*Proc. Natl. Acad. Sci. USA*
Vol. 77, No. 9, pp. 5370–5374, September 1980
Genetics

# Overlapping genes at the *cheA* locus of *Escherichia coli*

**(bacterial chemotaxis/genetic complementation/protein synthesis)**

ROBERT A. SMITH AND JOHN S. PARKINSON†

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

*Communicated by Julius Adler, June 3, 1980*

ABSTRACT **The *cheA* locus of *Escherichia coli*, which is required for chemotactic behavior, encodes two polypeptide products designated p[*cheA*]$_L$ and p[*cheA*]$_S$. The mode of synthesis of these two proteins was investigated by transferring various missense and nonsense mutations to a λ transducing phage and observing the mutant *cheA* products made after infection of ultraviolet-irradiated host cells. Missense mutations had no effect on either the size or the relative amounts of the two *cheA* polypeptides. Most nonsense mutations caused premature translational termination of both *cheA* products, indicating that p[*cheA*]$_L$ and p[*cheA*]$_S$ must be translated from the same coding sequence in the same reading frame. Two exceptional nonsense alleles at the promoter-proximal end of *cheA* made an intact p[*cheA*]$_S$ but no detectable p[*cheA*]$_L$. These findings show that the *cheA* locus may contain two different sites for initiation of translation. The synthesis of both proteins can be effected by the same promoter, but it is not yet clear whether both are translated from identical mRNA molecules. Complementation studies of *cheA* mutants provided evidence for two functional activities, one associated with the amino terminus of p[*cheA*]$_L$ and the other with the common portions of p[*cheA*]$_L$ and p[*cheA*]$_S$. It is possible that each *cheA* product has a different function required for chemotaxis. The possible roles of these two products and the functional significance of bacterial genes with overlapping coding sequences are discussed.**

Overlapping genes, which synthesize more than one polypeptide product from the same coding sequence, have been observed in a number of bacterial and animal viruses. Some overlapping gene pairs in bacteriophage—for example, *A* and *A** in φX174 or *C* and *Nu*3 in λ—are transcribed from a common promoter and translated in the same reading frame, but from different initiation sites, yielding two related proteins (1, 2). The $A_1$ and coat proteins of phage Qβ are also translated in the same phase but differ at their carboxyl ends due to an inefficient translation termination signal (3). Other overlapping genes in bacteriophages φX174 and G4 are transcribed from independent promoters and then translated in different reading frames, yielding essentially unrelated products from the same coding sequence (4, 5). Animal viruses such as polyoma virus and simian virus 40 utilize messenger RNA splicing to synthesize several proteins, either in the same or different reading frames, from one coding sequence (6, 7).

Regardless of their mechanism of expression, overlapping genes are generally thought to have evolved as a means of increasing the coding capacity of viral genomes without a concomitant increase in overall genome size, which in most viruses is limited by the capacity of the virion. According to this view, the evolutionary constraints that accompany the use of the same nucleotide sequence by two different genes are offset by selective pressures to retain a particular genome size. If this were the only factor favoring the existence of overlapping genes, it seems unlikely they would have arisen in bacteria or higher cells, in which limitations on the amount of genetic material are much less severe than in viruses.

In this report we describe an apparent case of overlapping genes at the *cheA* locus in *Escherichia coli*. Mutants defective in *cheA* function are viable but nonchemotactic due to an inability to change directions as they swim (8, 9). Silverman, Simon, and coworkers (10, 11) constructed λ phages carrying the *cheA* region and showed that *cheA* complementation activity was associated with the presence of two polypeptides.‡ In order to determine the nature of the relationship between these two *cheA* products, we examined the functional properties and the polypeptide products of a variety of *cheA* mutants. Our findings are consistent with a model in which both polypeptides are translated independently from the same coding sequence and in the same reading frame, but from different initiation sites. Moreover, each polypeptide may have a distinct function essential for chemotaxis. We suggest that overlapping genes may have evolved in bacteria for reasons that are totally unrelated to coding capacity considerations.

## MATERIALS AND METHODS

**Media.** Tryptone broth, plates, and swarm agar have been described (8). The sulfate-free medium used in labeling experiments contained 50 mM Tris·HCl, 25 mM $Na_2HPO_4$, 22 mM $KH_2PO_4$, 20 mM $NH_4Cl$, 3 mM $MgCl_2$, 0.1 mM $CaCl_2$, and 3 μM $FeCl_3$ supplemented with thiamin·HCl at 1 μg/ml, biotin at 1 μg/ml, 1% glycerol, and 0.5% maltose. Limiting concentrations of sulfate (30 μM $NaSO_4$) were added to support cell growth.

