Biochem. J. (1992) 286, 117–123 (Printed in Great Britain)

Nucleotide sequence and expression of a cDNA encoding rabbit liver cytosolic serine hydroxymethyltransferase

Paula C. BYRNE,*† Peter G. SANDERS* and Keith SNELL†‡

*Molecular Microbiology Group and †Receptors and Cellular Regulation Group, School of Biological Sciences, University of Surrey, Guildford, Surrey GU2 5XH, U.K.

A rabbit liver cDNA library in phage λgt10 was screened using a portion of the coding sequences for rabbit cytosolic serine hydroxymethyltransferase (amino acids 244–420) that had been amplified by PCR, with total rabbit liver RNA as a template. A clone of 2.3 kb (pUS1203) was isolated and the nucleotide sequence showed that it contained an open reading frame of 1452 bp, which coded for serine hydroxymethyltransferase and was flanked by 155 bp at the 5′ end and 653 bp at the 3′ end. The full-length cDNA was cloned into an expression vector and transfected into COS-1 cells. Serine hydroxymethyltransferase activity was increased by 33% in the transfected cells and a new protein band of the appropriate size was seen by SDS/PAGE analysis of proteins extracted from transfected cells. The protein sequence for rabbit cytosolic serine hydroxymethyltransferase derived from the cDNA nucleotide sequence was compared with three other derived or known prokaryotic and eukaryotic sequences. An overall sequence similarity of 34% was noted between all four sequences, whereas the similarity between the rabbit cytosolic and mitochondrial isoforms was 62%.

INTRODUCTION

Serine is an important amino acid used not only in protein synthesis but also for phospholipid biosynthesis and cysteine synthesis. Furthermore, it has a unique and essential role in the biosynthesis *de novo* of purine and pyrimidine nucleotides for DNA replication as it is the major source of one-carbon units in the cell. The utilization of serine for this purpose is initiated by the enzyme serine hydroxymethyltransferase (SHMT), which catalyses the transfer of the one-carbon unit from the side chain of serine to tetrahydrofolate forming 5,10-methylenetetrahydrofolate and glycine. The 5,10-methylenetetrahydrofolate thus formed participates directly in the synthesis of pyrimidines by acting as a carbon donor for thymidylate synthesis. In addition, 5,10-methylenetetrahydrofolate indirectly participates in the biosynthesis of purines by acting as a precursor for 10-formyltetrahydrofolate which provides carbon for purine biosynthesis. Furthermore, the glycine formed from serine in the SHMT reaction acts as a carbon and nitrogen donor for purine synthesis. In tissues such as the central nervous system, where glycine acts as a neurotransmitter, SHMT is likely to play an additional role in providing glycine for this purpose (Daly & Aprison, 1974).

The importance of SHMT in DNA synthesis is reflected by the high levels of activity observed in proliferating cells. Stimulation of resting lymphocytes by the mitogen phytohaemagglutinin is followed by a four- to nine-fold increase in SHMT activity (Thorndike *et al.*, 1979; Eichler *et al.*, 1981). Similarly during the proliferative phase of rat hepatoma 3924A cell growth in culture there is an increase in the specific activity of SHMT coincident with increased incorporation from [3-^{14}C]serine into DNA (Snell *et al.*, 1987). Increased SHMT activity has also been demonstrated in a variety of solid tumour tissues *in vivo*, including rat renal carcinoma (Snell, 1985), mammary carcinoma (Snell, 1989), sarcoma and human colon carcinoma (Snell *et al.*, 1988). The increased activity of SHMT observed in tumour tissues over normal tissues from the same origin points to it as a potential target for enzyme-directed anticancer chemotherapy (Snell *et al.*, 1988).

SHMT is a tetramer, consisting of identical subunits each with a bound pyridoxal 5′-phosphate which acts as a coenzyme. The activity is widely distributed in nature having been reported in prokaryotic, animal and plant cells. In mammals activity is found to be ubiquitous, being present in muscle, colon, liver, kidney, spleen, intestine, heart, lung (Snell, 1984) and the central nervous system (Daly & Aprison, 1974). In eukaryotic cells both mitochondrial and cytosolic isoforms of the enzyme exist. Both isoforms have been purified from rabbit and rat liver (Nakano *et al.*, 1968; Schirch & Peterson, 1980) and the cytosolic form alone from bovine and sheep liver (Jones & Priest, 1976; Ulevitch & Kallen, 1977), and from plants and bacteria (Fujioka, 1969; Rao & Rao, 1982). The cytosolic and mitochondrial enzymes have similar substrate specificities and kinetic constants (Schirch & Peterson, 1980), and the precise cellular role of each isoenzyme form is not yet known. The cytosolic and mitochondrial SHMT proteins from rabbit liver have been purified and sequenced (Martini *et al.*, 1987, 1989). To date no eukaryotic cDNA coding for the enzyme has been isolated. In *Escherichia coli* and *Campylobacter jejuni* the nucleotide sequence of the *gly a* gene, which produces the prokaryotic counterpart of SHMT, has been reported (Plamann *et al.*, 1983; Chan & Bingham, 1990). Here we report the first eukaryotic sequence of cytosolic SHMT cDNA, which we have cloned from rabbit liver, and demonstrate the expression of rabbit SHMT in COS cells, after transfection with the cDNA. The deduced protein sequence is compared with other prokaryotic and eukaryotic sequences.

