ANALYSIS OF VARIANT CYTOSOLIC SERINE HYDROXYMETHYLTRANSFERASES

KAREN JUDY CHAVE

Submitted for the degree of
Doctor of Philosophy
January 1997

School of Biological Sciences
University of Surrey
Guildford
Surrey
GU2 5XH
UK

Copyright © Karen J. Chave, 1997. All rights reserved
ABSTRACT

Serine hydroxymethyltransferase (SHMT) is a major enzyme for serine, glycine and one carbon group metabolism and aberrant SHMT activity has been implicated in some diseases such as cancer and schizophrenia.

cDNAs encoding the cytosolic isozyme of human and rabbit SHMT (cSHMT) were expressed in *E.coli* using the pET vector system which produces SHMT proteins fused to an amino terminal motif of six histidine residues. This allows rapid purification of cSHMT protein by metal ion chromatography. Enzymatically active SHMT was purified.

Mutant SHMT proteins were produced by site directed mutagenesis of the cDNA coding for human cSHMT. The amino acids mutated had been implicated in SHMT activity by chemical modification. Mutant proteins were expressed in *E.coli*, purified as above and were assayed for enzyme activity. Whereas chemical modification had in all cases completely inactivated enzyme activity, none of the mutants created completely inactivated the protein suggesting that chemical modification of proteins only has a limited role in determining residues important in SHMT activity.

Two bacteriophage lambda clones containing human genomic DNA hybridising to cSHMT cDNA were characterised. Sequence data was obtained showing although the sequence had 90% homology to cSHMT cDNA there were many point mutations, deletions and insertions into the genomic DNA including a mutated translation initiation codon. No introns were present in the genomic sequence and a 16 base pair direct repeat was found either side of the cSHMT homologous sequence. It was concluded that the clones contained DNA encoding a processed pseudogene of cSHMT.

PCR amplification of human genomic DNA was performed and sequences from the cSHMT gene were identified. Five introns were identified when compared to the known cDNA sequence and a further intron may be present.
ACKNOWLEDGMENTS

I would like to thank my supervisor, Pete Sanders for his help and advice throughout the project and to all the researchers in Molecular Genetics laboratory for helpful discussions. Particular thanks go to Lynne Barratt, an undergraduate project student who generated two mutant SHMT cDNAs. This work could not have been done without the technical support of Sue Wall and Tina O’Shaughnessy who run the oligonucleotide synthesis service and Stephen Dyer who performs the automated DNA sequencing.

Thanks also go to Paula Byrne, Keith Snell, Suzy Whitehouse at the Institute of Cancer Research for helpful discussions and particularly Suzanne Renwick for providing purified wild type SHMT protein and for performing the SHMT assays with radiolabeled serine. I thank the Faculty of Science and School of Biological Sciences at the University of Surrey for providing the funding to perform the work presented here.

Finally I would like to thank my parents for their unending support and encouragement.
# TABLE OF CONTENTS

ABSTRACT ...................................................................................................................................2  
ACKNOWLEDGEMENTS ........................................................................................................3  
TABLE OF CONTENTS .............................................................................................................4  
TABLE OF FIGURES .................................................................................................................9  
TABLE OF TABLES ..................................................................................................................11  
ABBREVIATIONS .....................................................................................................................12  
1. CHAPTER 1 ...........................................................................................................................14  
   1.1 Structure and function of Serine Hydroxymethyltransferase .......................................15  
      1.1.1 SHMT isoymes ........................................................................................................16  
      1.1.2 Conservation of SHMT primary structure across species .....................................17  
      1.1.3 Reactions catalyse by SHMT ...................................................................................17  
   1.2 Proposed reaction mechanism for serine and glycine interconversion .......................18  
      1.2.1 Role of Tetrahydrofolate in SHMT activity ...........................................................20  
      1.2.2 Important amino acids for SHMT activity .............................................................21  
   1.3 Use of serine, glycine and one carbon units in the cell .................................................26  
      1.3.1 Biosynthesis of serine ...............................................................................................26  
      1.3.2 Serine catabolism .......................................................................................................26  
      1.3.3 Glycine catabolism ...................................................................................................29  
      1.3.4 Utilisation of one carbon units .................................................................................29  
   1.4 Control of SHMT activity ................................................................................................29  
      1.4.1 *Escherichia coli* and *Salmonella typhimurium* ......................................................32  
      1.4.2 Control of SHMT activity in *Neurospora crassa* ..................................................33  
      1.4.3 Control of rabbit cSHMT activity ............................................................................33  
      1.4.4 Control of Mouse SHMT activity ............................................................................34  
   1.5 Compartmentation of folate and one carbon metabolism ............................................34  
      1.5.1 One carbon group metabolism in yeast .................................................................35  
      1.5.2 Chinese hamster ovary cell metabolism .................................................................36
2.5.12 Digoxygenin labeling of DNA ................................................................. 64
2.5.13 Hybridisation and detection of DNA ......................................................... 64
2.5.14 5’ Phosphorylation of oligonucleotides ................................................... 65
2.6 Amplification of rabbit and human cSHMT cDNA by PCR ......................... 65
2.7 Modification of pUC18 to remove multiple cloning site ............................. 65
2.8 Site Directed mutagenesis of human cytosolic SHMT .................................. 66
  2.8.1 Production of single stranded DNA .................................................... 66
  2.8.2 Mutagenesis of human cSHMT cDNA ................................................ 66
2.9 Expression of cSHMT in E.coli ................................................................. 66
2.10 Purification of SHMT protein ..................................................................... 67
  2.10.1 Biorad Assay of protein concentration .............................................. 68
2.11 Assay of SHMT activity using serine and THF ............................................ 69
2.12 dl-allothreonine assay ................................................................................ 69
2.13 Bacteriophage lambda techniques ............................................................. 70
  2.13.1 Preparation of plating bacteria ............................................................. 70
  2.13.2 Plating Bacteriophage lambda ............................................................ 70
  2.13.3 Plaque purification ............................................................................. 70
  2.13.4 Rapid small scale isolation of Lambda DNA ....................................... 70
2.14 Mapping of Pseudogene clones ................................................................ 71
2.15 Analysis of the human genomic cytosolic SHMT gene ................................ 72
3. CHAPTER 3 ......................................................................................................... 73
  3.1 Summary .................................................................................................... 74
  3.2 Results ......................................................................................................... 75
    3.2.1 Subcloning of rabbit cSHMT into pET14b ........................................... 75
    3.2.2 Expression of rabbit cytosolic SHMT ................................................. 75
    3.2.3 Subcloning of human cytosolic SHMT into pET14b ............................ 79
    3.2.4 Expression of human cytosolic SHMT ................................................. 79
    3.2.5 Sequencing of PCR modified SHMT sequences ............................... 80
  3.3 Discussion ................................................................................................... 83
4. CHAPTER 4 ............................................................................................................................88
  4.1 Summary ............................................................................................................................89
  4.2 Results .................................................................................................................................90
    4.2.1 Insertion of human cSHMT orf into the pALTER-1 vector ........................................90
    4.2.2 Site directed mutagenesis of human cSHMT orf to revert amino acid 182 to asparagine ................................................................. 90
    4.2.3 Synthesis of new vectors to aid site directed mutagenesis ........................................91
    4.2.4 Mutagenesis of human cSHMT cDNA .....................................................................97
    4.2.5 Reformation of whole human cSHMT orf for mutants generated in pUS1255102 .........
    4.2.6 Subcloning of the wild-type cSHMT orf into pET14b for expression in E.coli104 ......
    4.2.7 Subcloning of mutant human cSHMT orfs into pET14b ...........................................104
    4.2.8 Expression and purification of mutant human cSHMT protein ................................110
    4.2.9 Assay of mutant cSHMT protein activity ...............................................................117
  4.3 Discussion ........................................................................................................................117

5. CHAPTER 5 ...........................................................................................................................129
  5.1 Summary ...........................................................................................................................130
  5.2 Results ...............................................................................................................................131
    5.2.1 Sequencing of restriction digest fragments hybridising to cSHMT cDNA .................131
    5.2.2 Mapping of genomic clones ......................................................................................135
  5.3 Discussion ........................................................................................................................141
    5.3.1 Sequencing of genomic DNA homologous to cSHMT cDNA .............................141

6. CHAPTER 6 ...........................................................................................................................153
  6.1 Summary ...........................................................................................................................154
  6.2 Results ...............................................................................................................................155
    6.2.1 Amplification of human cSHMT genomic DNA using sequences from the 3’ end of cSHMT cDNA sequence ........................................ 155
    6.2.2 Amplification of human cSHMT genomic DNA from the 5’ end of cSHMT cDNA towards promoter region .............................................. 159
6.2.3 Amplification of human cSHMT genomic DNA to amplify across the intron junction boundary at nt 1231 ................................................................. 160

6.2.4 Amplification of the intron present between nt 696-697 of cSHMT cDNA .. 165

6.3 Discussion .................................................................................................................. 166

7. CHAPTER 7 ................................................................................................................... 173

7.1 Concluding remarks and further study .................................................................... 174

8. CHAPTER 8 ................................................................................................................... 177

9. APPENDICES ............................................................................................................... 192
TABLE OF FIGURES

FIGURE 1.1. Proposed mechanism for SHMT catalysis ..............................................................19
FIGURE 1.2. Pathway for the biosynthesis of serine from glucose via the glycolytic
intermediate glyceraldehyde 3-phosphate .........................................................................25
FIGURE 1.3. Utilisation of serine by the cell ........................................................................27
FIGURE 1.4. Diagram of the purine ring indicating the contributions of glycine and 10-
formylTHF to the structure .........................................................................................28
FIGURE 1.5. Structure of tetrahydrofolate (THF) ..............................................................30
FIGURE 1.6. Diagram demonstrating the interaction of enzymes utilising folates as
coenzymes .....................................................................................................................31
FIGURE 3.1. Vectors used for the expression of rabbit and human cSHMT proteins ........76
FIGURE 3.2. Coomassie brilliant blue stained SDS-PAGE gel showing a time course of
rabbit cSHMT protein expressed in E.coli .......................................................................77
FIGURE 3.3. Comparison of soluble/insoluble fractions of rabbit cSHMT protein expressed
in E.coli .............................................................................................................................78
FIGURE 3.4. Coomassie brilliant blue stained SDS-PAGE gel showing a time course of
human cSHMT protein expression in E.coli ....................................................................81
FIGURE 3.5. Comparison of soluble/insoluble fractions of human cSHMT protein expressed
in E.coli .............................................................................................................................82
FIGURE 3.6. Crystals of human cSHMT ..............................................................................86
FIGURE 4.1. Formation of the vector for site directed mutagenesis ....................................92
FIGURE 4.2. Construction of pUS1255 for site directed mutagenesis of the 0.6kb
KpnI/BamHI fragment of the cSHMT cDNA .................................................................94
FIGURE 4.3. Schematic of pUS1248 ..................................................................................95
FIGURE 4.4. Agarose gel of restriction digests of four clones obtained when kjc7 was ligated
into digested pUC18 ......................................................................................................96
FIGURE 4.5. Diagram to show the formation of pUS1249 from pUS1248 and human cSHMT
cDNA .................................................................................................................................98
FIGURE 4.6. Diagram to show the formation of expression vector containing the human
cSHMT orfs ...................................................................................................................105
FIGURE 4.7. Expression of C204A cSHMT protein at 37°C, 30°C and 25°C ....................111
FIGURE 4.8. Purification of C204A cSHMT using a nickel ion column ..............................113
FIGURE 4.9. Purification of H256QK257A cSHMT .........................................................114
FIGURE 4.10. Expression of K446stop cSHMT protein at 20°C with sorbitol and betaine
added to the growth media .............................................................................................115
FIGURE 4.11. Expression of C380stop cSHMT protein at 20°C with sorbitol and betaine
added to the growth media ............................................................................................116
FIGURE 5.1. Diagram of the 3kb fragment isolated from clone 4 by digestion with
EcoRI ...............................................................................................................................132
FIGURE 5.2. Diagram of the 1.7kb fragment from clone 3 isolated by digestion of the clone
with BamHI ..................................................................................................................133
FIGURE 5.3. Relationship between the sequenced fragments of human genomic
DNA .................................................................................................................................134
FIGURE 5.4. Sequence of the human genomic DNA fragments merged to produce a 4.3kb
fragment ..........................................................................................................................136
FIGURE 5.5. Hybridisation of human cSHMT (pUS1213) to digests of lambda
clones ............................................................................................................................138
FIGURE 5.6. Hybridisation of human cSHMT (pUS1213) to digested lambda clones........139
FIGURE 5.7. Hybridisation of a 1.4kb EcoRI/KpnI fragment from clone 4 to the digested lambda clones..................................................................................................................................................140
FIGURE 5.8. Mapping of clone 3 by partial digestion and hybridisation to oligonucleotides complementary to the cos ends.................................................................142
FIGURE 5.9. Mapping of clone 4 by partial digestion and hybridisation to oligonucleotides complementary to the cohesive ends of the lambda arms........................................143
FIGURE 5.10. Mapping of clone 4 by partial digestion and hybridisation to the cohesive ends of the lambda arms..............................................................................................................144
FIGURE 5.11. Diagram of lambda clones 3 and 4..........................................................................................................................145
FIGURE 5.12. Mapping of clone 4 by partial digestion and hybridisation to the cohesive ends of the lambda arms....................................................................................................................146
FIGURE 5.13. Diagram of lambda clones 3 and 4..................................................................................................................................................................................147
FIGURE 6.1. Putative intron locations in the cSHMT gene derived from alternatively spliced and incompletely spliced cSHMT cDNAs.................................................................156
FIGURE 6.2. Agarose gel showing results of PCR amplification of the human genomic libraries in the human PromoterFinder kit using nested primers gsp1 and gsp2.....157
FIGURE 6.3. Sequence of the products obtained using primers gsp1 and gsp2 in conjunction with the human PromoterFinder kit.........................................................................................158
FIGURE 6.4. Agarose gel of the PCR amplification of human genomic libraries 1 to 5 of the human PromoterFinder kit using kjc22 and kjc23..........................................................161
FIGURE 6.5. Sequence of the 1kb product produced by amplification of library 4 of the human PromoterFinder kit with primers kjc22 and kjc23.................................................................162
FIGURE 6.6. Agarose gel of the amplification of the human genomic libraries in the human PromoterFinder kit libraries with kjc8 and kjc21........................................................................163
FIGURE 6.7. Sequence of the products isolated from using the human PromoterFinder kit with primers kjc8 and kjc21........................................................................................................164
FIGURE 6.8. Agarose gel of PCR products from human genomic DNA amplified using kjc9/kjc16 and kjc20/kjc16 primer pairs and subsequent hybridisation with DIG labeled probe..........................................................................................167
FIGURE 6.9. Illustration of the sequences isolated in this chapter in relation to human cDNA and putative intron boundaries illustrated in figure 6.1.........................................................168
FIGURE 6.10 Illustration of the human genomic DNA sequences isolated compared to human cSHMT cDNA..................................................................................................................168a
TABLE OF TABLES

TABLE 4.1. Mutations of human cSHMT generated ...............................................................100
TABLE 4.2. Results of the ligations of 0.6kb KpnI/BamHI mutant fragments into
  pUS1249 ............................................................................................................................103
TABLE 4.3. University of Surrey plasmid assignments for mutant cSHMT orfs in
  pET14b..................................................................................................................................109
TABLE 4.4. Results of experiments to determine the extent of solubility of mutant cSHMT
  proteins when expressed in E.coli ....................................................................................112
TABLE 4.5. Results of assays on wild-type and mutant cSHMT proteins using either
dl-allothreonine or radiolabeled serine and THF as substrates ........................................118
## ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADP</td>
<td>adenosine diphosphate</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CHO</td>
<td>Chinese hamster ovary</td>
</tr>
<tr>
<td>CH$_2$THF</td>
<td>N$<em>5$N$</em>{10}$-methylene tetrahydrofolate</td>
</tr>
<tr>
<td>CIP</td>
<td>calf intestinal alkaline phosphatase</td>
</tr>
<tr>
<td>CNS</td>
<td>central nervous system</td>
</tr>
<tr>
<td>cos</td>
<td>cohesive ends</td>
</tr>
<tr>
<td>CPC</td>
<td>cross pathway control</td>
</tr>
<tr>
<td>CSPD</td>
<td>Disodium 3-(4-methoxySpiro[1,2-dioxetane-3,2'-(5'-chloro)-tricyclo[3.3.1.1$^{3,7}$]decan]-4-yl) phenyl phosphate</td>
</tr>
<tr>
<td>cSHMT</td>
<td>cytosolic serine hydroxymethyltransferase</td>
</tr>
<tr>
<td>dB6</td>
<td>4-deoxypyridoxine</td>
</tr>
<tr>
<td>DEPC</td>
<td>diethylpyrocarbonate</td>
</tr>
<tr>
<td>DHF</td>
<td>dihydrofolate</td>
</tr>
<tr>
<td>DHFR</td>
<td>dihydrofolate reductase</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>dNTP's</td>
<td>deoxynucleotide phosphate</td>
</tr>
<tr>
<td>DTNB</td>
<td>5,5'-dithiobis(2-nitrobenzoic acid)</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediamine tetraacetic acid</td>
</tr>
<tr>
<td>EMS</td>
<td>Ethylmethane sulphonate</td>
</tr>
<tr>
<td>eSHMT</td>
<td><em>E.coli</em> SHMT</td>
</tr>
<tr>
<td>FPGS</td>
<td>Folylpolyglutamyl synthase</td>
</tr>
<tr>
<td>GCS</td>
<td>glycine cleavage system</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-$\beta$-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase pairs</td>
</tr>
<tr>
<td>KD</td>
<td>kilodaltons</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MMTS</td>
<td>methyl methanethiosulphonate</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>mSHMT</td>
<td>mitochondrial serine hydroxymethyltransferase</td>
</tr>
<tr>
<td>MY</td>
<td>million years</td>
</tr>
<tr>
<td>NAD+</td>
<td>Nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>NADH</td>
<td>Nicotinamide adenine dinucleotide, reduced form</td>
</tr>
<tr>
<td>NBS</td>
<td>N-bromosuccinimide</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>orf</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PHA</td>
<td>phytohaemagglutinin</td>
</tr>
<tr>
<td>Pi</td>
<td>inorganic phosphate</td>
</tr>
<tr>
<td>PLP</td>
<td>pyridoxal-5'-phosphate</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>SSC</td>
<td>salt sodium citrate buffer</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl phosphate</td>
</tr>
<tr>
<td>SHMT</td>
<td>serine hydroxymethyltransferase</td>
</tr>
<tr>
<td>ssDNA</td>
<td>single stranded DNA</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate-EDTA electrophoresis buffer</td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA buffer</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>THF</td>
<td>tetrahydrofolate</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>TS</td>
<td>thymidylate synthase</td>
</tr>
<tr>
<td>utr</td>
<td>untranslated region</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-bromo-4-chloro-3-indolyl-β-galactopyranoside</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION
1.1 Structure and function of Serine Hydroxymethyltransferase

SHMT is a pyridoxal 5’ phosphate dependent enzyme responsible for the reversible conversion of L-serine to L-glycine and formaldehyde which is complexed with tetrahydrofolate (THF) to produce N5,N10-methylenetetrahydrofolate (methyleneTHF). SHMT activity was first demonstrated \textit{in vivo} by Shemin, (1946) and has been extensively studied since that date. SHMT is ubiquitous, found in all prokaryotes and eukaryotes, and in many different organs within mammals e.g. liver, kidney, spleen, intestine, heart, and lung (Snell 1984).

The SHMT enzyme is a homotetramer in eukaryotes (Martinez-Carrion \textit{et al.}, 1972). SHMT protein is assumed to be tetrameric in prokaryotes as it is in eukaryotes. However, SHMT has been demonstrated to be a homodimer in \textit{E.coli} (Schirch \textit{et al.}, 1985) and it follows that other prokaryotes may share this structure. A SHMT subunit has a molecular weight of 53KD in eukaryotes and 46KD in prokaryotes. Purified enzyme is yellow with an absorption maximum at 428nm due to bound pyridoxal 5’-phosphate (PLP) bound as a Schiff base to the 6-amino group of a lysine residue, forming the holoenzyme. The colourless “apo” form can be generated by removing the PLP molecule from SHMT by incubation with L-cysteine (Schirch and Mason, 1962).

\textit{E.coli} SHMT (eSHMT) contains 3 tryptophan residues. Site directed mutagenesis of eSHMT has been performed to produce 3 enzymes, each with only 1 tryptophan residue (Cai and Schirch, 1996a; 1996b). The folding of eSHMT was followed by using fluorescence quenching using these mutant proteins. It was found that 2 domains fold rapidly between residues 55-224 and 276-417 as fluorescence from tryptophan residues 183 and 385 was quenched in seconds. Tryptophan residue 16 does not form part of any ordered structure during the first stages of folding and only reaches the native state when the enzyme is in a
catalytically active form. The segment between the two rapidly folding domains contains the PLP binding lysine and remains in an unordered state even after dimerisation. Following dimerisation there is a slow rate determining conformational change during which the interdomain region containing the PLP binding lysine becomes ordered. PLP does not bind until folding is complete. Recent analysis of the cytosolic isozyme of sheep SHMT suggests that the subunits form tetramers before PLP binds to form the holoenzyme. The holoenzyme dissociates into its individual subunits with increasing pH of buffer suggesting that the subunits are associated electrostatically (Brahatheeswaran et al., 1996).

1.1.1 SHMT isozymes

In most eukaryotic organisms there are two isozymes of SHMT, namely the cytosolic (cSHMT) and mitochondrial (mSHMT) enzymes. There is evidence of a chloroplast isozyme in plants (Besson et al., 1995) and in Euglena gracilis (Sakomoto et al., 1993). SHMT isozymes have different primary structures and are encoded by separate genes. For example, the genes coding for human cytosolic (glyC) and mitochondrial (glyM) SHMT proteins are located on 17p11.2 and 12q13.2 respectively (Garrow et al., 1993). Prokaryotes have a single gene (glyA) encoding a single SHMT enzyme. The two SHMT isozymes most extensively studied are those isolated from rabbit liver. With the exception that rabbit liver mSHMT has a four fold greater affinity for glycine than its cytosolic counterpart (Schirch and Peterson, 1980), the SHMT isozymes have similar substrate specificity's and kinetic constants (Akhtar et al., 1975; Schirch and Peterson, 1980). mSHMT protein is located in the mitochondrial matrix (Cybulski and Fisher, 1976).