**Strains.** *E. coli* strain 159 was described by Hendrix (12) and was obtained along with its derivative 159 (λ*ind*$^-$) from C. Georgopoulos. Bacterial strains used for complementation and deletion analysis were all derived from RP495 [F$^-$ *rpsL thi thr*(am) *leu his met*(am) *eda srl*:Tn10 (λ$^+$)]. Various *che* mutations were introduced into this strain by transduction with P1 bacteriophage, using the *eda* locus as a selectable marker as described (8). RP495 derivatives were made *recA*$^-$ for complementation tests by mating them with KL1699 [Hfr *thi recA*] (13), selecting for streptomycin-resistant sorbitol-positive recombinants, and subsequently testing individual colonies for sensitivity to 0.2 mM methyl methanesulfonate.

The transducing phage λ*fla52* was obtained from M. Silverman. Its construction and genetic content have been described (see Fig. 1) (10). Deletion mutants of λ*fla52* were obtained by selecting phage capable of forming plaques on tryptone plates containing 1 mM EDTA. Individual plaques were cycled twice on these selective plates and stocks were prepared by confluent lysis on RP3098, which carries a deletion

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

† To whom reprint requests should be addressed.

‡ Previous reports (10, 11) estimated the molecular weights of these two polypeptides to be 76,000 and 66,000. However, in our hands they appear closer to 78,000 and 69,000. These are the values we use in this report. The larger polypeptide is henceforth designated p[*cheA*]$_L$, and the smaller one p[*cheA*]$_S$.

of the *flaG* to *motA* region that prevents recombination between the bacterial chromosome and host sequences carried by λ*fla52*. The construction of λ*fla52* derivatives carrying *cheA* point mutations is described in the legend to Fig. 2. Such phage were also grown on RP3098.

**Complementation and Mapping with λ*fla52* Derivatives.** The use of λ*fla* transducing phage to complement *che* mutants has been described (9, 11). Tryptone swarm plates containing λ*fla* phage were prepared prior to inoculation with the bacterial strains being tested. For recombinational assays, the phage were first treated with ultraviolet light (≈2000 ergs/mm$^2$) and added to swarm agar at a final concentration of approximately $1 \times 10^8$ particles per ml. For complementation assays, the swarm plates contained nonirradiated phage at a concentration of about $1 \times 10^9$ per ml. All complementation tests were done with *recA* bacterial strains to preclude formation of wild-type recombinants. Bacterial testers for both sorts of tests were lysogenic for λ in order to prevent killing by the λ*fla* present in the swarm agar.

**Radiolabeling of Polypeptides and Polyacrylamide Gel Electrophoresis.** Cells were grown overnight with limiting sulfate in the salts medium described above. They were then diluted 1:4 with fresh medium, grown an additional 90 min, harvested by centrifugation, and resuspended at $2 \times 10^9$ cells per ml in 10 mM $MgCl_2$. Irradiation was performed 20 cm from a germicidal ultraviolet lamp for 5 min at a flux of 380 μW/cm$^2$. Cells were then infected with phage at a multiplicity of infection of 10, incubated for 20 min at 35°C, then diluted to a concentration of $2 \times 10^8$ per ml with sulfate-free medium containing carrier-free $H_2{}^{35}SO_4$ (New England Nuclear) at 100 μCi/ml (1 Ci = $3.7 \times 10^{10}$ becquerels). After a 90-min incubation at 35°C, the cells were centrifuged and resuspended in 1/10 vol of electrophoresis sample buffer [5% (vol/vol) 2-mercaptoethanol, 3% sodium dodecyl sulfate, 62.5 mM Tris·HCl, pH 6.8] and placed in a boiling water bath for 1 min. Electrophoresis was carried out in discontinuous polyacrylamide gels as described by Laemmli (14).

