Proc. Natl. Acad. Sci. USA
Vol. 77, No. 10, pp. 5730–5733, October 1980
Biochemistry

Nucleotide sequence of the *thrA* gene of *Escherichia coli*

(bifunctional enzymes/gene fusion)

M. KATINKA*, P. COSSART*, L. SIBILLI†, I. SAINT-GIRONS†, M. A. CHALVIGNAC†, G. LE BRAS†, G. N. COHEN†, AND M. YANIV*

*Département de Biologie moléculaire; and †Département de Biochimie et Génétique moléculaire, Institut Pasteur, 28, rue du Dr. Roux, 75015 Paris, France

Communicated by Roger Y. Stanier, July 1, 1980

ABSTRACT The *thrA* gene of *Escherichia coli* codes for a single polypeptide chain having two enzymatic activities required for the biosynthesis of threonine, aspartokinase I and homoserine dehydrogenase I. This gene was cloned in a bacterial plasmid and its complete nucleotide sequence was established. It contains 2460 base pairs that encode for a polypeptide chain of 820 amino acids. The previously determined partial amino acid sequence of this protein is in good agreement with that predicted from the nucleotide sequence. The gene contains an internal sequence that resembles the structure of bacterial ribosome-binding sites, with an AUG preceded by four triplets, each of which can be converted to a nonsense codon by a single mutation. This suggests that the single polypeptide chain was formed by the fusion of two genes and that initiation of translation may occur inside the gene to give a protein fragment having only the homoserine dehydrogenase activity.

The *thrA* gene is the first structural gene of the threonine operon of *Escherichia coli* K-12 (1, 2). It is composed of two parts, *thrA1* and *thrA2* and codes for a bifunctional enzyme, aspartokinase I-homoserine dehydrogenase I (EC 2.7.2.4 and EC 1.1.1.3). The native enzyme (3) is a tetramer with each chain carrying, on discrete domains, the aspartokinase I and homoserine dehydrogenase I activities, which are regulated allosterically by L-threonine. Limited proteolysis of the native enzyme leads to a homodimeric fragment having the same COOH-terminal sequence as the native enzyme having only the dehydrogenase activity and no longer inhibited by threonine (3). On the other hand, a polypeptide chain synthesized by an ochre mutant that has the same NH_2 terminus as the native enzyme assembles as a tetramer having only the aspartokinase activity, still regulated by threonine (3). The determination of the primary structure of aspartokinase I homoserine-dehydrogenase I seemed warranted for a number of reasons. Sequence information was important to understand enzyme structure–function relationships and to elucidate the allosteric properties of the enzyme. It should permit the study of possible evolutionary relationships between the different proteins coded by the threonine operon and the homology with the isofunctional enzymes in *E. coli*, aspartokinase II-homoserine dehydrogenase II, coded by *metL*, and aspartokinase III coded by *lysC*.

The determination of the amino acid sequence of the aspartokinase I–homoserine dehydrogenase I was in progress (4–9) when the chemical and enzymatic DNA sequence determination techniques became available (10, 11). It then seemed advantageous to clone the gene and determine its sequence. Determination of the nucleotide sequence of *thrA* has now been completed and is presented here.

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.

MATERIALS AND METHODS

Molecular Cloning of the *thrA* Gene. The pBR322 hybrid plasmid containing the *thrA* and *thrB* genes (pIPII) was constructed as described (12).

Enzymes and Materials. Restriction endonucleases *Eco*RI, *Hpa* II, *Hin*dII + *Hin*dIII, and *Bgl* I were purified according to published procedures (13), *Sau*3A, *Hha* I, *Alu* I, *Msp* I, *Sal* I, and *Taq* I were purchased from New England BioLabs or from Bethesda Research Laboratories (Rockville, MD). Phage T4 polynucleotide kinase was purified according to the method of Richardson (14). Acrylamide was from either Serva (Heidelberg, West Germany) (twice crystallized) or BDH (Poole, England), urea was from Schwarz/Mann or Merck, and dimethyl sulfate was from Aldrich. All other chemicals were analytical grade or purer, mostly from BDH or Merck. Phenol, formamide, and piperidine were distilled, and acrylamide was deionized.

