JOURNAL OF BACTERIOLOGY, Mar. 1991, p. 2116–2119
0021-9193/91/062116-04$02.00/0

Vol. 173, No. 6

NOTES

Tandem Translation Starts in the *cheA* Locus of *Escherichia coli*

ERIC C. KOFOID AND JOHN S. PARKINSON*

Biology Department, University of Utah, Salt Lake City, Utah 84112

Received 9 October 1990/Accepted 6 January 1991

The *cheA* locus of *Escherichia coli* encodes two protein products, $CheA_L$ and $CheA_S$. The nucleotide sequences of the wild-type *cheA* locus and of two nonsense alleles confirmed that both proteins are translated in the same reading frame from different start points. These start sites were located on the coding sequence by direct determination of the amino-terminal sequences of the two CheA proteins. Both starts are flanked by inverted repeats that may play a role in regulating the relative expression rates of the CheA proteins through alternative mRNA secondary structures.

The *cheA* locus of *Escherichia coli*, which is required for chemotactic behavior, encodes two cytoplasmic proteins, $CheA_L$ and $CheA_S$, of apparent molecular weights 78,000 and 69,000, respectively (7, 8). Nonsense mutations throughout most of the *cheA* coding region truncate both proteins, demonstrating that $CheA_L$ and $CheA_S$ are made in the same reading frame. However, several nonsense mutations at the promoter-proximal end truncate only $CheA_L$, indicating that they lie outside of the coding sequence for $CheA_S$. These observations led to the suggestion (8) (Fig. 1) that $CheA_L$ and $CheA_S$ were made by initiating translation of the *cheA* mRNA at two different in-frame start sites, which we denote as start(L) and start(S).

To test the two-start model of *cheA* expression, we determined the nucleotide sequence of the wild-type *cheA* locus and the N-terminal amino acid sequences of its two protein products. Our findings not only support the model, but also imply that competitive interactions between start(L) and start(S) may be important in regulating the relative expression levels of the two CheA proteins.

Nucleotide sequence of the *cheA* locus. The *cheA* locus lies in the middle of an operon containing several other genes, *motA* and *motB* upstream, and *cheW* downstream. A restriction fragment spanning the entire *cheA* operon was obtained by *Xma*I-*Xba*I digestion of λche22 DNA (5) and cloned into the corresponding sites of plasmid pUC118 (11), yielding pEK46 (Fig. 1). Single- or double-stranded DNA from pEK46 or one of its derivatives (not shown) was used as the template for dideoxy sequencing reactions (6). Synthetic oligonucleotides complementary to the cloned insert were used as primers. Sequence was determined on both strands from the *Acc*I site in *motB* through the *Xba*I site at the end of the insert (Fig. 1). The sequence of the *cheA* coding region and pertinent flanking features is shown in Fig. 2.

An open reading frame of 654 codons begins at a potential start triplet (ATG) 5 bases downstream from the *motB* stop codon. To confirm that this was the proper *cheA* reading frame, and to delineate the regions in which the start sites should lie, we determined the sequence changes in two *cheA* nonsense mutations. According to the two-start model (Fig. 1), amber mutation *cheA169* must lie between start(L) and start(S) because it produces $CheA_S$ molecules of normal size, whereas *cheA140* must lie downstream of both start sites because it produces amber fragments of both CheA proteins (8). Both mutations create TAG triplets in the 654-codon open reading frame: *am169* at codon 10 and *am140* at codon 107 (Fig. 2). Thus, start(L) should be located between codons 1 and 10 of this open reading frame, and start(S) should be between codons 10 and 107.

Location of *cheA* start sites. Visual inspection of the pertinent portions of the *cheA* coding region revealed two potential translation starts at ATG triplets located at codons 3 and 98. Both are preceded by purine-rich sequences that could represent Shine-Dalgarno sites for initiating ribosome binding. However, when scanned with the W71 perceptron matrix used by Stormo et al. (10), which scores a variety of sequence features characteristic of orthodox translational starts, only the site at codon 98 had a positive score (+31).