Abbreviations used: SHMT, serine hydroxymethyltransferase (EC 2.1.2.1); DMEM, Dulbecco's modified minimal essential medium; PBS, phosphate-buffered saline, pH 7.5; 20 × SSC, 3 M-NaCl/0.3 M-sodium citrate.

‡ Present address and address for correspondence and reprint requests: Institute of Cancer Research, Block X, 15 Cotswold Road, Sutton, Surrey SM2 5NG.

The nucleotide sequence of SHMT cDNA will appear in the EMBL/Genbank/DDBJ Nucleotide Sequence Databases under the accession no. Z11846.

MATERIALS AND METHODS

PCR amplification of rabbit SHMT RNA and cDNA cloning

Total RNA (1 μg), extracted from rabbit liver tissue using RNAzol (Biogenesis Ltd., Bournemouth, Dorset, U.K.) was reverse-transcribed by Rous-associated virus 2 reverse transcriptase using 600 ng of oligo 3 (see the Results section and Fig. 3) to prime cDNA synthesis. The reaction was carried out at 42 °C in 50 mM-KCl/10 mM-Tris/HCl (pH 8.3)/1.5 mM-$MgCl_2$/0.01% (w/v) gelatin. The RNA–DNA hybrid product was denatured by boiling for 6 min and then cooled on ice. Oligo 1 (600 ng) (Fig. 3) and 2.5 units of Taq polymerase were then added, and 25 successive rounds of denaturation (1 min at 94 °C), annealing (1 min at 42 °C) and extension (1 min at 72 °C) of the DNA were carried out. After electrophoresis of the PCR products through a 1.2% agarose gel, the DNA was eluted from the gel using NA45 paper (Schleicher & Schuell, Anderman, Kingston-upon-Thames, Surrey, U.K.). The eluted fragment was treated with Klenow enzyme to fill in the 3′ ends, and the 5′ termini were treated with T4 polynucleotide kinase before ligation into *Sma*I-digested dephosphorylated pUC13. Putative recombinant clones were analysed by agarose-gel electrophoresis of the purified DNA. Double-stranded plasmid sequencing of the insert was carried out using the dideoxy chain-termination method as described for Sequenase (USB, Cleveland, OH, U.S.A.). The recombinant pUC13 containing the SHMT sequence was designated pUS1201.

Northern-blot hybridization

Total rabbit liver RNA (5 μg) was denatured in 2.2 M-formaldehyde/50% (v/v) formamide at 70 °C for 10 min before electrophoresis through a 1.5% agarose gel containing 2.2 M-formaldehyde. [^{32}P]dCTP-labelled SHMT cDNA from pUS1201 (Multiprime, Amersham, Aylesbury, Bucks., U.K.) was used to probe a Northern blot of the gel (Sambrook *et al.*, 1989). The membrane was then washed three times in 2 × SSC (20 × SSC = 3 M-NaCl/0.3 M-sodium citrate)/0.1% SDS at 42 °C for 20 min, and finally in 2 × SSC for 10 min at room temperature, before exposure to X-ray film at −70 °C.

Isolation of full-length rabbit SHMT cDNA

A rabbit liver cDNA library (Clontech, Cambridge Biosciences, Cambridge, U.K.) in phage λgt10 was plated out on L-agar, using host strain C600. Plaques were lifted on to Nylon membrane (Hybond N, Amersham) in duplicate. They were denatured for 5 min in 1.5 M-NaCl/0.5 M-NaOH, neutralized for 5 min in 1.5 M-NaCl/0.5 M-Tris and rinsed in 2 × SSC, before being dried and fixed by baking for 2 h at 80 °C.

[^{32}P]dCTP-labelled SHMT cDNA from pUS1201 was hybridized to these filters by standard protocols (Sambrook *et al.*, 1989). The filters were washed three times in 2 × SSC/0.1% SDS (65 °C, 20 min), and rinsed once in 2 × SSC (65 °C, 15 min), before exposure to X-ray film at −70 °C.

Purification of positive phage and subcloning of SHMT cDNAs

Positive plaques were transferred into 1 ml of SM solution [0.1 M-NaCl, 0.01 M-$MgSO_4$,$7H_2O$, 0.05 M-Tris/HCl, pH 7.5, 0.01% (w/v) gelatin] containing two drops of chloroform and stored at 4 °C. Dilutions (10^{-2} and 10^{-3}) of these stocks were plated out on L-agar using host strain C600 and plaques were screened as before. Positive plaques were again identified and four rounds of this purification were carried out until the different phage isolates were in homogeneous populations. The DNA was purified from λ clones containing putative SHMT cDNA and digested with *Eco*RI to release the insert by using standard protocols (Sambrook *et al.*, 1989). The inserts were isolated by Geneclean (Stratatech Scientific, Luton, Beds., U.K.) and ligated into *Eco*RI-digested and dephosphorylated pUC18 and M13mp18. The ligation reactions were transformed into *E. coli* DH5 α cells and positive colonies and plaques were identified by agarose-gel electrophoresis of purified DNA from individual clones. After digestion with *Kpn*I to orientate the inserts, single-stranded DNA was prepared from M13mp18 (Amersham). Single-stranded dideoxy sequencing was performed using Sequenase, and both strands of cDNA were completely sequenced.