Distribution of the SHMT isozymes varies according to the organism and the tissue. For example, 75-80% of the SHMT activity in rabbit or monkey liver is predominantly from the cytosolic isozyme (Schirch and Peterson, 1980; Snell, 1984) whereas in rat mSHMT and
cSHMT contribute equally to the total liver SHMT activity (Snell, 1984). In Chinese hamster ovary cells (Chasin et al., 1974), *Saccharomyces cerevisiae* (Zelikson et al., 1977), the central nervous system (Daly and Aprison, 1974) and plant leaves (Bourguignon et al., 1988) the total SHMT activity is attributed mainly to the mitochondrial isozyme.

### 1.1.2 Conservation of SHMT primary structure across species

29 amino acid sequences of SHMT protein have been either directly determined or translated from nucleic acid data. This data shows that the primary structure of SHMT protein is highly conserved across all species. For instance, human cSHMT protein has 92% identity to rabbit cSHMT and 63% identity to *E.coli* SHMT. Alignments of the sequences with the PILEUP and PRETTY programs (Genetics Computer Group) (appendix 2) indicates there are 43 amino acids conserved in all SHMT protein sequences. Divergences between SHMT sequences primarily occur at the amino and carboxyl termini of the proteins. In addition the eukaryotic SHMT’s are larger than the prokaryotic forms with 4 conserved insertions compared to the prokaryotic sequence.

### 1.1.3 Reactions catalyse by SHMT

*In vitro*, SHMT has a broad substrate specificity catalysing the aldol cleavage of a number of 3-hydroxyamino acids to form L-glycine and the corresponding aldehyde. Only those reactions producing formaldehyde require THF as a cofactor. SHMT also catalyses decarboxylation, transamination and racemisation (for reviews see Schirch, 1982; 1984). SHMT also acts as a threonine aldolase (L-threonine acetaldehyde lyase, EC 4.1.2.5) (Schirch and Gross, 1968; Akhtar and El Obeid, 1972). This activity is mainly limited to the cytosolic isozyme in rat liver (Snell, 1984).

SHMT is in equilibrium between an “open” form where substrates enter and leave the active site, and a “closed” conformation where catalysis occurs (Schirch et al., 1991).
Racemisation and transamination occur only in the open form suggesting that one reason the closed structure is formed is in order to confer reaction specificity.

1.2 Proposed reaction mechanism for serine and glycine interconversion

Addition of saturating levels of L-glycine to SHMT gives a spectrum with 3 distinct peaks appearing in the order 343nm, 425nm and 495nm. Incubation of SHMT with excess L-serine leads to additional spectrum maxima at 343, 427 and 495nm (Schirch, 1984). One reaction mechanism which has been derived from the absorption maxima evidence and homology to other PLP dependent enzymes involves the formation of a quinoid intermediate either from glycine (by loss of $\alpha$ proton) or from serine (by cleavage of the 2-3 carbon-carbon bond). Figure 1.1 illustrates the mechanism proposed by Schirch, (1984). The key steps of this mechanism are:

1. The enzyme-PLP undergoes nucleophilic attack from the NH$_2$ group of glycine at c-4' to form the geminal diamine (structure II).

2. A second nucleophilic attack by unpaired electrons on the unprotonated glycine amine to form a Schiff base (external aldimine, structure III). This reaction displaces the 6-amino group of the enzyme lysine.

3. A base on the enzyme removes a proton from the $\alpha$ carbon of the bound glycine to form a resonance stabilised quinoid intermediate (structure IV).

4. In the presence of formaldehyde the $\alpha$ carbon of glycine attacks the formaldehyde to form the Schiff base of serine (structure V).

5. The transamination procedure is reversed to complete the reaction.
Figure 1.1 Proposed mechanism for SHMT catalysis (Schirch, 1984)
Other reaction mechanisms have been proposed (Matthews and Drummond, 1990). One possibility is the conversion of the serine external aldimine (structure V, figure 1.1) to a quinoid intermediate and a thiohemiacetal. This mechanism requires a thiol group on the enzyme. A further possible mechanism is the direct nucleophilic attack on N5 of THF by the β carbon of serine aldimine with the loss of H₂O. Both of these mechanisms would yield methyleneTHF with the correct stereospecificity but there is no direct evidence to support either mechanism.

An insight into the physiological reaction would be gained by determining the X-ray crystal structure of SHMT bound to various substrates.

1.2.1 Role of Tetrahydrofolate in SHMT activity

The reaction mechanism proposed above (Schirch, 1984) does not indicate how THF is involved in the reaction. It was proposed by Jordan and Akhtar, (1970) that the complexing of formaldehyde with THF follows the same path as for the non-enzymatic reaction of these two compounds described by Kallen and Jencks, (1966). This pathway would lead to the racemisation of methyleneTHF but the high stereospecificity of the methyleneTHF produced by SHMT rules out this possibility.

The primary role of THF in the reaction is, therefore, to remove formaldehyde from the active site as in the absence of THF, serine is broken down to glycine and formaldehyde by SHMT but the reaction then stops as formaldehyde is slow to leave the active site. (Schirch, 1984). The one carbon group is transferred to glycine to form serine via an enzyme-formaldehyde complex (Chen and Schirch, 1973a). It was suggested that formaldehyde is bound as an imine at the active site and THF appeared to catalyse the formation or breakdown of this intermediate (Chen and Schirch, 1973b). However, the SHMT catalysed reaction is highly stereospecific and the formation of an enzyme-formaldehyde complex has several
steps where the stereospecificity could be lost. Matthews and Drummond, (1990) state that there is no evidence for the formation of the imine.

In the absence of THF the reaction is ordered with formaldehyde released before glycine. In the presence of THF, however, the reaction becomes a sequential random mechanism with either of the substrates binding first and either product being released first (Schirch, 1984).

The presence of THF also increases the rate of exchange of the α proton of glycine with solvent by more than 2 orders of magnitude (Schirch and Jenkins, 1964). It is thought that THF slows down the reverse step of protonation of the quinoid intermediate i.e. the rate of which the quinoid is protonated to form structure III (figure 1.1) (Schirch, 1984).

The reaction mechanism and role of THF in SHMT activity will only be elucidated by the determination of the tertiary and quaternary structures of SHMT complexed with various substrates.

1.2.2 Important amino acids for SHMT activity

In the absence of a determined SHMT tertiary and quaternary structure, residues important for SHMT activity have been identified by chemical modification and site directed mutagenesis. SHMT isolated from several organisms have been studied. As the protein is of a different length in each organism, the amino acids mentioned below refer to the equivalent amino acid position in human cSHMT (appendix 1) as identified from the alignments presented in appendix 2.

By homology to other PLP dependent enzymes, a lysine residue in E.coli SHMT was identified as being the amino acid to which PLP binds to form the internal aldimine (section 1.2). This was confirmed by site directed mutagenesis as being the PLP binding site (Schirch et al., 1993) and is equivalent to lysine 257 (K257) in human cSHMT. Therefore, this lysine
is critical for SHMT activity. Unlike other PLP dependent enzymes this lysine residue does not also remove the α proton (see section 1.2) (Schirch et al., 1993).

A cysteine residue was implicated as being necessary for catalysis following modification with D-fluoroalanine, which resulted in inactivation of lamb cSHMT and loss of a cysteine residue (Wang et al., 1981). This residue was found to be equivalent to C335 in rabbit and human cSHMT (Gavilanes et al., 1982) but further studies suggested that C335 was not necessary for enzyme activity in rabbit cSHMT (Gavilanes et al., 1982).

Extensive work with agents which modify cysteine amino acids, namely 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB), iodoacetate and methyl methanethiosulphonate (MMTS), has shown that there are at least 2 cysteine residues at the active site of rabbit cSHMT (Gavilanes et al., 1982). In rabbit mSHMT there is 1 cysteine residue at the active site but this residue is not necessary for enzyme activity (Gavilanes et al., 1983). None of the cysteine residues in either isozyme form disulphide bonds.

In sheep cSHMT, cysteine residues equivalent to human cSHMT C68 and C204 are protected by PLP when treated with iodoacetate. This study inferred that cysteines equivalent to C248 and C262 are buried in the protein (Usha et al., 1994). In rabbit cSHMT cysteine 204 was also modified by iodoacetate. PLP does not bind to rabbit cSHMT post modification due to steric hindrance so the activity of modified enzyme could not be determined and implies that C204 is located at the PLP binding site (Schirch et al., 1980).

Whereas human cSHMT has 10 cysteine residues (Garrow et al., 1993), E.coli SHMT has only 3 cysteine residues (Schirch et al., 1985) one of which when modified by DTNB and MMTS inhibits the enzyme. This residue is protected from modification by the presence of serine. Cysteine 410 in eSHMT was modified by site directed mutagenesis and found to play a role in the structure of the enzyme but not catalytic activity (Joshi-Tope and Schirch, 1990).
A histidine in eSHMT equivalent to H256 in human cSHMT is totally conserved in all SHMT sequences and has been modified by site directed mutagenesis. This residue was found not to be essential for catalytic activity (Hopkins and Schirch, 1986). However, there was no change from the open to closed conformation on substrate binding. H256 plays a critical role in determining substrate specificity by controlling the size of the one carbon binding pocket and the orientation of the hydroxymethyl group of serine with respect to the PLP ring. This residue is not the base which catalyses the removal of the $\alpha$ proton (section 1.2) (Stover et al., 1992).

Modification of sheep cSHMT with diethylpyrocarbonate (DEPC) suggests that a histidine residue is essential for catalytic activity and that inactivation is protected by serine (B.Venkatesha, personal communication). The modified histidine was found to be equivalent to H135 in human cSHMT. However, the modified histidine is not conserved in other SHMT proteins (appendix 2). Histidine residues have been modified with DEPC in human and sheep cSHMT leading to enzyme inactivation but the specific residues modified have not been identified. Inactivation by DEPC was prevented by serine and THF (Vijayalaksmi and Rao, 1989; Manohar and Rao, 1984).

Arginine residues in eSHMT equivalent to R402 and R411 in human cSHMT have been modified by site directed mutagenesis. R402 plays a role in binding the substrate carboxyl group and has a role in conversion of SHMT from an open to a closed conformation (Delle Fratte et al., 1994). R411 is totally conserved in SHMT sequences (appendix 2) but does not appear to play a critical role in catalysis (Delle Fratte et al., 1994).

Usha et al., (1992) determined that THF prevented the chemical modification of arginine residues in sheep cSHMT equivalent to R270 and Q462 in human cSHMT by phenylglyoxal implying that these residues are at the THF binding site. However, amino acid 462 in human
cSHMT is not an arginine residue so the arginine residues modified in sheep cSHMT may not play a role in THF binding but steric hindrance produced by modifying agents prevents THF binding to the correct residues. Phenylglyoxal modification of arginine residues was also prevented by serine and THF in human cSHMT (Vijayalaksmi and Rao, 1989).

Human cSHMT has been chemically modified with N-bromosuccinimide (NBS). The loss of 1 unidentified tryptophan is correlated to the loss of enzyme activity. This inhibition is prevented by the presence of either serine or THF (Vijayalaksmi and Rao, 1989).

An aspartic acid residue in eSHMT equivalent to D228 in human cSHMT has been compared to an equivalent residue in aspartate aminotransferase (AAT) in which this residue hydrogen bonds to N1 of PLP (Pascarella and Bossa, 1993). To date no work has been published to determine whether this is the case in SHMT.

There is a highly conserved motif in all SHMT protein sequences between amino acids 252-258 around the invariant lysine which binds PLP. This sequence is 4 serine or threonine residues followed by histidine, lysine and another serine or threonine with eSHMT and human cSHMT only using threonine and not serine. All of these threonines have been modified in eSHMT by site directed mutagenesis (Angelaccio et al., 1992). The threonine residues equivalent to T252, T253, T255 and T258 in human cSHMT do not appear to play a critical role in the mechanism of the enzyme and none of the threonine residues play an important role in PLP binding. T253 and T254 are important in defining the orientation of the external aldimine with respect to the base which removes the $\alpha$ carbon. The hydroxyl group of T254 plays an important role in the conversion of the geminal diamine to the external aldimine (Angelaccio et al., 1992).
Figure 1.2 Pathway for the biosynthesis of serine from glucose via the glycolytic intermediate glyceraldehyde 3-phosphate
1.3 Use of serine, glycine and one carbon units in the cell

The substrates and products of SHMT are used in a variety of biosynthetic reactions.

1.3.1 Biosynthesis of serine

Serine is synthesized from the glycolytic intermediate 3-phosphoglycerate (figure 1.2). The rate limiting step in the pathway is the last irreversible step of serine synthesis and is catalysed by phosphoserine phosphatase which is regulated by feedback inhibition. The rate of serine synthesis is not influenced by the rate of glycolysis (Snell, 1980; Snell and Fell, 1990). A second proposed pathway of serine biosynthesis by the reverse of the serine catabolism pathway by serine aminotransferase has been shown not to be physiologically used (Snell, 1986).

1.3.2 Serine catabolism

Serine is utilised in the cell for biosynthesis of the amino acids glycine and cysteine. Glycine is formed by the action of SHMT and cysteine is formed from serine and methionine in a mutienzyme pathway. Serine is also used in the metabolic pathways of several lipids by donating a hydroxyl group in the biosynthesis of phosphatidyl serine which in turn is used to produce phosphatidyl ethanolamine and phosphatidyl choline. Sphingosine and ceramide, lipids found particularly in the nervous system are also utilise serine in their biosynthesis.

Apart from SHMT two other enzymes of serine utilisation are present in the cell, namely, serine dehydratase and serine aminotransferase (figure 1.3). Serine dehydratase is a cytosolic enzyme present only in the liver in rat and catalyses the conversion of serine to pyruvate for lipogenesis and gluconeogenesis (Snell, 1984). Serine aminotransferase is localised to the mitochondria in rat liver, kidney and heart and converts serine to hydroxypyruvate for gluconeogenesis (Snell, 1984). The enzymes of serine utilisation vary in activity during various stages of development in rats (Snell, 1984).
Figure 1.3 Utilisation of serine by the cell. Enzymes of serine catabolism are shown in blue.
Figure 1.4 Diagram of the purine ring indicating the contributions of glycine (red) and 10-formylTHF (blue) to the structure. 10-formylTHF is derived from methyleneTHF generated by SHMT.
1.3.3 Glycine catabolism

Glycine acts as an inhibitory and excitatory neurotransmitter and is also used in many biosynthetic reactions. For example, glycine is a precursor in the production of threonine, bile acids, porphyrin rings, purine rings (figure 1.4), collagen and β keratin. Along with arginine, glycine is a precursor for the formation of creatine used to store phosphate bond energy. Glycine is also used in the biosynthesis of glutathione. Glutathione is present at a concentration of 5mM in cells and helps keep the cysteine residues of the cell in a reduced state, and is essential for maintaining the normal structure of red blood cells. Glutathione also acts as an amino acid transporter and helps to detoxify the cell by reacting with H2O2. Glycine is catabolised in mitochondria to form ammonia and carbon dioxide by the glycine cleavage system (GCS). During this reaction a molecule of THF is converted to methyleneTHF (Stryer, 1988).

1.3.4 Utilisation of one carbon units

The one carbon units produced by the conversion of serine to glycine by SHMT are also used in a variety of reactions such as the biosynthesis of lipids, hormones and methionine. These one carbon units are complexed with THF in a variety of ways (figure 1.5) and are utilised by many different enzymes (figure 1.6). One carbon units are essential in DNA biosynthesis as they are required for the biosynthesis of the purine ring (figure 1.4) and for converting deoxyuridine monophosphate to deoxythymidine monophosphate.

1.4 Control of SHMT activity

SHMT activity is controlled at different levels in different organisms.
Figure 1.5 Structure of tetrahydrofolate (THF). The various forms in which a one carbon group can be complexed with the molecule are shown. Also shown is the location of the polyglutamate tail.
1 10-formylTHF dehydrogenase EC 1.5.1.6
2 5,10-methyleneTHF cyclohydrase EC 3.5.4.9
3 methenylTHF synthase EC 6.3.3.2
4 5,10-methyleneTHF dehydrogenase EC 1.5.1.5
5 SHMT EC 2.1.2.1
6 10-formylTHF synthase EC 6.2.4.3
7 dihydrofolate reductase EC 1.5.1.3
8 Thymidylate synthase EC 2.1.1.45
9 5-formiminoTHF cyclodeaminase EC 4.3.1.4
10 5,10-methylene reductase EC 1.1.1.88
11 homocysteine transferase

Figure 1.6 Diagram demonstrating the interaction of enzymes utilising folates as coenzymes. Reactions 2, 4 and 6 make up the Cl-THF synthase enzyme. Reactions 5, 7 and 8 make up the thymidylate synthase cycle.
1.4.1 *Escherichia coli* and *Salmonella typhimurium*

The rate of synthesis of eSHMT is repressed by its end products. Activity is decreased by growth in media containing high glycine concentrations and moderate concentrations of one carbon metabolites particularly adenine and guanine (Miller and Newman, 1974).

In *E.coli* and *S.typhimurium*, SHMT activity is controlled at the level of transcription. MetR protein (the product of the *metR* gene) is required for the activation of the *glyA* gene (Plamann and Stauffer, 1989; Lorenz and Stauffer, 1995). There are two binding sites for MetR in the *glyA* control region in *E.coli* and both sites are required for the normal activation of expression. DNA footprinting analysis of the *glyA* control region suggests that RNA polymerase and MetR proteins interact in some way to increase the affinity of RNA polymerase for the *glyA* promoter (Lorenz and Stauffer, 1996) Homocysteine (a precursor of methionine) increases MetR binding. S-adenosylmethionine inhibits *metR* transcription and hence *glyA* transcription (Urbanowski *et al.*, 1987).

The purine repressor protein, the gene product of *purR* also binds to the *glyA* control region and is a negative regulator of *glyA* transcription (Steiert *et al.*, 1990). PurR protein prevents RNA polymerase from binding to the *glyA* promoter. Hypoxanthine acts as a co-repressor with PurR and increases the ability of PurR to prevent RNA polymerase binding (Lorenz and Stauffer, 1996).

Sequences downstream of the *glyA* gene have been shown to be important in the stability of *glyA* mRNA. A *glyA* mutant with phage DNA inserted between the end of the coding sequence of the *glyA* gene and the proposed transcription termination site had only 30% SHMT activity of wild-type (Plamann and Stauffer, 1985). Therefore, sequences distal to the *glyA* structural gene play an important role in SHMT expression. In *E.coli* the *glyA* gene is followed by two REP sequences and a rho dependent transcription terminator. If all three
sequences are removed, there is a 7 fold reduction in SHMT expression (Plamann and Stauffer, 1990). These three sequences are necessary for mRNA stability. It has been demonstrated that two unidentified trans acting factors regulate SHMT expression in E.coli (Lorenz et al., 1996).

1.4.2 Control of SHMT activity in Neurospora crassa

The for locus of N.cassa codes for cSHMT. There are several different mRNA start and stop sites giving 4 different mRNA lengths of 2600, 2300, 2450 and 2050 nt. Glycine supplementation of growth medium produces an increase in cSHMT mRNA. There are 5 potential binding sites for the transcriptional activator CPC-1 upstream of the transcriptional start site and one sequence which matches the consensus sequence exactly in the first intron of the cSHMT gene (McClung et al., 1992). The CPC-1 protein is involved in the cross pathway control (CPC) system in N.cassa similar to a system of amino acid biosynthesis control in yeast. It is not yet been demonstrated whether CPC-1 binds for DNA. However, supplementation with formate at 1mM and 10mM gives a two fold increase in for mRNA levels and supplementation with 10mM glycine gives a 2.8 fold increase in for mRNA abundance but only in the presence of a functional CPC-1 protein (McClung et al., 1992).

1.4.3 Control of rabbit cSHMT activity

Rabbit cSHMT has been shown to be regulated at the level of translation in when expressed in COS-1 cells (Byrne et al., 1995). The cDNA for rabbit cSHMT has an upstream open reading frame and overlapping coding sequence (Byrne et al., 1992). Removing the upstream AUG stops translational repression approximately 100 fold demonstrated by an increase in expressed cSHMT protein but not in cSHMT mRNA levels.

There is also evidence for a pathway for rabbit cSHMT protein degradation. Two asparagine residues equivalent to 6 and 221 in human cSHMT are deamidated to isoaspartyl
residues (Artigues et al., 1990; 1995). It has been proposed that the deamidation of asparagine residues targets proteins for degradation (Robinson and Rudd, 1974). Purified rabbit liver cSHMT is degraded in 24 hours when injected into *Xenopus laevis* oocytes. Removing the first 14 amino acids of cSHMT protein reduced the extent and rate of cSHMT degradation implying deamidation of N6 may be a signal for cSHMT degradation. The sequence asparagine-glycine (the signal for deamidation of asparagine) is present at equivalent positions in human and sheep cSHMT but is absent in the protein sequences of their mitochondrial counterparts and all other SHMT proteins (Artigues et al., 1995).

1.4.4 Control of mouse SHMT activity

Recently Nakshatri et al., (1996) have demonstrated using subtractive hybridisation to isolate genes regulated by retinoic acid, that mouse cSHMT gene in the cell line P19 down regulated by retinoic acid. Retinoic acid is involved in the control of cell growth arrest and differentiation. These authors suggest that inhibition of SHMT expression by retinoic acid is an important step in growth arrest.

1.5 Compartmentation of folate and one carbon metabolism

There is evidence that one carbon group metabolism is compartmentalised with most enzymes involved in folate metabolism having cytosolic and mitochondrial isozymes to enable the conversion of folates to the state required in the intracellular compartments of each organism. The reason for this duplication of enzymes in the cytosol and mitochondria is that many folates do not cross the mitochondrial membrane. Cybulski and Fisher, (1976) first determined that serine and glycine are rapidly taken up by mitochondria. The same authors showed that only oxidised folates are taken up by rat liver mitochondria (e.g. folate, dihydrofolate (DHF), methotrexate). Reduced folates such as THF, 5-methylTHF and methyleneTHF penetrate only as far as the inter membrane space. DHF is taken up by mitochondria by a carrier mediated mechanism (Cybulski and Fisher, 1981). However, the
study was not done under physiological conditions. It has since been demonstrated using physiological conditions that 5-formylTHF is taken up by a carrier mediated mechanism by mitochondria and that this uptake is inhibited by 5-methylTHF suggesting there is a system for the uptake of reduced folate monoglutamates (Horne et al., 1992).

