys Ala Arg Ser Gly Ile Tyr Arg His Glu Glu Leu Lys Asn Leu Thr Pro Gln Gln Leu

573  CAA CGG TAT TTC ATG CGA GGG ACG GGG CCG CAT GAA GGG CTG GTA CGC GTG CGT CAG GAG CTG GCG AAC TAT GTT GAT TTT   653
     Gln Arg Tyr Phe Met Arg Gly Thr Gly Pro His Glu Gly Leu Val Arg Val Arg Gln Glu Leu Ala Asn Tyr Val Asp Phe

654  GCC CCG CTG AAT CTA CTG GCG AAA CAG TAC ACC GTG CCG GGG CCG TTT GAT GCG ATC TTC TGT CGT AAC GTC ATG ATC TAC   734
     Ala Pro Leu Asn Leu Leu Ala Lys Gln Tyr Thr Val Pro Gly Pro Phe Asp Ala Ile Phe Cys Arg Asn Val Met Ile Tyr

735  TTC GAT CAA ACT ACC CAG CAG GAG ATT TTG CGC CGC TTT GTT CCG CTC CTT AAA CCC GAC GGA TTG CTG TTT GCG GGT CAC   815
     Phe Asp Gln Thr Thr Gln Gln Glu Ile Leu Arg Arg Phe Val Pro Leu Leu Lys Pro Asp Gly Leu Leu Phe Ala Gly His

816  TCT GAA AAC TTT AGC CAC CTT GAG CGC CGC TTC ACG CTG CGT GGT CAG ACG GTG TAT GCG CTA AGT AAG GAT TAA   890
     Ser Glu Asn Phe Ser His Leu Glu Arg Arg Phe Thr Leu Arg Gly Gln Thr Val Tyr Ala Leu Ser Lys Asp

891  CGATGAGCAA AATCAGGGTG TTATCTGTCG   920


     che B
865  CGGTGTATG CGCTAAGTAA GGATTAA CG ATG AGC AAA ATC AGG GTG TTA TCT GTC GAT GAT TCG GCA CTG ATG CGC CAG ATC ATG   949
                                     Met Ser Lys Ile Arg Val Leu Ser Val Asp Asp Ser Ala Leu Met Arg Gln Ile Met

950  ACA GAA ATC ATC AAC AGC CAT AGC GAC ATG GAA ATG GTG GCG ACC GCG CCT GAT CCG CTG GTC GCG CGT GAC TTG ATT AAG  1030
     Thr Glu Ile Ile Asn Ser His Ser Asp Met Glu Met Val Ala Thr Ala Pro Asp Pro Leu Val Ala Arg Asp Leu Ile Lys

1031 AAA TTC AAT CCC GAT GTG CTG ACG CTG GAT GTT GAA ATG CCG CGG ATG GAC GGA CTG GAT TTC CTC GAA AAA TTA ATG CGT  1111
     Lys Phe Asn Pro Asp Val Leu Thr Leu Asp Val Glu Met Pro Arg Met Asp Gly Leu Asp Phe Leu Glu Lys Leu Met Arg

1112 TTG CGT CCA ATG CCC GTT GTG ATG GTT TCT TCC CTG ACC GGC AAA GGG TCA GAA GTC ACG CTG CGC CCG CTG GAC CTG GGG  1192
     Leu Arg Pro Met Pro Val Val Met Val Ser Ser Leu Thr Gly Lys Gly Ser Glu Val Thr Leu Arg Ala Leu Glu Leu Gly

1193 GCG ATA GAT TTT GTC ACC AAA CCG CAA CTG GGT ATT CGC GAA GGT ATG CTG GCG TAT AAC GAA ATG ATT GCT GAA AAG GTG  1273
     Ala Ile Asp Phe Val Thr Lys Pro Gln Leu Gly Ile Arg Glu Gly Met Leu Ala Tyr Asn Glu Met Ile Ala Glu Lys Val

1274 CGT ACG GCA GCA AAG GCG AGC CTT GCA GCA CAT AAG CCA TTG TCG GCA CCG ACA ACG CTG AAG GCC GGC CCG TTC TTG AGT  1354
     Arg Thr Ala Ala Lys Ala Ser Leu Ala Ala His Lys Pro Leu Ser Ala Pro Thr Thr Leu Lys Ala Gly Pro Leu Leu Ser