Expression of cloned full-length SHMT cDNA in COS-1 cells

The full-length cDNA was ligated into *Eco*RI-digested and dephosphorylated pUS1000 where expression is driven by the human cytomegalovirus promoter (Jackson *et al.*, 1991). The ligation reaction was transformed into *E. coli* DH5 α cells and transformants were checked for correct orientation with respect to the promoter by digestion with *Kpn*I. Plasmids were purified on a caesium chloride density gradient (Sambrook *et al.*, 1989). The expression unit is designated pUS1202. Control transfections were carried out by using caesium chloride-purified pUS1000.

Semiconfluent COS-1 cells were first washed with 2 ml of phosphate-buffered saline (PBS), then incubated with a solution containing 1 μg of DNA, 20 μl of DEAE-dextran (10 mg of PBS/ml) and 380 μl of PBS for 30 min at 37 °C. Dulbecco's modified minimal essential medium (DMEM) containing 80 μM-chloroquine was added, and incubation continued for 3 h. The medium was removed by aspiration and the cells were shocked with dimethyl sulphoxide for 2.5 min with 2 ml of DMEM containing 10% dimethyl sulphoxide. This solution was removed and the cells were incubated in 8 ml of DMEM for 72 h.

Cells required for the assay of protein activity were harvested after treatment with trypsin, sonicated in the presence of Triton X-100 and stored on ice for use in the SHMT assay. Cells for RNA extraction were harvested using RNAzol and the extraction was carried out according to the manufacturer's instructions (Biogenesis Ltd.).

For the analysis of proteins by SDS/PAGE, the cells were first radiolabelled by using [^{35}S]methionine. The medium was aspirated off, and cells were washed with Eagle's medium (Eagle, 1955) without methionine. [^{35}S]Methionine (80 μCi) was added with Eagle's medium and cells were incubated for 4 h. The cells were then harvested using RIPA buffer (150 mM-NaCl, 1% Nonidet P40, 0.5% (w/v) deoxycholate, 0.1% (w/v) SDS, 50 mM-Tris) and stored at −70 °C for PAGE.

Reduced proteins were electrophoresed through an SDS/9% (w/v) polyacrylamide gel (Laemmli, 1970) and processed for autoradiography by using Enhance according to the manufacturer's instructions (du Pont, Boston, MA, U.S.A.).

Serine hydroxymethyltransferase assay

SHMT activity was assayed by a method (Snell, 1989) based on the radiometric method of Taylor & Weissbach (1965). Activity is expressed as nmol of [^{14}C]formaldehyde trapped/h per 10^6 cells.

RESULTS

Preparation of rabbit SHMT probe

The sequences of the primers used to clone, by PCR amplification, a portion of the rabbit SHMT RNA were based on the published protein sequence of rabbit liver cytosolic SHMT (Martini *et al.*, 1987). Degenerate primers biased to the rabbit

codon usage (Maruyama *et al.*, 1986) were designed from amino acid residues 244–249 for oligo 1 on the sense strand, and amino acid residues 414–420 for oligo 3 on the antisense strand (for positions see Fig. 3).

Oligo 1: 5′TTCGAA$_G$CAT$_C$TGT$_C$CAT$_C$GT3′

Oligo 3: 5′ACC$_T$TTCTGGAAA$_G$TCT$_C$TTT$_C$TC3′

These primers were used to amplify RNA extracted from rabbit liver. The primers span a region of 176 amino acids, which includes several residues that contribute to the putative active site of the enzyme and also the binding site for pyridoxal 5′-phosphate. When the products of the PCR reaction were electrophoresed on an agarose gel, a band of approx. 500 bp was observed (Fig. 1). This band was purified and blunt-end-cloned into pUC13 to produce pUS1201. The nucleotide sequence of this fragment (Fig. 3, nucleotide 888–1415), obtained by the dideoxy chain-termination method, confirmed that the amplified region coded for amino acids 244–420 of the cytosolic form of SHMT. This fragment of DNA was subsequently used to screen a rabbit liver cDNA library.

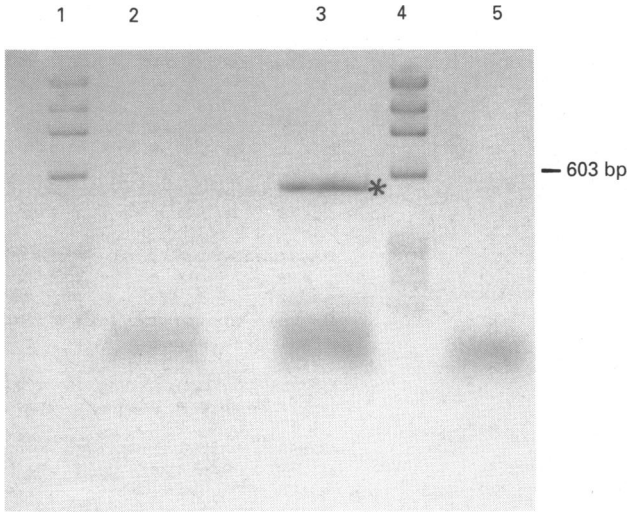

Fig. 1. PCR amplification of rabbit liver RNA

Total RNA extracted from rabbit liver was used as a template for reverse transcription followed by amplification using PCR as described in the Materials and methods section. Products of the PCR reaction were electrophoresed on a 1.2% agarose gel. Lane 1, DNA size markers, ϕX174 digested with *Hae*III; lane 2, negative control, no RNA; lane 3, amplified product of PCR reaction; lane 4, DNA size markers, ϕX174 digested with *Hae*III; lane 5, negative control. * indicates the amplified band.