## RESULTS

**Deletion Analysis of the *cheA* Region.** In previous studies of *E. coli* chemotaxis, the *cheA* locus was defined on the basis of mapping and complementation tests with a large number of independent point mutations (8, 9). We chose a set of approximately 60 of these mutations in order to construct a fine-structure map of the *cheA* region. Deleted derivatives of λ*fla52*, a plaque-forming transducing phage that carries the *cheA* locus and several neighboring *E. coli* genes (Fig. 1), were obtained by selecting for particles resistant to inactivation by chelating agent. Independent deletions were then tested for ability to yield wild-type recombinants when crossed to hosts with mutations in the *cheA* locus or in the adjacent *motA*, *motB*, *cheW*, *tar*, or *cheX* loci. These tests served to identify a number of deletion endpoints within the region and to orient the *cheA* map with respect to the adjacent markers. In Fig. 1 the *cheA* region has been subdivided into 10 deletion segments whose physical sizes have not yet been measured directly. We refer to this map in presenting subsequent results because it unambiguously establishes the relative positions of mutations within the *cheA* region.

The *motA*, *motB*, *cheA*, and *cheW* loci form a cotranscribed cluster (the "*mocha*" operon) whose promoter is located adjacent to *motA* (15). Several of the λ*fla52* deletions mentioned above, for example λ*fla52*Δ*105* (see Fig. 1), failed to recombine with mutant alleles in *motA* and *motB* but did recombine with all mutant sites in *cheA* and *cheW*. Nevertheless, λ*fla52*Δ*105* was unable to complement *cheA* or *cheW* mutants, suggesting

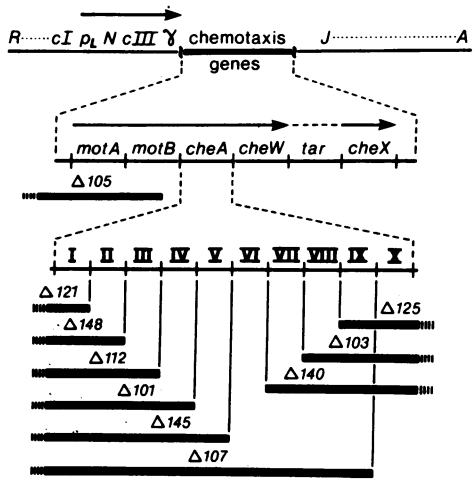


FIG. 1. Deletion map of the *cheA* locus. The structure of λ*fla52* is shown in the top line. The bacterial sequences carried by this phage are located in the central, dispensable, segment of the λ chromosome and are shown in expanded form in the second line. Arrows indicate the orientation and extent of transcriptional units. The inserted *E. coli* material in λ*fla52* is transcribed independently of the λ $p_L$ promoter. The location of material missing in various deletion derivatives of λ*fla52* is indicated by the horizontal bars. Deletion endpoints within the *cheA* locus are shown in the bottom line and serve to divide this region into 10 deletion segments (shown by roman numerals). The phage and bacterial sequences, genes, and deletion segments are not drawn to scale.

that the *cheA* locus may remain intact, but that a site essential for *cheA* complementation activity, presumably the "*mocha*" promoter, has been deleted.

**Mechanism of Synthesis of p[*cheA*]$_L$ and p[*cheA*]$_S$.** The relationship between the two *cheA* products observed by Silverman and Simon (11) was investigated by using λ*fla52* and its derivatives to program protein synthesis in host cells that had been irradiated with ultraviolet light to suppress synthesis from the host chromosome. Bacteria lysogenic for a noninducible strain of λ produce only a few proteins upon infection with λ*fla52* because, due to the ultraviolet treatment and the presence of functional λ repressor in the cell, the only gene products made are those encoded by bacterial sequences carried on the transducing phage. As shown in Fig. 3, lane a, parental λ*fla52* directs the synthesis of two *cheA* proteins with sizes of approximately 78,000 (p[*cheA*]$_L$) and 69,000 (p[*cheA*]$_S$) daltons detectable by electrophoresis on sodium dodecyl sulfate/polyacrylamide gels. In order to prove that both of these polypeptides were encoded by the *cheA* locus, a series of λ*fla52* derivatives carrying various types of *cheA* point mutations was constructed. Host mutations were recombined onto λ*fla52* by stepwise integration and excision of a λ*fla52* prophage mediated by homologous recombination between the bacterial sequences in the phage and host (Fig. 2). Examples of programming experiments using these mutant phage are shown in Fig. 3, lanes b–g. The data from all such experiments are summarized in Fig. 4, which also presents our model of how the *cheA* locus synthesizes two products from the same coding sequence. The results of these studies and our conclusions are discussed below.

