Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 4914–4918, August 1983
Biochemistry

Primary structure of the *Escherichia coli thyA* gene and its thymidylate synthase product

(overlapping BAL-31 deletions/M13-dideoxy sequence analysis/protein sequence determination/fluorodeoxyuridylate binding site/*Lactobacillus casei* sequence comparison)

MARLENE BELFORT, GLADYS MALEY, JOAN PEDERSEN-LANE, AND FRANK MALEY

Center for Laboratories and Research, New York State Department of Health, Empire State Plaza, Albany, New York 12201

Communicated by Henry A. Lardy, May 9, 1983

ABSTRACT **The nucleotide sequence of a 1,163-base-pair fragment that encodes the entire *thyA* gene of *Escherichia coli* K-12 was determined. The strategy involved sequence determination of both DNA strands by using overlapping deletions that had been generated *in vitro* from the two ends of the fragment with BAL-31 nuclease. The amino-terminal sequence of thymidylate synthase (5,10-methylenetetrahydrofolate:dUMP *C*-methyltransferase, EC 2.1.1.45), the product of the *thyA* gene, located the 792-base-pair open reading frame, which codes for the 264 amino acid residues of this enzyme. The amino acid sequence deduced from the nucleotide data was confirmed to the extent of 40% by partial sequence analysis of the enzyme purified from extracts of the amplified cloned gene. Transcriptional and translational control areas were apparent in the regions flanking the structural gene. The 5-fluorodeoxyuridylate-binding residue of the active site was identified as cysteine-146. Comparison of the *E. coli* and *Lactobacillus casei* synthase sequences reveals consistent homology (62%) over extensive regions. This homology is particularly striking in a very hydrophobic region bordering cysteine-146. In the two enzymes, this region, which probably defines the active site, is 82% homologous. However, a dramatic difference between the two sequences is reflected by the surprising finding that a 51-amino-acid stretch, present midway through the *L. casei* sequence, is completely absent from the *E. coli* enzyme.**

Because of the key role that thymidylate synthase (5,10-methylenetetrahydrofolate:dUMP *C*-methyltransferase; EC 2.1.1.45) plays in DNA synthesis, it has been an important chemotherapeutic target enzyme in the treatment of various proliferative diseases (1, 2). To develop a rational approach for the development of even more effective chemotherapeutic agents than those currently directed against the synthase, and to better understand the mechanism of action of this unique enzyme, we undertook an analysis of the primary sequence of the *Lactobacillus casei* synthase (3) and the loci of its substrate binding sites (4, 5). These studies were successfully completed primarily because of the relative abundance of this enzyme in methotrexate-resistant mutants of *L. casei* (6, 7).

The advent of recombinant DNA genetics has facilitated an extension of this approach to enzymes that are expressed weakly in nature. A case in point is the thymidylate synthase of *Escherichia coli* K-12, which contains only about 250 molecules per actively growing cell. By manipulating the *thyA* gene, it was possible to construct enzyme-overproducing strains (8) with techniques similar to those used in amplifying T4 phage thymidylate synthase expression (9). This was accomplished by resecting a 7.8-kilobase (kb) *Hin*dIII DNA fragment of *E. coli* to one of 1.16 kb, which on insertion into the expression plasmid pKC30 resulted in synthase overproduction (8). In this report we describe the complete nucleotide sequence of the 1.16-kb insert, which includes transcriptional and translational control elements and the entire enzyme-coding region. The deduced amino acid sequence was confirmed by protein sequence data, which also permitted identification of the putative active site or fluorodeoxyuridylate (FdUMP) binding region.

The availability of the sequence of the *E. coli* synthase provides the basis for a detailed structure–function analysis of the *E. coli thyA* region and facilitates phylogenetic studies with other synthases. A preliminary comparison between the *E. coli* and *L. casei* enzymes is presented, revealing both extensive homologies and interesting structural differences.

MATERIALS AND METHODS

Bacterial Strains, Phage Strains, and Plasmids. *E. coli* strain JM103 (Δ*lac-pro*l, *Sup*E, *thi*, *end*A, *str*A, *sbc*B15, *hsd*R4, F′*tra*D36, *pro*AB$^+$, *lac*I^{q}ZΔM15), used for propagation of M13 phage, was obtained from Bethesda Research Laboratories, whereas strain MB241 [N99*hfl*$^-$*recA*(λ*int*6c1857*cro*27*Pam*3)/pKTAH] was constructed in this laboratory and used for synthase overproduction (8). The double-stranded replicative form of M13 strains mp8 and mp9 (10) were purchased from P-L Biochemicals. The recombinant plasmid pBTAH contains the *thyA* gene on a 1.16-kb fragment cloned with linkers into the *Hin*dIII site of pBR322 (8).