Nucleotide Sequence Determination. The nucleotide sequences were determined mostly by the chemical method of Maxam and Gilbert (10). Labeling the 5′ extremities of DNA fragments with [γ-^{32}P]ATP (Amersham) and T4 polynucleotide kinase was done by the exchange reaction of Berkner and Folk (15). Sequencing acrylamide-urea gels were made and run as described in ref. 10 or alternatively were the thin gels of Sanger and Coulson (16). One of the sequences was established by using the dideoxynucleoside triphosphate terminator technique (11), after cloning of a *Sau*3A fragment in the single-stranded phage vector M13mp2Bam (17). The primer used was an *Eco*RI 96-base-pair fragment from phage M13mp2962 (18).

Amino Acid Sequence Determination. The purification of the enzyme, the determination of its molecular weight and amino acid composition, the isolation of the cysteine- and tryptophan-containing tryptic peptide, and the purification and sequence of part of the cyanogen bromide fragments have already been described (4–9).

Computer Analysis. Analysis of the nucleotide sequences was done with the programs of Staden (19–21), and F. Schaeffer (personal communication). The protein secondary structure simulation was performed with the program of Garnier *et al.* (22).

RESULTS AND DISCUSSION

Nucleotide Sequence of the *thrA* Gene. A restriction fragment, containing a large fraction of the threonine operon, isolated from a $\lambda dthr$ transducing phage by cleavage with *Eco*RI and *Hin*dIII, was cloned in pBR322 (12). Genetic and biochemical analysis showed that this fragment contained the entire *thrA* gene (Fig. 1*A*). We then undertook the determination of the nucleotide sequence of the region presumably coding for *thrA*. The DNA fragments and the restriction sites

