Sequence of a cDNA for Mouse Thymidylate Synthase Reveals Striking Similarity with the Prokaryotic Enzyme

Sylvia M. Perryman, Cindy Rossana, Tiliang Deng, Elio F. Vanin, and Lee F. Johnson

Department of Biochemistry, The Ohio State University

We report the nucleotide sequence of a cloned cDNA, pMTS-3, that contains a 1-kb insert corresponding to mouse thymidylate synthase (E.C. 2.1.1.45). The open reading frame of 921 nucleotides from the first AUG to the termination codon specifies a protein with a molecular mass of 34,962 daltons. The predicted amino acid sequence is 90% identical with that of the human enzyme. The mouse sequence also has an extremely high degree of similarity (as much as 55% identity) with prokaryotic thymidylate synthase sequences, indicating that thymidylate synthase is among the most highly conserved proteins studied to date. The similarity is especially pronounced (as much as 80% identity) in the 44-amino-acid region encompassing the binding site for deoxyuridylic acid. The cDNA sequence also suggests that mouse thymidylate synthase mRNA lacks a 3' untranslated region, since the termination codon, UAA, is followed immediately by a poly(A) segment.

Introduction

Thymidylate synthase (E.C. 2.1.1.45) converts dUMP to dTMP by reductive transfer of a methyl group from 5,10-methylenetetrahydrofolate. The enzyme is necessary for the de novo synthesis of TMP and is essential in rapidly proliferating cells. For this reason, it is an important target enzyme for chemotherapeutic drugs (Hartmann and Heidelberger 1961; Danenberg 1977). Thymidylate synthase activity is much greater in exponentially growing cells than in nondividing cells (Maley and Maley 1960; Conrad and Ruddle 1972; Navalgund et al. 1980; Rode et al. 1980) and is synthesized primarily, if not exclusively, during the S phase of the cell cycle (Storms et al. 1984; Jenh et al. 1985a).

To facilitate the analysis of thymidylate synthase and its regulation, an FdUrd-resistant mouse 3T6 cell line (LU3-7) was isolated. The LU3-7 cells overproduce thymidylate synthase by a factor of 50–100 compared with the parental cells (Rossana et al. 1982). By several different criteria, the overproduced enzyme appears to be the same as the wild-type enzyme. However, minor differences cannot be ruled out at present. The overproduction of thymidylate synthase in LU3-7 cells is due to a corresponding increase in the amount of thymidylate synthase mRNA (Geyer and Johnson 1984) and the number of copies of the thymidylate synthase gene (Jenh et al. 1985a). Northern and Southern blot analyses indicate that the size of the mRNA and the physical map of the gene are the same in the overproducing and parental cell

1. Key words: mouse thymidylate synthase, sequence conservation, cDNA sequence, poly(A).

Address for correspondence and reprints: Dr. Lee F. Johnson, Department of Biochemistry, The Ohio State University, Columbus, Ohio 43210.
lines. Since thymidylate synthase gene expression is regulated in LU3-7 cells in a manner similar to that in 3T6 cells (Navalgund et al. 1980; Jenh et al. 1985c), this cell line has been used as a convenient model system for investigating the content and metabolism of thymidylate synthase (Jenh et al. 1985c) and its mRNA (Jenh et al. 1985b).

Previously, we have isolated a number of cDNA clones corresponding to thymidylate synthase mRNA from the overproducing cell line (Geyer and Johnson 1984). The cDNAs were inserted into pBR322 at the PstI site using the poly(dG)/poly(dC) tailing procedure. Several cDNA inserts appeared to be of sufficient length to encode the entire protein. In this paper, we report the DNA sequence of one of these thymidylate synthase cDNA plasmids, pMTS-3. Analysis of this sequence revealed a striking degree of similarity with previously reported sequences of prokaryotic and human thymidylate synthase as well as the apparent absence of a 3' untranslated region on the thymidylate synthase mRNA.