Enzymes. T4 DNA ligase and BAL-31 nuclease were purchased from New England BioLabs. The Klenow fragment of *E. coli* DNA polymerase was obtained from Bethesda Research Laboratories, and restriction enzymes *Bam*HI, *Cla* I, *Eco*RV, *Eco*RI, *Hin*dIII, *Pst* I, and *Pvu* II were from either of these suppliers and used according to their instructions. Carboxypeptidase A was a gift from T. Plummer, Jr.

Biochemicals. [α-^{32}P]dATP for DNA sequence determination was purchased from Amersham (400 Ci/mmol, 10 mCi/ml; 1 Ci = 3.7 × 10^{10} Bq) and [2-^{14}C]FdUMP for ternary complex formation from Moravek Biochemicals (Brea, CA) (52 mCi/mmol). Oligonucleotide linkers were from New England BioLabs, and dideoxynucleotide triphosphates and the heptadecanucleotide universal primer were from P-L Biochemicals. Cyanogen bromide was supplied by Pierce, and chemicals for protein sequence analysis were obtained from Pierce or Beckman. High-performance liquid chromatography solvents were HPLC grade, prepared with "nanopure" water (Barnstead) and filtered before use through an Ultipor NX 0.45-μm membrane. Bio-Gel P-60 was supplied by Bio-Rad.

Construction of Overlapping Deletion Libraries. The overall scheme for generating the two BAL-31 deletion series that

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.

Abbreviations: kb, kilobase pair(s); FdUMP, fluorodeoxyuridylate; bp, base pair(s).