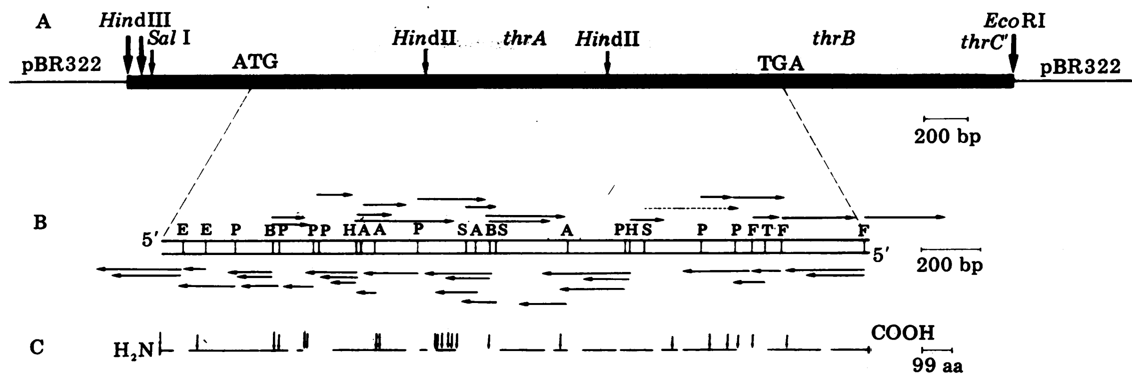

FIG. 1. The pBR322-*thr* (pIPII) hybrid plasmid and the sequencing strategy for the *thrA* gene. (*A*) Restriction map with the *Hin*dIII and *Eco*RI sites used for the cloning of the 4-kilobase-pair *thr* fragment containing the *thrA*, *thrB*, and part of the *thrC* genes. bp, Base pairs. (*B*) Sequencing strategy for the *thrA* gene, the arrows indicating the sites used for 5′ labeling as well as the direction and extent of the sequence (H, *Hin*dII; B, *Bgl* I; F, *Hin*fI; E, *Hae* III; P, *Hpa* II; A, *Alu* I; S, *Sau*3A; T, *Taq* I). The dashed arrow indicates the sequence determined with the dideoxynucleotide-terminator technique. (*C*) Partial sequence of aspartokinase I-homoserine dehydrogenase I with the sequenced peptides given in the full line. The downward arrows point to the methionines of the protein. aa, amino acids.

```
MET ARG VAL LEU LYS PHE GLY GLY THR SER LEU ALA ASN ALA GLU ARG PHE LEU ARG VAL ALA ASP ILE LEU GLU SER ASN ALA ARG GLN GLY GLN VAL
ATG CGA GTG TTG AAG TTC GGC GGT ACA TCA TTG GCA AAT GCA GAA CGT TTT CTG CGT GTT GCC GAT ATT CTG GAA AGC AAT GCC AGG CAG GGG CAG GTG
100
ALA THR VAL LEU SER ALA PRO ALA LYS ILE THR ASN HIS LEU VAL ALA MET ILE GLU LYS THR ILE SER GLY GLN ASP ALA LEU PRO ASN ILE SER ASP
GCC ACC GTC CTC TCT GCC CCT GCC AAA ATC ACC AAC CAC CTG GTG GCG ATG ATT GAA AAA ACC ATT AGC GGC CAG GAT GCT TTA CCC AAT ATC AGC GAT
199
ALA GLU ARG ILE PHE ALA GLU LEU LEU THR GLY LEU ALA ALA ALA GLN PRO GLY PHE PRO LEU ALA GLN LEU LYS THR PHE VAL ASP GLN GLU PHE ALA
GCC GAA CGT ATT TTT GCC GAA CTT TTG ACG GGA CTC GCC GCC GCC CAG CCG GGG TTC CCG TTG GCG CAA TTG AAA ACA TTC GTC GAT CAG GAA TTT GCC
298
GLN ILE LYS HIS VAL LEU HIS GLY ILE SER LEU LEU GLY GLN CYS PRO ASP SER ILE ASN ALA ALA LEU ILE CYS ARG GLY GLU LYS MET SER ILE ALA
CAA ATA AAA CAT GTC CTG CAT GGC ATT AGT TTG TTG GGG CAG TGC CCG GAT AGC ATC AAC GCT GCG CTG ATT TGC CGT GGC GAG AAG ATG TCG ATC GCC
397
ILE MET ALA GLY VAL LEU GLU ALA ARG GLY HIS ASN VAL THR VAL ILE ASP PRO VAL GLU LYS LEU LEU ALA VAL GLY HIS TYR LEU GLU SER THR VAL
ATT ATG GCC GGC GTA TTA GAA GCG CGC GGT CAC AAC GTT ACT GTT ATC GAT CCG GTC GAA AAA CTG CTG GCA GTG GGG CAT TAC CTC GAA TCC ACC GTC
496
ASP ILE ALA GLU SER THR ARG ARG ILE ALA ALA SER ARG ILE PRO ALA ASP HIS MET VAL LEU MET ALA GLY PHE THR ALA GLY ASN GLU LYS GLY GLU
GAT ATT GCC GAG TCC ACC CGC CGT ATA GCG GCA AGC CGC ATT CCG GCT GAT CAC ATG GTG CTG ATG GCA GGT TTC ACC GCC GGT AAT GAG AAA GGC GAA
595
LEU VAL VAL LEU GLY ARG ASN GLY SER ASP TYR SER ALA ALA VAL LEU ALA ALA CYS LEU ARG ALA ASP CYS CYS GLU ILE TRP THR ASP VAL ASN GLY
CTG GTG GTG CTT GGA CGC AAC GGT TCC GAC TAC TCT GCT GCG GTG CTG GCT GCC TGT TTA CGC GCC GAT TGT TGC GAG ATT TGG ACG GAC GTT AAC GGG
694
VAL TYR THR CYS ASP PRO ARG GLN VAL PRO ASP ALA ARG LEU LEU LYS SER MET SER TYR GLN GLU ALA MET GLU LEU SER TYR PHE GLY ALA LYS VAL
GTC TAT ACC TGC GAC CCG CGT CAG GTG CCC GAC GCG AGG TTG TTG AAG TCG ATG TCC TAC CAG GAA GCG ATG GAG CTT TCC TAC TTC GGC GCT AAA GTT
793
LEU HIS PRO ARG THR ILE THR PRO ILE ALA GLN PHE GLN ILE PRO CYS LEU ILE LYS ASN THR GLY ASN PRO GLN ALA PRO GLY THR LEU ILE GLY ALA
CTT CAC CCC CGC ACC ATT ACC CCC ATC GCC CAG TTC CAG ATC CCT TGC CTG ATT AAA AAT ACC GGA AAT CCT CAA GCA CCA GGT ACG CTC ATT GGT GCA
892
SER ARG ASP GLU ASP GLU LEU PRO VAL LYS GLY ILE SER ASN LEU ASN ASN MET ALA MET PHE SER VAL SER GLY PRO GLY MET LYS GLY MET VAL GLY
AGC CGT GAT GAA GAC GAA TTA CCG GTC AAG GGC ATT TCC AAT CTG AAT AAC ATG GCA ATG TTC AGC GTT TCT GGT CCG GGG ATG AAA GGG ATG GTC GGC
991
MET ALA ALA ARG VAL PHE ALA ALA MET SER ARG ALA ARG ILE SER VAL VAL LEU ILE THR GLN SER SER SER GLU TYR SER ILE SER PHE CYS VAL PRO
ATG GCG GCG CGC GTC TTT GCA GCG ATG TCA CGC GCC CGT ATT TCC GTG GTG CTG ATT ACG CAA TCA TCT TCC GAA TAC AGC ATC AGT TTC TGC GTT CCA