Northern-blot analysis of rabbit liver RNA

Total RNA extracted from rabbit liver was electrophoresed and transferred on to Nylon membrane for probing with SHMT cDNA from pUS1201. ^{32}P-labelled cDNA from pUS1201 hybridized to three bands of sizes of 1.5 kb, 2.2 kb and very faintly 3.4 kb (Fig. 2). The cytosolic SHMT protein is reported to be 483 amino acids long (Martini *et al.*, 1987) and therefore would require 1449 nucleotides of coding sequence. The relationship between the different RNA species has not yet been determined and cross-hybridization with RNA coding for the mitochondrial form of the enzyme is possible. The degree of sequence similarity between the cytosolic and the mitochondrial enzyme at the amino acid level is 61.9% (Martini *et al.*, 1989). The band that hybridized most intensely to the SHMT cDNA probe is the 2.2 kb band. Subsequently a clone of 2.2 kb was isolated from a rabbit liver cDNA library and shown to contain all the coding sequences necessary to code for cytosolic SHMT as well as 155 bp of 5′ non-coding and 653 bp of 3′ non-coding sequences.

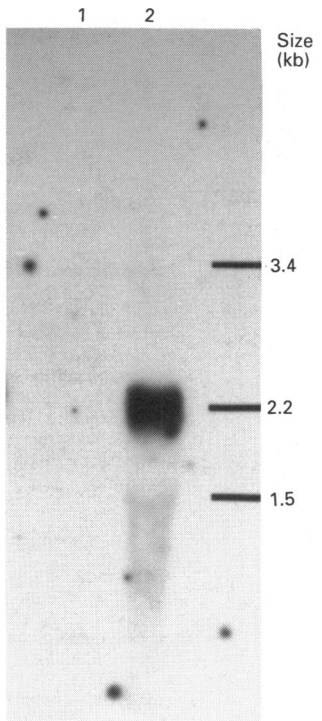

Fig. 2. Analysis of SHMT mRNA in rabbit liver

Northern-blot analysis of rabbit liver RNA using ^{32}P-labelled SHMT cDNA from pUS1201 as a probe was carried out as described in the Materials and methods section. Lane 1, control, no RNA; lane 2, RNA extracted from rabbit liver tissue.

Isolation of cDNA coding for rabbit cytosolic SHMT

Some 300000 clones of a rabbit liver cDNA library were screened using [^{32}P]dCTP-labelled SHMT cDNA from pUS1201 as a probe. On a first round of screening, 67 positive clones were identified, giving a frequency of approx. 1 in 4000 clones being SHMT. Of these clones, 20 were taken through four successive rounds of purification until each clone was in a homogeneous population. The DNA was extracted from these λ-clones and digested with *Eco*RI to release the inserted cDNA. The inserts ranged in size from 1.4 kb to 2.3 kb. Three clones (pUS1203, pUS1204, pUS1205) were analysed further and the inserts cloned into pUC18 and M13mp18 for sequencing and characterization. The inserts of these clones were 2.3 kb, 2.2 kb and 1.8 kb respectively and each could potentially have contained all the SHMT coding sequences. A *Kpn*I digestion of pUS1203 indicated that it most likely contained all of the coding sequences.

Nucleotide sequencing of SHMT

Both strands of cDNA were sequenced by using synthetic primers to 'walk' along the entire cDNA sequence in both directions. Fig. 3 shows that the insert contains 2.261 kb with an open reading frame of 1452 bp. This open reading frame is flanked by 155 bp of 5′ untranslated sequences of 653 bp of 3′ untranslated sequences. The ATG initiation codon is found at position 156. The deduced amino acid sequence obtained from the DNA sequence of pUS1203 corresponds exactly to the published protein sequence of cytosolic SHMT obtained by characterization of tryptic and CNBr peptides of the protein