As mentioned in section 1.3.3, glycine is catabolised in the mitochondria by the glycine cleavage system (GCS) yielding CO$_2$ and ammonia, and converting of NAD$^+$ to NADH and THF to methyleneTHF. The GCS is a multiprotein complex comprising of multiple copies of H, P, L and T proteins (for review of the GCS see Oliver and Raman, 1995). mSHMT is coupled to the GCS to recycle THF by converting glycine and methyleneTHF to serine and THF.

Another important complex is C1-THF synthase, made up of 5,10-methyleneTHF dehydrogenase, 5,10-methyleneTHF cyclohydrase and 10-formylTHF synthetase (figure 1.6). In human cells the three activities are located on a single peptide but in other organisms the three activities are located on either two or three peptides (Pelletier and Mackenzie, 1995). The C1-THF synthase complex coupled to SHMT converts serine to 10-formylTHF for purine biosynthesis. In the reverse direction the multifunctional enzyme generates THF and formate (figure 1.6).

Work has been performed on several organisms and the roles played by folate dependent enzymes has found to be different in each organism studied.

1.5.1 One carbon group metabolism in yeast

The yeast *S. cerevisiae* has been extensively studied as mutants lacking individual folate metabolism enzymes are relatively easy to obtain and grow. In yeast mitochondria it has been shown that C1-THF synthase converts formate to methyleneTHF (Mckenzie and Jones, 1977).
Several research groups have used $^{13}$C NMR to follow serine production in *S. cerevisiae*. Pasternak *et al.*, (1994) demonstrated that 25% of one carbon units utilised in cytoplasmic purine synthesis are derived from formate in the mitochondria. McNeil *et al.*, (1996) determined that the remaining 75% of one carbon units in the cytoplasm were derived from serine generated in the mitochondria by the activities of the GCS and mSHMT.

### 1.5.2 Chinese hamster ovary cell metabolism

Chinese hamster ovary cells (CHO) have also been used to study folate metabolism, particularly the roles of SHMT isozymes. The results of the studies below infer that in CHO cells mSHMT converts serine to glycine and cSHMT converts glycine to serine.

In wild type CHO cells the specific activity of mSHMT is approximately 20 times higher than cSHMT. A glycine requiring mutant cell line named 51-11 was produced by mutagenesis with ethylmethane sulphonate (EMS). 51-11 cells have lost all mSHMT activity and half of the cSHMT activity leaving only 10% of overall wild-type SHMT activity (Kao and Puck, 1968). A 51-11 revertant was also produced by EMS mutagenesis to obtain the cell line R13 which has 50% of wild type mSHMT activity and normal cSHMT activity (Chasin *et al.*, 1974).

Pfender and Pizer, (1980) used K-1 (wild-type), 51-11 and R13 cells to measure serine and glycine levels. They found a correlation between the amount of mSHMT and the intracellular glycine concentration and the extent to which extracellular serine can be utilised to increase the glycine pool. K-1 and 51-11 have comparable activities for serine biosynthesis. However, it was found that for glycine synthesis K-1 had an appreciably higher glycine concentration than the other two cell lines. 51-11 was unable to expand the glycine pool when serine was added to the media.
Narkewicz et al., (1996) measured the serine and glycine utilisation and production in wild-type (K-1) and 51-11 cells. They found that the rate of glycine production was lower in 51-11 cells but the rate of glycine utilisation was similar in both cell lines. The rate of serine production was similar in both cell lines but the rate of serine utilisation was lower in 51-11. Increasing the serine concentration of the growth media lead to an increase in glycine production in K-1 but not in 51-11. The activity of the GCS was low in both cell lines.

Over expression of human cSHMT in GlyA CHO cells lacking mSHMT activity did not eliminate the requirement of glycine in these cells. Therefore, the glycine requirement of GlyA CHO cells was not due to the reduced levels of cSHMT limiting the rate of glycine synthesis (Choi and Shane, 1996).

The lack of other folate utilising enzymes can lead to glycine auxotrophy in CHO cells. Cells with the mutation GLYB have a 10 fold lower rate for serine to glycine conversion but have wild-type levels of mSHMT and cSHMT. GLYB auxotrophy was found to be due to these cells containing 1/5 of the parental level of folylpolyglutamates which may be due to the impaired transport between the mitochondria and the cytosol or impaired recycling of folates (Taylor and Hanna, 1982).

It was found that CHO cell line AUXB1 lacks folylpolyglutamyl synthase (FPGS) activity and is, therefore, unable to retain folates within the cell. FPGS is the enzyme which adds glutamate residues to the THF molecule (figure 1.5). The AUXB1 cell line is auxotrophic for methionine, glycine, purines and thymidine. By transfecting cells with the E.coli or human FPGS gene it was found that folates could by polyglutamylated by up to 3 residues (Osborne et al., 1993). Purine and pyrimidines could be synthesised by the transfected cells but they remained auxotrophic for glycine. These cells lacked mitochondrial folates but had normal cytosolic folate pools (Lowe et al., 1993). When a mitochondrial target sequence was added.
to the E.coli FPGS gene and was transfected into AUXB1 mitochondrial folates began to accumulate and the cells were no longer auxotrophic for glycine (Lin and Shane, 1993). This evidence suggests that mitochondrial folates supply one carbon metabolism in the cytosol and that this supply might be via formate.

1.5.3 Plant metabolism

In C3 metabolism plants the GCS plays an important role in photorespiration. Rebeille et al., (1994) showed that the GCS activity in pea leaf mitochondria is greater than that of mSHMT. As a result high levels of methyleneTHF are maintained directing mSHMT to synthesise serine and recycle the THF for GCS use. Mitochondrial and chloroplast isozymes of pea leaf SHMT have been purified (Besson et al., 1995). mSHMT accounts for 50% of total SHMT activity in pea leaves with cytosolic and chloroplast isozymes accounting for 20-25% each. Total SHMT activity determined from potato tubers is 5-10 times lower suggesting a correlation between SHMT and the presence of the photorespiratory system in leaf tissue. Chloroplasts are the major site of amino acid biosynthesis (Shah and Cossins, 1970). Therefore, chloroplast SHMT may have a role in serine or glycine synthesis (Somerville and Ogren, 1981). Somerville and Ogren, (1981) also showed that in C3 plants glycine decarboxylation via the GCS was the only source of CO₂ release. In wild type Arabidopsis thaliana plants mSHMT was required for this process and was not required for any other function.

A large complex has been isolated from the pea leaf mitochondrial matrix which catalyses the rapid transformation of glycine to serine (Neuberger et al., 1986). This complex was characterised and found to contain mSHMT and the four proteins of the GCS (Bourguignon et al., 1988).
Rebeille et al., (1994) deduced that there was no channeling of THF between the T protein of the GCS with mSHMT in vitro but the protein concentrations used in this experiment were 400 times lower than those present in plant leaf mitochondria. Prabhu et al., (1996) used $^{13}$C NMR to study serine production in *Arabidopsis thaliana*. The results suggested that methyleneTHF is not released from the T protein of the GCS but channeled to mSHMT. A second pathway for serine production was found to be from formate by the action of C1-THF synthase and SHMT. As one carbon groups produced by the GCS are used solely for photorespiration, it is proposed that this second pathway supplies one carbon groups for all the other metabolic requirements of the plant cell.

The mitochondrial folate pool of pea leaf contains folate with 4-5 glutamate residues bound and the glutamate chain length of THF has been studied with respect to affinity constants for THF and maximal velocities of SHMT and the GCS in pea leaf mitochondria (Besson et al., 1993). THF containing chains of 3 or more glutamate residues binds 10 fold more tightly to the GCS and mSHMT than THF with a smaller glutamate chain length but the maximum velocities remained the same.

In animal cells folate is provided by the exogenous medium but plant cells synthesise folate *de novo*. By measuring pea leaf mitochondrial folate pools, and folate biosynthesis enzymes, Neuberger et al., (1996) concluded that mitochondria are the major site for folate and thymidylate synthesis in plants. The mitochondrial folate pool represented approximately half of the total folate pool in pea leaves and contained all the enzymes necessary for THF polyglutamate synthesis whereas folate biosynthesis enzymes were absent from the cytosol. Potato tuber mitochondria also contained significant amounts of folate suggesting that this was a general feature of plant mitochondria rather than related solely to photorespiration.
In conclusion the role of SHMT isozymes depends on the organism and its individual pathways of one carbon metabolism. The role of the SHMT isozymes in human cells varies according to tissue type and the precise roles within each tissue type are still to be elucidated.

1.6 Effects of aberrant SHMT activity in humans

In humans altered SHMT activity has been implicated in several disorders.

1.6.1 Role of SHMT in cell proliferation and cancer

SHMT has been shown to play an important role in the biosynthesis DNA precursors. As mentioned previously MethyleneTHF is metabolised by the cell to form 10-formylTHF, 2 molecules of which are also required for purine ring synthesis (see figure 1.4). MethyleneTHF is also used by thymidylate synthase to add a methyl group to deoxyuridine monophosphate to form deoxythymidine monophosphate (figure 1.6). A molecule of glycine is used in the synthesis of purine rings.

The following evidence suggests that SHMT activity and uptake of serine from exogenous media play a significant role in proliferating cells. Addition of radiolabelled serine to medium in which peripheral blood lymphocytes are growing and stimulating these cells to divide with the mitogenic agent phytohaemagglutinin (PHA) leads to the incorporation of radiolabel into DNA and the elevation of SHMT activity. Thorndike et al., (1979) and Eichler et al., (1981) have shown that SHMT activity is elevated 4-9 fold in PHA stimulated peripheral blood lymphocytes compared with non stimulated cells. SHMT activity increases and incorporation of radiolabel occurs before an increase in cell number. Inhibitors of DNA synthesis block incorporation of radiolabel into DNA. SHMT activity in the uterus or prostate has been shown to be increased after stimulation with oestrogen or testosterone suggesting that cells undergoing rapid proliferation require an increased supply of one carbon units (Schirch, 1982). Snell, (1984) has also shown that there is an increased capacity for
serine biosynthesis from glycolytic intermediates in the fetal-perinatal period in the rat. During this time there is a rapid phase of liver growth and it is thought that serine is needed for incorporation into proteins and DNA or RNA precursors during this phase. Serine is classified as a non-essential amino acid, as it can be synthesised by cells. However, Eagle, (1959) demonstrated that serine is required in the culture media if cells are undergoing rapid proliferation. 3-phosphoglycerate dehydrogenase catalyses the first committed step of serine biosynthesis and the activity of this enzyme is higher in cells of high renewal capacity e.g. kidney, thymus, testis, spleen and in neonatal or regenerating liver (Snell and Weber, 1986) with serine dehydratase and serine aminotransferase activities almost absent, meaning all serine synthesised can be channeled into utilisation by SHMT.

As a link between increased SHMT activity and cell proliferation has been demonstrated it follows that in cancers, where cell proliferation is amplified, there should also be an increase in SHMT activity. Studies have been performed on the enzymes of serine biosynthesis and utilisation in a variety of solid tumours and tumour cell lines. Snell et al., (1988) studied the enzymes 3-phosphoglycerate dehydrogenase, SHMT, serine aminotransferase and serine dehydratase. Serine aminotransferase and serine dehydratase activities were essentially absent from normal skeletal muscle and normal colon. This pattern was repeated in human colon carcinomas and transplantable rat sarcomas. However, the activity of 3-phosphoserine dehydrogenase was elevated 30 fold and 10 fold and SHMT activity was elevated 6 fold and 5 fold in rat sarcomas and human colon carcinomas respectively when compared to normal tissue. In another study (Snell and Weber, 1986) hepatomas 20 and 3294A were studied (slow growing and fast growing hepatomas respectively). In normal liver serine dehydratase and serine aminotransferase activities are present. However, in the hepatomas the activity of these two enzymes were absent. 3-
phosphoglycerate activity was elevated approximately 10 fold and 60 fold and SHMT activity was retained, although at a lower level than in control liver. In a more malignant hepatoma (5123tc) SHMT activity was increased 3 fold over control liver SHMT activity. Again in 1985, Snell found that in rat tumours there were appreciable SHMT and phosphoserine aminotransferase activities, but no serine aminotransferase or serine dehydratase activities. It has also been demonstrated (Snell et al., 1987) that 3-phosphoglycerate dehydrogenase and SHMT activities are increased markedly during the transition of hepatoma 3924A cells in culture from a resting non-proliferating phase into a proliferating, growth phase. This increase in activity was concomitant with incorporation of radiolabel from exogenous serine into nucleic acid. The activities of the two enzymes dropped when cells reached confluency and entered the plateau phase.

These studies indicate a pattern of enzyme activities in tumours. If serine aminotransferase and serine dehydratase are absent in the normal cell then its corresponding tumour will show an increase in enzymes of serine biosynthesis and SHMT. If the activities of serine aminotransferase and serine dehydratase are normally present, then these activities are removed with the retention of SHMT activity and an increase in serine biosynthesis pathway enzymes. In either case there is an increase in serine biosynthesis and a channeling of this serine into nucleotide biosynthesis by the action of SHMT.

1.6.2 Schizophrenia and Psychosis

There is considerable debate as to whether SHMT activity contributes to psychotic disorders including schizophrenia. Evidence is conflicting as to whether SHMT activities and serine or glycine concentrations are higher or lower in brains and plasma of schizophrenic patients. Some of the evidence is cited below.
Studies by Carl et al., (1978) found significant decreases in levels of SHMT and methionine adenosyltransferase in blood cells from normal patients and those diagnosed with schizophrenia. Waziri et al., (1984) demonstrated plasma serine levels were significantly higher in patients with psychoses than in control patients and SHMT activities were significantly lower in psychotic patients. It was proposed that altered plasma serine concentrations could, therefore, be an indicator of susceptibility to psychosis.

SHMT activities and serine and glycine concentrations were measured in medial and lateral temporal lobes of frozen autopsied brains of schizophrenics and control brains. It was found that the Km of SHMT and concentrations of serine and glycine were higher in the medial temporal lobe in schizophrenic patients (Waziri et al., 1990).

Devor and Waziri, (1993) identified a single major gene locus regulating plasma serine and glycine concentrations and it is proposed that this locus is SHMT.

Contrary to this evidence a paper by Lucca et al., (1993) demonstrated no difference between schizophrenics and a control group when given oral serine and conversion to glycine measured. Therefore, SHMT activity could not be deficient.

It is conceivable that an SHMT imbalance could lead to high glycine concentrations. There is no active transport of glycine across the blood/brain barrier (Smith et al., 1987) and the main source of glycine in the CNS has been demonstrated to be from serine by SHMT action (Shank and Aprison, 1970). Glycine is an absolute requirement for activation of the NMDA receptor producing an excitatory response in neurones (Kemp and Leeson, 1993). A high glycine concentration could potentially have a toxic effect at the NMDA site (Waziri et al., 1990), which would in turn have a degenerative effect on neurones leading to abnormal brain functions.
1.6.3 Smith Magensis Syndrome

Smith Magensis syndrome is a mental retardation disorder exhibiting a complex phenotype. Indicators include developmental and growth delay, facial abnormalities and unusual behaviours. The syndrome may be due to a haploinsufficiency of several contiguous genes localised to chromosome 17p (Elsea et al., 1995). cSHMT has been shown to be localised to 17p11.2 (Garrow et al., 1993). In a study of 26 patients with Smith Magensis syndrome, all had one copy of the cSHMT gene deleted and SHMT activity in patient lymphoblasts was only 50% of unaffected parents. Loss of cSHMT activity may account for some of the traits exhibited by these patients (Elsea et al., 1995).

1.7 Inhibitors of SHMT activity

As the increased activity of SHMT has been demonstrated in cancer cells, the enzyme is a potential target for drugs to combat some forms of this disease. Along with dihydrofolate reductase (DHFR) and thymidylate synthase (TS), SHMT is part of the thymidylate synthase cycle. (figure 1.6). Both TS and DHFR have inhibitors in therapeutic use in cancer chemotherapy. However, the potential for inhibitors of SHMT activity to be used in the same way has often been overlooked. Inhibitors of SHMT activity studied to date fall into two classes; natural inhibitors present in a cell and those designed to specifically inhibit SHMT.

The first class includes the compounds 5-methylTHF (Matthews et al., 1982) and 5-formylTHF (Stover and Schirch, 1990; 1991) which are competitive inhibitors of SHMT. These 2 compound bind the enzyme-glycine complex with similar affinities but the triglutamate form of 5-formylTHF is released much more slowly from the enzyme-glycine complex than the triglutamate form of 5-methylTHF. Inhibition is increased with increasing addition of glutamate residues. Recently Lin et al., (1996) have shown that the plant amino acid mimosine inhibits SHMT in CHO cells although the mechanism of inhibition in this case is not clear.
Several different approaches have been used to chemically inhibit SHMT activity. These involve either a 3-hydroxy amino acid analogue to mimic serine, antifolates to mimic THF, or PLP analogues.

Wang et al., (1981) investigated the activity of D-fluoroalanine and related amino acids on lamb and rabbit liver cSHMT. The SHMT catalysed reaction lead to the elimination of hydrogen fluoride with the concomitant formation of aminoacrylate, an enzyme-PLP intermediate. The reaction inhibited the enzyme with the loss of one cysteine residue. The Km for D-fluoroalanine was 50mM suggesting this compound would form an effective drug treatment.

The compound 4-chloro-L-threonine was tested by Webb and Matthews, (1995). SHMT cleaves this compound to form glycine and chloracetaldehyde, the latter of which is responsible for inhibiting the enzyme. SHMT was protected from inhibition by serine and glycine.

Other compounds tested include analogues of serine (Tendler et al., 1987). α-vinylserine, α-allylserine, α-prop-2-ynylserine and N-(dichloroacetyl)serine were designed as irreversible inhibitors and tested for activity. All compounds had little inhibitory effect at 0.25-4mM. At 2.5-10mM α-vinylserine was found to be a competitive reversible inhibitor. It could be that these compounds were too bulky to act as inhibitors or that removal of the β proton necessary for irreversible inhibition does not occur. So far no amino acid analogue has been found which inhibits SHMT at a level required for potential drug use.

The properties of NSC127753, a triazine antifolate, were investigated by Snell and Riches, (1989). This compound already has an established anti-tumour activity (Corbett et al., 1982) associated with impaired thymidylate synthase activity and decreased cell viability. The compound was found to inhibit SHMT in mouse myeloma X63 cells in culture with an
IC$_{50}$ of 5x$10^{-9}$ M. Incorporation of [6-$^3$H] deoxyuridine incorporation was also inhibited although this result could be due to impaired TS activity.

The effects of antimetabolites of PLP were investigated in vivo by Bukin et al., (1979) on mouse spleen SHMT. Compounds were administered to live animals via interperitoneal injections. D-cycloserine was found to inhibit SHMT and when pyridoxal azine was administered at 75mg/kg body weight injected 1 hour before the injection of D-cycloserine there was a slight increase in the effect of D-cycloserine inhibition. Pyridoxal azine also exhibited moderate inhibition when administered alone. 4-vinylpyridoxal administered at 0.5-1.0mg/kg did not reduce SHMT activity alone but increased the effect of D-cycloserine and its dimer 2-2.5 fold, leading to more extensive and prolonged inhibition of SHMT.

Trakatellis et al., (1995) used SHMT activity in mitogen stimulated cells to test antiproliferative and immunosuppressive compounds, namely D-actinomycin, cytarabine, L-asparaginase and cyclosporin. It was found that increasing concentrations of these enzymes increased inhibition of SHMT activity and by combining these agents with 4-deoxypyridoxine (dB6) the inhibitory effect of dB6 on SHMT was increased. Therefore, by combining these drugs with dB6, smaller doses could be used to achieve the same effect.

An alternative strategy for drug design would be to tailor an inhibitor to the structure of SHMT. This process is known as rational drug design. However, in order to undertake rational design the tertiary and quaternary structure of SHMT is required, which to date has not been determined.

1.8 Crystallography of SHMT protein

Unsuccessful attempts have been made to determine the tertiary and quaternary structure of SHMT by X-ray crystallography. SHMT isolated from both prokaryotic and eukaryotic
organisms has been crystallised but crystals were either not of a high enough quality or disintegrated when subjected to X-radiation.

Miyazaki et al., (1986) crystallised *Hyphomicrobium methylovorum* SHMT using ammonium sulphate. Rhomboid crystals of the apo form of SHMT were produced within 15 days at 5°C. It is not clear whether any attempts were made to determine the X-ray crystal structure of SHMT from these crystals.

*E.coli* SHMT has also been crystallised in order to perform X-ray crystallography. Two crystal forms were obtained, orthorhombic and monoclinic. Both crystal forms diffracted to 2.8 angstroms. (Stover et al., 1993). However, no results have been published and recent work published by the authors suggests that they do not have access to an X-ray structure.

Rabbit mitochondrial and cytosolic isozymes of SHMT protein have been crystallised (Schirch et al., 1984). mSHMT crystals were tetragonal bipyramids, diffracted to no more than 12 angstroms resolution and were sensitive to X-radiation (Stover et al., 1993). The cytosolic isozyme crystallised as fine microneedles but no crystals which diffracted could be obtained on a regular basis (Stover et al., 1993).

1.9 Heterologous expression of protein  
Please refer to chapter 3

Heterologous expression is the expression of a protein in an organism different to its origin. In many cases mammalian DNA is expressed in *E.coli* and kits are now commercially available to allow the expression and rapid affinity purification of the protein of interest. Many of these kits involve the expression of the protein as a fusion with another protein e.g. maltose binding protein (Riggs et al., 1989) or Glutathione-S-transferase (Smith and Johnson, 1988). The fusion product can then be divided into its individual components with proteases such as thrombin or factor X.
The pET system (Novagen) was developed by Studier et al., (1990). With this system the gene of interest is placed under the control of T7 transcription and translation sequences. The plasmid is transformed into a strain of *E.coli* containing a chromosomal copy of the T7 RNA polymerase gene under the control of the lacUV5 promoter. Induction of T7 RNA polymerase production is by addition of IPTG to exponentially growing cells. T7 RNA polymerase then transcribes the gene of interest. This method of indirect expression prevents the build up of the gene product too early in the growth of the culture which is especially important if the product is toxic to the cell. Before addition of IPTG there is a small amount of T7 RNA polymerase produced. This can be counteracted by the presence of a second plasmid in the expression strain. pLysS contains the gene for T7 lysozyme which is expressed at a low level and inhibits any T7 RNA polymerase which may be produced before induction. There is not enough lysozyme to significantly effect the T7 RNA polymerase after induction. T7 lysozyme also weakens the *E.coli* cell wall, easing lysis of the bacteria after expression.