1090
GLN SER ASP CYS VAL ARG ALA GLU ARG ALA MET LEU GLU GLU PHE TYR LEU GLU LEU LYS GLU GLY LEU LEU GLU PRO LEU ALA VAL ALA GLU ARG LEU
CAA AGC GAC TGT GTG CGA GCT GAA CGG GCA ATG CTG GAA GAG TTC TAC CTA GAA CTG AAA GAA GGC TTA CTG GAG CCG CTG GCA GTG GCG GAA CGG CTG
1189
ALA ILE ILE SER VAL VAL GLY ASP GLY LEU ARG THR LEU ARG GLY ILE SER ALA LYS PHE PHE ALA ALA LEU ALA ARG ALA ASN ILE ASN ILE VAL ALA
GCC ATT ATC TCG GTG GTA GGT GAT GGT TTG CGC ACC TTG CGT GGG ATC TCG GCG AAA TTC TTT GCC GCA CTG GCC CGC GCC AAT ATC AAC ATT GTC GCC
1288
ILE ALA GLN GLY SER SER GLU ARG SER ILE SER VAL VAL VAL ASN ASN ASP ASP ALA THR THR GLY VAL ARG VAL THR HIS GLN MET LEU PHE ASN THR
ATT GCT CAG GGA TCT TCT GAA CGC TCA ATC TCT GTC GTG GTA AAT AAC GAT GAT GCG ACC ACT GGC GTG CGC GTT ACT CAT CAG ATG CTG TTC AAT ACC
1387
ASP GLN VAL ILE GLU VAL PHE VAL ILE GLY VAL GLY GLY VAL GLY GLY ALA LEU LEU GLU GLN LEU LYS ARG GLN GLN SER TRP LEU LYS ASN LYS HIS
GAT CAG GTT ATC GAA GTG TTT GTG ATT GGC GTC GGT GGC GTT GGC GGT GCG CTG CTG GAG CAA CTG AAG CGT CAG CAA AGC TGG CTG AAG AAT AAA CAT
1486
ILE ASP LEU ARG VAL CYS GLY VAL ALA ASN SER LYS ALA LEU LEU THR ASN VAL HIS GLY LEU ASN LEU GLU ASN TRP GLN GLU GLU LEU ALA GLN ALA
ATC GAC TTA CGT GTC TGC GGT GTT GCC AAC TCG AAG GCT CTG CTC ACC AAT GTA CAT GGC CTT AAT CTG GAA AAC TGG CAG GAG GAA CTG GCG CAA GCC
1585
LYS GLU PRO PHE ASN LEU GLY ARG LEU ILE ARG LEU VAL LYS GLU TYR HIS LEU LEU ASN PRO VAL ILE VAL ASN CYS THR SER SER GLN ALA VAL ALA
AAA GAG CCG TTT AAT CTC GGG CGC TTA ATT CGC CTC GTG AAA GAA TAT CAT CTG CTG AAC CCG GTC ATT GTT AAC TGC ACT TCC AGC CAG GCA GTG GCG
1684
ASP GLN TYR ALA ASP PHE LEU ARG GLU GLY PHE HIS VAL VAL THR PRO ASN LYS LYS ALA ASN THR SER SER MET ASP TYR TYR HIS GLN LEU ARG TYR
GAT CAA TAT GCC GAC TTC CTG CGC GAA GGT TTC CAC GTT GTC ACG CCG AAC AAA AAG GCC AAC ACC TCG TCG ATG GAT TAC TAC CAT CAG TTG CGT TAT
1783
ALA ALA GLU LYS SER ARG ARG LYS PHE LEU TYR ASP ILE ASN VAL GLY ALA GLY LEU PRO VAL ILE GLU ASN LEU GLN ASN LEU LEU ASN ALA GLY ASP
GCG GCG GAA AAA TCG CGG CGT AAA TTC CTC TAT GAC ATC AAC GTT GGG GCT GGA TTA CCG GTT ATT GAG AAC CTG CAA AAT CTG CTC AAT GCA GGT GAT
1882
GLU LEU MET LYS PHE SER GLY ILE LEU SER GLY SER LEU SER TYR ILE PHE GLY LYS LEU ASP GLU GLY MET SER PHE SER GLU ALA THR ARG LEU ALA
GAA TTG ATG AAA TTC TCC GGC ATT CTT TCT GGT TCG CTT TCT TAT ATC TTC GGC AAG TTA GAC GAA GGC ATG AGT TTC TCC GAG GCG ACC CGG CTG GCG
1981
ARG GLU MET GLY TYR THR GLU PRO ASP PRO ARG ASP ASP LEU SER GLY MET ASP VAL ALA ARG LYS LEU LEU ILE LEU ALA ARG GLU THR GLY ARG GLU
CGG GAA ATG GGT TAT ACC GAA CCG GAC CCG CGA GAT GAT CTT TCT GGT ATG GAT GTG GCG CGT AAA CTA TTG ATT CTC GCT CGT GAA ACG GGA CGT GAA
2080
LEU GLU LEU ALA ASP ILE GLU ILE GLU PRO VAL LEU PRO ALA GLU PHE ASN ALA GLU GLY ASP VAL ALA ALA PHE MET ALA ASN LEU SER GLN LEU ASP
CTG GAG CTG GCG GAT ATT GAA ATT GAA CCT GTG CTG CCC GCA GAG TTT AAC GCC GAG GGT GAT GTT GCC GCT TTT ATG GCG AAT CTG TCA CAA CTC GAC
2179
ASP LEU PHE ALA ALA ARG VAL ALA LYS ALA ARG ASP GLU GLY LYS VAL LEU ARG TYR VAL GLY ASN ILE ASP GLU ASP GLY VAL CYS ARG VAL LYS ILE
GAT CTC TTT GCC GCG CGC GTG GCG AAG GCC CGT GAT GAA GGA AAA GTT TTG CGC TAT GTT GGC AAT ATT GAT GAA GAT GGC GTC TGC CGC GTG AAG ATT
2278
ALA GLU VAL ASP GLY ASN ASP PRO LEU PHE LYS VAL LYS ASN GLY GLU ASN ALA LEU ALA PHE TYR SER HIS TYR TYR GLN PRO LEU PRO LEU VAL LEU
GCC GAA GTG GAT GGT AAT GAT CCG CTG TTC AAA GTG AAA AAT GGC GAA AAC GCC CTG GCC TTC TAT AGC CAC TAT TAT CAG CCG CTG CCG TTG GTA CTG
2377
ARG GLY TYR GLY ALA GLY ASN ASP VAL THR ALA ALA GLY VAL PHE ALA ASP LEU LEU ARG THR LEU SER TRP LYS LEU GLY VAL ***
CGC GGA TAT GGT GCG GGC AAT GAC GTT ACA GCT GCC GGT GTC TTT GCT GAT CTG CTA CGT ACC CTC TCA TGG AAG TTA GGA GTC TGA
```