Material and Methods
Strains, Plasmids, and Reagents

The cDNA plasmids pMTS-3 and pMTS-4 (Geyer and Johnson 1984) were maintained in E. coli HB101. M13 derivatives were maintained in JM101. M13 sequencing kits were purchased from New England Biolabs. Restriction endonucleases were obtained from Bethesda Research Laboratories and International Biotechnologies, Inc. Calf intestinal alkaline phosphatase was obtained from Boehringer Mannheim. T4 polynucleotide kinase was obtained from P-L Biochemicals and Amersham. Isotopes were obtained from New England Nuclear.

DNA Sequence Analysis

Sequence analysis of 5' end-labeled fragments was carried out essentially as described by Maxam and Gilbert (1977) using reactions specific for G, G + A, A + C, C + T, and C bases. The nucleotide sequence was confirmed by the dideoxy chain-termination method of Sanger et al. (1977). Restriction fragments of pMTS-3 or pMTS-4 were subcloned into M13 (mp8, mp9, mp18, or mp19), and the single-stranded viral DNA was used as a template for sequence analysis. Sequencing gels were 6%, 8%, or 20% polyacrylamide gels with an acrylamide:bis acrylamide ratio of 19:1. The gels contained 47% urea and 100 mM Tris-borate, pH 8.3, 2 mM EDTA and were run as described by Maxam and Gilbert (1977) and Sanger et al. (1977). Sequences were analyzed using the DNASATR computer program (provided by F. Blattner and J. Schroeder, University of Wisconsin—Madison).

Results and Discussion
Sequence Determination

The longest of the cDNA clones obtained by Geyer and Johnson (1984), pMTS-4, had a cDNA insert of ~1.1 kb. However, preliminary sequence analysis revealed that 130 nucleotides at the 5' end of pMTS-4 were an inverted repeat of the sequence at the 3' end of pMTS-4 (data not shown). This was most likely the result of a reverse transcriptase cloning artifact. The same type of artifact was observed previously for other cloned cDNAs (e.g., Weaver et al. 1981). For this reason, we focused our analysis on pMTS-3, which had a cDNA insert of ~1.0 kb. The cDNA sequences of pMTS-3 and pMTS-4 are the same except for the 5' end.
Figure 1 shows the restriction map and the strategy for DNA sequence determination of pMTS-3. The nucleotide sequence of the coding strand of the insert DNA is shown in figure 2. Computer analysis of all six possible reading frames revealed a single long open reading frame, which is shown below the nucleotide sequence. The first ATG start codon is shown at position 1. The TAA stop codon is indicated by asterisks below the codon. The stop codon is followed immediately by a stretch of A nucleotides, which probably represents the poly(A) tail of the mRNA (see below). The 5' and 3' ends of the cDNA insert are flanked by oligo(dG) and oligo(dC), respectively, which were added during the cloning of the cDNA (Geyer and Johnson 1984).

The 921-nucleotide sequence from the first ATG to the TAA stop codon predicts a protein with 307 amino acids and a molecular mass of 34,962 daltons. This is slightly lower than the molecular mass of mouse thymidylate synthase protein as estimated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (38,000 daltons) (Rode et al. 1979; Rossana et al. 1982) but similar to that for the human enzyme (35,706 daltons) as predicted by the cDNA sequence (Takeishi et al. 1985).

At present we cannot be certain that the cDNA represents the entire coding region of the enzyme. The cDNA appears to be complete at the 3' end, since poly(A) is present. However, it is possible that the cDNA does not extend to the initiation codon, since the open reading frame continues upstream of the first AUG codon. Thus a small portion of the amino acid sequence at the amino terminal region of the enzyme may be missing. Several attempts were made to determine the amino acid sequence at the amino terminus of mouse thymidylate synthase. Unfortunately, these attempts were unsuccessful, suggesting that the terminal amino acid of the mouse enzyme may be blocked (J. Freisheim and L. F. Johnson, unpublished data). However, the similarity between the mouse and human thymidylate synthase sequences at the amino terminus of the human protein (see below) does suggest that the entire mouse encoding region is represented.