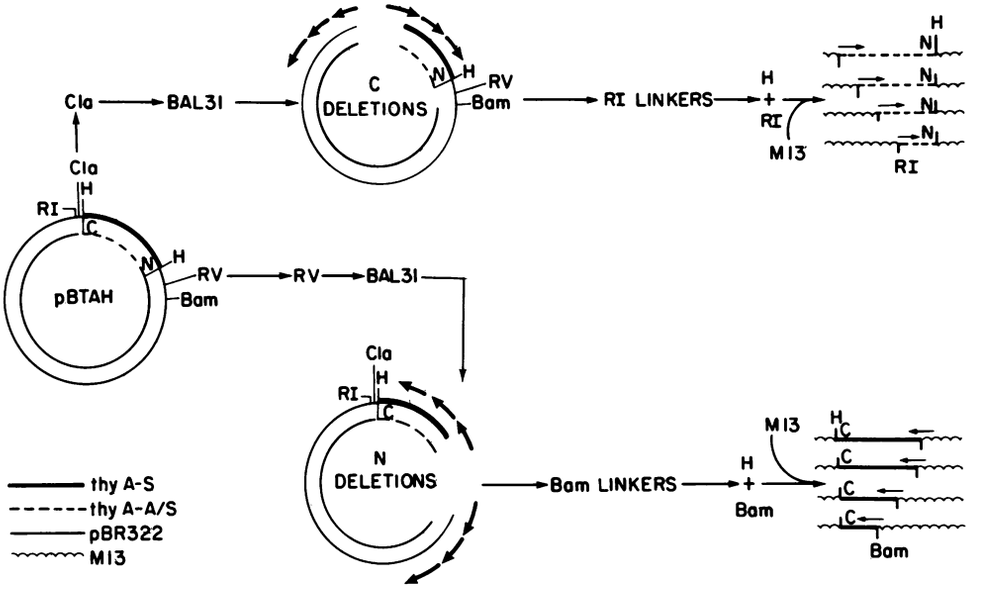

FIG. 1. Construction of overlapping deletion mutant libraries for DNA sequence determination. pBTAH was linearized near the amino-terminal (N) end of the *thyA* insert with *Eco*RV (RV) or at the carboxy-terminal (C) end with *Cla* I (Cla). In each case bidirectional exonucleolytic digestion with BAL-31 was allowed to proceed in from the ends for progressively increasing distances (heavy arrows). Oligonucleotide linkers were used to add restriction enzyme sites to the flush ends left by BAL-31. The *Bam*HI (Bam) linkers ligated to the N-deletion series and the *Eco*RI (RI) linkers of the C-deletion series had their 5′ extensions exposed by digestion with the corresponding enzyme. The *thyA* deletion fragments were subsequently liberated from pBR322 sequences by *Hin*dIII (H) treatment. These fragments were then force-cloned into double-stranded M13mp9 DNA, which had been appropriately cleaved with either a *Hin*dIII/*Bam*HI combination or with *Hin*dIII/*Eco*RI. The location of the sequences homologous to the universal primer in the M13 vector is such that single-stranded recombinant phage DNA serving as template in the Sanger sequence determination reaction will allow primed synthesis (light arrows) to occur from the *thyA* sense strand (S) in the N-deletion series, with the newly synthesized DNA corresponding effectively in sequence to mRNA. The C-deletions were used to generate the complementary sequence from the antisense (A/S) strand.

were used in determining the sequences of the overlapping fragments of both the sense and antisense strands is presented in Fig. 1. Cloning procedures used in these studies, including plasmid DNA preparation, restriction, ligation, linker addition, and transformation have been described (8, 9). BAL-31 digestion was carried out at 30°C on 10 μg of linearized DNA with 10 units of enzyme in a total volume of 100 μl of the recommended buffer. Aliquots were removed at 3-min intervals and added to an equal volume of ice-cold 100 mM EDTA (pH 7.5). These conditions resulted in the removal of about 50 nucleotides per min from each end. Nuclease-treated DNAs were ethanol-precipitated from aliquots containing deletions of appropriate size. After *Bam*HI or *Eco*RI linker addition, the remaining portion of the truncated *thyA* fragment was released from the vector portion with *Hin*dIII and cloned into appropriately cleaved M13mp9.

DNA Sequence Analysis. The sequence of both strands of the DNA was determined by the dideoxy-chain-termination method of Sanger (11, 12) with single-stranded DNA templates of clones of either overlapping BAL-31-generated deletions or restriction fragments (Fig. 2) in vectors M13mp8 or M13mp9 (10).

Protein Sequence Analysis. Thymidylate synthase from *E. coli* K-12 was purified to homogeneity as described (8). The ternary complex was prepared by reacting 300 nmol of enzyme with 6 μmol of [2-^{14}C]FdUMP (diluted to 6,100 cpm/nmol) in the presence of Mg^{2+} and 5,10-methylene tetrahydrofolate at room temperature for 1 hr. Ternary complex formation was measured by a filter assay (13), which indicated that 1.6 mol of FdUMP was fixed per mol of enzyme. The complex was dialyzed vs. 5% acetic acid, after which it was lyophilized and carboxymethylated (14). The carboxymethylated protein was used for amino terminal sequence analysis and conversion to peptides by reaction with CNBr (15). Peptides were partially separated by Bio-Gel P-60 chromatography and purified by HPLC with a C_{18} μBondapak column (Waters Associates) and

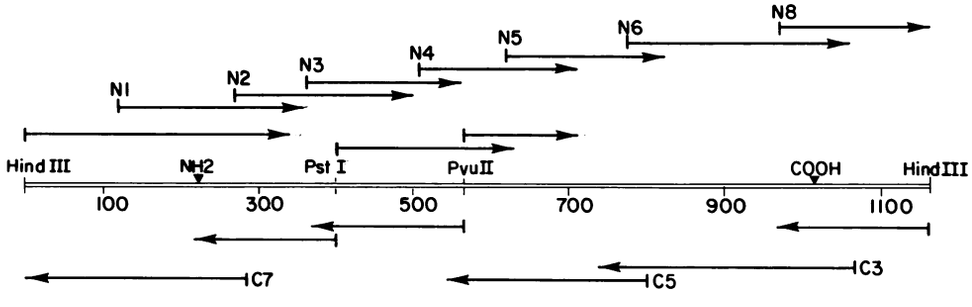

FIG. 2. DNA sequence determination strategy. The double line represents a map of the 1.16-kb *thyA* fragment containing *Hin*dIII linkers at its termini (8). The limits of the structural gene as defined by the amino- (NH_2) and carboxyl-terminal (COOH) residues are shown. The arrows show the direction of sequence analysis and the length of sequence determined either from a particular restriction enzyme site (arrows originating from the *Hin*dIII, *Pst* I, or *Pvu* II sites) or from the end points of selected deletions of either the amino-terminal (N1–N6 and N8) or carboxyl-terminal (C3, C5, and C7) series.