FIG. 2. Sequence of the *thrA* gene. The given sequence is that of the noncoding strand. The deduced amino acid sequence for the correct reading frame is shown.

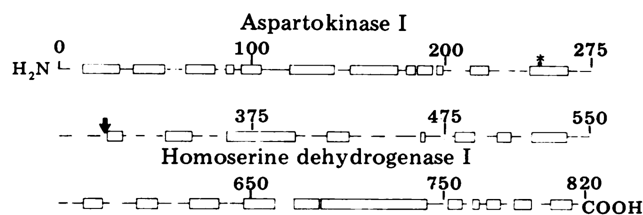

FIG. 3. Predicted secondary structure of the aspartokinase I-homoserine dehydrogenase I. The boxes correspond to the possible α-helix structures, the lines to the possible extended regions of the protein. The coordinates above are the amino acids from 1 to 820. The arrow indicates the position of limited proteolysis cleavage generating an active homoserine dehydrogenase fragment. The asterisk corresponds to the position of methionine-249.

used in the sequence determination are indicated in Fig. 1*B*. The sequence was determined for over 70% of both strands. In Fig. 1*C* are indicated the regions where the corresponding amino acid sequence had been previously determined; the nucleotide and amino acid sequences are shown in Fig. 2. The gene is 2460 nucleotides long and encodes a single polypeptide chain of 820 amino acids. Except for a few explainable discrepancies (e.g., Glu for Gln), the predicted amino acid sequence is in good agreement with the partially determined amino acid sequence. Examination of the sequence homologies between *thrA1* and *thrA2* cistrons did not show any large duplications. However, several tetrapeptides are reported twice (Leu-Arg-Thr-Leu, positions 406 and 811; Leu-Ala-Arg-Glu, positions 659 and 686; Ala-Ala-Arg-Val, positions 322 and 730). One tetrapeptide (Ile-Ser-Val-Val) is present three times at positions 344, 399, and 439.