1355 TCT GAA AAA CTG ATT GCG ATT GGT GCT TCA ACG GGT GGA ACT GAG GCA ATT CGT CAC GTA CTG CAA CCG TTG CCG CTT TCC  1435
     Ser Glu Lys Leu Ile Ala Ile Gly Ala Ser Thr Gly Gly Thr Glu Ala Ile Arg His Val Leu Gln Pro Leu Pro Leu Ser

1436 AGC CCG GCA CTG TTA ATT ACC CAG CAT ATG CCG CCC GGT TTC ACC CGC TCT TTT GCC GAC AGA CTT AAT AAG CTT TGC CAG  1516
     Ser Pro Ala Leu Leu Ile Thr Gln His Met Pro Pro Gly Phe Thr Arg Ser Phe Ala Asp Arg Leu Asn Lys Leu Cys Gln

1517 ATC GGG GTT AAA GAA GCC GAA GAC GGA GAA CGT GTC TTG CCG GGG CAT GCC TAT ATT GCG CCG GGC GAT CGG CAT ATG GAG  1597
     Ile Gly Val Lys Glu Ala Glu Asp Gly Glu Arg Val Leu Pro Gly His Ala Tyr Ile Ala Pro Gly Asp Arg His Met Glu

1598 CTG TCG CGT AGT GGC GCA AAT TAC CAA ATC AAA ATT CAC GAT GGC CCG GCG GTT AAC CGT CAT CGG CCT TCG GTA GAT GTG  1678
     Leu Ser Arg Ser Gly Ala Asn Tyr Gln Ile Lys Ile His Asp Gly Pro Ala Val Asn Arg His Arg Pro Ser Val Asp Val

1679 TTG TTC CAT TCT GTC GCC AAA CAG GCG GGG CGT AAT GCG GTT CGG GTG ATC CTG ACC GGT ATG GGC AAC GAC GGC GCG GCG  1759
     Leu Phe His Ser Val Ala Lys Gln Ala Gly Arg Asn Ala Val Gly Val Ile Leu Thr Gly Met Gly Asn Asp Gly Ala Ala

1760 GGA ATG TTG GCG ATG CGT CAG GCG GGG GCA TGG ACC CTT GCG CAA AAC GAA GCA AGT TGC GTG GTG TTC GGC ATG CCG CGC  1840
     Gly Met Leu Ala Met Arg Gln Ala Gly Ala Trp Thr Leu Ala Gln Asn Glu Ala Ser Cys Val Val Phe Gly Met Pro Arg

1841 GAG GCC ATC AAT ATG GGT GGT GTC TGC GAA GTG GTC GAT CTT AGC CAG GTA AGC CAG CAA ATG TTG GCA AAA ATT AGT GCC  1921
     Glu Ala Ile Asn Met Gly Gly Val Cys Glu Val Val Asp Leu Ser Gln Val Ser Gln Gln Met Leu Ala Lys Ile Ser Ala

1922 GGA CAG GCG ATA CGT ATT TAA ATCAGGAG TGTGAAATGG CGGATAAAGA   1970
     Gly Gln Ala Ile Arg Ile
```

FIG. 3. Nucleotide sequence of the *cheR* and *cheB* genes. The TAA codon that corresponds to the termination of the *cheB* gene is underlined to show that the ribosome-binding site for the *cheB* gene overlaps with the translated sequence of the *cheR* gene.

stream from the *cheZ* open reading frame and did not discover a long stretch of T residues or an inverted repeat sequence that could correspond to transcription signals similar to those found at the end of the Mocha operon. On the other hand, we did not find any other long, extensive open reading frames. We suggest that the *cheZ* gene may represent that last gene product in the cotranscribed unit corresponding to the chemotaxis genes.