a linear gradient from 0.1% phosphoric acid to 60% acetonitrile/0.1% phosphoric acid (9/1). The amino acid compositions of the enzyme and its CNBr peptides were determined on a Beckman 119CL amino acid analyzer after hydrolysis in constant boiling HCl for 24 hr at 110°C. The amino-terminal sequence of the synthase was determined on an automated Beckman 890B sequencer with 70 nmol of carboxymethylated enzyme using a 0.1 M Quadrol program (16). The thiazolinone derivatives obtained from the sequencer were converted to their respective phenylthiohydantoin-amino acids by treatment with 1 M HCl for 10 min at 80°C and were characterized qualitatively by TLC (17) and quantitatively by HPLC (18). Partial amino-terminal sequences of the CNBr peptides were determined; 25–50 nmol of each peptide and the same methods as described above were used. The carboxyl-terminal amino acids of the synthase were identified by carboxypeptidase A digestion (3).

Data Storage and Analysis. Both DNA and protein sequence data were stored and analyzed by the programs (versions 1 and 2) of Larson and Messing (19).

RESULTS AND DISCUSSION

Nucleotide Sequence Determination. The entire sequence of a 1.16-kb fragment that encodes the *thyA* gene of *E. coli* K-12 was determined in both strands by using the Sanger chain-termination method and the sequence strategy depicted in Fig. 2. This procedure involved sequence analysis of about 200 residues from the end points of two series of overlapping DNA deletion fragments, which extended progressively inward (at about 150 nucleotide intervals) from both ends of the gene. Parts of the sequence also were determined outward from *Hin*dIII, *Pst* I, and *Pvu* II restriction enzyme sites. All fragments were cloned into the M13 vectors mp8 or mp9 (10) to generate single-stranded DNA for sequence analysis.

The two deletion libraries were generated *in vitro* from the *thyA* gene cloned into pBR322 (pBTAH) by using BAL-31 nuclease (Fig. 1). A striking and useful aspect of recloning the truncated *thyA* fragments (0.3–1.0 kb) into M13mp9 after cleavage from pBR322 is their preferential representation in the M13 recombinant population by a factor of at least 10:1 over the released pBR322 fragments (3.0–4.0 kb). This bias probably reflects the preference of these M13 vectors for propagating small DNA inserts. Insert size was verified on agarose gels by the appropriate restriction analyses by using rapid DNA preparations (20) of the double-stranded replicative form of recombinant phage.

This nonrandom approach to sequence determination is similar in many respects to that of Poncz *et al.* (21) and has the advantage of rapidity and simplicity, obviating detailed restriction analysis, laborious gel purification of DNA fragments, and the repetitive sequence assay of more random approaches (12, 22, 23). Further, the deletions divide the gene into distinct domains, which may facilitate the localization of point mutations to a specific deletion interval of the gene.

Features of the *thyA* Sequence and Gene Expressivity. We were guided to the initiator ATG* codon of the structural gene (Fig. 3) by the amino-terminal protein sequence of pure *E. coli* K-12 thymidylate synthase obtained from an overproducing strain constructed in this laboratory (8). The 792-nucleotide open reading frame corresponds to the 264-residue enzyme with the initiator methionine residue being retained as the amino-terminal residue of the synthase. The molecular weight of the enzyme subunit calculated from these data is 30,441. The deduced amino acid sequence was corroborated by protein sequence analysis, and a domain of the sequence associated with enzyme function was delineated as described below.

Sequences flanking the coding region of the *thyA* gene (Fig. 3) include a 206-nucleotide leader segment between the 5′ *Hin*dIII linker residues and the ATG start codon. In addition there are 146 nucleotides between the TAA stop codon and the *Hin*dIII linker sequence at the 3′ end of the gene. Analysis of putative control regions within these flanking sequences, as well as codon usage within the structural gene, may provide a rational basis for the low level of thymidylate synthase expression, which amounts to only about 250 enzyme molecules per