Conservation of Amino Acid Sequences

In figure 3, the predicted amino acid sequence of mouse thymidylate synthase is compared with the sequences of thymidylate synthase from human (Takeishi et al. 1985) and rat (Bargon et al. 1984). The amino acid sequence of rat thymidylate synthase is slightly shorter than that of the human enzyme, and the two sequences differ in six positions.

FIG. 1.—Restriction map and sequencing strategy for pMTS-3. For clarity, only restriction enzymes with 6-base recognition sequences are shown on the map. Restriction sites are represented as follows: B, BamHI; C, ClaI; G, BglII; P, PstI; S, SphI. The arrows indicate the direction and extent to which the sequences were determined. Sequences determined by the Maxam and Gilbert (1977) procedure are indicated below the restriction map, while sequences determined by the M-13/Sanger procedure are shown above the map. The thick lines at both ends of the cDNA insert represent pBR322. Sequences were read across all restriction sites. Approximately 80% of the sequence was determined on both DNA strands.
FIG. 2.—Nucleotide sequence and predicted amino acid sequence of the cDNA insert of pMTS-3. Numbering of both sequences begins at the first AUG codon. The termination codon TAA is indicated with asterisks. The poly(C) and poly(G) at the ends of the sequence were added during the cloning of the cDNA and are immediately adjacent to the PstI sites (not shown) that flank the cDNA insert.

Table of Nucleotide and Amino Acid Sequences:

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATG CTG GTG CTG GCC TCC GAG CTG CAG</td>
<td>Met Leu Val Val Gly Ser Glu Leu Glu Ser Asp Ala Glu Glu Leu Ser Ala Glu</td>
</tr>
<tr>
<td>GCC CCA CGG CAT GGA GAA CTC CAG TAC CTG AGG CAG GAC GAG GTC AGC</td>
<td>Ala Pro Arg Gly Gln Cys Arg Asp</td>
</tr>
<tr>
<td>TGC GGC TTC AAG AAG GAG GAC CGC AGC ACC CTG TCG CTG GTC TGG TTC</td>
<td>Cys Gly Phe Gly Leu Leu Ser Val</td>
</tr>
<tr>
<td>ATG CAG GCA CGA TAC AGC CTG AGA GAT GAA TTT CCT CTG CTC ACA ACC AAA CCA</td>
<td>Met Glu Ala Arg Tyr Ser Ala Arg Asp Glu</td>
</tr>
<tr>
<td>GTG TTC TGG AAG GGT GTT TTG GAG GAG TTT TAT TGG TTT TAT ATC TGG GAT</td>
<td>Val Phe Trp Leu Gly Leu Leu Tyr Val</td>
</tr>
<tr>
<td>AAT GCT AAA GGA TTG TCC TCA AAG GGA GTG AAT GTA GAT GAA TTT CAA TGG AGG GAC</td>
<td>Asn Ala Lys Glu Leu Ser Glu Leu Leu</td>
</tr>
<tr>
<td>CCA GTT TAT GGT TTC CAA TGG AAG CAT TTT GGA GAG TAC AAA GAT ATG</td>
<td>Pro Val Tyr Gly Phe Glu Leu</td>
</tr>
<tr>
<td>TCA GAT TAC TCG GAA GGA GAG CAG AAT CCA CAT CCA AAG</td>
<td>Asp Glu Leu Ser Asp Glu</td>
</tr>
<tr>
<td>AAA ACC AAC CCT GAT GAC AGA AGA ATC AGT TGG GCC TGG AAG CCA AAA</td>
<td>Lys Thr Asn Asn Pro Arg</td>
</tr>
<tr>
<td>CTG ATT GCC AGC TAT GCT CTG CTC ACC TAC ATG</td>
<td>Leu Pro Met Ala Leu Pro Pro</td>
</tr>
<tr>
<td>AAT GGG GAA CTG TCT TGC CAG CTT GAT CAG AAG TCA GGA GAT ATG GTG GGC</td>
<td>Asn Gly Leu Ser Cys Glu</td>
</tr>
<tr>
<td>GTG CCC TTC AAC ATT GCC AGC TAT GCT CTG CTC ACC ATG</td>
<td>Val Pro Phe Asn Ala Ser Tyr</td>
</tr>
<tr>
<td>ACA GGC CTG CAG CGG GGT GAT TTG GTC CAC ACT TTG GGA GAT GCA ATT TAC</td>
<td>Ala Gln Leu Gly Asp Phe Val</td>
</tr>
<tr>
<td>CTG AAT CAT ATA GAG CCG CTG AAA ATT CAG CTA CAG</td>
<td>Leu Asn His Ile Glu Pro Leu</td>
</tr>
<tr>
<td>CCA AAG CTC AAA ATC CTT CAG AAA GTA GAC ACA ATG GAT TAC AAG</td>
<td>Leu Pro Leu Lys Ile Leu Arg</td>
</tr>
<tr>
<td>GAC TTT CAG ATT GAA GGG TAT AAT CCA CAT CCA ACG ATT AAA ATG GAA ATG</td>
<td>Asp Phe Glu Ile Glu Tyr Asn Pro</td>
</tr>
</tbody>
</table>