The vector pET14b also incorporates a sequence coding for a run of 6 histidine residues (histidine tag) and a thrombin cleavage site before the point where the gene of interest is inserted. As a result the gene is expressed to form protein with these sequences at the N-terminus. The expressed protein can then be purified form the cell lysate by passing the lysate through a resin containing either nickel ions (Novagen) or cobalt ions (Clontech). The histidine tag binds to the resin and can then be eluted by addition of imidazole which has a higher affinity for the metal ions than histidine. Thrombin can then be used to cleave off the histidine residues and leave just glycine, serine and histidine on the amino terminus of the protein.
The advantage of cobalt over nickel is that cobalt ions have 4 sites of attachment to the resin as opposed to 3 sites for the nickel ions, leading to less metal ion loss from the column and an increased capacity for histidine binding. In addition to this E.coli proteins are less likely to bind to the column (for example, a 21KD E.coli protein which has 20 histidine residues in 196 amino acids has been shown to bind to nickel ion columns, but binds less tightly to cobalt ion columns (Wulfing et al., 1994)). Cobalt columns also bind the histidine tag slightly less strongly than nickel ones. As a result less washing of the column is required and the protein can be eluted using less stringent conditions.

1.10 Site directed mutagenesis

Site directed mutagenesis is a method of changing specific amino acid residues by replacing them with another amino acid. This is achieved by changing the desired codon within the gene for the protein. Coupled with a system for expressing the mutant protein, this method is a powerful tool for investigating the role of particular amino acids within the protein. This method can complement the data obtained from X-ray crystal structures by modifying the specific residues implicated in activity to determine the accuracy of the predictions. In the absence of X-ray crystal data site directed mutagenesis can be used to determine whether educated guesses of important residues warrant further investigation. Site directed mutagenesis also differs from chemical modification in that the latter is generally used to modify classes of amino acids e.g. iodoacetate reacts with exposed cysteine residues, DEPC targets histidine residues. However, modification by chemical agents lacks specificity in targeting a precise residue. In addition chemical modification adds a chemical group to the residues concerned and, therefore, the enzyme can be inhibited by steric hindrance preventing substrates binding rather than through a critical catalytic residue being modified. Site directed mutagenesis overcomes the specificity and steric hindrance problems and another
The advantage of site directed mutagenesis is that the residue of interest can be changed to retain the charge (e.g. change glutamic acid to aspartic acid), reverse the charge (e.g. change glutamic acid to lysine) or remove the side group by changing to alanine to determine whether the charge of the amino acid side chain is necessary for catalysis.

There are several different ways of performing site directed mutagenesis. These include, PCR based methods (e.g. Stratagene ExCite) and methods using ssDNA as a template such as the Promega Altered Sites kit. Using the Promega kit, the gene of interest is placed into the supplied vector (pALTER-1) and selected using blue/white selection. pALTER-1 contains two antibiotic resistance genes, ampicillin and tetracycline. The ampicillin gene has been inactivated by a small deletion. Mutagenesis is performed by annealing the mutagenic oligo (containing the desired codon change), an oligonucleotide to repair the ampicillin gene and a third to knock out the tetracycline gene to the single stranded pALTER-1 containing the gene to be modified. A second strand is the synthesised with T4 DNA polymerase and the three pieces of synthesised DNA ligated with T4 DNA ligase to form a complete second strand. The double stranded plasmid is then transformed into a DNA repair deficient strain to ensure the mismatched bases are not recognized and replaced. After a few replications there should be plasmid where both strands have been mutated. The plasmid is then transformed into a standard *E.coli* strain (e.g. JM109, DH5α) to prevent creation of spontaneous mutations in the mismatch repair deficient strain. Colonies are selected on L-agar containing ampicillin and tested for tetracycline sensitivity. Sequencing is then performed to confirm the mutation. Further site directed mutagenesis can be performed on the mutated DNA by annealing another mutagenic oligonucleotide along with ampicillin knockout and tetracycline repair oligonucleotides.
1.11 Mapping of restriction endonuclease sites in bacteriophage lambda

Restriction sites present in DNA inserted into bacteriophage lambda clones can be rapidly mapped by a method involving partial endonuclease digestion (Rackwitz et al 1984).

Radioactively labeled probes complementary to the left and right cohesive (cos) ends are then added, each one to half of the digested DNA. The fragments are separated by agarose gel electrophoresis and fragments hybridising to the probes are detected using X-ray film. This method has been adapted to use of Digoxygenin (DIG) (Boehringer Mannheim) labeled probes to be used with each batch of digested DNA being split into two, separated by agarose gel electrophoresis and transferred to a nylon membrane. DIG labeled cos probes are then hybridised to the blotted DNA.

The lengths of the DNA fragments hybridising to the probes can be determined by comparison to size standards and each length represents the distance of the restriction site from the left or right cos end of the lambda clone. Please refer to chapter 5.

1.12 Structure of eukaryotic genes encoding proteins

Many eukaryotic genes consist of regions coding for protein (exons) interspersed with non coding sequence (introns). At the junctions between intron and exon there are highly conserved motifs. The 5' intron splice junction (donor site) consensus is C/A AG\\text{I}/G\text{R}AGT where R is a purine and I represents the exon-intron boundary. AT the 3' end the intron splice junction (acceptor site) is Y_1NYAGIR where Y is a pyrimidine and N is any base (Ohshima and Gotoh, 1987). Introns also have a branch site near the 3' end which has the consensus YNYRAY. When RNA polymerase II transcribes a gene all of the exons and introns are transcribed to form the primary transcript. The 5' end of the mRNA precursor is modified by the addition of a deoxyguanine molecule in a 5'-5' linkage to form a “cap”. The added guanine is then methylated and adjacent nucleotides may also be methylated. Most eukaryotic mRNA molecules contain a polyA tail at their 3' end. The cap at the 5' end is
important for the stability of the molecule and for subsequent intron splicing and translation of the mRNA into protein.

To form mRNA the introns need to be spliced out of the sequence. Intron splicing is catalysed by small nuclear ribonucleoprotein particles (snRNPs). The snRNP U1 binds to the intron donor site and protects a 15 nucleotide region at the 3' end of the upstream exon. Another snRNP (U2) binds to the pyrimidine tract at the 3' end of the intron and the splice acceptor site is recognised by snRNP U5. U4 and U6 are also required in the process. The assembly of all these snRNPs is called the splicosome. The mechanism of splicing is thus. The phosphodiester bond between the upstream exon (exon 1) and the intron is broken by an attack by the 2'-hydroxy group of the invariant adenine residue in the branch site forming a 2',5'-phosphodiester bond between this adenine and the phosphate at the 5' terminal of the intron. The branch generated at this site is called a lariat. The 3' hydroxy terminus of exon 1 attacks the phosphodiester bond between the intron and exon 2. Exons 1 and 2 are joined and the intron is released. For reviews on intron splicing see Legrand and Chanfreau, 1994; Sharp, 1987.

1.13 Amplification of human genomic DNA by the polymerase chain reaction
The polymerase chain reaction (PCR) is used to amplify DNA where the 5' and 3' ends of the sequence to be amplified are known. Isolation of the cDNA for a particular gene provides the mRNA but will not contain the sequence of the introns or any 5' or 3' sequence outside of the transcribed region. Therefore, if isolation of genomic DNA from the known cDNA sequence into the genome is required then the cDNA can only provide the sequence for one of the primers. In order to enable the PCR amplification of specific genes from human genomic DNA, systems have been designed to overcome the absence of the second PCR primer. The PromoterFinder kit (Clontech) is a commercial form of 5 libraries of human genomic DNA, each digested with a different blunt ended restriction enzyme and with an adaptor ligated to each end of the fragments. Two rounds of amplification are performed on each library, using
for the first round one gene specific primer and one primer within the adaptor. The products from the first round of PCR are diluted and used as templates for a second round using primers located within the product from the first round of amplification (nested PCR). This method reportedly yields a single PCR product in at least four out of the five libraries (Clontech). A mixture of proof reading and non-proof reading thermostable DNA polymerases can be in order to allow amplification of large templates (Barnes, 1994). The use of five libraries ensures that the length of DNA between the gene specific primers and the end of the DNA fragments is not too long to enable amplification or too short to amplify only known sequence in at least one library.
CHAPTER 2

MATERIALS AND METHODS
2.1 Reagents

Agarose gel sample buffer: 0.25% (w/v) bromophenol blue, 15% (w/v) ficol in water

Luria Broth (LB): 10g tryptone, 5g yeast extract, 10g NaCl in 1 litre water pH7.0

L-Agar/Agarose: LB containing 15g agar or agarose per litre of broth (7g for top agar/agarose)

SM: 5.8g NaCl, 2g MgSO₄.7H₂O, 50ml 1M Tris/HCl pH 7.5, 5ml 2% (w/v) gelatin, water to 1 litre

SOB: 20g tryptone, 5g yeast extract, 0.5g NaCl, water to 950ml, 10ml 250mM KCl, pH to 7.0

and make up to 1 litre. After autoclaving add 5ml 2M MgCl₂.

SOC: SOB + 20mM glucose

50 x TAE: 242g Tris base, 57.1ml glacial acetic acid, 100ml EDTA pH8.0

TE: 10mM Tris/HCl pH8.0, 1mM EDTA pH8.0

Xgal: 20mg/ml in dimethylformamide, stored in dark in glass container at -20°C

IPTG: 100mM stock in H₂O for blue white selection, 1M stock for induction of expression

stored at -20°C

Buffers for nucleic acid hybridisation

denaturing buffer: 0.5M NaOH, 1.5M NaCl (+ 0.1% (w/v) SDS for colony hybridisation)

neutralisation buffer: 0.5M Tris/HCl pH7.2, 3M NaCl (1M Tris/HCl pH7.5, 1.5M NaCl for colony hybridisation)

prehybridisation buffer: 5xSSC, 1.0% (w/v) blocking solution, 0.1% (w/v) N-lauroylsarcosine, 0.02% (w/v) SDS

hybridisation buffer: prehybridisation solution + 5-25ng/ml Digoxygenin labeled probe

20 x SSC: 3M NaCl, 300mM tri-sodium citrate pH7.2

10% (w/v) blocking solution: powdered blocking reagent was obtained from Boehringer Mannheim
Preparation of plasmid DNA

solution 1: 50mM glucose, 25mM tris/HCl pH 8.0, 10mM EDTA

solution 2: 0.2M NaOH, 1% (w/v) SDS

solution 3: 60ml 5M potassium acetate, 1.5ml glacial acetic acid, 28.5ml water

Buffers for SDS-PAGE

4x (w/v) resolving gel buffer: 1.5M Tris/HCl pH 8.8, 0.4% (w/v) SDS

4x (w/v) stacking gel buffer: 0.5M Tris/HCl pH 6.8, 0.4% (w/v) SDS

12% (w/v) resolving gel: 3.45ml water, 2.5ml 4x resolving gel buffer, 4.0ml 30% (w/v) acrylamide, 50μl 10% (w/v) ammonium persulphate, 5μl TEMED

4% (w/v) stacking gel: 6.2ml water, 2.5ml 4x resolving gel buffer, 1.3ml 30% (w/v) acrylamide, 50μl 10% (w/v) ammonium persulphate, 10μl TEMED

10x running buffer: 0.5M Tris, 1% (w/v) SDS, 0.55M glycine

10% (w/v) SDS, 1ml glycerol, 7ml water, 5mM

2x sample buffer: 1ml Tris/HCl pH 6.8, 1ml 10% (w/v) SDS, 1ml glycerol, 7ml water, 5mM DTT, trace bromophenol blue

Coomassie stain: 45% (v/v) methanol, 5% (v/v) acetic acid, 0.1% (w/v) Coomassie blue

Destain: 7.5% (v/v) acetic acid, 5% (v/v) methanol

Antibiotics

Ampicillin: 50mg/ml in H₂O. Filter sterilised stored at -20°C

Tetracycline: 12.5mg/ml in 80% (v/v) ethanol stored at -20°C

Chloramphenicol: 34mg/ml in ethanol stored at -20°C

Buffers for cSHMT protein purification

8x Binding buffer: 160mM Tris/HCl pH 7.9, 4M NaCl, 40mM imidazole. Dilute to 1x buffer for use.
Wash buffer: 160mM Tris/HCl pH 7.9, 4M NaCl, 480mM imidazole

Sonication buffer: 20mM Tris/HCl pH 7.9, 100mM NaCl

### 2.2 Bacterial strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH5α</td>
<td>supE44, ΔlacU169, (φ80 lacZA15M15), hsdR17, recA1, endA1, gyrA96, thi-1, relA1 (Hanahan 1983)</td>
</tr>
<tr>
<td>JM109</td>
<td>endA1, recA1, gyrA96, thi, hsdR17 (rK-,mK+), relA1, supE44, λ-Δ(lac-proAB), [F’, traD36, proA⁺B⁺, lacZDA15] (Yanisch-Perron et al., 1985) (Promega)</td>
</tr>
<tr>
<td>ES1301 mutS</td>
<td>lacZ53, mutS201::Tn5, thyA36, rha-5, metB1, deoC. IN(rrnD-rrnE) (Siegel et al., 1982) (Promega)</td>
</tr>
<tr>
<td>BL21(DE3)</td>
<td>F’, ompT rB-mB-, (DE3) (Studier and Moffatt, 1986) (Novagen)</td>
</tr>
<tr>
<td>BL21(DE3)pLysS</td>
<td>F’, ompT rB-mB-, (DE3), pLysS Cm’ (Studier, 1990) (Novagen)</td>
</tr>
<tr>
<td>C600</td>
<td>thi-1, thr-1, leuB6, lacY1, tonA21, supE44 (Jendrisak et al., 1987)</td>
</tr>
<tr>
<td>INVαF’</td>
<td>F’, endA1 recA1, hsdR17 (rK-,mK+), supE44, thi-1, gyrA96, relA1 φ80lacZDA15M15, Δ(lacZYA-argF)U169 (Invitrogen)</td>
</tr>
</tbody>
</table>

### 2.3 Vectors used

- pUC18: general purpose cloning vector (Yanisch-Perron et al., 1985)
- pCR™TII: PCR product cloning vector (Invitrogen)
- pT7blue(R): PCR product cloning vector (Novagen)
- pET14b: *E. coli* expression vector (Studier et al., 1990) (Novagen)
pALTER-1®: site directed mutagenesis vector (Promega)

pUS1203: rabbit cSHMT cDNA containing full orf in pUC18 (Byrne, 1992)

pUS1206: human cSHMT in pUC18 containing cSHMT from nt 697 (appendix 1) and 850bp of upstream intron sequence (Byrne, 1992)

pUS1213: human cSHMT cDNA (appendix 1 nt 265-3016) lacking first 29 amino acids in orf. subcloned into pUC18 (Byrne, personal communication)

pUS1217: human cSHMT cDNA (appendix 1 nt 1-1867) containing the whole orf in pUC18 (Byrne, personal communication)

EMBL3A: lambda replacement vector (Frischauf et al., 1983)

2.4 Oligonucleotides used

All oligonucleotide sequences are in a 5' to 3' orientation

Human and rabbit orf PCR

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>kjc5</td>
<td>AACCAGTCATATGACGAT</td>
</tr>
<tr>
<td>kjc6</td>
<td>CCAGGTGGGTCCATATGGGG</td>
</tr>
<tr>
<td>C</td>
<td>CAACCAGCCATATGGCGAC</td>
</tr>
<tr>
<td>A</td>
<td>ACTCATATGCCGGCCCCTCTA</td>
</tr>
</tbody>
</table>

Human cSHMT cDNA sequencing primers

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>kjc8</td>
<td>TGTGCCCTTGGGAACGGAGAT</td>
</tr>
<tr>
<td>kjc9</td>
<td>GGCCTATAAGCTGGACCAC</td>
</tr>
<tr>
<td>kjc10</td>
<td>TGGTGCCGGCTGGCGTGGTG</td>
</tr>
<tr>
<td>kjc11</td>
<td>GGCAGGGAAACAGCACAGAA</td>
</tr>
<tr>
<td>kjc12</td>
<td>GCCAGGGAAACAGCAGAA</td>
</tr>
<tr>
<td>kjc13</td>
<td>ATCTCCGTTCAAAAGCCACA</td>
</tr>
<tr>
<td>kjc16</td>
<td>GGCGTTGCTTCTTCCATCCAGC</td>
</tr>
<tr>
<td>kjc20</td>
<td>TTTGAATCTATGCCCTACA</td>
</tr>
</tbody>
</table>

Site directed mutagenesis

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ampr</td>
<td>GTTGCCATTGCTGCGGATCGATGCTGGTG</td>
</tr>
<tr>
<td>ampko</td>
<td>GTTGCCAATTGCGGATCGATGCTGGTG</td>
</tr>
<tr>
<td>tetr</td>
<td>GCCGGGCCTCTTGGCGGGAATATCGTCCA</td>
</tr>
<tr>
<td>tetko</td>
<td>GCCGGGCCTCTTGGCGGGAATATCGTCCA</td>
</tr>
<tr>
<td>kjc7</td>
<td>AGCTCGAATTCG</td>
</tr>
<tr>
<td>D182N</td>
<td>TGGCTACATCAACTATGACC</td>
</tr>
<tr>
<td>W15A</td>
<td>GCTGACCTGCGTCCTCCACA</td>
</tr>
<tr>
<td>W15F</td>
<td>GCTGACCTGATTCCTCCACA</td>
</tr>
<tr>
<td>W15Y</td>
<td>GCTGACCTGATTCCTCCACA</td>
</tr>
<tr>
<td>C68A</td>
<td>TAGGCTCTGGCCCTTAAATAAC</td>
</tr>
<tr>
<td>W111F</td>
<td>CCCACAGTGGCTTGGCGTGCAAC</td>
</tr>
<tr>
<td>W111A</td>
<td>CCCACAGTGGCGGCGTGCAAACG</td>
</tr>
</tbody>
</table>
C204A    GAACCATGCCTACTCCCGA
K257A    CCACTCACGCAACCCCTGCGA
C380stop AAGCCTGATCTATTGCCTG
F424stop GCCCACTAGATTCACAGAGG
K446stop AGAGCCACCCTGTAAGAGTT

Pseudogene sequencing

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>kjc1</td>
<td>CTGAGCCCACCTAGTCATTG</td>
</tr>
<tr>
<td>kjc2</td>
<td>ACATCTTACACTCTCAG</td>
</tr>
<tr>
<td>kjc3</td>
<td>GGATGCTATATTATTGCTCCA</td>
</tr>
<tr>
<td>kjc4</td>
<td>CGACCATTTTTAATAAGCACAAT</td>
</tr>
<tr>
<td>ON-L</td>
<td>AGGTCGCGCCCC</td>
</tr>
<tr>
<td>ON-R</td>
<td>GGGCGGCGACCT</td>
</tr>
</tbody>
</table>

Human genomic PCR

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ap1</td>
<td>GTAATACGACTCAGTATAGGGC</td>
</tr>
<tr>
<td>ap2</td>
<td>ACTATAGGGCCAGCGGTGTT</td>
</tr>
<tr>
<td>gsp1</td>
<td>GTGACATTTTGCTCTTCGTATTACGG</td>
</tr>
<tr>
<td>gsp2</td>
<td>TGGATAACAAAGCTGAGGCGGCGAGT</td>
</tr>
<tr>
<td>kjc21</td>
<td>AGGATCAAATGTTTGTCAG</td>
</tr>
<tr>
<td>kjc22</td>
<td>CAAAACGCTTGGCTGCAGGAAT</td>
</tr>
<tr>
<td>kjc23</td>
<td>CAATCGCTCCTCAATCACAC</td>
</tr>
<tr>
<td>pfk1</td>
<td>TCTTTCAGGCTTGTGCAAGCATCT</td>
</tr>
<tr>
<td>pfk2</td>
<td>GTGACGAGCCTCAGCCTAAACC</td>
</tr>
<tr>
<td>pfk3</td>
<td>GCGCTAGGCGCACTCAGGTCCTA</td>
</tr>
<tr>
<td>22/234B6P</td>
<td>ACCTCCAAGAAAGGTCAGAGC</td>
</tr>
<tr>
<td>309INT-</td>
<td>GTCTCATTTTTCACAGGCG</td>
</tr>
<tr>
<td>INT1</td>
<td>CAGGCCAACAAATTGCTGAATG</td>
</tr>
<tr>
<td>INT2</td>
<td>TTCCAAAAAGTACCCACT</td>
</tr>
<tr>
<td>INT3</td>
<td>GGGCATTAGTCAGTACACA</td>
</tr>
<tr>
<td>8/21 24N</td>
<td>CCTAACAGACCCAAACAG</td>
</tr>
<tr>
<td>8/2124P</td>
<td>GGGCGTGTGGGTGCAATG</td>
</tr>
</tbody>
</table>

2.5 Basic molecular biology techniques

Most molecular biology techniques were performed according to Sambrook et al (1989). 1xTAE was used for agarose gel electrophoresis. Restriction endonucleases were purchased from Boehringer Mannheim Mannheim except cloned NdeI (New England Biolabs). Calf intestinal alkaline phosphatase was supplied by Boehringer Mannheim and New England Biolabs. Shrimp alkaline phosphatase was purchased from Boehringer Mannheim and T4 DNA ligase from Life Technologies.
2.5.1 Small scale preparation of Plasmid DNA

Plasmid DNA was routinely purified using the method of Birnboim and Doly (1979), although Wizard (Promega) or Wizard Plus DNA (Promega) purification systems were occasionally used.

The alkaline lysis method was adjusted slightly to produce DNA suitable for use in the automated DNA sequencer. All vortexing steps were replaced with rapid inversion to prevent the shearing of genomic and plasmid DNA.