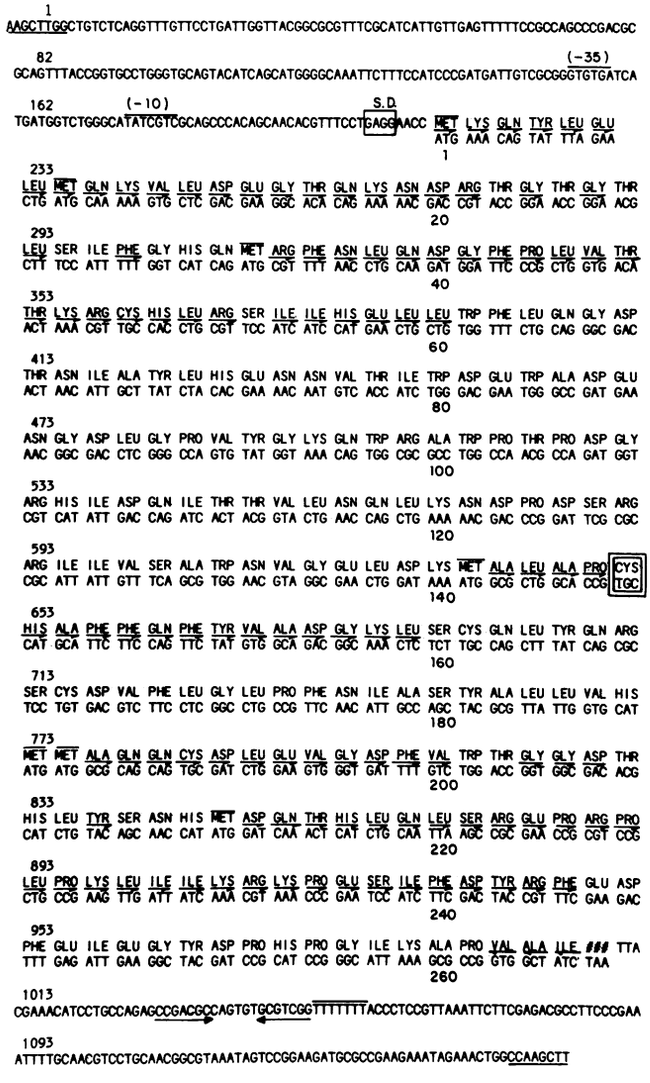

FIG. 3. Sequence of the *thyA* gene and *E. coli* thymidylate synthase. The nucleotides are numbered above the sequence from the 5′ end of the antisense strand. Each number represents the left-most nucleotide on its line. The terminal *Hin*dIII linker sequences are underscored. Putative transcriptional regulatory sequences are represented by bars above the sequences. These include the suspected "−35" and "−10" promoter sequences in the 5′ portion of the sequence and the poly(T) termination sequence in the 3′ region. The preceding seven-nucleotide inverted repeat sequence is represented by converging arrows below the bases. The putative Shine–Dalgarno sequence (S.D.) of the ribosome binding site immediately preceding the structural gene is boxed in. The amino acids of the synthase are numbered below the corresponding codon, beginning at the amino-terminal methionine residue. Those residues confirmed by protein sequence analysis are underscored by arrows indicating the direction of sequence determination. The seven methionines are marked with bars above the residues, whereas the active site cys-146, which binds FdUMP, is boxed in.

* Because the antisense strand is represented in Fig. 3, T in DNA and U in mRNA are used interchangeably in the text.

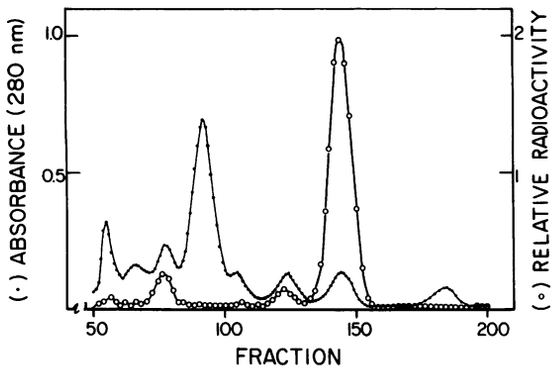

FIG. 4. Separation of CNBr-cleaved peptides obtained from *E. coli* K-12 [FdUMP-5,10-methylenetetrahydrofolate-thymidylate synthase] ternary complex. After cleavage of the modified protein with CNBr, the preparation was lyophilized, dissolved in a small volume of 30% acetic acid, and chromatographed on a column of Bio-Gel P-60 (2 × 115 cm). The volume of each fraction was 1.8 ml.

rapidly dividing *E. coli* K-12 cell.