(i) *Both* cheA *products are read in phase from the same coding sequence.* None of the 10 missense mutations tested in the λ*fla52* programming system affected the appearance of either of the two *cheA* products on gels. However, *cheA* nonsense mutations (with the two exceptions discussed below) appear to affect the synthesis of both p[*cheA*]$_L$ and p[*cheA*]$_S$: neither wild-type product was made, and, in several cases, two new

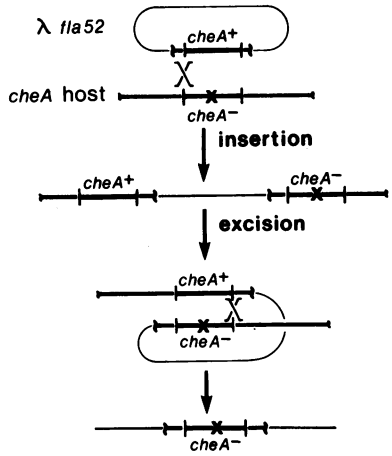


FIG. 2. Transfer of *cheA* mutations to λ*fla52*. Two reciprocal exchanges, one on each side of the mutant site, are required to recombine mutations from the host chromosome onto λ*fla52*. Exchanges were obtained in stepwise fashion by insertion and subsequent excision of a λ*fla52* chromosome, as shown. The *cheA* host strain was first lysogenized with λ*fla52* by selecting λ-immune chemotactic transductants after infection. Because λ*fla52* lacks a functional attachment system, it can form stable lysogens only by utilizing the homology provided by its bacterial insert for integrative exchanges. Lysogens were induced at 42°C and individual plaques from the resultant lysate were tested for ability to complement and recombine with various *cheA* mutations. Stocks of phage that failed to recombine with the original host mutation, but that recombined with *cheA* mutations at other sites, were prepared on RP3098.

bands were observed (Fig. 3, arrows). We feel these represent nonsense fragments of p[*cheA*]$_L$ and p[*cheA*]$_S$ and conclude that a single nonsense codon is sufficient to terminate translation of both polypeptides. This could happen only if both were translated from the same DNA sequence in the same reading frame.

In general, the sizes of the nonsense fragments listed in Fig. 4 are consistent with the relative positions of the nonsense mutations; alleles at the promoter-proximal end of *cheA* make shorter fragments than those at the promoter-distal end. It is also interesting to note (see Fig. 4) that nonsense mutations that map in the middle of the *cheA* locus made no detectable protein

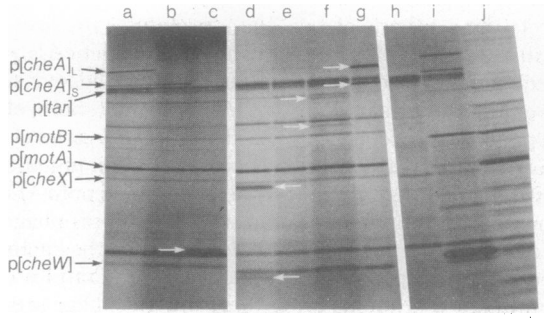


FIG. 3. Polypeptides made by different *cheA* mutations. Cells were irradiated with ultraviolet light, infected with phage, and incubated in radioactive medium. Cell extracts were subjected to electrophoresis in a 14% polyacrylamide gel containing sodium dodecyl sulfate. In lanes a–h, the bacterial host was 159(λind$^-$); in lanes i and j, the nonlysogenic host 159 was used. The infecting phage were: lane a, λ*fla52cheA*$^+$; lane b, λ*fla52cheA169*(am); lane c, λ*fla52-cheA116*(am); lane d, λ*fla52cheA133*(am); lane e, λ*fla52cheA-107* (am); lane f, λ*fla52cheA120*; lane g, λ*fla52cheA115*(am); lanes h and i, λ*fla52Δ105*; lane j, wild-type λ. The six mutations are shown in their correct relative map order. Arrows indicate nonsense fragments of p[*cheA*]$_L$ and p[*cheA*]$_S$. The products of bacterial genes carried by λ*fla52* are labeled p[*tar*], p[*motB*], etc., according to the assignments of Silverman and Simon (11).