Alternatively plasmids to be sequenced were purified by a method recommended by Applied Biosystems. The solutions were the same as for standard alkaline lysis but a PEG precipitation step was incorporated. Up to 4.5ml of culture was pelleted, the cells resuspended in 200μl solution 1 by pipetting up and down. 300μl freshly prepared solution 2 was added, the contents mixed by inversion and placed on ice for 5 minutes. 300μl of solution 3 was added, mixed by inversion and again incubated on ice for 5 minutes. Precipitate was removed by centrifuging at 13000g for 10 minutes and the supernatant was transferred to a clean tube. RNaseA was added to a final concentration of 20μg/ml and the tube was incubated at 37°C for 20 minutes. Following 2 extractions with chloroform an equal volume of isopropanol was added and immediately centrifuged at 13000g for 10 minutes. The pellet was washed with 70% ethanol and dried. The pellet was resuspended in 32μl of water and the DNA was precipitated by adding 8μl 4M NaCl and 40μl 13% PEG8000. The sample was incubated on ice for 20 minutes, centrifuged (13000g) at 4°C for 15 minutes and the pellet was washed in 70% ethanol, dried and resuspended in water. The DNA was stored at -20°C.
2.5.2 Large scale preparation of Plasmid DNA

Plasmid DNA from 50-100ml of bacterial culture was purified using commercially available columns (Qiagen, Promega). The alkaline lysis method was also scaled up to use 5 times the amounts of solutions for the purification of plasmid DNA from 50ml of bacterial culture.

2.5.3 Purification of DNA fragments from an agarose gel

The band of interest was excised from the agarose gel and purified in one of several ways; Geneclean (BIO101, Amersham), Crystal Gelex (Stratatech) or PCR pure (Clontech) using the manufacturer’s protocols. These methods are based on that of Vogelstein and Gillespie, (1979). Alternatively the DNA was spun out of the gel through 3mM filter paper (Whatman) using the method of Weichenhan, (1991).

2.5.4 Rapid lysis of bacterial colonies to ascertain plasmid size

Part of a bacterial colony from an agar plate was added to 30µl of 50mM NaOH, 0.5% SDS, 5mM EDTA pH8.0 followed by incubation at 68°C for 45 minutes in a heated water bath. 10µl of sample buffer was added and the DNA was separated by agarose gel electrophoresis.

2.5.5 Oligonucleotide synthesis and deprotection

Oligonucleotides were synthesised on an Applied Biosystems 381A synthesiser. Before use, oligonucleotides needed to be deprotected and removed from the synthesis column. 1.5ml of 35% ammonia was passed through the column at a rate of 0.2ml every 20 minutes and the eluted ammonia was incubated overnight at 55°C. DNA was precipitated overnight by the addition of 1/10 volume of 3M sodium acetate pH5.2 and 3 volumes of ethanol. DNA was harvested by centrifuging (13000g) for 20 minutes at 4°C. The pellet was washed twice with 80% ethanol and resuspended in water.
2.5.6 Insertion of PCR products into vectors

Two vectors were used to clone polymerase chain reaction (PCR) products (Invitrogen pCRII™ and Novagen pT7blue(R)). Both vectors contain single overhanging thymidine residues which can be ligated to the adenine residues added to the 3' end of the PCR product by non proof reading enzymes such as *Taq* (Marchuk *et al.*, 1991). PCR products were ligated into these vectors according to the manufacturers protocols and transformed into INVαF’ (Invitrogen) or DH5α competent cells.

2.5.7 Automated sequencing

DNA sequencing was performed on an Applied Biosystems 373A automated sequencer by Stephen Dyer.

2.5.8 Preparation of competent cells

DHα competent cells were already available in the laboratory. Competent INVαF’ and ES1301 mutS were purchased from Invitrogen and Promega respectively.

Competent bacterial cells of strains JM109, BL21(DE3), BL21(DE3)pLysS and ES1301 mutS were prepared as follows. A 10ml overnight culture was prepared from a single colony of the bacterial strain. 1ml was used to inoculate 50ml of SOB. This culture was incubated at 37°C, 250rpm until the OD<sub>600nm</sub> was 0.6. The cells were harvested by centrifuging at 1000g for 15 minutes at 4°C, and were resuspended in 25ml 50mM CaCl₂ and left on ice for 30 minutes. The cells were again harvested by centrifuging as before and resuspended in 2ml ice cold 50mM CaCl₂. If not used immediately, ice cold 50% glycerol was added to 15% and the cells were frozen in dry ice or liquid nitrogen.

2.5.9 Transformation of competent cells

Routinely, 20μl of competent cells were defrosted on ice. 10ng of DNA was added and the cells incubated on ice for 30 minutes, heat shocked at 42°C for 45 seconds, and returned
to ice for 2 minutes. 80μl of SOC was added and the cells were incubated at 37°C for 60 minutes before plating on L-agar containing the appropriate antibiotics.

A different protocol was used for transformation of competent ES1301 mutS cells. 100μl competent cells were placed in pre-chilled Falcon 2059 tubes on ice. 1.5μl of mutagenesis reaction was added, and the tubes incubated on ice for 10 minutes. After heat shocking the cells as before and incubating on ice for 2 minutes, 900μl of LB was added and the cells were incubated at 37°C for 30 minutes. 0.5ml of this culture was added to 4.5ml LB containing relevant antibiotics and the culture was incubated overnight.

2.5.10 Southern Blotting
Agarose gels containing DNA bands greater than 10kb were depurinated in 250mM HCl for 5 minutes at room temperature prior to blotting. After rinsing in water the gel was placed in denaturing buffer for 2 x 15 minutes and neutralising buffer for 2 x 15 minutes. The agarose gel was then placed face down on a piece of 3mM (Whatmann) filter paper soaked in 20xSSC and trailing into a reservoir of 20xSSC. A positively charged nylon membrane (Boehringer Mannheim) was then placed on the agarose gel, followed by 3 pieces of 3mM (Whatmann) filter paper soaked in 20xSSC, and dry paper towels. A weight was than placed on top and this apparatus was left overnight. The membrane was removed and baked in an oven for 30 minutes at 120°C or 2 hours at 80°C to fix the DNA.

2.5.11 Colony lifts
Transformed bacterial colonies were replated onto a second plate, incubated overnight and then placed in at 4°C for 1 hour. Positively charged nylon membrane was laid onto the colonies for 1 minute for a primary filter and 3 minutes for a secondary filter. The membranes were specifically cut allowing subsequent orientation with the agar plate. Membranes were placed colony side up on filter paper soaked with denaturing solution for 15
minutes, neutralisation solution for 5 minutes and finally 2xSSC for 15 minutes. DNA was fixed to the membrane by heating for 2 hours at 80°C or for 30 minutes at 120°C. Prior to hybridisation the membrane was incubated in 3xSSC, 0.1% SDS in the hybridisation tubes at 68°C and then wiped with a damp tissue to remove cellular debris.

2.5.12 Digoxygenin labeling of DNA

This was performed according to the manufacturers protocol (Boehringer Mannheim). Two different methods were employed depending on the DNA to be labeled. Oligonucleotides were 3'-end labeled by 1 molecule of DIG-11-dUTP being added to the end of each oligonucleotide. DNA 500bp-2.5kb in length was labeled by random priming. Single stranded DNA was created by boiling and rapid cooling and random hexanucleotides were annealed to prime the second strand synthesis using Klenow enzyme supplied with dNTP's and DIG-11-dUTP. The yield of labeled nucleotides was estimated by diluting the labeled DNA, spotting 1μl of each dilution onto positively charged nylon membrane along with labeled control DNA, and baking of the membrane to fix the DNA. This membrane was then placed through the detection procedure and X-ray film was exposed. The intensity of the exposed film for the test samples could then be compared to the test samples.

Labeled probes were boiled for 10 minutes and snap-frozen before use.

2.5.13 Hybridisation and detection of DNA

Membranes were placed into hybridisation bottles (Hybaid) and incubated in hybridising ovens (Hybaid). The membranes were incubated in prehybridisation solution for 1 hour (2 hours for colony hybridisations) at 68°C. Southern blots and colony hybridisations were then incubated in hybridisation solution overnight at 68°C. The membrane was then washed twice in 2xSSC, 0.1% SDS for 5 minutes at room temperature and then twice in 0.1xSSC, 0.1% SDS for 15 minutes at 68°C. Labeled DNA still attached to the membrane after the washes was detected according to the manufacturers protocol (Boehringer Mannheim).
procedure involved the blocking of the membrane to remove any extra binding sites, the 
incubation of the membrane with Anti-Digoxygenin Fab fragments coupled to alkaline 
phosphatase and the addition of Lumigen PPD and latterly CSPD. These are substrates for 
alkaline phosphatase and when cleaved emit light which can be detected on X-ray film.

2.5.14 5' Phosphorylation of oligonucleotides

100pmol of oligonucleotide was phosphorylated using T4 polynucleotide kinase (Gibco). 
ATP was added to the reaction buffer to a final concentration of 1mM. The reaction was 
incubated at 37°C for 30 minutes and then the enzyme was inactivated by incubating at 70°C 
for 10 minutes.

2.6 Amplification of rabbit and human cSHMT cDNA by PCR

PCR reactions were performed using 20ng of template, 100pmol of each primer, 100µM 
each dNTP and 5 units of Taq polymerase (Boehringer Mannheim) in 10mM Tris/HCl pH 
8.3, 1.5mM MgCl₂, 50mM KCl, 0.1mg/ml gelatin. The PCR reactions were performed at 
95°C for 3 minutes followed by 25 cycles of 95°C for 1 minute, 42°C for 1 minute and 72°C 
for 2 minutes. A final extension time of 72°C for 5 minutes was performed. PCR products 
were analysed by electrophoresis through a 1% agarose gel, bands were excised and purified 
(section 2.5.3) before ligation into pCRII™ (section 2.5.6).

2.7 Modification of pUC18 to remove multiple cloning site

An oligonucleotide was synthesised (kjc7) to replace the multiple cloning site between the 
HindIII and EcoRI sites of pUC18. 1µg of pUC18 was digested with HindIII, phosphatased 
and ligated with 100pmoles phosphorylated kjc7. After transformation into competent 
DH5α, one plasmid (pUS1248) was digested with EcoRI, the digested plasmid was purified, 
phosphatased and ligated with the EcoRI fragment from pCRII™ containing the human 
cSHMT cDNA.
2.8 Site Directed mutagenesis of human cytosolic SHMT

Site directed mutagenesis was performed using the Promega Altered Sites II kit and protocol.

2.8.1 Production of single stranded DNA

Single stranded DNA was created using the Altered Sites kit II protocol (Promega).

2.8.2 Mutagenesis of human cSHMT cDNA

Mutagenesis was performed according to the manufacturers protocol except that 10pmol of mutagenic oligonucleotide was added to the reaction instead of 1.25pmol (Piechocki and Hines, 1994) and the annealing reactions were heated to 95°C instead of 75°C. After addition of the T4 DNA polymerase and DNA ligase, the reaction was incubated at room temperature for 5 minutes before incubating at 37°C (Piechocki and Hines, 1994). The cotransformation of ES1301 mutS and JM109 cells yielded no colonies. Therefore, the resynthesised plasmid was transformed directly into ES1031 mutS cells, according to the manufacturers alternative protocol (Promega) (section 2.5.9). The plasmid DNA from these cells was the extracted and used to transform JM109 or DH5α.

2.9 Expression of cSHMT in E.coli

To assess whether cSHMT protein was soluble when expressed at various temperatures the following protocol was followed. 50 ml of LB was inoculated with 0.4ml of an overnight culture of BL21(DE3)pLysS containing pET14b with the gene of interest. The culture was grown at the required temperature until the OD₆₀₀nm was 0.4-0.6. 50µl of culture was removed, pelleted and resuspended in 20µl SDS-PAGE sample buffer. IPTG was added to a final concentration of 0.4mM and the culture was grown for either a further 2 hours or overnight. 20µl of cells were removed, pelleted and resuspended in 20µl SDS-PAGE sample buffer. The remaining culture was harvested by centrifuging at 4000g for 10 minutes. The
pellet was resuspended in 5ml 50mM Tris/HCl pH8.0. 2mM EDTA. Lysozyme was added to 100μg/ml and 0.5ml of 1% Triton X-100 were added. The mixture was incubated at 30°C for 15 minutes and then microfuged at 4°C for 15 minutes. The supernatant (soluble fraction) was removed and an equal volume of SDS-PAGE sample buffer was added. The pellet (insoluble fraction) was resuspended in 0.5ml SDS-PAGE sample buffer. The protocol was modified by the addition of DNAsel and RNaseA were added to final concentrations of 20μg/ml and 40μg/ml at the same stage as lysozyme to reduce the viscosity of the solution by removing the DNA and RNA.

C380stop and K446stop mutants were grown as above but were also tested for production of soluble protein by incubation of the culture at 20°C in LB containing 2.5mM betaine and 660mM sorbitol (Blackwell and Horgan, 1991).

Protein production was monitored by SDS-PAGE electrophoresis to ascertain whether SHMT had been expressed and whether any SHMT protein was present in the supernatant. 20μl of sample was analysed except for the insoluble protein in 0.5ml sample buffer where 5-10μl was analysed.

Large scale cultures (4x250ml of culture in four 1 litre flasks) were inoculated with 4ml of overnight culture per flask. The culture was grown at the temperature at which each SHMT was soluble (usually 25°C). Expression of SHMT protein was induced as above and cells harvested 2 hours later. The culture was harvested by centrifugation at 4000g for 15 minutes at 4°C and the pellets were stored at -80°C.

2.10 Purification of SHMT protein

SHMT was purified using 2 different manufacturer’s protocols from Novagen and Clontech. Cell pellets from 1 litre of induced culture were thawed and resuspended in 20ml binding buffer (Novagen) or sonication buffer (Clontech). DNaseI and RNaseA were added
**Amount of protein loaded on SDS-PAGE gels**

Overnight culture: 0.04% of culture volume

Before induction with IPTG: 0.1% of culture volume

Post induction with IPTG: 0.04% of culture volume

(uninduced culture has 1/3 of protein of induced culture (Novagen, personal communication)

Pellet post lysis: 0.1% of culture volume

Supernatant post lysis: 0.4% of culture volume

Pellet post sonication: 0.1% of culture volume

Supernatant post sonication: 0.025% of culture volume

Flowthrough from column: 0.025% of culture volume

Other lanes cannot be expressed as percentage of culture volume, expressed instead as percentage of sample

Flowthrough of wash buffer: 0.05% of sample

Eluate: 0.013% of sample
to final concentrations of 20μg/ml and the samples incubated for 30-60 minutes at 4°C. The cell slurry was then sonicated for 6 periods of 10 seconds and centrifuged at 20000g for 30 minutes at 4°C. The supernatant was passed through a 0.45μM filter to remove any remaining viscous material before being applied to the affinity purification column.

Wild type human cSHMT, the C204A mutant and the rabbit cSHMT were purified on the Novagen His Bind resin according to the manufacturer’s protocol except that the wash buffer step was replaced with a further wash with binding buffer, and the protein was eluted from the purification column with binding buffer containing 200mM imidazole instead of 1M.

All other cSHMT mutants were purified using Talon resin (Clontech) using the manufacturer’s batch/gravity flow column protocol for purification of native protein except, the sonication buffer was pH7.9. Protein was eluted from the affinity column using sonication buffer containing 100mM imidazole.

10μl of SDS sample buffer was added to 10μl of the column flowthrough after application of the cell lysate and after washing. 15μl of SDS sample buffer was added to 5μl of fractions containing protein eluted from the column with imidazole. Samples were analysed by SDS-PAGE using 20μl of sample except for the protein eluted from the column where 2-5μl was analysed.

After SHMT elution, protein was concentrated and desalted using Centriplus columns (Amicon). The buffer was changed to 50mM Tris/HCl pH7.5, 2mM Pyridoxal 5’-phosphate, 2mM serine, 10mM Dithiothreitol.

2.10.1 Biorad Assay of protein concentration

This is a commercially available protein concentration assay based on the method of Bradford, (1976). 0.8ml of water and up to 50μl of sample was placed in a cuvette and 0.2ml of Biorad dye reagent (Biorad) was added and the solution mixed thoroughly by inversion.
Two enzyme assays were performed on wild-type and mutant cSHMT proteins to assess whether the mutations had any affect on cSHMT activity. The radioactive assay was used as serine and THF are the physiological substrates of SHMT. The dl-allothreonine assay does not use THF as a substrate so any mutation affecting THF binding can be determined by differing enzyme activities exhibited in the two assays.
The OD$_{595\text{nm}}$ was measured and compared to a standard curve created using known concentrations of BSA.

2.11 Assay of SHMT activity using serine and THF

This assay was based on the method of Taylor and Weissbach (1965). SHMT was diluted to various concentrations in 80μl of 310mM Hepes pH 7.4, 2mM THF, 12mM 2-mercaptoethanol, 0.2mM PLP. The reaction mixture was incubated for 5 minutes at 37°C before the addition of 20μl 1.25mM L-[3-$^{14}$C]-serine. After 10 minutes incubation at 37°C the reaction was stopped by the sequential addition of 60μl 1M sodium acetate pH 4.5, 40μl 0.1M formaldehyde and 60μl 0.4M dimedone. The samples were boiled for 5 minutes and then placed on ice for 5 minutes. 1ml of toluene was added and the samples were shaken for 60 minutes at room temperature. After vortexing and centrifuging at 13000g for 10 minutes, 800μl of the supernatant was added to 3ml scintillant and counts per minute recorded. The results were expressed as nmoles formaldehyde formed/min/nmol SHMT and the percentage activity compared to wild type SHMT was calculated. A graph of SHMT concentration verses percentage activity of the wildtype was drawn and the amount of enzyme equating to half of the maximum activity of the wild type SHMT (EC$_{50}$) was calculated.

2.12 dl-allothreonine assay

This assay was based on the method of Schirch and Gross (1968). 950μl 50mM potassium phosphate buffer, 1mM EDTA pH 7.4, 90units alcohol dehydrogenase, 128nM NADH and 0.5mM to 50mM dl-allothreonine was placed in a 1ml cuvette and incubated at 30°C for 10 minutes. SHMT was added in 50μl to a final concentration of 1.06x$10^{-8}$M and the OD$_{340\text{nm}}$ was measured for 10 minutes. The rate of absorbance change per minute was calculated and plotted according to (Hanes, 1932) to obtain a Km for dl-allothreonine.
2.13 Bacteriophage lambda techniques

2.13.1 Preparation of plating bacteria

50ml of LB containing 0.2% maltose was inoculated with a single colony of *E.coli* strain C600 and grown overnight at 37°C with shaking at 250rpm. The cells were harvested by centrifuging at 4000g for 10 minutes at room temperature and the pellet was resuspended in 20ml 0.01M MgSO4. The OD600nm was taken and the cells diluted so that OD600nm was 2.0. Cells were stored at 4°C.

2.13.2 Plating Bacteriophage lambda

10 fold dilution’s of stock phage were made in SM. 0.1ml of each dilution was added to 0.1ml of plating bacteria and incubated at 37°C for 20 minutes. 3ml of molten (47°C) top L-agar or agarose was added to each tube, mixed and poured onto a petri dish containing 30ml hardened L-agar/agarose. Plates were allowed to stand for 5 minutes to allow the top layer to set before inverting and incubating at 37°C overnight.

2.13.3 Plaque purification

Plaques were sucked into a pasteur pipette along with the underlying L-agar/agarose and pushed out into 1ml SM containing 1ml chloroform. Phage were titrated using serial dilutions and following the protocol in section 2.13.2.

2.13.4 Rapid small scale isolation of Lambda DNA

50-100μl of 10^5 pfu/ml phage suspension was added to 0.1ml plating bacteria. The procedure for plating lambda (section 2.13.2) was followed using L-agarose. After overnight incubation, 5ml SM was added to the plates which were then gently shaken at room temperature for 1-2 hours. The SM was poured into a centrifuge tube together with the top L-agarose. The tubes were centrifuged at 7500g for 10 minutes at 4°C and the supernatant was decanted. DNase and RNase were added to the supernatant to a final concentration of
1µg/ml. and incubated at 37°C for 30 minutes. An equal volume of SM containing 20% PEG and 2M NaCl was added followed by incubation at 0°C (ice water) for 1 hour. Precipitated phage were pelleted by centrifuging for 20 minutes at 10000g, 4°C. The supernatant was removed and excess supernatant drained off. The pellet was resuspended in 0.5ml SM by vortexing and centrifuged (13000g) for 2 minutes at 4°C to remove debris. SDS was added to 0.1%, EDTA to 5mM and the solution incubated at 68°C for 15 minutes. The solution was phenol extracted, phenol/chloroform extracted, chloroform extracted and isopropanol precipitated. Following incubation at -70°C for 20 minutes and centrifuging (13000g) for 15 minutes at 4°C, the pellet was washed with 70% ethanol, dried and resuspended in 50µl TE (pH8.0).

2.14 Mapping of Pseudogene clones

Two primers ON-L and ON-R complementary to the left and right cos end sequences respectively were end labeled with digoxygenin (section 2.5.12). Partial digests of lambda clones were performed by incubating 1µg of DNA with 1-10 units of restriction enzyme at room temperature for 10 minutes. EDTA was added to a final concentration of 10mM, the reaction tubes were incubated at 68°C for 10 minutes to disrupt the cos ends, divided into 2 and phage DNA fragments separated by agarose gel electrophoresis through a 0.5 or 0.6% gel. DNA was visualised by staining by ethidium bromide prior to Southern blotting (section 2.5.10). Half of the samples were hybridised to ON-L and the other half to ON-R at 42°C for 4 hours (section 2.5.13). Sizes of DNA fragments hybridising to the probes were calculated by comparison to DNA length standards.
2.15 Analysis of the human genomic cytosolic SHMT gene

This was performed using the Human PromoterFinder kit (Clontech) with either DNA Advantage Tth polymerase mix (Clontech) or Expand™ Long template polymerase mix (Boehringer Mannheim). PCR reactions were performed according to the PromoterFinder kit protocol using the Clontech polymerase. When Expand™ long template polymerase (Boehringer Mannheim) was used 2 master solutions were made; one containing water, dNTP’s and oligonucleotides, the other containing water, polymerase, template and 10x buffer. These 2 solutions were mixed on ice prior to amplification.

The first PCR was performed as follows. 7 cycles of 94°C, 25 seconds, 72°C, 4 minutes, then 32 cycles of 94°C, 25 seconds and 67°C, 4 minutes and one cycle of 68°C for 4 minutes.

The second PCR conditions were 5 cycles of 94°C, 25 seconds and 72°C, 4 minutes followed by 22 cycles of 94°C, 25 seconds, 67°C, 4 minutes and a final cycle of 68°C for 4 minutes.