The region of presumed transcriptional termination indicated by the poly(T) tract following a G·C-rich region of hyphenated dyad symmetry (Fig. 3) conforms well to *E. coli* terminator regions that have been characterized (24). In the promoter region of the gene, there exists homology in four of the seven nucleotides between the putative "−10" sequence, T-A-T-C-G-T-C, and the canonical "−10" Pribnow box heptamer, T-A-T-A-A-T-G. The sequence G-T-G-T-G-A, located 17 nucleotides upstream, shares homology in three of its six positions with the "−35" consensus sequence T-T-G-A-C-A. Whether the low level of *thyA* expression can be attributed to these deviations from the ideal cases in the −10 and −35 RNA polymerase interaction regions must await transcriptional and mutational analysis of the *thyA* gene and its promoter.

Control of enzyme copy number may well be governed also by a translational component, and although a Shine–Dalgarno sequence is evident immediately preceding the coding region, its close proximity (four nucleotides) to the ATG start codon may reduce ribosome binding and, thereby, also translational initiation (25). Further, an analysis of codon usage suggests a bias in the choice of degenerate codons ending in U or C which may lead to a reduction in translational elongation rate. Strong expression has been correlated with the choice of U or C in the "wobble" position such that codon–anticodon interaction energy is optimized, with C preferentially following an AU, UA, AA, or UU doublet and U following CG, GC, CC, or GG (26, 27). In six of these eight cases, the *thyA* gene is biased to an energetically unfavorable codon choice, where isoleucine, tyrosine, phenylalanine, alanine, proline, and glycine are preferentially coded by AUU, UAU, UUU, GCC, CCC, and GGC, respectively. Although other factors undoubtedly come in to play, this feature is characteristic of weakly expressed genes (26, 27).

Protein Sequence Determination and Identification of the FdUMP Binding Residue. To determine the protein sequence at the active site, the synthase was labeled with [2-^{14}C]FdUMP in the presence of 5,10-methylenetetrahydrofolate. The residual sulfhydryl groups were blocked by *S*-carboxymethylation, and the resulting modified protein was cleaved with CNBr to yield six peptides, in contrast to the five that were found for the *L. casei* synthase (3). The six peptides were partially separated by chromatography in 30% acetic acid on a Bio-Gel P-60 column (Fig. 4), where the peptide fractions were identified by scanning in the 280-nm region and the FdUMP-containing peptide was detected by radioactivity measurement. The latter peptide was eluted between fractions 140 and 152 (Fig. 4). Each peak was pooled and further purified by HPLC as described. The amino acid compositions of each of the six peptides are in Table 1. Their purity enabled the partial sequence determination of peptides CB3 to CB6 in duplicate, starting with 25–50 nmol per sequence run. The amino-terminal sequence of the intact modified protein included the complete sequence of CB1 and part of CB2. Amino acid residues with known sequence are indicated by the arrows in Fig. 3 and amount to 40% of that established by nucleotide sequence determination. In all cases

Table 1. Amino acid compositions of peptides obtained by CNBr cleavage of *E. coli* K-12 thymidylate synthase

Amino acid	CB1	CB2	CB3	CB4	CB5	CB6	*S*Cm-enzyme
CmCys			0.7 (1)	2.7 (3)	0.9 (1)		4.6 (4)
Asp		3.0 (3)	16.6 (18)	3.0 (3)	4.2 (4)	4.3 (1)	31.8 (32)
Thr		3.8 (4)	6.5 (7)		2.0 (2)	1.3 (1)	14.2 (14)
Ser		1.0 (1)	2.8 (3)	2.8 (3)	1.0 (1)	2.1 (2)	10.2 (10)
Glu	1.6 (2)	3.6 (4)	9.3 (10)	3.2 (3)	3.3 (3)	7.2 (7)	29.4 (29)
Pro			4.6 (5)	1.9 (2)		6.8 (7)	13.9 (14)
Gly		4.2 (4)	6.8 (7)	2.0 (2)	3.9 (3)	2.2 (2)	18.5 (18)
Ala			3.5 (4)	5.8 (6)	1.1 (1)	1.9 (2)	13.8 (13)
Val		1.0 (1)	5.4 (6)	2.7 (3)	2.1 (2)	1.0 (1)	14.4 (13)
Met							6.3 (7)
Ile		1.0 (1)	6.9 (8)	0.9 (1)		4.8 (6)	15.7 (16)
Leu	1.8 (2)	2.1 (2)	10.7 (11)	6.6 (7)	1.9 (2)	4.1 (4)	28.1 (28)
Tyr	1.0 (1)		1.7 (2)	2.8 (3)	1.0 (1)	2.0 (2)	6.5 (9)
Phe		1.0 (1)	2.6 (3)	4.7 (5)	1.0 (1)	2.9 (3)	12.0 (13)
Lys	1.0 (1)	2.4 (2)	4.3 (4)	0.9 (1)		3.8 (4)	11.6 (12)
His		0.9 (1)	4.3 (4)	1.9 (2)	1.9 (2)	2.0 (2)	10.1 (11)
Arg		1.0 (1)	6.9 (7)	1.0 (1)		3.1 (4)	12.2 (13)
Hse	0.9 (1)	0.7 (1)	0.6 (1)	0.7 (1)	0.6 (1)		
Trp			ND (6)		ND (1)		5.7 (7)
Residue no.	2–8	9–34	35–141	142–187	189–213	214–264	1–264