Val ***
FIG. 3.—Comparison of amino acid sequences of mammalian and prokaryotic thymidylate synthases. The published sequences of human (H), *L. casei* (L), *E. coli* (E), T4 phage (T), and the φ51 phage of *Bacillus subtilis* (P) are aligned with the amino acid sequence of the mouse (M) enzyme from fig. 2. Alignment was performed by computer analysis as well as by visual inspection. Gaps (blank spaces) were introduced to increase identities. Dots indicate that the amino acid at that position is the same as that in the mouse sequence. The position of the amino acid at the right end of each row is indicated by the numbers to the right of the sequences. Region 1 encompasses the folate binding site, region 2 the dUMP binding site (cys at position 189 of the mouse sequence).
1985), *Lactobacillus casei* (Maley et al. 1979), *E. coli* (Belfort et al. 1983), T4 phage (Chu et al. 1984), and the φ3T phage of *Bacillus subtilis* (Kenny et al. 1985). The sequences are aligned so as to increase the regions of identity. The percentages of amino acids that are identical are shown in table 1. As expected, the mouse and human sequences show the highest overall degree of similarity, 90%. The region of lowest similarity is the amino terminal portion of the protein, at amino acids 1–42 of the mouse sequence. If this region is excluded, the mouse and human protein sequences are ~95% identical.

The mouse and bacterial protein sequences are also quite similar. The overall extent of conservation between the mouse sequence and the *E. coli* sequence (55%) is almost as great as that between *L. casei* and *E. coli* (60%). Even the phage and mouse sequences are 40%–47% identical. This striking similarity between mammalian and bacterial thymidylate synthase sequences was noted previously by Takeishi et al. (1985) in their analysis of the human sequence. Table 1 shows that region 1 (amino acids 44–70 of the *L. casei* sequence), corresponding to the folate binding site (Maley et al. 1982), is conserved to approximately the same extent as the entire protein. The most conserved domain of the protein is region 2, comprising the 44 amino acids (193–236 of the *L. casei* sequence) surrounding the dUMP binding site (Chu et al. 1984). In this region, the mouse and bacterial sequences are ~80% identical.

These data show that thymidylate synthase is among the most conserved proteins that have been examined. Because thymidylate synthase is essential for the biosynthesis of TMP, it was probably present in the earliest organisms having DNA genomes. The fact that the enzyme has the same basic structure in widely divergent organisms suggests that the structure of the enzyme was optimized very early during evolution and implies that many of the conserved amino acids play an important structural or functional role. The amino terminal region is probably less important for the maintenance of structure or function, since it is the least conserved region of the enzyme.