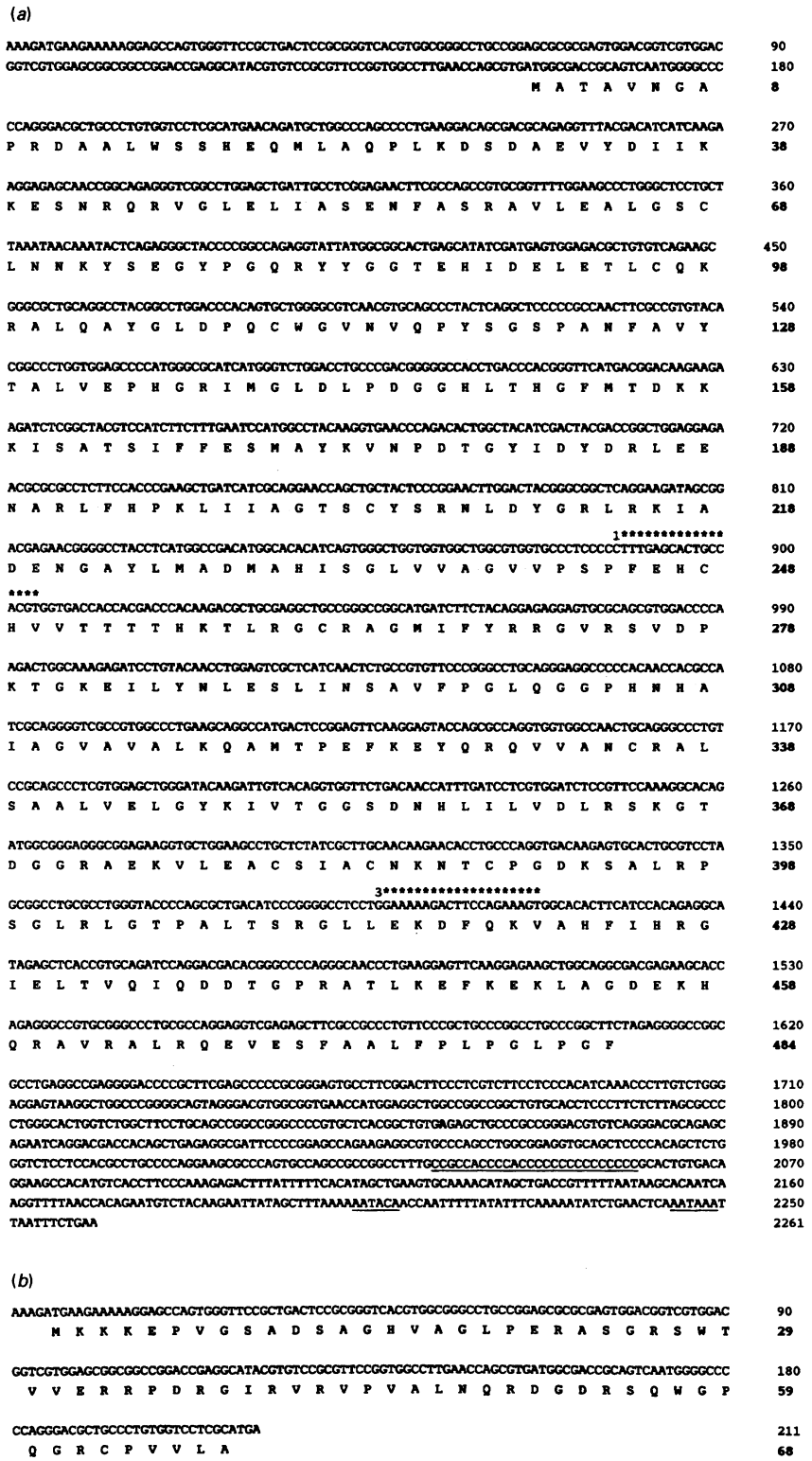

Fig. 3. Complete nucleotide sequence and deduced amino acid sequence of rabbit cytosolic SHMT

(*a*) Sequences for polyadenylation and a long C-rich region in the 3′ tail are underlined. The sequence positions corresponding to the oligonucleotides, oligo 1 and oligo 3, for PCR cloning are marked 1**** and 3**** respectively. (*b*) Nucleotide and deduced amino acid sequence for the open reading frame in the SHMT cDNA which starts upstream of, and overlaps, the region coding for SHMT.

(Martini *et al.*, 1987). The open reading frame terminates with a TAG stop codon at position 1608, and this is followed by 653 bp of 3′ untranslated sequence. Two consensus sequences for polyadenylation are present within this 3′ untranslated region at position 2204–2209 AAUACA and position 2244–2249 AAUAAA. No poly(A) tail was observed. The clone ends 12 bp downstream of the main poly(A) addition sequence at position 2261 with the *Eco*RI adaptor sequence.

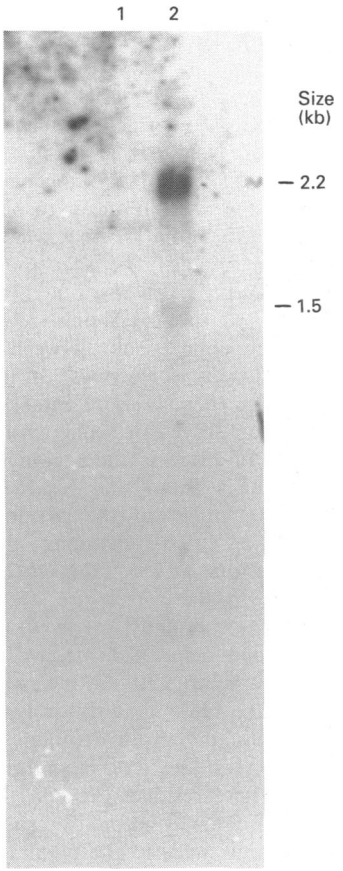

Fig. 4. Northern-blot analysis of RNA extracted from transfected COS-1 cells

Northern-blot analysis of RNA extracted from transfected COS-1 cells using ^{32}P-labelled SHMT cDNA from pUS1201 as a probe was carried out as described in the Materials and methods section. Lane 1, RNA extracted from control vector pUS1000-transfected cells; lane 2, RNA extracted from pUS1202-transfected cells. RNA sizes were determined by comparison with the mobility of marker RNAs (BRL, Uxbridge, Middx., U.K.) of known size.

Expression of SHMT in COS-1 cells

pUS1202 and control vector PUS1000 were used to transfect COS-1 cells as described in the Materials and methods section. After 72 h of incubation at 37 °C the cells were harvested and assayed for transcription of SHMT mRNA, production of SHMT protein and SHMT enzyme activity.