Numbers in parenthesis are residues per mol and were obtained from the nucleic acid sequence data. Other values in residue numbers were obtained from 24-hr hydrolysis. The yield of the peptides ranged from 24% to 57%. Tryptophan was determined after hydrolysis in 4 nM methanesulfonic acid. ND, not determined; SCm, *S*-carboxymethyl.

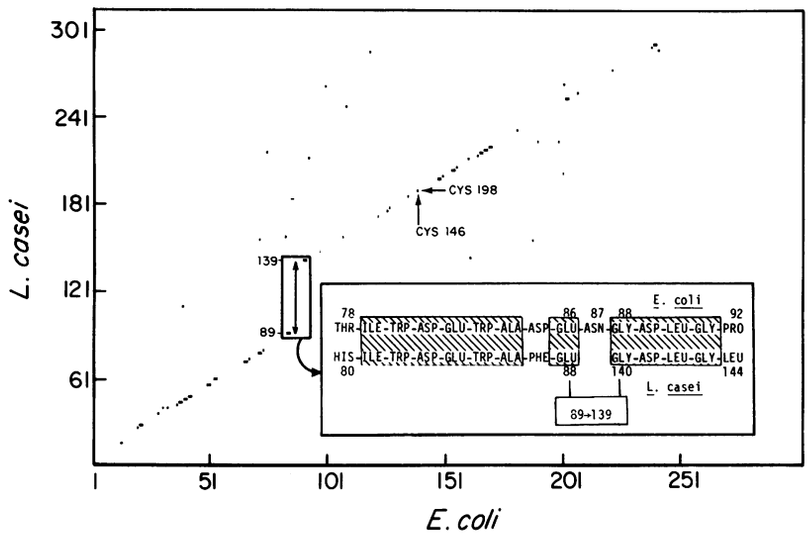

FIG. 5. Matrix comparison of the thymidylate synthase sequences of *E. coli* and *L. casei*. Three consecutive amino acid matches between the 264-residue sequence of the *E. coli* enzyme and the 316-residue *L. casei* sequence have been plotted by using the matrix comparison option of version 2 of the Larson and Messing sequence program (19). The 51-residue *E. coli* "deletion" is represented by a corresponding break in the diagonal homology plot. The inset is an amino acid comparison of the two sequences in the area immediately surrounding the 51-residue disparity.

complete agreement between the two methods was obtained. As in the case of the *L. casei* synthase (4), FdUMP was located at the amino terminus of CB4, which on sequence determination revealed that this nucleotide was fixed at cysteine-146. Sequence analysis of CB1 and carboxypeptidase A treatment of the intact protein clearly established the amino- and carboxyl-terminal ends of the synthase structural gene within its 1.16-kb DNA fragment.

Comparison Between the Synthases of *E. coli* and *L. casei*. The *E. coli* synthase shows 62% homology to the amino acid sequence of the *L. casei* enzyme, which was determined previously in this laboratory (28). The strong sequence conservation is evident from the matrix comparison shown in Fig. 5. Another feature evident from this plot is a block of 51 amino acids of the *L. casei* sequence that is absent from the otherwise similar *E. coli* synthase. Although this region of the *E. coli* protein sequence was deduced strictly from nucleotide sequence, we believe the "deletion" to be real on the following grounds: (*i*) The sequence of the region was determined several times in both strands by using a number of different clones (Fig. 2); (*ii*) the area of interest falls in the *Pst* I–*Pvu* II interval of the *thyA* gene, which, when sized on polyacrylamide gels, agrees exactly with that deduced from the DNA sequence data (not shown); (*iii*) the *E. coli* enzyme subunit is about 5,000 daltons smaller than the *L. casei* monomer (3, 8), corresponding to a 50-amino-acid difference; and (*iv*) the amino acid analysis of both the whole *E. coli* protein and its CNBr peptides agrees with that obtained from the sequence deduced from the nucleotide data (Table 1). The functional significance of this surprising finding is currently unknown.