PCR of human genomic DNA was also performed using Expand™ long distance polymerase (Boehringer Mannheim). The cycles used were 94°C for 2 minutes followed by 10 cycles of 94°C for 10 seconds, 65°C for 30 seconds and 68°C for 8 minutes, then a further 15 cycles of the above but extending the elongation element by an extra 20 seconds each cycle.

One cycle of 68°C for 7 minutes completed the amplification.
CHAPTER 3

CONSTRUCTION OF RABBIT AND HUMAN CYTOSOLIC SHMT EXPRESSION VECTORS AND EXPRESSION AND PURIFICATION OF PROTEIN
3.1 Summary

Please refer to section 1.9

The precise tertiary and quaternary structures of SHMT are required in order to design inhibitors of SHMT activity on a rational basis. To date no X-ray crystal structure has been published of SHMT protein isolated from any source (section 1.8) and, therefore, the structure and reaction mechanism of SHMT protein is subject to speculation. Some information as to critical residues for activity can be provided by modification of residues by chemicals and use of reaction intermediates but this method does not reveal the environment in which these essential residues lie. Highly pure SHMT protein is required in order for an X-ray crystal structure to be determined. In addition, pure SHMT protein can be used to assay potential inhibitory compounds designed using any potential X-ray structure or to test existing compounds. The orfs encoding human and rabbit cSHMT were subcloned into pET14b and were expressed in an E.coli in order to produce large quantities of protein for crystallography studies and for inhibitor screening by collaborators at the Institute of Cancer Research.

cDNA's containing the full open reading frames for rabbit and human cytosolic SHMT's had previously been isolated (Byrne, 1992). The cDNA's had been subcloned into pUC18 at the EcoRI site and designated pUS1203 and pUS1217 respectively. To enable the rapid purification of expressed SHMT protein from E.coli, the expression vector pET14b (Novagen) was chosen (section 1.9) Modifying the translational start codons to introduce an NdeI site (CATATG) and incorporating a second NdeI site downstream of the stop codons allowed the human and rabbit cSHMT orfs to be cloned into pET14b in frame with the histidine tag encoding sequence. PCR amplification was chosen for the modification, with NdeI sites being incorporated into the oligonucleotide primers of the PCR reaction. Expression was followed over a three hour time course after induction with IPTG and it was
determined that both human and rabbit cSHMT proteins are expressed in a soluble form enabling rapid purification of protein.

3.2 Results

3.2.1 Subcloning of rabbit cSHMT into pET14b

The cDNA coding for rabbit cSHMT (pUS1203) was amplified by PCR using primers C and A (see section 2.6) to produce NdeI restriction sites overlapping the start codon, and downstream of the stop codon. The PCR product was purified from an agarose gel and ligated into the vector pCRU™ (Invitrogen). After transformation of the ligated plasmid into INVαF’ cells, 40 white and 14 blue colonies resulted. Of 5 white colonies analysed, 4 contained the PCR product. One of the plasmids containing the PCR product (pUS1246) was digested with NdeI and the 1.5kb band removed and purified. The 1.5kb fragment containing the rabbit cSHMT orf was then ligated into NdeI digested pET14b. DH5α cells were transformed to ampicillin resistance with the ligation mixture and 21 colonies were produced. Of 6 colonies analysed, 5 contained plasmids with the insert and one contained plasmid alone. Four plasmids contained the inserted rabbit cSHMT orf in the right orientation for expression (figure 3.1).

3.2.2 Expression of rabbit cytosolic SHMT

One of the clones (pUS1250) containing the rabbit cSHMT orf in the correct orientation was used in subsequent experiments. The plasmid was extracted from DH5α cells and used to transform the expression strains BL21(DE3) and BL21(DE3)pLysS to ampicillin resistance. After checking that the ampicillin resistant colonies contained the plasmid, 1 colony of BL21(DE3) containing the plasmid was amplified and protein expression was induced by the addition of IPTG to the medium. Samples were removed at 30 minute intervals for analysis of protein expression by SDS-PAGE. A strain containing the gene for
Figure 3.1 Vectors used for the expression of rabbit and human cSHMT proteins. pET 14b (Novagen) is the expression vector. pUS 1250 contains the PCR amplified rabbit cSHMT in pET 14b. pUS 1251 contains the amplified human cSHMT orf in pET 14b.
Figure 3.2 Coomassie brilliant blue stained SDS-PAGE gel showing a time course of rabbit cSHMT protein expression in *E.coli*. Cells containing pUS1250 were grown and protein expression was induced in log phase cells by the addition of IPTG. Samples were taken every 30 minutes after induction for 3 hours.

Lane M  MW markers
Lane 1  Cells before addition of IPTG
Lanes 2 to 7  Samples taken from 30 minutes to 3 hours post induction

Please refer to p67a for protein loadings
Figure 3.3 Comparison of soluble/insoluble fractions of rabbit cSHMT protein expressed in *E.coli*. Cells containing pUS1250 were induced with IPTG and harvested 2 hours post induction. Soluble and insoluble fractions of lysed cells were separated by centrifugation and analysed by SDS-PAGE.

<table>
<thead>
<tr>
<th>Lanes M</th>
<th>MW markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lane 1</td>
<td>Overnight culture</td>
</tr>
<tr>
<td>Lane 2</td>
<td>Before induction with IPTG</td>
</tr>
<tr>
<td>Lane 3</td>
<td>Whole cells 2 hours post induction</td>
</tr>
<tr>
<td>Lane 4</td>
<td>Pellet post centrifugation of lysed cells</td>
</tr>
<tr>
<td>Lane 5</td>
<td>Supernatant post centrifugation of lysed cells</td>
</tr>
</tbody>
</table>

Please refer to p67a for protein loadings
\( \beta \)-galactosidase in pET14b was used as a control. The induced SHMT from pUS1250 had a molecular weight of 53KD, the correct weight for SHMT (figure 3.2). SHMT was seen 30 minutes post induction and the yield of protein increased with time to approximately 15% total cell protein 3 hours post induction equivalent to approximately 25\( \mu \)g/ml of culture.

To determine whether rabbit cSHMT was soluble in the cytoplasm or was sequestered into inclusion bodies, expression was induced in BL21(DE3)pLysS cells growing at 37°C. After 2 hours induction the cells in the culture were collected, resuspended in 50mM Tris/HCl pH 8.0, 2mM EDTA and treated with lysozyme and Triton X-100 to lyse the cells. The soluble and insoluble components were separated by centrifugation. SDS-PAGE showed that the 53KD band was present both in the pellet and supernatant showing that the protein was at least partly soluble (figure 3.3).

### 3.2.3 Subcloning of human cytosolic SHMT into pET14b

The human cSHMT was cloned into the pET14b vector in a manner identical to that used for the rabbit. PCR was performed on pUS1217 using oligonucleotides kjc5 and kjc6. The PCR product was ligated into pCRII\(^\text{TM}\) and 26 colonies were produced by transforming INV\( \alpha \)F\(^\text{r}\) with the ligation reaction. One plasmid containing the human orf PCR product (pUS1247) was digested with \(NdeI\) to release the 1.5kb PCR product fragment which was purified from an agarose gel slice and ligated to \(NdeI\) digested pET14b. After transforming DH5\( \alpha \) cells with the ligation reaction, 6 colonies were analysed and 2 contained plasmids with the human cSHMT orf in the correct orientation for expression. One of these plasmids was designated pUS1251 (figure 3.1).

### 3.2.4 Expression of human cytosolic SHMT

pUS1251 was used to transform BL21(DE3)pLysS to ampicillin resistance. BL21(DE3)pLysS containing pUS1251 was grown in culture at 37°C and induction of human cytosolic
SHMT expression was initiated by addition of IPTG to the medium. Samples were removed for analysis of protein expression by SDS-PAGE at 30 minute intervals (figure 3.4). SHMT protein was present 30 minutes post induction. The gels show appreciably easily detectable expression of cSHMT by 3 hours post induction.

To determine whether human cSHMT was soluble when expressed at 37°C, an experiment identical to that used to determine the solubility of rabbit cSHMT was performed. SDS-PAGE analysis identified a 53KD band in both the supernatant and the pellet fractions of induced and lysed cells. From the bands seen on SDS-PAGE gels it was demonstrated that expressed human cSHMT protein is at least partly soluble (figure 3.5).

### 3.2.5 Sequencing of PCR modified SHMT sequences

Amplification of DNA fragments by PCR can lead to mutations in the fragment as the polymerase may incorporate the wrong nucleotide. This error is then carried through future PCR cycles. The modified SHMT orfs were to be used to generate proteins to be used to solve the X-ray crystal structure of SHMT and, therefore, it was vital that there are no errors in the coding sequence that may affect protein structure. Boehringer Mannheim Taq polymerase has an error rate of 2.5x10⁻⁵ (Boehringer Mannheim, personal communication), although PCR amplifications performed by colleagues suggested a higher error rate. The PCR products of the rabbit and human cSHMT orfs were, therefore, sequenced to determine whether any mutations had occurred.

Sequencing of the human cSHMT orf PCR product indicated three mutations. Two of these mutations at nucleotide (nt) 513 (C to G) and nt 1638 (G deletion) (appendix 1) did not affect the coding sequence as codon 112 remained coding for glycine and the second mutation was downstream of the stop codon. However, the third mutation was at nt 721 changing an
Figure 3.4 Coomassie brilliant blue stained SDS-PAGE gel showing a time course of human cSHMT protein expression in *E.coli*. Cells containing pUS1251 were grown and protein expression was induced in log phase cells by the addition of IPTG. Samples were taken every 30 minutes post induction for 3 hours.

Lane M  MW markers
Lane 1  Cells before addition of IPTG
Lanes 2 to 7  Samples taken from 60 minutes to 3 hours post induction

Please refer to p67a for protein loadings
Figure 3.5 Comparison of soluble/insoluble fractions of human cSHMT protein expressed in E.coli. Cells containing pUS1251 were induced with IPTG and harvested 2 hours post induction. Soluble and insoluble fractions of lysed cells were separated by centrifugation and analysed by SDS-PAGE.

<table>
<thead>
<tr>
<th>Lanes M</th>
<th>MW markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lane 1</td>
<td>Overnight culture</td>
</tr>
<tr>
<td>Lane 2</td>
<td>Before induction with IPTG</td>
</tr>
<tr>
<td>Lane 3</td>
<td>Whole cells 2 hours post induction</td>
</tr>
<tr>
<td>Lane 4</td>
<td>Pellet post centrifugation of lysed cells</td>
</tr>
<tr>
<td>Lane 5</td>
<td>Supernatant post centrifugation of lysed cells (soluble fraction)</td>
</tr>
</tbody>
</table>

Please refer to p67a for protein loadings
adenine to a guanine. This mutation changed codon 182 from asparagine to aspartic acid. To check that the error was genuine and not a sequencing error in the original pUS1217, sequencing of pUS1217 was repeated in this region. The codon was confirmed as AAC (asparagine) in pUS1217.

The sequence of the modified rabbit SHMT orf PCR product is incomplete. However, two mutations have been found at nt 1484 and nt 1581 (pUS1203 sequence) changing adenine to cytosine and guanine to cytosine respectively. Both of these mutations were silent and did not affect the amino acid sequence of the rabbit cSHMT protein.

3.3 Discussion

The open reading frames encoding human and rabbit cSHMT have been subcloned into the prokaryotic expression vector pET14b. Both plasmids pUS1250 and pUS1251 direct the production of cSHMT protein when induced with IPTG. Microgram amounts of cSHMT protein are produced per millilitre of culture. At least 50% of protein expressed at 37°C is soluble and, therefore, assumed to be in a native state. The solubility of the proteins enables them to be rapidly purified compared to the process involving denaturation of insoluble protein aggregated in inclusion bodies, and refolding of the protein by removing the denaturant (see Fischer et al., 1993 for review). Proteins produced in a soluble state are more likely to have the native structure, whereas refolding denatured proteins can lead to different structures being formed. Consequently activity assays of proteins produced from the soluble fraction are more reliable than those from refolded proteins. It also follows that the X-ray crystal structure would be more accurate from a natively folded structure than from a refolded one.

From the results presented here, a method for the large scale expression of cSHMT protein has been devised. The strain used for expression was BL21(DE3)pLysS as this cell
line was more easy to lyse than the strain without the pLysS plasmid. After the cells reach log phase and protein production induced with IPTG, cells are grown for a further 2 hours and then harvested (S. Renwick, personal communication). This time course gives the maximum amount of expression without giving the chance for excessive degradation of the expressed protein. There is evidence that if expressed protein reaches 30% of total cell protein, protein synthesis and cell growth stops, ribosomal RNA is degraded and the cell dies (Kurland and Dong, 1996).

PCR amplification of the human cSHMT orf generated an A to G mutation, compared to pUS1217, changing codon 182 to aspartic acid. Sequencing of the original pUS1217 clone determined that nt 721 was an adenine residue and, therefore, codon 182 encoded an asparagine residue. This data was also confirmed by the publication of the coding sequence of human cSHMT cDNA (Xu, 1992; Garrow et al., 1993). However, an alignment of SHMT protein sequences from different organisms using the PILEUP and PRETTY programs (Genetics Computer Group) (appendix 2) showed that in some other SHMT proteins there is an aspartic acid residue at this position instead of an asparagine so the mutation generated was conservative. There could be some genetic variation within the population for the cSHMT gene leading to sequences different to those isolated to date. However, all of the human cSHMT protein primary structures translated from the cDNAs isolated so far show 100% identity. Therefore, the mutation in the amplified human cSHMT cDNA may affect the activity and X-ray crystal structure of the protein.

Attempts to sequence the PCR amplified product from rabbit cSHMT template produced relatively poor sequence. It is possible that the primers, synthesised in 1990-92 have degraded and new ones need to be synthesised to allow the accurate determination of the rabbit cSHMT orf PCR product. Out of the good sequence obtained, two silent mutations
were found in the space of less than 100bp. The sequence of the amplified rabbit cSHMT cDNA needs to be examined further to ascertain whether there are any mutations affecting the primary structure of the protein.

The error rate of Taq polymerase can be determined by the method in Hayes, (1965) as used by Saiki et al., (1988). The error rate is \( (m) = 2(f/d) \) where \( f \) is the observed error frequency in the PCR product and \( d \) is the number of doublings. The error rate for the human cSHMT orf PCR product can be calculated as \( 2((3/1490)/25) \) which works out at \( 1.6 \times 10^{-4} \). However, the number of effective PCR cycles is only 60% of the number of cycles when amplifying a product of less than 3kb (Boehringer Mannheim, personal communication). Therefore, this would lead to an error rate of \( 2.6 \times 10^{-4} \). Both error rates are at least 10 fold greater than those advertised by Boehringer Mannheim.

It can be concluded from the error rates shown in the PCR products discussed here that if it is critical for a PCR product to have no errors then an alternative DNA polymerase to Taq should be used. There are now several proofreading thermostable DNA polymerases available such as Pfu polymerase from Stratagene and Pwo polymerase from Boehringer Mannheim which have error rates of \( 1.6 \times 10^{-6} \) and \( 3 \times 10^{-6} \) respectively.

Once the PCR induced error had been corrected by mutagenesis (chapter 4), milligram quantities of wild-type human cSHMT protein were subsequently produced and purified on a column containing nickel ions (section 1.9) (S. Renwick, personal communication). Purification of cSHMT protein can be achieved in one day. Twelve milligrams of human cSHMT can be purified from 1 litre of culture. However, purified enzyme rapidly precipitates out of solution if left solely in 50mM Tris/HCl pH7.5. This problem was solved by adding pyridoxal phosphate, serine and dithiothreitol (DTT) to the purified enzyme (S. Renwick, personal communication).
Figure 3.6 Crystals of human cSHMT. Human cSHMT cDNA was subcloned into pET14b and expressed in *E.coli*. Purified cSHMT protein was crystallised at the Institute of Cancer Research (Courtesy of S.Renwick)
Since performing this work, methods for the bacterial expression and purification of sheep and human cSHMT have been published (Jagath-Reddy et al., 1995; Kruschwitz et al., 1995). Both of the cDNA’s were put under the control of the T7 promoter and produced 10mg/l and 11.25mg/l protein respectively. However, neither system had an affinity purification method and purification of cSHMT protein was laborious involving two purification columns and ammonium sulphate precipitations. The expression system described here produces a higher quantity of purified cSHMT more quickly.

Crystallography studies have been initiated at the Institute of Cancer Research (S. Renwick, personal communication) and human cSHMT crystals grown (Figure 3.6). The first crystals produced were sensitive to X-rays and disintegrated. Further studies are in progress and more stable human cSHMT crystals for X-ray diffraction have been obtained. Determination of an X-ray crystal structure is in progress.

Human cSHMT has also used to assay a panel of antifolate compounds synthesised at the Institute of Cancer Research for inhibition of SHMT activity (S. Renwick, personal communication). The highest inhibition exhibited by one of these compounds was an IC\textsubscript{50} of 1.25±0.2mM. For an inhibitor to be potentially useful an IC\textsubscript{50} in the nM to low μM range is required. Therefore, none of the compounds tested would be regarded as useful inhibitors of SHMT. However, only a small range of compounds was tested. There are two potential routes to follow with inhibitor testing, wait until a tertiary and quaternary SHMT structure is available to allow the design of inhibitors based on the SHMT structure or continue testing banks of compounds until one with a lower IC\textsubscript{50} is found and then modifying that particular compound to improve its inhibitory properties.
CHAPTER 4

SITE DIRECTED MUTAGENESIS OF HUMAN CYTOSOLIC SHMT
4.1 Summary

Previous attempts to determine which amino acids are required for the catalytic activity of SHMT have involved the use of chemical modification and site directed mutagenesis (section 1.2.2). Chemical modification adds bulky groups and may inactivate the enzyme by steric hindrance preventing the binding of substrates or cofactors rather than altering an amino acid side chain critical for activity. For example, modification of a cysteine residue with iodoacetate (C204) produces an inactive cSHMT protein due to the fact that the added side chain prevents PLP binding to the critical lysine residue (K257) (Schirch et al., 1980). Therefore, some of the residues implicated in the catalytic reaction of SHMT may not in fact be involved.

The mutations chosen to produce variant human cSHMTs were based on the residues present in the literature which had been inactivated by chemical modification, in order to assess whether the modified amino acids were important for cSHMT activity. Reduction of the size of the cSHMT protein at the amino and carboxyl termini would aid in the identification of the regions essential for enzyme activity. Previous truncation studies have been performed using proteases (Schirch et al., 1984) but this technique is limited to truncation at sites recognised by the protease used. Precise truncation of the carboxyl terminus can be achieved by introducing early stop codons into the coding sequence by site directed mutagenesis. Generating a mutant cSHMT protein without the critical pyridoxal phosphate binding lysine (K257) would inactivate the enzyme producing a negative control by which activities of the mutant cSHMT proteins could be gauged. Site directed mutagenesis was performed to retain the charge and side chain structure or to remove the selected amino acid side chain by changing to alanine.
The mutations performed converted; amino acid 68 from cysteine to alanine (C68A) (Usha et al., 1994), C204A (Schirch et al., 1980), W15A, W15F, W15Y, W111A, W111F (Vijayalaksmi et al., 1989), K257A (Schirch et al., 1993), C380stop, F424stop and K446stop.

Once pure SHMT protein had been obtained for each mutant, two assays were chosen to assay the enzyme activity. One assay was a spectrophotometric method involving the conversion of dl-allothreonine to glycine and ethanal. Ethanal production is coupled to a second reaction where it is converted by alcohol dehydrogenase to ethanol, with the concomitant conversion of NADH to NAD⁺. The oxidation of NADH is followed by the decrease in absorption at 340nm (Schirch and Gross, 1968). The second assay measures the rate of transfer of the one carbon group from 3-[¹⁴C]-L-serine to THF (Taylor and Weissbach, 1965). The second assay was performed at the Institute of Cancer Research by S.Renwick.

4.2 Results

4.2.1 Insertion of human cSHMT orf into the pALTER-1 vector

The vector pCRU™ containing the human cSHMT orf PCR product (pUS1247) was digested with EcoRI, the 1.5kb fragment containing human cSHMT cDNA was purified from a 1% agarose gel and ligated into EcoRI digested and phosphatased pALTER-1®. Two clones containing pALTER-1® with the desired insert were obtained. Digestion with KpnI and BamHI indicated that the inserts were in opposite orientations. The orientation used for site directed mutagenesis was the one where the cSHMT orf ran in the same direction as the tetracycline gene (figure 4.1). Single stranded DNA was produced for site directed mutagenesis.

4.2.2 Site directed mutagenesis of human cSHMT orf to revert amino acid 182 to asparagine

The cSHMT orf PCR product used in the generation of the site directed mutagenesis vector contained the N182D mutation. In order to revert the orf to wild-type, site directed
mutagenesis was performed on the single stranded DNA produced above using the oligonucleotide D182N (section 2.4), the ampicillin repair and tetracycline knockout oligonucleotides to reverse the PCR generated mutation. ES1301 mutS cells were transformed to ampicillin resistance and grown in culture overnight. Plasmid DNA was extracted from the culture by alkaline lysis and was used to transform JM109 cells to ampicillin resistance. The transformation resulted in over 300 ampicillin resistant colonies. Fifty colonies were subcultured onto tetracycline plates and forty eight of these colonies were tetracycline sensitive indicating that the antibiotic knockout and repair oligonucleotides had been incorporated. Plasmid DNA was isolated from two colonies (1 and 10) and sequenced. Both plasmids were found to contain the cSHMT orf with the D182N mutation. The entire cSHMT orf was sequenced in the plasmid DNA isolated from colony 10 and found not to contain any spurious mutations. The pALTER-1® containing the modified cSHMT orf was renamed pUS1252 (figure 4.1) and used in subsequent experiments.

4.2.3 Synthesis of new vectors to aid site directed mutagenesis

Every mutant orf produced should be sequenced after mutagenesis to ensure that only the desired mutation has been incorporated. The human cSHMT PCR product inserted into pALTER-1® is just under 1.5kb and, therefore, if the entire sequence was used for mutagenesis, each mutation would require several sequencing reactions to confirm that only the required mutation had been incorporated. In the human cSHMT orf there are unique Kpnl and BamHI sites bordering a 0.6kb fragment between nucleotides 404-1005 (appendix 1), containing codons Y77 to V276. This 0.6kb fragment codes for a number of potentially critical amino acids which could be altered by site directed mutagenesis. Subcloning and mutating this fragment would mean that only two sequencing reactions would be required for
Figure 4.1 Formation of the vector for site directed mutagenesis. The human cSHMT orf (red) was subcloned into pALTER-1 using the EcoRI sites highlighted in red and then site directed mutagenesis was performed to revert the error generated by PCR to form pUS1252. Highlighted are the ampicillin resistance gene in blue and the tetracycline resistance gene in green. Partial shading of the antibiotic resistance genes indicates inactivated genes.