RNA extracted from both control- and SHMT-transfected cells was separated by electrophoresis and transferred by Northern blotting to Nylon membrane to which radiolabelled SHMT cDNA from pUS1201 was hybridized. Two bands of 2.2 kb and 1.5 kb were detected in the SHMT transfection, whereas no bands hybridized in the control transfection (Fig. 4). The 528 bp fragment of cDNA from pUS1201 when used as a probe hybridizes only weakly to SHMT RNA from other species which would explain the lack of hybridization signal in the control cells (P. C. Byrne, P. G. Sanders & K. Snell, unpublished work). The RNA species detected were similar in size to those seen in the RNA extracted from rabbit liver. The occurrence of a 1.5 kb band in both rabbit liver and transfected COS-1 cells may indicate further processing of the 2.2 kb SHMT mRNA produced from the cloned cDNA.

[^{35}S]Methionine-radiolabelled proteins extracted from transfected COS-1 cells were separated by electrophoresis on a 9% polyacrylamide gel. A band of the expected size (approx. 53 kDa) was observed in the SHMT transfection that was not seen in the control pUS1000 transfection (results not shown).

To establish that the SHMT produced from the transfected DNA was biologically active, the transfected cells were assayed for SHMT enzyme activity. pUS1000 control-transfected cells showed an activity of 33 nmol of formaldehyde produced/h per 10^6 cells. pUS1202-transfected cells showed 49 nmol of formaldehyde/h per 10^6 cells, giving an increase of approx. 33% in the pUS1202-transfected cells.

DISCUSSION

This paper has reported the isolation, characterization and expression of a full-length cDNA for SHMT. The SHMT cDNA was isolated from a rabbit liver cDNA library using a probe that had been cloned by PCR amplification of rabbit liver RNA. The oligonucleotides used to prime amplification of RNA extracted from rabbit liver were based on the published protein sequence of the cytosolic form of SHMT (Martini *et al.*, 1987). Cloning of the amplified fragment into pUC13 to produce pUS1201 and subsequent sequencing of the insert confirmed that it corresponded to the expected region of the protein (amino acid 244–420). Before using the SHMT cDNA in pUS1201 to screen the cDNA library it was hybridized to RNA extracted from rabbit liver tissue. Discrete bands ranging in size from 1.5 kb to 3.4 kb were detected. pUS1201 was then used to screen the rabbit liver cDNA library and positive clones were identified and subsequently cloned into M13mp18 for sequencing.

Sequencing of one clone, pUS1203, showed it contained all of the coding sequences necessary to code for the entire rabbit cytosolic SHMT. The ATG that is used to initiate translation of SHMT is surrounded by sequences that are in good agreement with the consensus sequence (A/GCCATGG) for eukaryotic initiation (Kozak, 1987). Located 150 nucleotides upstream from this ATG, however, is another ATG codon, located at position 5, which potentially could initiate translation of an open reading frame that is 69 amino acids long. Although this ATG lacks a G at position +4, the sequence surrounding the ATG is still in good agreement with Kozak's consensus sequence. The sequence surrounding this ATG contains an A at position −3. The purine at position −3 (usually an A) is the most highly conserved nucleotide in eukaryotic mRNA initiation sequences, and it is only in the absence of this purine that the G at position +4 is essential for efficient translation (Kozak, 1989). It is possible that this upstream ATG may play a role in regulating the expression of SHMT at the level of ribosome binding, as would be predicted by the scanning model for ribosomal movement along the RNA (Kozak, 1989). A comparison of the open reading frame protein product with the 'swiss-prot' database (Altschul *et al.*, 1990) reveals a weak similarity to the S3 ribosomal protein of *E. coli*. The significance of this, if any, is not known.

Translation from the ATG at position 156 proceeds along the RNA until the ribosome reaches the termination codon TAG at position 1608. This is followed by a long 3′ tail that is 58.8% G-C-rich, and contains a continuous stretch of 15 cytosine nucleotides (Fig. 3). Although no poly(A) tail was included in this clone a consensus sequence required for polyadenylation, AAUAAA, is found at position 2244 (Proudfoot, 1991), with a related poly(A) site, AATACA, at position 2204. Cleavage of the mRNA and polyadenylation usually occur between 20 and 30 nucleotides downstream from the consensus sequence (Proudfoot, 1991).

The only nucleotide sequence available for SHMT at the eukaryotic level is the sequence presented here for the rabbit liver SHMT cDNA. At the prokaryotic level, the sequence of the *gly a*