```
     cheY
1928 GCGATACGTA TTTAAATCAG GAGTGTGAA ATG GCG GAT AAA GAA CTT AAA TTT TTG GTT GTG GAT GAC TTT TCC ACC ATG CGA CGC  2013
                                     Met Ala Asp Lys Glu Leu Lys Phe Leu Val Val Asp Asp Phe Ser Thr Met Arg Arg

2014 ATA GTG CGT AAC CTG CTG AAA GAG CTG GGA TTC AAT AAT GTT GAG GAA GCG GAA GAT GGC GTC GAC GCT CTC AAT AAG TTG  2094
     Ile Val Arg Asn Leu Leu Lys Glu Leu Gly Phe Asn Asn Val Glu Glu Ala Glu Asp Gly Val Asp Ala Leu Asn Lys Leu

2095 CAG GCA GGC GGT TAT GGA TTT GTT ATC TCC GAC TGG AAC ATG CCC AAT ATG GAT GGC CTG GAA TTG CTG AAA ACA ATT CGT  2175
     Gln Ala Gly Gly Tyr Gly Phe Val Ile Ser Asp Trp Asn Met Pro Asn Met Asp Gly Leu Glu Leu Leu Lys Thr Ile Arg

2176 GCG GAT GGC GCG ATG TCG GCA TTG CCA GTG TTA ATG GTG ACT GCA GAA GCG AAG AAA GAG AAC ATC ATT GCT GCG GCG CAA  2256
     Ala Asp Gly Ala Met Ser Ala Leu Pro Val Leu Met Val Thr Ala Glu Ala Lys Lys Glu Asn Ile Ile Ala Ala Ala Gln

2257 GCG GGG GCC AGT GGC TAT GTG GTG AAG CCA TTT ACC CCC GCG ACG CTG GAG GAA AAA CTC AAC AAA ATC TTT GAG AAA CTG  2337
     Ala Gly Ala Ser Gly Tyr Val Val Lys Pro Phe Thr Ala Ala Thr Leu Glu Glu Lys Leu Asn Lys Ile Phe Glu Lys Leu

2338 GGC ATG TGA GGAT GCGACTATGA TGCAACCATC AATCAAACCT  2380
     Gly Met

     cheZ
2328 TGAGAAACTG GGCATGTGAG GATGCGACT ATG ATG CAA CCA TCA ATC AAA CCT GCT GAC GAG CAT TCA GCT GGC GAT ATC ATT GCG  2413
                                     Met Met Gln Pro Ser Ile Lys Pro Ala Asp Glu His Ser Ala Gly Asp Ile Ile Ala

2414 CGC ATC GGC AGC CTG ACG CGT ATG CTG CGC GAC AGT TTG CGG GAA CTG GGG CTG GAT CAG GCC ATT GCC GAA GCG GCG GAA  2494
     Arg Ile Gly Ser Leu Thr Arg Met Leu Arg Asp Ser Leu Arg Glu Leu Gly Leu Asp Gln Ala Ile Ala Glu Ala Ala Glu

2495 GCC ATC CCC GAT GCG CGC GAT CGT TTG TAC TAT GTT GTG CAG ATC ACC GCC CAG GCT GCG GAG CGG GCG CTG AAC AGT GTT  2575
     Ala Ile Pro Asp Ala Arg Asp Arg Leu Tyr Tyr Val Val Gln Met Thr Ala Gln Ala Ala Glu Arg Ala Leu Asn Ser Val

2576 GAG GCG TCA CAA CCG CAT CAG GAT CAA ATG GAG AAA TCA GCA AAA GCG TTA ACC CAA CGT TGG GAT GAC TGG TTT GCC GAT  2656
     Glu Ala Ser Gln Pro His Gln Asp Gln Met Glu Lys Ser Ala Lys Ala Leu Thr Gln Arg Trp Asp Asp Trp Phe Ala Asp

2657 CCG ATT GAC CTT GCC GAC GCC CGT GAA CTG GTA ACA GAT ACA CGA CAA TTT CTG CCA GAT GTA CCC GCG CAT ACC AGC TTT  2737
     Pro Ile Asp Leu Ala Asp Ala Arg Glu Leu Val Thr Asp Thr Arg Gln Phe Leu Ala Asp Val Pro Ala His Thr Ser Phe

2738 ACT AAC GCG CAA CTG CTG GAA ATC ATG ATG GCG CAG GAT TTT CAG GAT CTC ACC GGC CAG GTC ATT AAG CGG ATG ATG GAT  2818
     Thr Asn Ala Gln Leu Leu Glu Ile Met Met Ala Gln Asp Phe Gln Asp Leu Thr Gly Gln Val Ile Lys Arg Met Met Asp