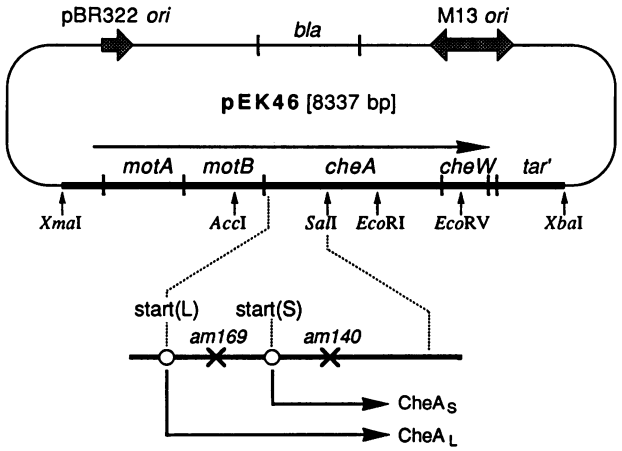

FIG. 1. Physical and genetic organization of the *cheA* region. Plasmid pEK46 is typical of those used in this project. Genes *motA* through *cheW* make up an operon of motility- and chemotaxis-related genes. Also shown are the relative positions of the two *cheA* translational start sites and two amber mutations (*cheA140* and *cheA169*) used to determine their correct reading frames. Other features indicated are as follows: *ori*, replication origin; *bla*, β-lactamase gene conferring resistance to ampicillin.

* Corresponding author.

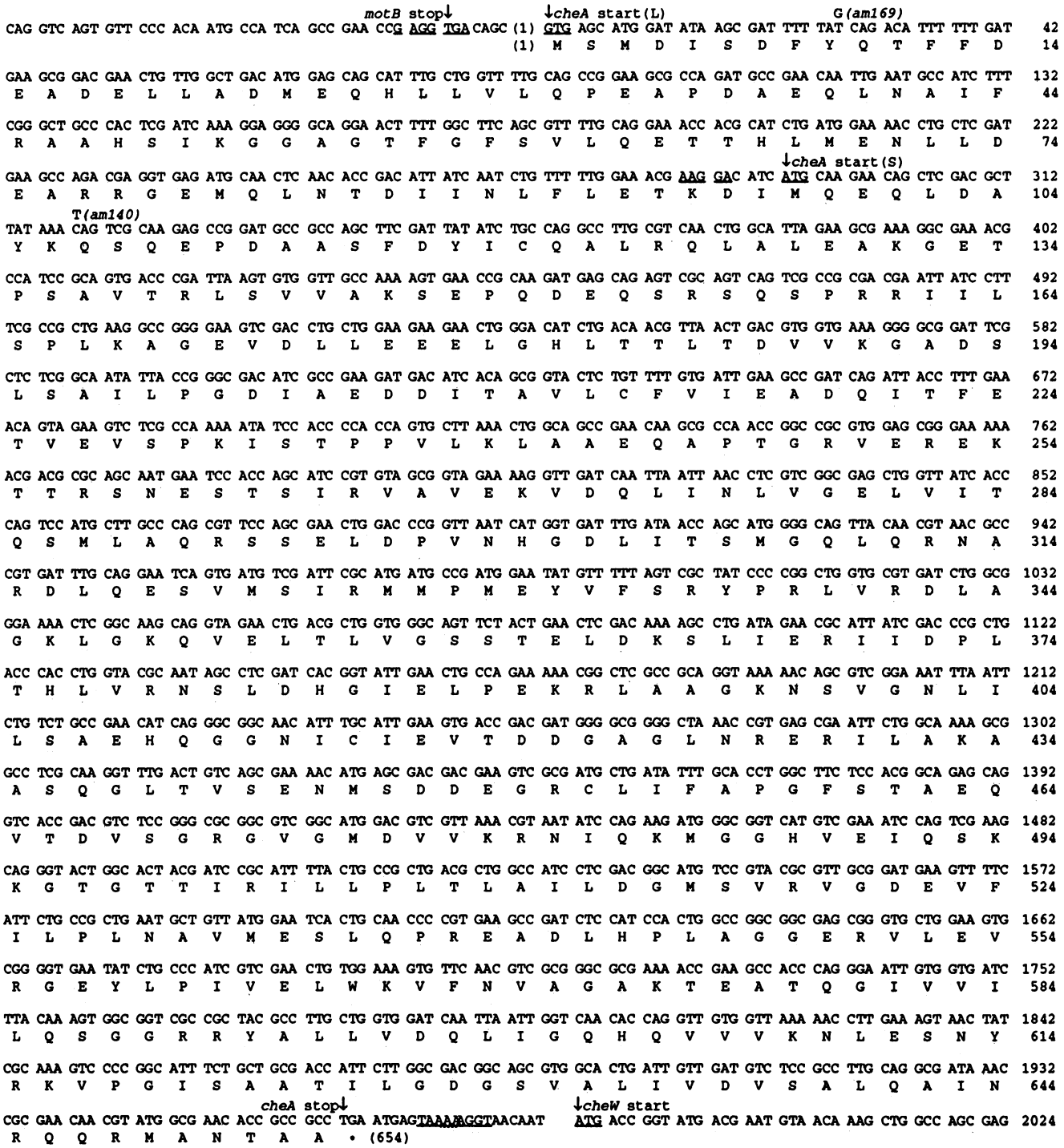

FIG. 2. Sequence of the *cheA* gene. Important features of the nucleotide sequence are shown above the DNA sequence. Shine-Dalgarno sites and initiation codons are underlined; stop codons are indicated by a black bullet. The predicted primary structure of the *cheA* proteins is listed below the sequence, using the single-letter amino acid code. The nucleotide sequence is numbered from the first base of the first *cheA* codon. The amino acid residues in $CheA_L$ are also numbered.