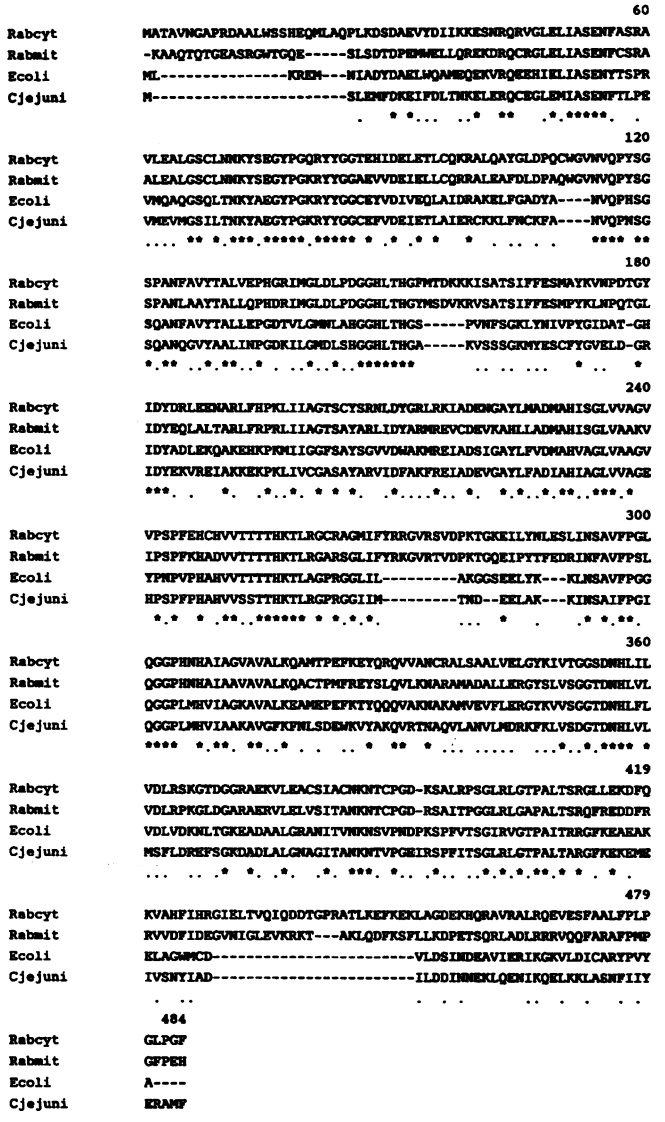

Fig. 5. Comparison of rabbit, *E. coli* and *C. jejuni* SHMT amino acid sequences

Rabcyt, rabbit cytoplasmic SHMT; rabmit, rabbit mitochondrial SHMT; Ecoli, *E. coli* SHMT; Cjejuni, *C. jejuni* SHMT. Numbering is for rabbit cytosolic SHMT. * indicates identical amino acids in all proteins; . indicates conservative substitutions. Alignment was performed using the Clustal program (Higgins & Sharp, 1988, 1989).

gene from *E. coli* and *C. jejuni* have been published (Plamann *et al.*, 1983; Chan & Bingham, 1990). A comparison of the codon usage of these sequences shows the higher G-C content of the eukaryotic rabbit sequence. The coding sequences of rabbit cDNA are 62.6% G-C-rich compared with 53.5% for *E. coli* and 33% for *C. jejuni*. This difference in G-C content is clearly seen in the example of the amino acid alanine. Of the four possible codons that code for alanine, GCT, GCA, GCC and GCG, only 10.9% are either GCC or GCG in *C. jejuni* and 36% in *E. coli* compared with 72.3% in the rabbit sequence.

The deduced amino acid sequence of rabbit liver cytosolic SHMT is compared in Fig. 5 with that deduced from the *gly a* gene from *E. coli* (Plamann *et al.*, 1983) and *C. jejuni* (Chan & Bingham, 1990) as well as with the published protein sequence of the rabbit liver mitochondrial isoenzyme (Martini *et al.*, 1989). The rabbit cytosolic protein sequence is the largest at 484 amino acids, followed by the mitochondrial isoenzyme at 475 amino acids, with the bacterial proteins being somewhat smaller (417, *E. coli*; 414, *C. jejuni*). There is extensive sequence similarity between the two mammalian isoenzymes (61.9%) and between the two bacterial enzymes (55.6%), whereas the similarity between all four protein sequences is much lower (34%). It is noteworthy that of those residues that are conserved between the two bacterial species and one or other of the mammalian proteins, more are uniquely conserved in the cytosolic isoenzyme (21 residues) than in the mitochondrial isoenzyme (13 residues).

The lesser sequence similarity between the prokaryotic and eukaryotic enzymes can be largely attributed to the considerable variation in the *N*- and the *C*-terminal regions (and this is also the case for the comparison between the mammalian mitochondrial and cytosolic isoenzymes). In the central regions of the protein sequence there is more extensive similarity and numerous regions of conserved or semiconserved amino acids. The most significant of these is the region from amino acid 247–261, which includes the lysine residue (257) which is implicated in the attachment of the pyridoxal 5′-phosphate coenzyme (Bossa *et al.*, 1976; Plamann *et al.*, 1983). The conserved histidine residue at 256, adjacent to this lysine, has also been implicated in the active site of the enzyme and, although apparently not directly involved in catalysis, may interact with the amino acid substrate, or the pyridoxal 5′-phosphate (Hopkins & Schirch, 1986). It has been suggested by Schirch (Martini *et al.*, 1987) that the polyglutamate-binding domain for the polyglutamyltetrahydrofolate cofactor may be located at three lysine residues (corresponding to amino acids 191, 194 and 196). These residues are conserved between the bacterial proteins and either conserved or substituted by other positively charged amino acids in the mammalian isoenzymes. On the other hand, an essential thiol group in the mammalian cytosolic isoenzyme (Cys-204; Schirch & Peterson, 1980) must be a unique feature of that protein since it is substituted for by alanine in all the other protein sequences, as is also the case for other cysteine residues (248 and 335) in the region of the active-site lysine.