Computer-Predicted Secondary Structures of Aspartokinase I-Homoserine Dehydrogenase I. The predicted secondary structure of the protein was calculated according to Garnier (22) and is shown in Fig. 3. This analysis gave a high percentage of extended regions (34%) and α helices (52%). This percentage of α helix is not in agreement with the value of 31% calculated from circular dichroism experiments (23). The site of limited proteolysis seems to be located in a nonstructured region.

Codon Utilization. The codon usage in *thrA* gene was found to be highly nonrandom (Table 1). The *thrA* gene shows a preference for the use of only two codons (CGU and CGC) out of the six codons for arginine, CUG for leucine, GGY for glycine (Y, pyrimidines), AUY for isoleucine, GAA for glutamic acid, AAA for lysine, and CCG for proline. For each of these cases, the codons preferentially used are those recognized efficiently by the most abundant tRNA species (24–30). However, the alanine codon GCC is used the most frequently and is recognized by a minor $tRNA^{Ala}$ species (31, 32). Nonrandom use of the codons for asparagine and threonine was also noted, but there are not enough data available to link this to the abundance of the tRNA species. One can only notice that the codons mostly used for threonine and isoleucine in the protein are those mostly used in the leader peptide of the attenuator region of the threonine operon, where seven of the eight threonine codons are ACC and three of the four isoleucine codons are AUU (33).

The overall codon usage is similar to that found in the *trpA* gene of *E. coli* (34). The frequency of nucleotides employed in the third positions of the codons is also similar for the two genes (in *thrA*, A = 15.7%, U = 26%, G = 28.4%, C = 29.7%; in *trpA*, A = 18.6%, U = 25%, G = 28.7%, C = 27.6%).

An interesting feature of the sequence is the cluster of methionine residues in the middle part of the protein: 10 out of the 22 methionine residues lie in a region corresponding to 16% of the protein, starting at methionine-249. We have looked for a Shine and Dalgarno sequence (35) before the corresponding ATGs in the DNA sequence. In fact, only methionine-249 is actually preceded by the sequence G-A-G-G-U, which is complementary to the 3′ end of the ribosomal 16S RNA. As shown in Fig. 4*A*, the sequence surrounding this region can be folded into a stem and loop structure that resembles in its properties the functional intergenic ribosomal binding sites identified in *E. coli* (36). The calculated ΔG value for formation of this structure is about −13 kcal (−54 kJ)/mol (37). In Fig. 4*B* is presented one of the other more energetically stable potential mRNA secondary structures, which also would have a functional ribosomal binding site.

The methionine in position 249 could correspond to a start codon of the dehydrogenase part of the polypeptide chain. This residue is preceded by the sequence TTG TTG AAG TCG *ATG*. These four triplets can be converted into termination codons by a single base change. Two separate base changes could have abolished two termination triplets frequently found at the end of the genes and hence created a bifunctional polypeptide chain. It was previously shown that the NH_2-terminal sequence of the proteolytic fragment carrying only the dehydrogenase activity starts at serine-297 (6). This finding is not in contradiction with the idea that a gene fusion has indeed occurred during evolution. The presumptive kinase would have

Table 1. Codon usage in *thrA*

	U			C			A			G		
U	UUU	Phe	11	UCU	Ser	10	UAU	Tyr	12	UGU	Cys	3
	UUC	Phe	19	UCC	Ser	11	UAC	Tyr	8	UGC	Cys	9
	UUA	Leu	10	UCA	Ser	6	UAA	Ochre		UGA	Opal	1
	UUG	Leu	16	UCG	Ser	9	UAG	Amber		UGG	Trp	4
C	CUU	Leu	8	CUU	Pro	4	CAU	His	8	CGU	Arg	18
	CUC	Leu	13	CCC	Pro	5	CAC	His	6	CGC	Arg	19
	CUA	Leu	3	CCA	Pro	2	CAA	Gln	11	CGA	Arg	3
	CUG	Leu	42	CCG	Pro	18	CAG	Gln	18	CGG	Arg	5
A	AUU	Ile	29	ACU	Thr	4	AAU	Asn	22	AGU	Ser	3
	AUC	Ile	16	ACC	Thr	18	AAC	Asn	18	AGC	Ser	12
	AUA	Ile	2	ACA	Thr	3	AAA	Lys	22	AGA	Arg	0
	AUG	Met	22	ACG	Thr	6	AAG	Lys	12	AGG	Arg	2
G	GUU	Val	19	GCU	Ala	14	GAU	Asp	29	GGU	Gly	22
	GUC	Val	18	GCC	Ala	36	GAC	Asp	13	GGC	Gly	22
	GUA	Val	5	GCA	Ala	15	GAA	Glu	38	GGA	Gly	9
	GUG	Val	26	GCG	Ala	27	GAG	Glu	15	GGG	Gly	10