The next highest value, −8, occurred at the GTG 6 bases upstream from the ATG at codon 3. When this GTG was replaced by ATG, the perceptron score increased to +83, implying that it might be a valid translational start. Few GTG initiation signals were used to train the W71 matrix, which may account for the improvement following this substitution.

To establish the precise locations of the *cheA* translational initiation sites, we determined the N-terminal amino acid sequences of $CheA_L$ and $CheA_S$. These proteins were pre-

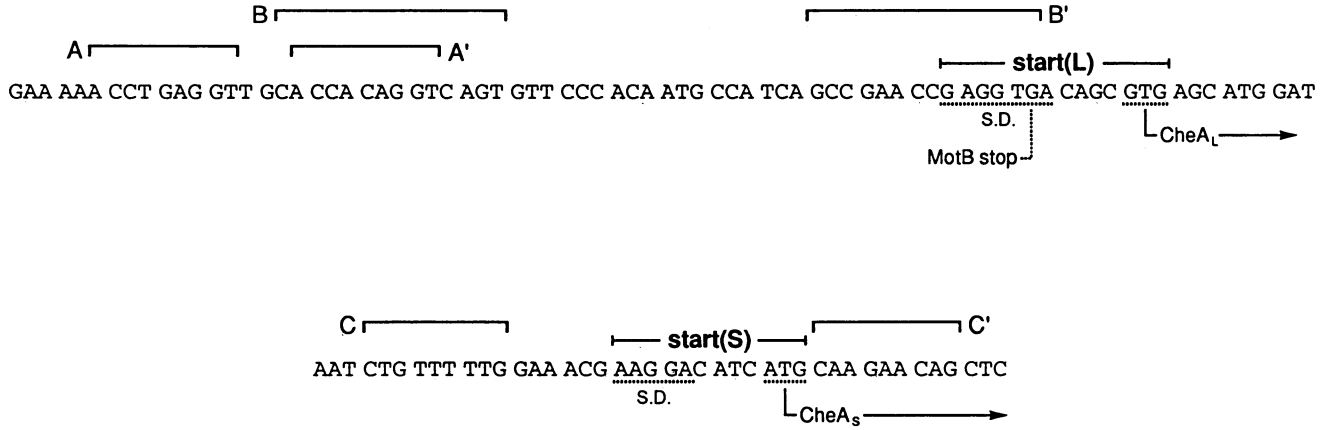

FIG. 3. Potential mRNA secondary structures and other sequence features in the vicinity of the two *cheA* start sites. Segments A/A', B/B', and C/C' make up inverted near repeats that could conceivably base pair to form hairpin structures in the mRNA. B/B' pairing would place the Shine-Dalgarno site of start(L) in a stem, whereas C/C' pairing would place the Shine-Dalgarno site and initiation codon of start(S) in a loop.

pared from plasmids pDV4 and pDV41, respectively, both of which contain the *cheA* coding region under the transcriptional control of an inducible *trp* promoter (3). A tract containing the translation initiation region and the first few codons of the *cheY* gene, which has an unusually efficient start site (4), precedes the insert. In pDV4, the *motB* fragment upstream of *cheA* is frame-shifted with respect to the *cheY* start and terminates early, whereas in pDV41 it is properly positioned and translated in frame. Upon induction with 100 μg of 3-β-indoleacrylic acid per ml, $CheA_L$ and $CheA_S$ are made from pDV4 in a ratio of roughly 20:1, whereas from pDV41 they are made in a ratio of about 5:1. Expression levels were determined by scanning densitometry of sodium dodecyl sulfate-polyacrylamide gel electrophoresis bands visualized with anti-CheA antiserum coupled to [^{35}S]protein A. Although the absolute expression levels from the two plasmids are not strictly comparable owing to possible differences in copy number, promoter strength, and polarity effects, the $CheA_L/CheA_S$ ratio should be insensitive to these variables. Thus, the different expression ratios in the two plasmids must reflect differences in the relative rates of initiation at the two start sites, which could be subject to physiological regulation, as discussed next.