$A_{Pr}$ Ampicillin resistance

$A_{Ps}$ Ampicillin sensitivity

$T_{Cr}$ Tetracycline resistance

$T_{Cs}$ tetracycline sensitivity

$F1$ ORI $F1$ origin of replication

ORI origin of replication

PMB1 origin of replication derived from PMB1

LAC promoter for the $\beta$-galactosidase $\alpha$ peptide

SP6 P Sp6 promoter

T7 P T7 promoter

mcs multiple cloning site
After 1 round of mutagenesis to revert sequence to wild type
each mutant. This 0.6kb KpnI/BamHI fragment was subcloned from pUS1252 into KpnI/BamHI digested pALTER-1®.

pUS1252 and pALTER-1® were digested with KpnI and BamHI, and the linear pALTER-1® vector and 0.6kb bands were purified from a 1% agarose gel and ligated together. Three white colonies on L-agar containing X-gal and IPTG were produced when JM109 cells were transformed to ampicillin resistance with ligation mix. Two of these colonies (2 and 3) contained plasmids with a 0.6kb KpnI/BamHI fragment. The recombinant vector isolated from colony 2 was renamed pUS1255 (figure 4.2).

Following mutagenesis of the 0.6kb KpnI/BamHI fragment the whole human cSHMT orf needs to be reconstructed to enable expression and purification of mutant cSHMT protein. To achieve this, pUS1248, a modified pUC18 vector was produced to replace the HindIII site of the pUC18 multiple cloning site with EcoRI (figure 4.3). To create pUS1248, pUC18 was digested with HindIII and phosphatased. An oligonucleotide kjc7 (section 2.4) was synthesised which formed dimers to produce a linker with a site that would ligate to a HindIII generated site but could not be redigested. Kinased kjc7 was ligated to the digested vector. Transformation of DH5α cells by the ligation mixture resulted in greater than 300 colonies, with EcoRI, Ndel, BamHI and HindIII were used to digest plasmid DNA, extracted from four cultures grown overnight. All four plasmids produced the same restriction pattern (figure 4.4). HindIII did not digest the plasmid DNA showing that the linker had been incorporated, BamHI and Ndel linearised the plasmids but EcoRI digested plasmid ran more quickly on the agarose gel than BamHI and NdeI digested plasmid showing that there was more than one copy of the linker present in each plasmid. One plasmid (1) was designated pUS1248.

To remove the multiple cloning site, pUS1248 was digested with EcoRI and purified from an agarose gel. The human cSHMT orf also purified from an agarose gel after EcoRI
Figure 4.2 Construction of pUS1255 for site directed mutagenesis of the 0.6kb KpnI/BamHI fragment of the cSHMT cDNA (red). Indicated in red are the restriction sites used. Indicated in blue and green are the ampicillin and tetracycline resistance genes respectively. Hashed shading indicates inactivated antibiotic resistance genes.

- **AP<sub>r</sub>** Ampicillin resistance
- **AP<sub>s</sub>** Ampicillin sensitivity
- **TC<sub>r</sub>** Tetracycline resistance
- **TC<sub>s</sub>** Tetracycline sensitivity
- **F1 ORI** F1 origin of replication
- **ORI** origin of replication
- **PMB1** origin of replication derived from PMB1
- **LAC** promoter for the β-galactosidase α peptide
- **SP6 P** Sp6 promoter
- **T7 P** T7 promoter
- **mcs** multiple cloning site
Figure 4.3 Schematic of pUS1248. This vector was constructed as an intermediate in the removal of the multiple cloning site (mcs) from pUC18. The sequence of the multiple cloning site of pUS1248 is shown with the added linker DNA in bold type. There are potentially many copies of the linker present at this site in the vector. Underlined is the HindIII site removed during the process of the linker addition.

AP<sub>r</sub>  Ampicillin resistance
AP<sub>s</sub>  Ampicillin sensitivity
p(BLA)  promoter for β-lactamase gene
Figure 4.4  Agarose gel of restriction digests of four clones obtained when kjc7 was ligated into HindIII digested pUC18 to replace the HindIII site with an EcoRI site to form pUS1248.

Lanes 1-4  *EcoRI* digestion  
Lane 5-8  *NdeI* digestion  
Lane 9-12  *HindIII* digestion  
Lane 13-16  *BamHI* digestion  
Lane M  *λHindIII* markers  
Lanes 1,5,9,13  Plasmid 1  
Lanes 2,6,10,14  Plasmid 2  
Lanes 3,7,11,15  Plasmid 3  
Lanes 4,8,12,16  Plasmid 4

Approximately 200ng of DNA was loaded per track
digestion of (pUS1247). The vector and SHMT were ligated and transformation of DH5α to ampicillin resistance resulted in more than 300 colonies. Plasmid DNA was isolated from one clone and digested with EcoRI, KpnI and BamHI. EcoRI digested plasmid contained a 1.5kb band. KpnI and BamHI digests linearised the plasmid. Therefore, the human cSHMT had been cloned into a vector which only contained KpnI or BamHI restriction sites in the human cSHMT orf. This new plasmid was designated (pUS1249) (figure 4.5).

4.2.4 Mutagenesis of human cSHMT cDNA

The Altered Sites protocol was followed with the two modifications; the amount of mutagenic oligonucleotide used was increased to 10pmol from 1.25pmol to ensure a higher rate of oligonucleotide binding and the oligonucleotides were annealed to the single stranded plasmid by heating to 95-100°C to ensure all secondary structure was removed before slowly cooling to 45°C and then to more rapidly 20°C. These alterations to the protocol were recommended by Piechocki and Hines, (1994).

Test reactions using pALTER-1® showed that the recommended cotransformation procedure did not work effectively. This procedure involves transforming ES1301 mutS cells with both DNA from the mutagenesis reaction and helper phage DNA (e.g. R408 phage DNA). Cotransformation causes the mutated vector to be packaged as a bacteriophage which is then purified and used to infect an overnight culture of JM109 cells. The probable reason for this method not working was that a transformation efficiency of at least 1x10⁷ cfu/µg of DNA is required and “home made” ES1301 mutS competent cells produced only 2x10⁵ cfu/µg DNA. It is in fact difficult to get a higher transformation efficiency in ES1301 mutS cells (Promega). Therefore, the alternative protocol, where ES1301 mutS cells were transformed with DNA produced in the mutagenesis reaction and grown overnight in broth, was followed. ES1301 mutS cells grew slowly and were harvested after incubating for 16-20
Figure 4.5 Diagram to show the formation of pUS1249 from pUS1248 and human cSHMT cDNA. pUS1249 did not contain KpnI or BamHI sites except within the cSHMT orf (red) allowing the exchange of the original 0.6kb fragment with mutated 0.6kb fragments to produce variant cSHMT orfs. Highlighted in red are the EcoRI sites used for subcloning and in green the KpnI and BamHI sites lost in formation of pUS1249.

\begin{tabular}{ll}
AP$_r$ & Ampicillin resistance \\
ORI & origin of replication \\
p(BLA) & promoter for the β-lactamase gene \\
p(LAC) & promoter for the β-galactosidase α peptide \\
PRM & primer \\
\end{tabular}
hours with shaking at 37°C. Plasmid DNA was then extracted by alkaline lysis and used to transform competent JM109. This method had the disadvantage that as ES1301 mutS is mismatch repair deficient, there could be some spontaneous mutations generated. However, these would be screened out by sequencing of the mutated cSHMT orfs after mutation. The cotransformation protocol was later removed from the Altered Sites protocol (Promega) and a revised transformation protocol was developed (Promega). Commercially produced ES1301 mutS competent cells were also developed (Promega) and were used for subsequent transformations.

Using the Altered Sites kit protocol with the above modifications to the methodology on single stranded pUS1252 or pUS1255, mutants W15A, W15F, W15Y, W111A, W111F, C204A, C380stop, F424stop and K446stop were achieved (table 4.1 provides a summary of the mutagenesis results). Initially site directed mutagenesis with the W111A oligonucleotide yielded 6 colonies, 5 of these were wild type, the sixth had the desired mutation, but a nucleotide 3 bases downstream had been deleted. Restriction digest of plasmids isolated from colonies produced by transformation with DNA from the W15A mutagenesis reaction did not contain mutated pUS1252. This was curious as the reaction was performed at the same time as W15F and W15Y mutagenic reactions using the same plasmid preparation. W15F and W15Y were checked and found to contain mutated pUS1252. W111A and W15A mutagenesis reactions were repeated and resulted in mutations being obtained.

Two mutants, namely C68A and K257A had not been isolated after several mutagenesis attempts. In the case of K257A, the repair and knockout oligonucleotides had bound to single stranded pUS1255 indicated by ampicillin resistant and tetracycline sensitive colonies. However, sequences of the cSHMT orf in plasmids extracted from these colonies showed that the K257A mutagenic oligonucleotide had not bound. When using C68A as the mutagenic
<table>
<thead>
<tr>
<th>Mutation</th>
<th>Plasmid used for mutagenesis</th>
<th>Number of colonies sequenced</th>
<th>Number of mutants isolated</th>
</tr>
</thead>
<tbody>
<tr>
<td>W15A</td>
<td>pUS1252</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>W15F</td>
<td>pUS1252</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>W15Y</td>
<td>pUS1252</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C68A</td>
<td>pUS1252</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>W111A</td>
<td>pUS1255</td>
<td>10</td>
<td>4*</td>
</tr>
<tr>
<td>W111F</td>
<td>pUS1255</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C204A</td>
<td>pUS1255</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>K257A</td>
<td>pUS1255</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>C380stop</td>
<td>pUS1252</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>F424stop</td>
<td>pUS1252</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>K446stop</td>
<td>pUS1252</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

* One of the mutants contained a deletion in the sequence leading to a frame shift
† Mutants subsequently produced by L.Barratt.

Table 4.1 Mutations of human cSHMT generated. Plasmid templates pUS1252 or pUS1255 were used as templates in site directed mutagenesis reactions. After mutagenesis and confirmation that the antibiotic selection of the plasmids had been reversed by the mutagenesis reaction, plasmids were isolated and sequenced to determine whether the mutagenic oligonucleotide had been incorporated into the cSHMT sequence.
oligonucleotide ES1301 mutS cells did not grow when transformed with DNA from the mutagenesis reaction.

It was though that in the case of K257A the oligonucleotide was not properly phosphorylated. However, for the oligonucleotide C68A, it was possible there was some contaminant present in the oligonucleotide preparation which inhibited the synthesis of the second strand of the plasmid as all the other possible explanations were discounted. The presence of contamination in the K257A oligonucleotide may also have been a reason for the unsuccessful phosphorylation reaction.

Primers C68A and K257A were precipitated to remove any contaminants in the solution, rephosphorylated and the mutagenesis reaction undertaken by a final year undergraduate student. Mutations were generated at the first attempt. However, sequencing of the cSHMT orf mutants produced revealed extra mutations with both C68A and K257A primers. The C68A mutated cSHMT orf had a deletion of nt 381 and the K257A mutated cSHMT orf had a mutation of nt 945 from C to G changing the codon 256 from encoding histidine to glutamine. The experiment was repeated by the author with the same results.

The extra mutations generated in C68A and K257A cSHMT orfs were within the area where the mutagenic oligonucleotide annealed to the single stranded plasmid template and only plasmids containing the required mutation also had the additional mutations. Therefore, it can be assumed that the mutagenic oligonucleotides were responsible for these extra mutations and it follows that the sequence of these oligonucleotides was not that originally intended. The C68A mutation generated was not proceeded with further as the extra deletion introduces a frame shift and any protein produced would not have any homology to cSHMT downstream of amino acid 68. However, the extra mutation in the K257A cSHMT orf retained the correct open reading frame and as the critical lysine codon had been changed,
protein derived from this orf would still be able to act as a negative control. The cSHMT orf generated with the K257A oligonucleotide was renamed H256QK257A.

4.2.5 Reformation of whole human cSHMT orf for mutants generated in pUS1255

The 0.6kb cSHMT sequences in pUS1255, containing mutations generating W111A, W111F, C204A and H256QK257A had to be subcloned into the KpnI/BamHI sites of pUS1249 in order to regenerate the whole human cSHMT orfs. 1μg of pUS1249 was digested with KpnI and BamHI sequentially and the larger 3550bp fragment containing the vector and part of the cSHMT was purified from an agarose gel by either spinning the gel slice through filter paper or by Geneclean (BIO101). The vector was not treated with alkaline phosphatase as the ends were not compatible and religation to produce a functional vector in the absence of insert was not possible. For all the mutations, approximately 1μg of mutated pUS1255 was digested sequentially with KpnI and BamHI and the digested DNA electrophoresed through 1% agarose gels. The 0.6kb band was excised and DNA was purified from the agarose by the following methods; C204A, spun through filter paper, W111A, W111F by Geneclean (Bio 101) and H256QK257A using PCR Pure (Clontech) (section 2.5.3). The 0.6kb KpnI/BamHI fragment was then ligated into the KpnI/BamHI digested and purified pUS1249 vector. Controls of involving ligation of the vector alone were performed to assess any uncut or partially cut vector carried through. Ligation reactions were used to transform DH5α to ampicillin resistance. Control ligations yielded only 1 or 2 colonies. Table 4.2 gives a summary of the transformation results. Although purification of the KpnI/BamHI vector form an agarose gel had been performed to prevent any unmutated 0.6kb KpnI/BamHI from being carried over into the ligation, sequencing was performed to confirm that the cSHMT fragments ligated into pUS1249 were mutated.
Table 4.2 Results of the ligations of 0.6kb KpnI/BamHI mutant fragments into pUS1249.

pUS1255 containing the mutated 0.6kb KpnI/BamHI was digested with KpnI and BamHI to release the 0.6kb fragment which was subcloned into pUS1249 between the KpnI and BamHI sites. After transforming E.coli cells to ampicillin resistance with the ligation reaction, plasmids from the colonies were extracted by alkaline lysis and sequenced to ensure that the fragment inserted into pUS1249 contained the mutation.
4.2.6 Subcloning of the wild-type cSHMT orf into pET14b for expression in E. coli

pUS1252 was digested with NdeI to release the 1.5kb human cSHMT orf and was ligated with NdeI digested and dephosphorylated pET14b. DNA from the ligation reaction was used to transform DH5α competent cells to ampicillin resistance, and twelve colonies were obtained. Of six plasmid DNAs extracted, five consisted of pET14b with the cSHMT orf insert, two of which were in the correct orientation with respect to the T7 promoter and translation initiation signals required for expression (figure 4.6). The pET14b vector containing the cSHMT cDNA was renamed pUS1253.

4.2.7 Subcloning of mutant human cSHMT orfs into pET14b

pET14b was prepared by digesting 3μg with NdeI, followed by heat inactivation of the enzyme, and dephosphorylation. The digested pET14b was electrophoresed through an agarose gel, the linear vector DNA was cut out under UV light and the gel slice spun through filter paper.

The C204A mutation in pUS1249 was the first mutant cSHMT to be transferred into pET14b. 1μg of pUS1249 containing the C204A mutation was digested with NdeI and after electrophoresis through an agarose gel, the 1.5kb band containing the C204A cSHMT orf was removed and the DNA purified by spinning the gel slice through filter paper. 31 colonies were obtained after ligation into pET14b and transformation of DH5α competent cells to ampicillin resistance. Of 10 plasmids extracted, 9 had 1.5kb inserts when digested with NdeI and 3 of these plasmids with inserts produced a 0.8kb band when digested with BamHI indicating that the cSHMT orf was in the correct orientation relative to the T7 transcription and translation signals. One of these plasmids was designated pUS1254.
Figure 4.6 Diagram to show the formation of expression vector containing the mutant cSHMT orfs. The mutated orf (red) was subcloned into pET14b from either pUS1252 or pUS1249 using the NdeI sites highlighted in red. The ampicillin gene is highlighted in blue. The tetracycline gene is highlighted in green and shaded to indicate inactivity.

- **AP<sub>r</sub>** Ampicillin resistance
- **TC<sub>s</sub>** tetracycline sensitivity
- **F1 ORI** F1 origin of replication
- **ORI** origin of replication
- **PMB1** origin of replication derived from PMB1
- **p(BLA)** promoter for the β-lactamase gene
- **p(LAC)** promoter for the β-galactosidase α peptide
After the insertion of the C204A cSHMT orf into pET14b problems were encountered. Although ligation ratios were varied, a new batch of pET14b was purchased (Novagen), different methods of purification of DNA from agarose gels such as Geneclean, Crystal gelex and PCR Pure (section 2.5.3) were tested and ligations were performed without prior gel purification of plasmid and insert, no plasmids were obtained that contained pET14b with a mutant cSHMT orf insert.

The full length cSHMT orf mutants were subcloned from pUS1252 into pUC18 using the EcoRI sites to test all the protocols which had been used in the digestion, alkaline phosphatase treatment, gel extraction, ligation and transformation. This method indicates which steps were causing the problems and also produce larger amounts of mutant cSHMT orf DNA as pUC18 is a much higher copy number.

The first mutant tested was K446stop. The 1.5kb EcoRI fragment was purified from an agarose gel using Geneclean, ligated into EcoRI digested and CIP treated pUC18 and seven out of ten colonies resulting from the transformation of DH5α competent cells to ampicillin resistance contained pUC18 with a cSHMT orf inserted. The procedure was repeated for W15A, W15F, W15Y and C380stop mutations and the ratios of plasmids with inserts to plasmids isolated were 3/3, 3/3, 4/4 and 3/4 respectively. The success of this subcloning validated all of the methods used in the subcloning procedure.

Another attempt at cloning the mutant cSHMT cDNAs into pET14b was performed. All of the mutant cSHMT orf generated were now in pUC18 based plasmids (either in pUC18 or in pUS1249) and large amounts of all the mutant plasmids were purified using commercial columns (Promega, Qiagen) from 100ml cultures. Plasmids were digested with NdeI and 1.5kb bands were purified from agarose gels and ligated to NdeI digested and dephosphorylated pET14b.
The mutation K446stop was the first mutant cSHMT orf transferred. Of 92 colonies obtained in the transformation, 4 contained the 1.5kb cSHMT orf inserted into the vector, one of which was in the correct orientation with relation to the promoter and translation initiation signals. The correct plasmid was designated pUS1256.

The W15Y mutation transformation yielded 11 colonies, one of these colonies was mixed and contained both vector alone and vector with cSHMT insert. The other colonies contained only vector. The mixed colony was subcultured on a L-agar plate to isolate individual colonies and screening yielded a pure culture of pET14b with insert. The human cSHMT mutant insert was in the correct orientation for expression and was designated pUS1257.

Four colonies were obtained after transforming the ligation mixture of W15A, three of these contained inserted cSHMT, two of which were in the right orientation for cSHMT expression. One of the correct plasmids was designated pUS1258.

For the mutation C380stop, 1 colony was obtained after transforming DH5α competent cells to ampicillin resistance. Plasmid DNA extracted from this colony contained a 1.5kb insert and was in the correct orientation. The plasmid was designated pUS1259.

One colony was obtained when the ligation of W111A human cSHMT orf with pET14b was used to transform DH5α competent cells to ampicillin resistance. Plasmid DNA was extracted from this colony and it was found to contain the human cSHMT orf in the correct orientation. The vector was renamed pUS1261.

Transformations of DH5α cells to ampicillin resistance with ligation into pET14b of human cSHMT orfs W111F, and W15F yielded hundreds of colonies whereas the vector religation control yielded 20-50 colonies. Standard plasmid preparations by alkaline lysis yielded only pET14b. Therefore, 100 colonies of each mutant were replated and a colony hybridisation was performed using the 1.8kb EcoRI fragment of pUS1217 as a probe.
(sections 2.5.11, 2.5.13). Those colonies containing DNA hybridising to the probe, as indicated by a corresponding signal on X-ray film, were grown overnight and the plasmids were extracted by alkaline lysis. For the W111F cSHMT orf, 3 plasmids with inserts, all in the correct orientation were isolated. For the mutation W15F, one colony contained pET14b with the human cSHMT orf, again in the correct orientation. pET14b plasmids with inserts in the correct orientation were designated pUS1260 and pUS1262 for W111F and W15F respectively.

The mutation H256QK257A in the human cSHMT orf was the last subcloning to be performed out of all the mutants. Review of previous results indicated that there had previously been no problems ligating DNA which had been extracted from agarose gels by spinning through filter paper and the sudden failure of this method had led to the use of all the other gel purification systems and results had been patchy. The method was retried with a new batch of filter paper. 1μg of plasmid isolated by alkaline lysis containing the human cSHMT orf with the H256QK257A mutation in pUS1249 was digested with NdeI. The fragments of DNA were separated by electrophoresis through an agarose gel and the 1.5kb band was excised and spun through a new filter paper column. The eluate was used in a ligation with pET14b digested with NdeI and dephosphorylated. The digested pET14b had already been tested to ensure that dephosphorylation had been effective. Transformation of DH5α competent cells to ampicillin resistance yielded 14 colonies. Six colonies contained pET14b with a 1.5kb insert. Five of these colonies contained the inserted cSHMT orf in the correct orientation for expression. The designation pUS1265 was given to one of the plasmids with the cDNA in the correct orientation.