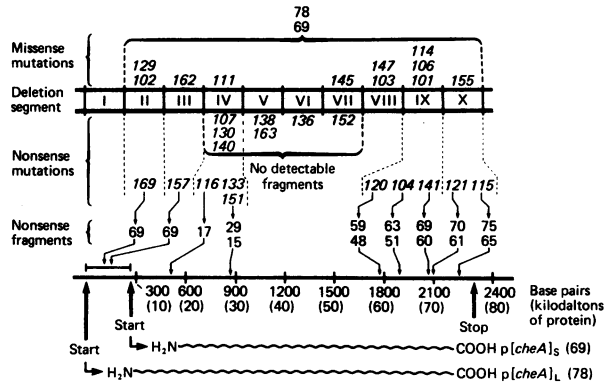


FIG. 4. (*Upper*) Polypeptides programmed by λ*fla52* carrying *cheA* point mutations. The mutations studied are shown directly above (missense alleles) or below (nonsense alleles) the deletion segment in which they map. Missense mutations made two products indistinguishable from wild-type p[*cheA*]$_L$ and p[*cheA*]$_S$. Some nonsense mutations located in deletion segments IV–VII produced no detectable *cheA* polypeptides. Other nonsense mutations made one or two nonsense fragments, whose apparent molecular masses (kilodaltons) are shown. (*Lower*) Model of the *cheA* locus. The *cheA* coding sequence appears to contain two in-phase start sites for translation. The position of nonsense alleles within this sequence, based on the sizes of polypeptides they produce, is colinear with their relative order based on deletion mapping. The exceptional alleles am*169* (segment II) and am*157* (segment III) may fall between the two start sites.

fragments, whereas promoter-proximal and promoter-distal nonsense mutations did. Evidently fragments of intermediate length are particularly susceptible to proteolytic degradation. Note that the *cheA120* allele (Fig. 3, lane f) makes detectable fragments, but in reduced amounts. This mutation maps in the first deletion segment promoter-distal to mutations for which no fragments were observed (Fig. 4).

(ii) *p*[cheA]$_L$ *and p*[cheA]$_S$ *are translated from different start sites.* Those nonsense mutations that synthesized detectable polypeptide fragments appeared to shorten both *cheA* products by the same amount (Fig. 4). The size difference between the larger and smaller fragments of each pair varied between 9000 and 14,000 daltons, but we feel this variation was due to inaccuracies in our molecular weight estimates. We conclude that the entire size difference of the wild-type *cheA* proteins probably resides at their amino termini. This may account for the behavior of am*116*, which made only one detectable fragment, of 17,000 daltons (Fig. 3, lane c). Because this mutation maps in segment IV, near the promoter-proximal end of the *cheA* region (see Fig. 4), it seems likely that a second, smaller fragment may have been synthesized, but was not seen, perhaps because fragments of less than about 10,000 daltons are not resolved in our gel system.

The key to understanding the relationship between p[*cheA*]$_L$ and p[*cheA*]$_S$ is provided by two exceptional nonsense mutations, am*169* and am*157*. These mutations map at the promoter-proximal end of the *cheA* region in segments II and III, respectively, and yet they both synthesize a 69,000-dalton product that appears to correspond to p[*cheA*]$_S$. In order to explain this result, we propose that the *cheA* locus contains two in-phase translation initiation sites separated by approximately 270 base pairs or enough to code for about 9000 daltons of protein (see Fig. 4). Translation of p[*cheA*]$_L$ begins at the promoter-proximal initiation site and proceeds to a termination site approximately 2400 nucleotides downstream. Translation of p[*cheA*]$_S$ begins at the promoter-distal initiation site and stops at the same termination signal. Thus the two proteins are identical except that p[*cheA*]$_L$ has approximately 90 additional

amino acid residues at its amino terminus. According to this model, the exceptional nonsense mutations am*157* and am*169* are thought to lie between the two translation start sites. They probably make very short fragments of p[*cheA*]$_L$ that we have been unable to detect, but they make a wild-type p[*cheA*]$_S$ because they fall outside the p[*cheA*]$_S$ coding sequence.