2819 GTC ATT CAG GAG ATC GAA CGC CAG TTG CTG ATG GTG CTG TTG GAA AAC ATC CCG GAA CAG GAG TCG CGT CCA AAA CGT GAA  2899
     Val Ile Gln Glu Ile Glu Arg Gln Leu Leu Met Val Leu Leu Glu Asn Ile Pro Glu Gln Glu Ser Arg Pro Lys Arg Glu

2900 AAC CAG AGT TTG CTT AAT GGA CCT CAG GTC GAT ACC AGC AAA GCC GGT GTG GTA GCC AGT CAG GAT CAG GTG GAC GAT TTG  2980
     Asn Gln Ser Leu Leu Asn Gly Pro Gln Val Asp Thr Ser Lys Ala Gly Val Val Ala Ser Gln Asp Gln Val Asp Asp Leu

2981 TTG GAT AGT CTT GGA TTT TGA TTTGTATTG CCTGATGTGG CGTGACCACG TCATATCAGG CGTTCTGATA AGGCGATGAC GCC  3063
     Leu Asp Ser Leu Gly Phe
```

FIG. 4. Nucleotide sequence of the *cheY* and *cheZ* genes. In the lower part of the figure, the TGA sequence corresponding to the termination of the *cheY* gene is underlined to indicate the relationship and spacing between the *cheY* and *cheZ* genes.

DISCUSSION

One of the best-characterized sensory transducing systems is the one involved in bacterial chemotaxis. Thus far, many of the experiments dealing with bacterial chemotaxis have focused on describing its components. However, the system is ideally suited to both genetic and biochemical manipulation. Experiments can be designed to modify genes and gene products and to study their interaction both in the whole organisms and in the test tube. The final stage in the characterization of the chemotaxis system is the description of the nucleotide sequence of the genes that encode the proteins involved in generating intracellular signals. The DNA sequence allows for precise gene manipulation, the preparation of overproducers, and site-specific mutagenesis. In this paper, we have described the DNA sequence of five genes that are centrally involved in regulating signal transduction in *E. coli* chemotaxis.

All of the sequences presented here have been tested for homology with an extensive collection of known protein sequences in the Bionet collection and by R. Doolittle of the University of California, San Diego. Thus far, no significant homology has been found between the *cheW*, *cheR*, *cheY*, or *cheZ* sequences and the other sequences in the libraries. However, the C-terminal fragment of the *cheA* gene did show a significant match with the C-terminal portion of the sequence corresponding to the *envZ* gene (11). When the stretch of amino acids between Phe 335 to Thr 391 of the *envZ* sequence was aligned with amino acids Phe 30 to Met 86 of the *cheA* C-terminal sequence shown in Fig. 2, without introducing any gaps, 15 amino acid identities were found. While this is not sufficient homology to suggest a direct relationship between the function of the *envZ* gene product and the product of the *cheA* gene, it is sufficient to suggest that these proteins may be ancestrally related. *envZ* is a component of a pathway of information processing. It is thought to be in part responsible for sensing changes in osmolarity and transducing these changes into changes in levels of different outer membrane proteins in *E. coli* (11). It is thus possible that the C-terminal portion of the EnvZ protein and the CheA protein may have evolved from a common ancestral protein that played some role in a primitive pathway for sensory transduction.

Hydropathy profiles of all of the sequences were determined, and they do not indicate any special regions that might be membrane associated. We would perhaps have predicted this result, since all of these gene products were found to be soluble cytoplasmic proteins in previous work (13). The codon usage in general was found to be fairly typical of that for *E. coli* proteins with the exception of the GGA and GGG codons, which were used significantly more frequently than expected (6).

The *cheR* and *cheB* gene products are involved in mediating the reversible methylation of the chemotaxis receptors. It is interesting that their coding regions apparently overlap so that the termination of the *cheR* translation may be