Despite the large extent of conservation between the proteins in their central regions, there are two extended insertions in the eukaryotic isoenzymes that are absent from the prokaryotic proteins (154–158 and 269–277). These insertion regions may represent amino acids that are involved in generating regulatory sites in the mammalian enzymes. Allosteric sites are unlikely since no allosteric effectors have been identified for the mammalian enzyme (Schirch, 1982). We have found that the rabbit liver cytosolic enzyme can be phosphorylated *in vitro* using the catalytic subunit of a cyclic AMP-dependent protein kinase (K. Snell, unpublished work). Therefore it is particularly noteworthy that the two insertion regions contain motifs with the potential for phosphorylating serine/threonine residues at positions 161 and 275. Insertion region 154–158 includes a lysine residue producing a basic doublet (KK, KR) on the *N*-terminal side of Ser-161; similarly, insertion region 269–277 contains a basic doublet (RR, RK) on the *N*-terminal side of either serine or threonine at position 275 (Fig. 5). Such a sequence motif is known to determine phosphorylation of eukaryotic proteins by cyclic AMP-dependent and other kinases (Kemp & Pearson, 1990; Roach, 1991). At present the regulatory significance of phosphorylation of SHMT is not known.

The roles of specific amino acids in the active site and putative regulatory sites of mammalian cytosolic SHMT can be tested directly by substituting other amino acids at the appropriate position in the protein using site-directed mutagenesis of the cloned cDNA. The nucleotide sequence that we have established now makes such an approach possible for the first time. The demonstration that the cloned SHMT cDNA can be expressed in

COS cells makes it feasible to produce sufficient amounts of wild-type and mutant proteins to carry out enzymological studies.

We are grateful to the Cancer Research Campaign for generous financial support for this work.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) J. Mol. Biol. **215**, 403–410
Bossa, F., Barra, D., Martini, F., Schirch, V. & Fasella, P. (1976) Eur. J. Biochem. **70**, 397–401
Chan, V. L. & Bingham, H. L. (1990) Gene **101**, 51–58
Daly, E. C. & Aprison, M. H. (1974) J. Neurochem. **22**, 877–885
Eagle, H. (1955) Science **122**, 501–504
Eichler, H. G., Hubbard, R. & Snell, K. (1981) Biosci. Rep. **1**, 101–106
Fujioka, M. (1969) Biochim. Biophys. Acta **185**, 338–349
Higgins, D. G. & Sharp, P. M. (1988) Gene **73**, 237–244
Higgins, D. G. & Sharp, P. M. (1989) Comput. Appl. Biosci. **5**, 151–153
Hopkins, S. & Schirch, V. (1986) J. Biol. Chem. **261**, 3363–3369
Jackson, T., Morris, B. A. & Sanders, P. G. (1992) Mol. Immunol. **29**, 667–676
Jones, C. W., III & Priest, D. G. (1976) Arch. Biochem. Biophys. **174**, 305–311
Kemp, B. E. & Pearson, R. B. (1990) Trends Biochem. Sci. **15**, 342–346
Kozak, M. (1987) Nucleic. Acids Res. **15**, 8125–8148
Kozak, M. (1989) J. Cell Biol. **108**, 229–241
Laemmli, U. K. (1970) Nature (London) **227**, 680–685
Martini, F., Angelaccio, S., Pascarella, S., Barra, D., Bossa, F. & Schirch, V. (1987) J. Biol. Chem. **262**, 5499–5509
Martini, F., Maras, B., Tanci, P., Angelaccio, S., Pascarella, S., Barra, D., Bossa, F. & Schirch, V. (1989) J. Biol. Chem. **264**, 8509–8519
Maruyama, T., Gojobori, T., Aota, S. & Ikemura, T. (1986) Nucleic Acids Res. **14**, 151–191 (Suppl.)
Nakano, Y., Fujioka, M. & Wada, H. (1968) Biochim. Biophys. Acta **159**, 19–25
Plamann, M., Stauffer, L. T., Urbanowski, M. L. & Stauffer, G. V. (1983) Nucleic. Acids Res. **11**, 2065–2075
Proudfoot, N. (1991) Gene **64**, 671–674
Rao, D. N. & Rao, A. N. (1982) Plant Physiol. **69**, 11–16
Roach, P. J. (1991) J. Biol. Chem. **266**, 14139–14142
Sambrook, J., Fritsch, E. F. & Maniatis, J. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Schirch, L. V. (1982) Adv. Enzymol. Relat. Areas Mol. Biol. **53**, 83–112
Schirch, V. & Peterson, D. (1980) J. Biol. Chem. **255**, 7801–7806
Snell, K. (1984) Adv. Enz. Regul. **22**, 325–400
Snell, K. (1985) Biochim. Biophys. Acta **843**, 276–281
Snell, K. (1989) in Liver Cell Carcinoma (Bannasch, P., Keppler, D. & Weber, G., eds.), pp. 375–387, Kluwer Academic Publisher, Dordrecht
Snell, K., Natsumeda, Y. & Weber, G. (1987) Biochem. J. **245**, 609–612
Snell, K., Natsumeda, Y., Eble, J. N., Glover, J. L. & Weber, G. (1988) Br. J. Cancer **57**, 87–90
Taylor, R. T. & Weissbach, H. (1965) Anal. Biochem. **13**, 80–84
Thorndike, J., Pelliniemi, T. & Beck, W. S. (1979) Cancer Res. **39**, 3435–3440
Ulevitch, R. J. & Kallen, R. G. (1977) Biochemistry **16**, 5342–5350

Received 27 November 1991/4 February 1992; accepted 14 February 1992