(iii) *Both* cheA *polypeptides can utilize the same promoter for their expression.* A programming experiment using λ*fla52Δ105*, which lacks the "*mocha*" promoter, is shown in Fig. 2, lane h. Although this phage most likely contains an intact *cheA* region, neither p[*cheA*]$_L$ nor p[*cheA*]$_S$ was produced, indicating that one or more promoters necessary for expression of these two polypeptides have been deleted in λ*fla52Δ105*. As shown in Fig. 1, the λ promoter $p_L$ is potentially able to control expression of the bacterial genes carried by λ*fla52*. Therefore, an attempt was made to utilize this promoter to achieve expression of the *cheA* sequences remaining in λ*fla52Δ105* by using this phage to program protein synthesis in a nonlysogenic host. The results of such an experiment are shown in Fig. 3, lane i. Under these conditions, λ*fla52Δ105* programmed the synthesis of two polypeptides identical in size to the wild-type *cheA* products. These bands were not observed after infection with wild-type λ (Fig. 3, lane j). Note that the *tar* and *cheX* genes, which are not part of the same transcriptional unit as *cheA* (see Fig. 1), are expressed under both lysogenic and nonlysogenic conditions, whereas the *motA* and *motB* genes, which are deleted in λ*fla52Δ105*, are not expressed under either condition. Note also that the *cheW* gene, which is located "downstream" from *cheA* and in the same operon, had the same pattern of expression as did *cheA*. We conclude that both *cheA* polypeptides can be expressed from the same promotor (i.e., λ$p_L$) and that they probably utilize a common promoter under normal conditions as well.

**Complementation Analysis of *cheA* Mutants.** In previous studies the *cheA* locus was assumed to be a single gene even though apparent "intragenic complementation" among *cheA* mutants had been observed. We reinvestigated this phenomenon by using λ*fla52* derivatives carrying mutant *cheA* alleles to complement various *cheA* host mutants. Examples are shown in Fig. 5, and the results of this work are summarized in Fig. 6. Three complementation classes of *cheA* mutants were observed: all mutants in class 1 complemented all mutants in class 2 and vice versa; mutants in class 3 failed to complement any *cheA* mutants.

The fact that *cheA* mutants could be divided into discrete complementation types provides evidence for two different *cheA* functions. It seems unlikely that these patterns are due to "intragenic complementation" in its usual form because this

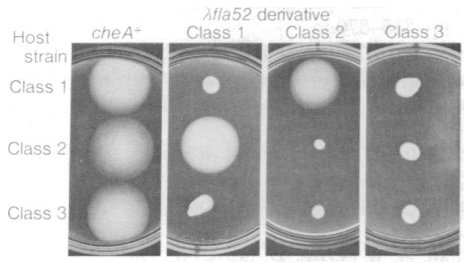


FIG. 5. Complementation among *cheA* alleles. Swarm plates containing λ*fla52cheA* derivatives were inoculated with individual colonies of *cheA recA* bacterial strains and incubated at 35°C for approximately 20 hr. Examples of the three *cheA* complementation patterns found are shown. The following alleles were used in this example. Bacterial strains: class 1, *cheA142*; class 2, *cheA155*; class 3, *cheA151*. λ*fla52* derivatives: class 1, *cheA167*; class 2, *cheA114*; class 3, *cheA107*.

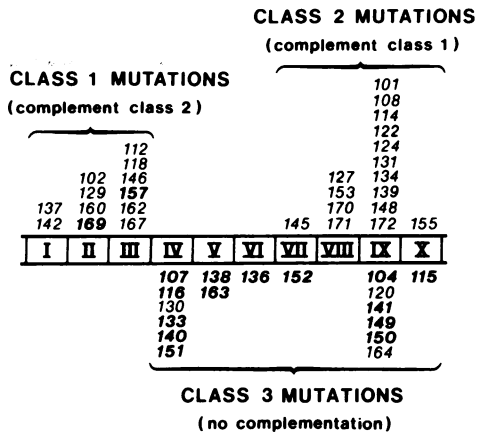


FIG. 6. Summary of *cheA* complementation behavior. Pairwise complementation tests, performed as shown in Fig. 5, were done for all the *cheA* mutations listed. Class 1 mutations map exclusively in deletion segments I–III and complement class 2 mutations in segments VII–X. The criterion for positive complementation was formation of a swarm with a diameter at least 1.5 times that of nonchemotactic control swarms. Class 3 mutations failed to show positive complementation with any *cheA* alleles. Mutations that are suppressed by amber suppressors are shown in bold face.

typically involves a complex pattern of interactions among a limited subset of mutations within a gene. Here the *cheA* complementation pattern seems to reflect two discrete functional entities and the clustering of mutant classes reinforces this impression (see Fig. 6). Class 1 mutations probably affect only the structure of the larger *cheA* polypeptide. This is clearly the case for amber mutations am*157* and am*169*, which synthesize p[*cheA*]$_S$ intact and therefore must map upstream from the second translation start site (see Fig. 4). All other class 1 mutations map in the same deletion segment as am*157* or in segments even more promoter proximal. Class 2 mutations map in segments of the *cheA* coding sequence common to both proteins and are predominately missense mutations. The class 3 mutations, on the other hand, which fail to exhibit complementation with any *cheA* mutants, are primarily nonsense mutations that map in the common coding sequences and have profound effects on the structures of both polypeptides.