The proteins were purified electrophoretically and their amino-terminal sequences were determined by automated Edman degradation (2). $CheA_S$ began with M Q E Q L D, which agrees unambiguously with the predicted sequence following the putative start(S) initiation site at codon 98. The terminus of $CheA_L$ resembled two superimposed sequences, the major being S M D I S D and the minor being M D I S D F. This result can be explained by initiation of translation at codon 1, followed by proteolytic processing at the N terminus. This would yield a mixed population lacking the initial methionine (encoded by GTG) and partially lacking the subsequent serine. We conclude that translation of $CheA_L$ is initiated primarily at codon 1, but these data do not preclude use of the ATG at codon 3 as a minor, unorthodox initiation site. Stock et al. (9) observed the same N-terminal heterogeneity in $CheA_L$ of *Salmonella typhimurium*. Note that this organization places the termination codon of *motB* within the Shine-Dalgarno tract of start(L).

Possible translational control of *cheA* expression. The different expression patterns of plasmids pDV4 and pDV41 (see above) imply that *cheA* may be subject to several kinds of translational control. First, $CheA_L$ expression is about fourfold lower from pDV41, in which ribosomes traverse the upstream *motB* sequence in frame, than from pDV4, in which translation of the sequence upstream of *cheA* is terminated by a shift in reading frame. This difference in $CheA_L$ expression implies that *motB* translation may interfere with initiation at start(L). Conceivably, the rather unusual placement of the *motB* stop codon within the Shine-Dalgarno tract of start(L) (Fig. 3) could lead to an inhibitory effect of this sort. This situation contrasts with the overlap of the *motA* stop codon and the *motB* initiation codon, which is thought to cause translational coupling of these two genes (1).

A second control mechanism may regulate the relative expression of $CheA_L$ and $CheA_S$, whose levels appear to be reciprocally related. In pDV4, $CheA_L$ expression is relatively high and the amount of $CheA_S$ is low. In pDV41, in which $CheA_L$ levels are reduced, $CheA_S$ expression is elevated. Since the mRNA transcribed from the *cheA* operon does not appear to undergo endonucleolytic processing (data not shown), both CheA products are probably translated from identical mRNA molecules. The sequences surrounding the two *cheA* start sites reveal potential mRNA secondary structures that could play a role in modulating their initiation rates (Fig. 3). Two mutually exclusive hairpins (A/A', B/B') could form near start(L), one of which (B/B') would obscure the *motB* stop and start(L) ribosome binding site. Another potential hairpin (C/C') embraces start(S), but, unlike the one at start(L), it should serve to expose the Shine-Dalgarno region and initiation codon. If start(L) normally captures most of the ribosomes moving down the message, translation at start(S) would depend on attraction of free ribosomes. However, ribosomes emanating from start(L) would be expected to disrupt the C/C' hairpin and occlude start(S), hindering expression of $CheA_S$. Thus, the efficiency of initiation at start(L) should dictate the rate of initiation at start(S), and the expression rates of $CheA_L$ and $CheA_S$ should be inversely related. Whether these regulatory effects play a role in fine-tuning the chemotaxis machinery of *E. coli* to different physiological conditions remains to be determined.

Nucleotide sequence accession number. The *cheA* sequence

has been submitted to GenBank under accession number M34669.

This work was supported by Public Health Service research grant GM28706 from the National Institutes of Health.

We thank Bob Bourret for carefully scrutinizing early versions of our *cheA* sequence.

REFERENCES

1. **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.
2. **Edman, P., and G. Begg.** 1967. A protein sequenator. Eur. J. Biochem. **1**:80–91.
3. **Matsumura, P.** Unpublished data.
4. **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.
5. **Parkinson, J. S., and S. E. Houts.** 1982. Isolation and behavior of *Escherichia coli* deletion mutants lacking chemotaxis functions. J. Bacteriol. **151**:106–113.
6. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
7. **Smith, R. A.** 1981. Detailed analysis of a genetic locus that contains a pair of overlapping genes and is involved in bacterial chemotaxis. Ph.D. thesis. University of Utah, Salt Lake City.
8. **Smith, R. A., and J. S. Parkinson.** 1980. Overlapping genes at the *cheA* locus of *E. coli*. Proc. Natl. Acad. Sci. USA **77**:5370–5374.
9. **Stock, A., T. Chen, D. Welsh, and J. Stock.** 1988. CheA protein, a central regulator of bacterial chemotaxis, belongs to a family of proteins that control gene expression in response to changing environmental conditions. Proc. Natl. Acad. Sci. USA **85**:1403–140.
10. **Stormo, G. D., T. D. Schneider, L. Gold, and A. Ehrenfeucht.** 1982. Use of the "perceptron" algorithm to distinguish translational initiation sites in *E. coli*. Nucleic Acids Res. **10**:2997–3011.
11. **Vieira, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153**:3–34.