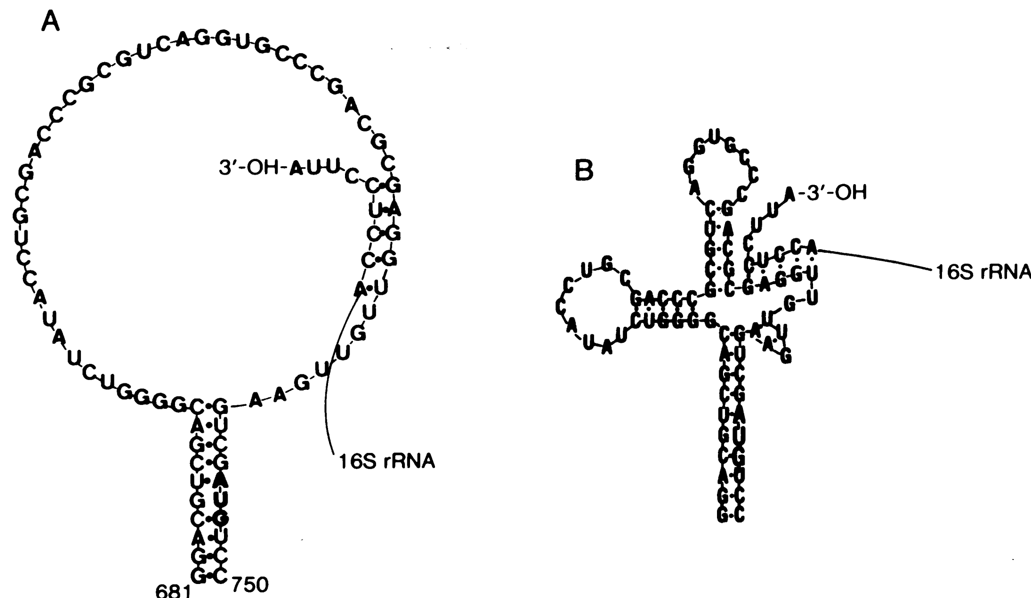

FIG. 4. Possible RNA secondary structures around an internal initiation site. (*A*) Stem and loop structure as described by Selker and Yanofsky (36) with the five base pairs complementary to the 3′ extremity of 16S ribosomal RNA, and the partially paired AUG initiation codon in boldface letters. The numbering system is that of Fig. 2. (*B*) A more thermodynamically stable conformation of the same sequence, in an almost perfect tRNA structure.

a molecular weight of about 25,000. The formation of a fused protein may have given the bacteria an advantage in controlling coordinatively both activities at the transcription and translation levels, in addition to the allosteric control of both activities by L-threonine.

E. coli has two proteins with aspartokinase activities that are fused to proteins with homoserine dehydrogenase activities and a third protein that has only aspartokinase activity. Further cloning and sequence studies are required to establish the evolutionary relationships among these homologous proteins.

We are grateful to R. Staden, F. Schaeffer, and R. Garnier for making available to us their computer programs. M.K. is extremely grateful to G. Winter and F. Sanger for the opportunity of learning the dideoxy-terminator sequencing technique. We thank S. Busby for valuable comments and C. Maczuka for help with the preparation of the manuscript. This work was supported by grants from the Centre National de la Recherche Scientifique (LA 270 and Action Thématique Programmée: Séquence des acides nucléiques informationnels), the Institut National de la Santé et de la Recherche Médicale (Action Thématique Programmée 77-82), and the Délégation Générale à la Recherche Scientifique et Technique.