Absence of 3′ Untranslated Region on mRNA

Surprisingly, the termination codon, UAA, was followed immediately by poly(A). The sequence of pMTS-4 was also determined in this region and was the same as that of pMTS-3 except that the poly(A) tail was 21 nucleotides long. If this is the real poly(A) tail of the mRNA, this is the first example of a eukaryotic nuclear-encoded

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Amino Acid Identity between Thymidylate Synthase Sequences of Mouse and Other Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPECIES</td>
<td>IDENTITY (%)</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
</tr>
<tr>
<td>Human</td>
<td></td>
</tr>
<tr>
<td><em>Lactobacillus casei</em></td>
<td>90</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>50</td>
</tr>
<tr>
<td>T4 phage</td>
<td>55</td>
</tr>
<tr>
<td>φ3T of <em>Bacillus subtilis</em></td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>40</td>
</tr>
</tbody>
</table>

NOTE.—Data are derived from fig. 3 and were determined by multiplying the no. of amino acid identities by 100 and dividing the product by the number of positions that are not gapped in either sequence.
mRNA that lacks a 3' noncoding region (except for the poly-A tail). Anderson et al. (1981, 1982) noted that in several human and bovine mitochondrial mRNAs, the UAA stop codon is followed immediately by the poly(A) tail. In these cases, the mitochondrial genes code for the T or TA of the TAA stop codon, and the addition of the poly(A) tail completes the TAA termination codon.

An alternative explanation for our observation is that the mouse thymidylate synthase mRNA may have a short oligo(A) region immediately after the UAA codon, followed by an untranslated region and the real poly(A) tail. During the cloning of thymidylate synthase cDNA, the oligo(dT) primer might have primed reverse transcription starting at the oligo(A) region rather than at the poly(A) tail. To address this possibility, it is necessary to analyze the sequence of the mouse thymidylate synthase gene in the vicinity of the poly(A) addition site. We have recently isolated the entire mouse thymidylate synthase gene from a BamHI genomic library (in preparation). Analysis of the 5' and 3' exons have confirmed the sequences shown in figure 2. Furthermore, an oligo(A) stretch was not found adjacent to the nonsense codon. Therefore it appears that the poly(A) region immediately adjacent to the UAA codon does represent the poly(A) tail of the mRNA.

The human thymidylate synthase mRNA has a 3' untranslated region of nearly 500 nucleotides (Takeishi et al. 1985). However, the human message is $\geq 1.6$ kb in length (Takeishi et al. 1985) compared with $\sim 1.3$ kb for the mouse message. This size difference may arise in part from the absence of a 3' untranslated region in the mouse message.

Although the mouse sequence does not have the AATAAA polyadenylation signal (Proudfoot and Brownlee 1976), there is a sequence ATTTAAA 21 bases upstream of the oligo(A) on the cDNA, in the coding region of the message. This sequence is known to function as a signal for polyadenylation in several other mRNAs and is located at the proper position to serve as the signal (Berget 1984; Birnstiel et al. 1985). The sequences of the mouse and human thymidylate synthase mRNAs are identical from the ATTTAAA up to the poly(A) stretch of the mouse mRNA and are extremely similar upstream from this region. Since the human mRNA is not polyadenylated at the same site as the mouse mRNA, this may indicate that sequences downstream of the poly(A) addition site are important for determining the site of cleavage and polyadenylation (Berget 1984; Birnstiel et al. 1985).

Since the content of thymidylate synthase mRNA in growth-stimulated mouse fibroblasts appears to be regulated in part by controlling the polyadenylation of this RNA species (Jehn et al. 1985b) and since poly(A) appears to be added immediately adjacent to the termination codon, analysis of the mechanism of poly(A) addition for this mRNA species is warranted.

Acknowledgments

We thank Drs. J. DeWille, C.-H. Jenh, and S. Pilistine for comments on the manuscript. This study was supported by Public Health Service grants GM29356, HL34129 (E.F.V.), and CA16058 and grant 83005 from the Ohio Cancer Research Associates. L.F.J. was supported by Faculty Research Award FRA210 from the American Cancer Society.

LITERATURE CITED


WALTER M. FITCH, reviewing editor

Received January 27, 1986.