coordinated with the initiation of translation of *cheB*. The *cheY* gene product is also produced at higher stoichiometric levels than the other gene products. This may be the result of the apparent ribosome-binding site preceding *cheY*. This sequence has very good correspondence to the canonical ribosome-binding sequence (14). There may be other structural features of the mRNA gene product that enhance translation. They are not apparent from our analysis thus far.

This work, together with previously published sequences, provides the complete nucleotide sequence of the Meche operon and the complete sequence of the *cheW* gene.

ACKNOWLEDGMENTS

This work was supported by a Public Health Service grant from the National Institute of Health. We thank R. Doolittle for analyzing the protein sequences presented in this paper for sequence homology with other proteins. His computer analyses found the homologies between the *cheA* and the *envZ* genes.

LITERATURE CITED

1. **Aswad, D., and D. E. Koshland, Jr.** 1975. Isolation, characterization and complementation of *Salmonella typhimurium* chemotaxis mutants. J. Mol. Biol. **97**:225–235.
2. **Boyd, A., A. Krikos, and M. Simon.** 1981. Sensory transducers of *E. coli* are encoded by homologous genes. Cell **26**:333–343.
3. **Clegg, D. O., and D. E. Koshland, Jr.** 1984. The role of a signaling protein in bacterial sensing: behavioral effects of increased gene expression. Proc. Natl. Acad. Sci. USA **81**:5056–5060.
4. **DeFranco, A. L., and D. E. Koshland, Jr.** 1981. Molecular cloning of chemotaxis genes and overproduction of gene products in the bacterial sensing system. J. Bacteriol. **147**:390–400.
5. **Goy, M. F., M. S. Springer, and J. Adler.** 1977. Sensory transduction in *Escherichia coli*: role of a protein methylation reaction in sensory adaptation. Proc. Natl. Acad. Sci. USA **74**:4964–4968.
6. **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.
7. **Heidecker, G., J. Messing, and B. Gronenbaum.** 1980. A versatile primer for DNA sequencing in the M13mp2 cloning system. Gene **10**:69–73.
8. **Krikos, A., N. Mutoh, A. Boyd, and M. I. Simon.** 1983. Sensory transducers of *E. coli* are composed of discrete structural and functional domains. Cell **33**:615–622.
9. **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.
10. **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.
11. **Mizuno, T., E. T. Wurtzel, and M. Inouye.** 1982. Osmoregulation of gene expression. II. DNA sequence of *envZ* gene of the *ompB* operon of *Escherichia coli* and characterization of its gene product. J. Biol. Chem. **257**:13692–13698.
12. **Parkinson, J. S.** 1978. Complementation analysis and deletion mapping of *Escherichia coli* mutants defective in chemotaxis. J. Bacteriol. **135**:45–53.
13. **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.
14. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
15. **Segall, J. E., A. Ishihara, and H. C. Berg.** 1985. Chemotactic signaling in filamentous cells of *Escherichia coli*. J. Bacteriol. **161**:51–59.
16. **Shine, J., and L. Dalgarno.** 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosomal binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
17. **Silverman, M., and M. Simon.** 1976. Operon controlling motility and chemotaxis in *E. coli*. Nature (London) **264**:577–580.
18. **Silverman, M., and M. Simon.** 1977. Identification of polypeptides necessary for chemotaxis in *Escherichia coli*. J. Bacteriol. **130**:1317–1325.
19. **Slocum, M. K., and J. S. Parkinson.** 1983. Genetics of methyl-accepting chemotaxis proteins in *Escherichia coli*: organization of the *tar* region. J. Bacteriol. **155**:565–577.
20. **Springer, W. R., and D. E. Koshland, Jr.** 1977. Identification of a protein methyltransferase as the *cheR* gene product in the bacterial sensing system. Proc. Natl. Acad. Sci. USA **74**:533–537.
21. **Stock, J. B., and D. E. Koshland, Jr.** 1978. A protein methylesterase involved in bacterial sensing. Proc. Natl. Acad. Sci. USA **75**:3659–3663.
22. **Tsui Collins, A. L., and B. A. D. Stocker.** 1976. *Salmonella typhimurium* mutants generally defective in chemotaxis. J. Bacteriol. **128**:754–765.
23. **Yonekawa, H., H. Hayashi, and J. S. Parkinson.** 1983. Requirement of the *cheB* function for sensory adaptation in *Escherichia coli*. J. Bacteriol. **156**:1228–1235.