1. Theze, J., Margarita, D., Cohen, G. N., Borne, F. & Patte, J. C. (1974) *J. Bacteriol.* **117**, 133–143.
2. Theze, J. & Saint Girons, I. (1974) *J. Bacteriol.* **118**, 990–998.
3. Cohen, G. N. & Dautry-Varsat, A. (1980) in *Multifunctional Proteins*, eds. Bisswanger, H. & Schmincke-Ott, E. (Wiley, New York), pp. 49–121.
4. Sibilli, L., Cossart, P., Chalvignac, M. A., Briley, P. A., Costrejean, J. M., Le Bras, G. & Cohen, G. N. (1977) *Biochimie* **59**, 943–945.
5. Cossart-Gheerbrant, P., Sibilli-Weil, L., Briley, P. A., Chalvignac, M. A., Le Bras, G. & Cohen, G. N. (1978) *Biochim. Biophys. Acta* **535**, 206–215.
6. Briley, P. A., Sibilli, L., Chalvignac, M. A., Cossart, P., Le Bras, G., DeWolf, A. & Cohen, G. N. (1978) *J. Biol. Chem.* **253**, 8867–8871.
7. Sibilli, L., Le Bras, G., Cossart, P., Chalvignac, M. A., Le Bras, G., Briley, P. A. & Cohen, G. N. (1979) *Biochimie* **61**, 733–739.
8. Sibilli, L. (1977) Dissertation (University of Paris, Paris, France).
9. Cossart, P. (1977) Dissertation (University of Paris, Paris, France).
10. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
11. Sanger, F., Nicklen, S. & Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
12. Cossart, P., Katinka, M., Yaniv, M., Saint-Girons, I. & Cohen, G. N. (1979) *Mol. Gen. Genet.* **175**, 39–44.
13. Greene, P. J., Heyneker, H. L., Bolivar, F., Rodriguez, R. L., Betlach, M. C., Covarrubias, A. A., Backman, K., Russel, D. J., Tait, R. & Boyer, H. (1978) *Nucleic Acids Res.* **5**, 2373–2380.
14. Richardson, C. C. (1971) in *Procedures in Nucleic Acids Research*, eds., Cantoni, G. L. & Davies, D. R. (Harper & Row, New York), Vol. 2, pp. 815–828.
15. Berkner, K. L. & Folk, W. R. (1977) *J. Biol. Chem.* **252**, 3176–3184.
16. Sanger, F. & Coulson, A. (1978) *FEBS Lett.* **87**, 107–110.
17. Schreier, P. H. & Cortese, R. (1979) *J. Mol. Biol.* **129**, 169–172.
18. Heidecker, G., Messing, J. & Gronenborn, B. (1980) *Gene* **10**, 69–73.
19. Staden, R. (1977) *Nucleic Acids Res.* **4**, 4037–4051.
20. Staden, R. (1978) *Nucleic Acids Res.* **5**, 1013–1015.
21. Staden, R. (1979) *Nucleic Acids Res.* **6**, 2601–2610.
22. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
23. Heck, H. d'A. & Truffa-Bachi, P. (1970) *Biochemistry* **9**, 2776–2783.
24. Celis, T. F. R. & Maas, W. J. (1971) *J. Mol. Biol.* **62**, 179–188.
25. Blank, H. U. & Söll, D. (1971) *J. Biol. Chem.* **246**, 4947–4956.
26. Fleck, E. W. & Carbon, J. (1975) *J. Bacteriol.* **122**, 492–501.
27. Yarus, M. & Barrell, B. G. (1971) *Biochem. Biophys. Res. Commun.* **43**, 729–734.
28. Ohashi, Z., Sanzoyoshi, M., Harada, F., Hara, M. & Nishimura, S. (1970) *Biochem. Biophys. Res. Commun.* **40**, 866–872.
29. Chakraburtty, K., Steinschneider, A., Case, R. V. & Mehler, A. H. (1975) *Nucleic Acids Res.* **2**, 2069–2075.
30. Söll, D., Cherayil, J. D. & Bock, R. M. (1967) *J. Mol. Biol.* **29**, 97–112.
31. Williams, R. J., Nagel, W., Roe, B. & Dudock, B. (1974) *Biochem. Biophys. Res. Commun.* **60**, 1215–1221.
32. Lund, E. & Dahlberg, J. E. (1977) *Cell* **111**, 247–262.
33. Gardner, J. F. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1706–1710.
34. Nichols, B. P. & Yanofsky, C. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5244–5248.
35. Shine, J. & Dalgarno, L. (1975) *Nature (London)* **254**, 34–38.
36. Selker, E. & Yanofsky, C. (1979) *J. Mol. Biol.* **130**, 135–143.
37. Tinoco, I., Jr., Borer, P. N., Dewgler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature (London)* **246**, 40–41.