## DISCUSSION

**Synthesis of p[*cheA*]$_L$ and p[*cheA*]$_S$.** Evidence has been presented that the two polypeptide products of the *cheA* locus in *E. coli* are translated from the same DNA sequence in the same reading frame but from different initiation sites. Our model predicts that these two polypeptides should be identical in primary structure, with the exception of approximately 90 additional amino acid residues at the amino terminus of the larger polypeptide. This is supported by previously published proteolytic digestion experiments, which indicated that p[*cheA*]$_L$ and p[*cheA*]$_S$ have similar primary structures (16). We believe that posttranslational processing is in no way involved in this phenomenon, primarily because we found that promoter-proximal nonsense mutations made an intact p[*cheA*]$_S$ but no p[*cheA*]$_L$. Moreover, because radioactive label cannot be chased from one *cheA* product to the other (M. Silverman, personal communication), there is no evidence for a precursor–product relationship between the two.

A single promoter is sufficient for expression of both polypeptides; however, it is not clear whether both are made from the same mRNA molecules. In φX174, for example, the A and A* proteins may be translated from newly synthesized and partially degraded mRNA, respectively (17). It is possible that

some mRNA processing event is involved in *cheA* expression. If mRNA processing is not involved and both *cheA* polypeptides are translated from the same mRNA species, one might expect that their relative proportions would be controlled by the relative accessibility of the two translation initiation sites to ribosomes. In this way, nonsense mutations that map between the two starts might be expected to enhance the accessibility of the downstream initiator and thus increase the amount of p[*cheA*]$_S$ produced. This hypothesis can be tested by quantitating the amount of p[*cheA*]$_S$ made by various *cheA* alleles.

**Functional Significance of p[*cheA*]$_L$ and p[*cheA*]$_S$.** The complementation behavior of *cheA* mutants provides evidence that this locus may encode two distinct functions involved in chemotaxis. The relationship between these two functions and the two *cheA* polypeptides is not yet clear, although the location of complementing mutations within the *cheA* coding sequence does indicate which portions of the sequence are functionally important. Class 1 mutations affect one of the proposed functions and all appear to map in the coding sequence unique to p[*cheA*]$_L$. Clearly, one of the *cheA* functions resides in the amino-terminal portion of this polypeptide. Class 2 mutations affect the other proposed function and are all missense mutations mapping in coding sequences common to both *cheA* products. Thus, this second function could reside in p[*cheA*]$_S$, in the carboxy-terminal portion of p[*cheA*]$_L$, or in both.

At least under some conditions, p[*cheA*]$_S$ is clearly capable of performing this second *cheA* function, because *cheA* nonsense mutants that made no p[*cheA*]$_L$ but an intact p[*cheA*]$_S$ were able to complement class 2 *cheA* mutants. This result could be ascribed to a type of fragment complementation similar to that seen with β-galactosidase mutants of *E. coli* (18). Note, however, that the polypeptides involved in *cheA* complementation seem to be the same size as the wild-type *cheA* products, unlike the mutant fragments involved in β-galactosidase complementation.

Because p[*cheA*]$_S$ and the carboxy-terminal portion of p[*cheA*]$_L$ have identical primary structures, only one of these proteins may be needed to perform the second *cheA* function. If both of the proposed *cheA* functions can be performed by p[*cheA*]$_L$, then p[*cheA*]$_S$ may not be needed for chemotaxis under normal conditions. However, the synthesis of two *cheA* products cannot be explained as an artifact of the ultraviolet programming system we describe here. Both polypeptides have also been observed by using plasmid templates in minicells (16) and two-dimensional gel analyses of total cell protein from wild-type bacteria (unpublished data). In all cases, p[*cheA*]$_L$ and p[*cheA*]$_S$ are made in roughly equal amounts, so neither can be considered a minor component. Furthermore, *Salmonella typhimurium* has a locus analogous to *cheA* that also encodes two polypeptides (19). It seems to us unlikely that both of these species would persist in synthesizing two *cheA* products if p[*cheA*]$_S$ were really superfluous.