The FdUMP binding residues are encompassed in the areas of greatest sequence homology between the two synthases and cover a very hydrophobic 42-residue region in which the two proteins are 83% homologous. This region spans residues 145–186 in the *E. coli* enzyme and amino acids 197–288 in the *L. casei* synthase and probably forms part of the active center of the synthases. The above studies provide the basis for a phylogenetic comparison of this enzyme from several species; also the complete nucleotide sequence of the *E. coli thyA* cistron will facilitate the delineation of functional domains at both the genetic and enzymic levels by systematic mutagenesis studies. This information should define more precisely those regions and residues involved in thymidylate synthase activity and expression.

We thank Don Guarino and Judy Reidl for their capable technical help and Rick Lipinskas for his assistance with the computer programs. We also thank Judy Valentino for her meticulous care in preparing the manuscript. This work was supported in part by grants from the National Science Foundation (PCM8118368) and the National Institutes of Health (GM26387 and GM26645).

1. Heidelberger, C. (1970) *Cancer Res.* **30**, 1549–1569.
2. Bertino, J. R. (1979) *Cancer Res.* **39**, 293–304.
3. Maley, G. F., Bellisario, R. L., Guarino, D. U. & Maley, F. (1979) *J. Biol. Chem.* **254**, 1288–1295.
4. Bellisario, R. L., Maley, G. F., Guarino, D. U. & Maley, F. (1979) *J. Biol. Chem.* **254**, 1296–1300.
5. Maley, G. F., Maley, F. & Baugh, C. M. (1982) *Arch. Biochem. Biophys.* **216**, 551–558.
6. Dunlap, R. B., Harding, N. G. L. & Huennekens, F. M. (1971) *Biochemistry* **10**, 88–97.
7. Leary, R. P. & Kisliuk, R. L. (1971) *Prep. Biochem.* **1**, 47–54.
8. Belfort, M., Maley, G. F. & Maley, F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1858–1861.
9. Belfort, M., Moelleken, A., Maley, G. F. & Maley, F. (1983) *J. Biol. Chem.* **258**, 2045–2051.
10. Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276.
11. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
12. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178.
13. Santi, D. V., McHenry, C. S. & Sommer, H. (1974) *Biochemistry* **13**, 471–480.
14. Kuhn, R. W., Walsh, K. A. & Neurath, H. (1974) *Biochemistry* **13**, 6720–6726.
15. Steers, E., Jr., Craven, G. R., Anfinsen, C. B. & Bethune, J. L. (1965) *J. Biol. Chem.* **240**, 2478–2484.
16. Brauer, A. W., Margolies, M. N. & Haber, F. (1975) *Biochemistry* **14**, 3029–3035.
17. Inglis, A. S. & Nicholls, P. W. (1973) *J. Chromatogr.* **79**, 344–346.
18. Downing, M. R. & Mann, K. G. (1976) *Anal. Biochem.* **74**, 298–319.
19. Larson, R. & Messing, J. (1983) *DNA* **2**, 31–35.
20. Holmes, D. S. & Quigley, M. (1981) *Anal. Biochem.* **114**, 193–197.
21. Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E. & Surrey, S. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4298–4302.
22. Heidecker, G., Messing, J. & Gronenborn, B. (1980) *Gene* **10**, 69–73.
23. Anderson, S. (1981) *Nucleic Acids Res.* **9**, 3015–3027.
24. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13**, 319–353.
25. Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S. & Stormo, G. (1981) *Annu. Rev. Microbiol.* **35**, 365–403.
26. Grosjean, H. & Fiers, W. (1982) *Gene* **18**, 199–209.
27. Gouy, M. & Gautier, C. (1982) *Nucleic Acids Res.* **10**, 7055–7074.
28. Maley, G. F., Bellisario, R. L., Guarino, D. U. & Maley, F. (1979) *J. Biol. Chem.* **254**, 1301–1304.