If p[*cheA*]$_S$ is necessary for chemotaxis, it must perform some function not provided by p[*cheA*]$_L$. Presumably this is one of the functions defined by complementation between *cheA* mutants. Thus, there could be a direct correspondence between the two products and the two functions encoded by *cheA*. While class 1 mutations affect only p[*cheA*]$_L$, class 2 mutations may affect only the function of p[*cheA*]$_S$. This implies that only the amino-terminal portion of p[*cheA*]$_L$ is absolutely critical to its function, because most missense mutations in the coding sequence common to both polypeptides affect only one of the *cheA* functions. However, the carboxy terminus of p[*cheA*]$_L$ is not completely dispensable. Nonsense mutations that produce stable, nearly full-sized fragments of p[*cheA*]$_L$, lacking as few as 30 or so amino acids at the carboxy terminus, have no complementation activity with other *cheA* mutations.

Assuming that both *cheA* products play a role in chemotaxis, there are several possible reasons why they may be encoded in such an unusual fashion. The shared structural domains of p[*cheA*]$_L$ and p[*cheA*]$_S$ may enable them to form complexes by binding to each other. Shaw and Murialdo have suggested just such a role for the C and Nu3 proteins of phage λ, which are encoded in a manner similar to p[*cheA*]$_L$ and p[*cheA*]$_S$ (2). Another possibility is that the common portions enable both polypeptides to carry out a similar function, but the amino terminus of p[*cheA*]$_L$ enables it to interact with a different class of proteins or other cell components. For example, in ultraviolet programming experiments, p[*cheA*]$_S$ is found exclusively in the cytoplasmic fraction of the cell, whereas a significant portion of p[*cheA*]$_L$ appears associated with the cytoplasmic membrane (20). The amino end of p[*cheA*]$_L$ may function as a signal sequence that directs this polypeptide to the cell membrane but, unlike other signal sequences, is not removed by proteolytic cleavage. We hope that studies of the two *cheA* gene products will lead to a more general understanding of the phenomenon of overlapping genes.

This work was supported by U.S. Public Health Service Grant GM-19559 from the National Institute of General Medical Sciences. R.A.S. was supported by a U.S. Public Health Service predoctoral traineeship (5-T32-GM-07464-04) from the National Institute of General Medical Sciences.

1. Linney, E. & Hayashi, M. (1973) *Nature (London) New Biol.* **245,** 6–8.
2. Shaw, J. E. & Murialdo, H. (1980) *Nature (London)* **283,** 30–35.
3. Weiner, A. M. & Weber, K. (1971) *Nature (London) New Biol.* **234,** 206–209.
4. Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison, C. A., III, Slocombe, P. M. & Smith, M. (1977) *Nature (London)* **265,** 687–695.
5. Godson, G. N., Barrell, B. G., Staden, R. & Fiddes, J. C. (1978) *Nature (London)* **276,** 236–247.
6. Deininger, P., Esty, A., LaPorte, P. & Friedmann, T. (1979) *Cell* **18,** 771–779.
7. Reddy, V. B., Dhar, R. & Weissman, S. M. (1978) *J. Biol. Chem.* **253,** 621–630.
8. Parkinson, J. S. (1976) *J. Bacteriol.* **126,** 758–770.
9. Parkinson, J. S. (1978) *J. Bacteriol.* **135,** 45–53.
10. Silverman, M., Matsumura, P., Hilmen, M. & Simon, M. (1977) *J. Bacteriol.* **130,** 877–887.
11. Silverman, M. & Simon, M. (1977) *J. Bacteriol.* **130,** 1317–1325.
12. Hendrix, R. W. (1971) in *The Bacteriophage Lambda*, ed. Hershey, A. D. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 355–370.
13. Low, K. B. (1972) *Bacteriol. Rev.* **36,** 587–607.
14. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–685.
15. Silverman, M. & Simon, M. (1976) *Nature (London)* **164,** 577–580.
16. Matsumura, P., Silverman, M. & Simon, M. (1977) *J. Bacteriol.* **132,** 996–1002.
17. Linney, E. & Hayashi, M. (1974) *Nature (London)* **249,** 345–348.
18. Ullman, A. & Perrin, D. (1970) in *The Lactose Operon*, eds. Beckwith, J. R. & Zipser, D. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 143–172.
19. DeFranco, A. L. (1979) Dissertation (Univ. California, Berkeley, CA).
20. Ridgeway, H. F., Silverman, M. & Simon, M. (1977) *J. Bacteriol.* **132,** 657–665.