The F424stop mutation was not proceeded with in the light of results obtained on the solubility of expressed protein for the other C-terminal truncated mutants (see section 4.2.8).
<table>
<thead>
<tr>
<th>Mutant in pET14b</th>
<th>plasmid name</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild-type (D182N)</td>
<td>pUS1253</td>
</tr>
<tr>
<td>W15A</td>
<td>pUS1258</td>
</tr>
<tr>
<td>W15F</td>
<td>pUS1262</td>
</tr>
<tr>
<td>W15Y</td>
<td>pUS1257</td>
</tr>
<tr>
<td>W111A</td>
<td>pUS1261</td>
</tr>
<tr>
<td>W111F</td>
<td>pUS1260</td>
</tr>
<tr>
<td>N182D</td>
<td>pUS1251</td>
</tr>
<tr>
<td>C204A</td>
<td>pUS1254</td>
</tr>
<tr>
<td>H256QK257A</td>
<td>pUS1264</td>
</tr>
<tr>
<td>C380stop</td>
<td>pUS1259</td>
</tr>
<tr>
<td>K446stop</td>
<td>pUS1256</td>
</tr>
</tbody>
</table>

Table 4.3 University of Surrey plasmid assignments for mutant cSHMT orfs in pET14b. After site directed mutagenesis, mutant cSHMT orfs were subcloned into pET14b for expression in *E.coli* and were assigned plasmid numbers.
4.2.8 Expression and purification of mutant human cSHMT protein

All of the pET14b plasmids containing the mutant cSHMT orf were transformed into BL21(DE3)pLysS. Before large scale protein expression and purification was performed, the solubility of expressed mutant human cSHMT proteins at growth temperatures for the cultures of 37°C, 30°C and 25°C was assessed. For mutants C204A and W15Y, cultures were tested at each temperature and cSHMT protein was only found to be soluble in a reasonable quantity at 25°C (figure 4.7). For the other mutant human cSHMTs, solubility of expressed protein was tested only at 25°C. All except the C-terminal truncated mutant cSHMT proteins were found to be soluble at 25°C (see table 4.4).

For the expression of mutants one litre cultures were grown at 25°C in 4 batches of 250ml culture medium and harvested 2 hours post induction with IPTG. The plasmid pUS1251 containing the N182D mutation generated in chapter 3 was also expressed. C204A cSHMT protein was purified on a nickel ion column (Novagen) using gravity flow and cSHMT protein was eluted from the column using 200mM imidazole (figure 4.8). Cobalt ion columns (Clontech) were used to purify the other mutants using a hybrid batch/gravity flow method and mutant cSHMT proteins were eluted from the column using 100mM imidazole as recommended in the manufacturers protocol (Clontech) (figure 4.9). On passing elution buffer through the metal ion columns a yellow eluate was produced due to bound PLP (section 1.1). Once eluted from the metal ion columns, all cSHMT proteins were desalted and concentrated.

C380stop and K446stop cSHMT proteins had calculated molecular weights of 41.6KD and 48.9KD respectively. SDS-PAGE analysis of these proteins showed that the protein expressed by pUS1259 and pUS1256 agreed with the calculated results (figure 4.10, figure 4.11). The truncated C-terminal cSHMT proteins were insoluble at 37°C, 30°C, 25°C and
Figure 4.7 Expression of C204A cSHMT protein at 37°C, 30°C and 25°C in order to determine whether decreasing the expression temperature, raised the level of soluble cSHMT protein in the cytoplasm of *E.coli*. Culture containing pUS1254 was expressed at the above temperatures. Two hours after induction with IPTG, the cells were harvested, lysed and centrifuged to obtain the soluble (supernatant) and insoluble (pellet) fractions.

Lane M Molecular weight markers
Lane 1 25°C pellet (insoluble fraction)
Lane 2 25°C supernatant (soluble fraction)
Lane 3 30°C pellet
Lane 4 30°C supernatant
Lane 5 37°C pellet
Lane 6 37°C supernatant

Please refer to p67a for protein loadings
Table 4.4 Results of experiments to determine the extent of solubility of mutant cSHMT proteins when expressed in *E.coli*. Cultures expressing the cSHMT protein were grown at the temperature indicated. After expression, cells were collected, lysed and centrifuged. The supernatant (soluble fraction) and pellet was analysed by SDS-PAGE to determine whether cSHMT was present.

| Protein         | Presence of cSHMT protein in the soluble fraction at the expression temperature |
|-----------------|---------------------------------------------------------------------------------
|                 | 37°C | 30°C | 25°C | 20°C * |
| Wild-type (D182N) | YES  | nd   | nd   | nd   |
| W15A            | nd   | nd   | YES  | nd   |
| W15F            | nd   | nd   | YES  | nd   |
| W15Y            | YES  | YES  | YES  | nd   |
| W111A           | nd   | nd   | YES  | nd   |
| W111F           | nd   | nd   | YES  | nd   |
| N182D           | YES  | nd   | YES  | nd   |
| C204A           | NO   | NO   | YES  | nd   |
| H256QK257A      | nd   | nd   | YES  | nd   |
| C380stop        | NO   | NO   | NO   | NO   |
| K446stop        | nd   | NO   | NO   | NO   |

* experiment performed at 20°C included sorbitol and betaine in the growth medium (Blackwell and Horgan, 1991).
Figure 4.8 Purification of C204A cSHMT using a nickel ion column. One litre of cells containing pUS1254 was grown at 25°C and expression of C204A cSHMT protein was induced by IPTG. Two hours after induction the cells were harvested, resuspended in binding buffer and sonicated. After centrifugation, the supernatant was passed through a nickel ion column. After washing the column with binding buffer and wash buffer, the cSHMT protein was eluted in two 4ml fractions by the addition of binding buffer containing 200mM imidazole.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>Molecular weight markers</td>
</tr>
<tr>
<td>1</td>
<td>Whole culture 2 hours post expression</td>
</tr>
<tr>
<td>2</td>
<td>Pellet after centrifuging post sonication</td>
</tr>
<tr>
<td>3</td>
<td>Supernatant after centrifuging post sonication</td>
</tr>
<tr>
<td>4</td>
<td>Flowthrough of supernatant applied to column</td>
</tr>
<tr>
<td>5</td>
<td>Flowthrough after washing column with binding buffer</td>
</tr>
<tr>
<td>6</td>
<td>Eluate</td>
</tr>
</tbody>
</table>

Please refer to p67a for protein loadings
Figure 4.9 Purification of H256QK257A cSHMT. Purification was on a cobalt ion column. The cSHMT protein was expressed in the same way as for C204A protein. The cells after expression were resuspended in sonication buffer, sonicated and centrifuged. The supernatant was added to cobalt ion resin in a large tube and mixed. After spinning the supernatant was removed and the pelleted resin was washed with sonication buffer three times. The resin was then placed in a column and protein was eluted in two 4ml fractions using sonication buffer containing 100mM imidazole.

Lane M Molecular weight marker
Lane 1 Pellet after centrifuging sonicated cells
Lane 2 Supernatant after centrifuging sonicated cells
Lane 3 Supernatant after binding to cobalt resin
Lane 4 Wash of cobalt resin
Lane 5 Eluate using 4ml sonication buffer containing 100mM imidazole

Please refer to p67a for protein loadings
Figure 4.10  Expression of K446stop cSHMT protein at 20°C with sorbitol and betaine added to the growth medium. A lower molecular weight for the truncated protein is apparent and match calculated values (see text). No cSHMT protein is seen in the supernatant fraction post lysis of *E.coli* indicating that the truncated mutant is insoluble.

Lane M  Molecular weight marker
Lane 1  Overnight culture used to inoculate media
Lane 2  culture prior to induction of expression
Lane 3  culture 2 hours after induction of expression
Lane 4  Pellet after centrifuging lysed cells
Lane 5  Supernatant after centrifuging lysed cells

Please refer to p67a for protein loadings
Figure 4.11 Expression of C380stop cSHMT protein at 20°C with sorbitol and betaine added to the growth medium. A lower molecular weight for the truncated protein is apparent and match calculated values (see text). No cSHMT protein is seen in the supernatant fraction post lysis of E.coli indicating that the truncated mutant is insoluble.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>Molecular weight marker</td>
</tr>
<tr>
<td>1</td>
<td>Overnight culture used to inoculate media</td>
</tr>
<tr>
<td>2</td>
<td>culture prior to induction of expression</td>
</tr>
<tr>
<td>3</td>
<td>culture 2 hours after induction of expression</td>
</tr>
<tr>
<td>4</td>
<td>Pellet after centrifuging lysed cells</td>
</tr>
<tr>
<td>5</td>
<td>Supernatant after centrifuging lysed cells</td>
</tr>
</tbody>
</table>

Please refer to p67a for protein loadings
20°C with the addition of 660mM sorbitol and 2.5mM betaine to the media (Blackwell and Horgan, 1991) and, therefore, could not be purified in a native state.

### 4.2.9 Assay of mutant cSHMT protein activity

The results of the assays of cSHMT mutants performed using dl-allothreonine (section 2.12) and 3-[14C]-L-serine (section 2.11) are shown in table 4.5. The results from the assay using dl-allothreonine as the substrate were expressed in terms of \( \text{Km} \) and for the assay using 3-[14C]-L-serine as \( EC_{50} \) (the amount of enzyme required to achieve half of the maximum activity).

### 4.3 Discussion

The orf encoding human cSHMT was subcloned into pALTER-1\(^\circ\) enabling site directed mutagenesis to be performed to revert the N182D mutation in the orf generated by PCR back to asparagine. Once the mutation had been performed to form pUS1252 and sequence verified, the orf was subcloned into pET14b to form pUS1253. The wild type cSHMT orf in pET14b was transformed into the expression strain BL21(DE3)pLysS and expression and purification of wild-type human cSHMT protein for X-ray crystallography and inhibitor studies was subsequently performed at the Institute of Cancer Research (S.Renwick, personal communication) (chapter 3). A sample of purified protein was provided for determination of cSHMT enzyme activity by the dl-allothreonine assay.

The plasmid pUS1252 was used as a template to produce the mutations W15A, W15F, W15Y, C380stop, F424stop and K446stop in the human cSHMT orf. The 0.6kb \( KpnI/BamHI \) fragment was subcloned from the wild-type cSHMT orf in pUS1252 into pALTER-1\(^\circ\) to form pUS1255. pUS1255 was used to as a template to generate W111A, W111F, C204A and H256QK257A mutants. A vector (pUS1249) was constructed to enable the reformation of the whole cSHMT orf once the 0.6kb \( KpnI/BamHI \) fragment had been mutated.
<table>
<thead>
<tr>
<th>SHMT protein</th>
<th>Km (mM)</th>
<th>p₁</th>
<th>Vmax (μmoles NADH oxidised/min)</th>
<th>p₂</th>
<th>EC₅₀ (nM)</th>
<th>p₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type (D182N)</td>
<td>1.22±0.44 (n=5)</td>
<td>-</td>
<td>2.31±0.55</td>
<td>-</td>
<td>25.7±14.4 (n=11)</td>
<td>-</td>
</tr>
<tr>
<td>Wild-type - HisTag</td>
<td>nd</td>
<td>-</td>
<td>nd</td>
<td>-</td>
<td>31.7±2.4 (n=3)</td>
<td>0.6634</td>
</tr>
<tr>
<td>W15A</td>
<td>3.51† (n=1)</td>
<td>-</td>
<td>0.87</td>
<td>-</td>
<td>100.0±16.3 (n=3)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>W15F</td>
<td>ncd</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>53.3±2.4 (n=3)</td>
<td>0.0098</td>
</tr>
<tr>
<td>W15Y</td>
<td>2.37±0.93 (n=3)</td>
<td>0.081</td>
<td>0.49±0.07</td>
<td>0.0015</td>
<td>93.3±9.4 (n=3)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>W111A</td>
<td>4.09, 3.26 (n=2)</td>
<td>-</td>
<td>0.40, 0.43</td>
<td>-</td>
<td>46.7±2.4 (n=3)</td>
<td>0.0383</td>
</tr>
<tr>
<td>W111F</td>
<td>ncd</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>73.3±4.7 (n=3)</td>
<td>0.0002</td>
</tr>
<tr>
<td>N182D</td>
<td>2.11±0.68 (n=4)</td>
<td>0.065</td>
<td>2.08±1.05</td>
<td>0.6828</td>
<td>40.0±12.2 (n=3)</td>
<td>0.1237</td>
</tr>
<tr>
<td>C204A</td>
<td>2.10±0.21 (n=5)</td>
<td>0.0047</td>
<td>0.50±0.30</td>
<td>0.0002</td>
<td>52.5±13.5 (n=4)</td>
<td>0.0099</td>
</tr>
<tr>
<td>H256QK257A</td>
<td>17.52, 18.46 (n=2)</td>
<td>-</td>
<td>0.34, 0.52</td>
<td>-</td>
<td>nd</td>
<td>-</td>
</tr>
</tbody>
</table>

† Two other assays gave inconsistent results
‡ Probability that the means of the mutant cSHMT activities are equal to the wild-type

Table 4.5 Results of assays on wild type and mutant cSHMT proteins using either dl-allothreonine or radiolabelled serine and THF as substrates. Assays were performed using methods described in sections 2.11 and 2.12. Results are expressed as means and standard deviations except where n=2 when the two experimental values are shown. EC₅₀ is the amount of cSHMT protein required to produce half of the maximum amount of wild-type SHMT activity.

n, number of assays performed
nd, not determined
ncd, no consistent data available
The mutated cSHMT orfs were subcloned into pET14b in order to express cSHMT protein. The reasons for the initial cloning problems have not been determined but nuclease contamination of the filter paper used to purify DNA fragments from agarose gel slices may have contributed as a new batch was used successfully. Recent work has shown that digested vector purified from an agarose gel using PCR Pure (Clontech) reduces the frequency of religation of digested vector molecules (L. Docherty, personal communication).

Expression of the mutant cSHMT proteins indicated that the solubility of individual proteins was affected, in some cases drastically, by the mutations created. This suggests that the amino acid residues that were mutated may be important for the correct formation of secondary or tertiary cSHMT protein structure. Some proteins expressed in E.coli aggregate in inclusion bodies in an insoluble form (Fischer et al., 1993). Mizukami et al., (1986) demonstrated that by reducing the temperature at which E.coli cells expressing human β-interferon are grown from 37°C to 20°C more protein is produced in an active state although the overall amount of β-interferon protein produced remained the same. Reducing the temperature at which culture expressing mutated cSHMT orfs was grown to 25°C increased the amount of cSHMT protein produced in a soluble form in all except the C-terminus truncated mutants. The histidine tag at the N-terminus of the mutant cSHMT added by the expression vector enabled the rapid purification of mutant cSHMT protein passing the cell lysates over an affinity column.

The C-terminus truncated mutants were insoluble when expressed at 25°C, 20°C and 20°C with the addition of sorbitol and betaine. The addition of sorbitol and betaine has been shown to increase the amount of soluble protein produced (Blackwell and Horgan, 1991; Brownbridge et al., 1993). However, although large amounts of truncated cSHMT were
produced, no soluble protein was visible on a SDS-PAGE gel and no cSHMT protein bound to a sample of the cobalt column matrix when used in a batch purification.

Methods have been devised to purify expressed protein aggregated in inclusion bodies (Fischer et al., 1993). Inclusion bodies in cells expressing recombinant protein are predominantly composed of the recombinant protein in an aggregated form. Inclusion bodies can be sedimented by low speed centrifugation and can be used as a first step of purification (Marston, 1986). The protein in the inclusion bodies can be solubilised in denaturants such as urea or guanidine hydrochloride in which the protein molecules are unfolded (Fischer et al., 1993). In the case of cSHMT mutants expressed here, the histidine tag can then be used to purify the cSHMT protein. Removal of the denaturant results in the refolding of the protein. However, methods used to refold denatured proteins depends on the structure of the protein of interest and often only one particular method will work (Fischer et al., 1993). Success of refolding of denatured wild-type proteins can be assessed easily as activity of the refolded protein should be the same as the native protein. *E.coli* SHMT can be denatured in 8M urea and refolded by performing a 10 fold dilution (Cai et al., 1995). However, refolding of denatured mutant proteins is not desirable. If mutant proteins purified in this way show no activity, it is possible that this is due to incorrect refolding rather than the mutation having affected enzyme activity. With the inherent uncertainty about the validity of any refolded protein no further work was performed with the C-terminal truncated mutants.

The inability of truncated cSHMT proteins to fold correctly in *E.coli* suggest that the last 37 amino acids are important in the folding pathway of human cSHMT. Recent publications (Cai and Schirch, 1996a; 1996b) on eSHMT folding adds credence to this theory. They found that eSHMT has two domains which fold within seconds. The second of these domains spans from amino acids 276 to 417, the last amino acid in eSHMT. Although eSHMT protein is
smaller than human cSHMT, alignment of SHMT protein sequences from different organisms shows that most of this difference in sequence length occurs internally rather than at the protein termini (appendix 2). Therefore, if the results from eSHMT are extrapolated to human cSHMT the truncations described here would occur well within the second domain. In the light of this information it is questionable whether the C-terminus truncations created here would fold into a native protein structure if refolding was attempted on protein purified under denaturing conditions.

*E.coli* SHMT contains three tryptophan residues. Refolding studies on eSHMT were obtained using site directed mutants of eSHMT protein to generate three mutants, each containing only one tryptophan residue so that the quenching of fluorescence of each tryptophan residue could be followed (Cai and Schirch, 1996a). Human cSHMT contains only two tryptophan residues. Mutations to W15 and W111 also yield cSHMT mutant proteins containing a single tryptophan. In principle, fluorescence quenching could be applied to the tryptophan mutants generated for the human cSHMT. However, there is no tryptophan residue past amino acid 111 in human cSHMT so folding of the C-terminal domain could not be followed using this method. Site directed mutagenesis could be used to generate a tryptophan in the second domain, allowing fluorescence quenching to be used.

Although the purification method with the nickel ion column was rapid, the cobalt column offered protein binding advantages discussed in section 1.9. The hybrid batch/gravity flow protocol was faster giving proteolytic enzymes less opportunity to work and gave more thorough washing of the resin, resulting in isolation of purer cSHMT protein. When the wild-type cSHMT protein was purified it was found that eluting the protein in 1M imidazole caused rapid aggregation of cSHMT protein. Reduction of the imidazole in the elution buffer to 200mM still resulted in elution of cSHMT protein, but did not cause protein aggregation
(S.Renwick, personal communication). All of the cSHMT mutant proteins were eluted with 100mM as cobalt columns bind the histidine tag less tightly. Protein purified in this way did not aggregate.

Enzyme assays were performed on wild-type and mutant cSHMT proteins to assess whether the mutations had any affect on cSHMT activity. The radioactive assay was used as serine and THF are the physiological substrates of SHMT. The dl-allothreonine assay does not use THF as a substrate so any mutation affecting THF binding can be determined by differing enzyme activities exhibited in the two assays. The results from the assays do not have enough replicates in order to determine whether the differences between the activities of the mutants are statistically significant. However, some preliminary observations can be made.

Where data is available for 3 or more assays a 2 sample t-test has been performed to test the null hypothesis that there is no difference between the means for the wild-type and mutant cSHMT proteins. More replicates need to be performed to get more accurate probability values.

The 6nM difference in the means for the EC50 between wild-type cSHMT with and without the histidine tag used for purification of the protein suggested that the addition of the histidine tag to the N-terminus of the protein did not significantly affect the activity or structure of the protein. The probability that the means of the two proteins were the same was high (0.6634), also suggesting that the difference between these proteins was not significant. The histidine tag was, therefore, left on all the mutant proteins purified.

The role of lysine 257 in the catalytic activity of SHMT has already been demonstrated (Schirch et al., 1993, Iurescia et al., 1996). The K257A mutation was performed to produce an inactive SHMT to act as a negative control to establish
whether any of the cSHMT mutant proteins produced were inactive. Due to an extra mutation generated during site directed mutagenesis, the actual mutant created was H256QK257A. The H256QK257A mutated cSHMT protein would also act as a negative control as the critical lysine was still removed from the protein. There was a trend towards a very small amount of activity of H256QK257A cSHMT protein shown using dl-allothreonine as a substrate. Purification of site directed mutants of eSHMT expressed in *E. coli* has shown that wild-type eSHMT produced by *E. coli* can contaminate the mutant proteins and mask any potential lack of activity (Schirch *et al.*, 1993; Iurescia *et al.*, 1996). However, eSHMT and human cSHMT proteins are different in their sizes and quaternary structure (section 1.1) and the eSHMT mutants were not purified by an affinity purification method whereas the cSHMT mutants purified here a purified on the basis of a histidine tag present only on the N-terminus of the expressed protein. If contamination with wild-type protein was a problem with the purified cSHMT proteins then a band corresponding to eSHMT would be seen on SDS-PAGE gels at 46KD as well as the cSHMT band at 53KD. In all mutants purified there is only a band at 53KD seen suggesting that contamination with active eSHMT protein is not a problem. However, it cannot be ruled out that there may be very small amount of eSHMT present in the purified mutant proteins. This may account for the small amount of activity seen with the H256QK257A protein. If contamination was a problem then it could be identified by purifying more cSHMT protein and determining whether the activity between the two batches is different. However, to ensure that eSHMT cannot be a contaminant the expression of mutant cSHMT proteins could be performed in a strain of *E. coli* which has the glyA gene encoding eSHMT deleted. To date no such strain is available which also has the DE3
genotype containing the gene for T7 RNA polymerase required for expression of pET plasmids.

The assay using serine and THF showed that the activity of N182D was lower than the wild-type enzyme but the probability that the mean was equivalent to the wild-type was high suggesting that the change was not significant. Comparing the Km and Vmax for N182D to wild-type values in the dl-allothreonine assay also suggested that the differences between the N182D and wild-type values were not significant at the 95% confidence level. At amino acid 182 the only SHMT proteins containing an asparagine residue are human, mouse and sheep cSHMT, all the other SHMT proteins have an aspartic acid residue in the equivalent position (appendix 2). The side chain of aspartic acid is negatively charged whereas that of asparagine has no charge. As the change in activity between wild-type and N182D protein was not significant, it can be implied that a negative charge at this position does not play an important role in SHMT activity.

For the C204A mutation, both assays produced results which were statistically significant. The Km and Vmax for allothreonine were lower than the wild-type and in the radiolabelled serine assay, more C204A cSHMT protein was required to obtain half of the maximum activity of wild-type protein. However, comparing the Km for C204A cSHMT to H256QK257A cSHMT suggests that the C204A protein still has a relatively high activity. This change in activity for the C204A mutant cSHMT protein is different to that gained by chemical modification using iodoacetate (Schirch et al., 1980). Incubating rabbit cSHMT with iodoacetate abolished SHMT activity as PLP could not bind to the active site, establishing C204 was as being in the PLP binding pocket but its potential role in catalysis could not be established. Using site directed mutagenesis the activity of cSHMT lacking this cysteine residue has been ascertained.
and the results shown above suggest that although some activity is lost, this particular
cysteine residue is not essential for catalytic activity of cSHMT. An alignment of all
the known cSHMT protein sequences (appendix 2) shows that there is a cysteine
residue at position 204 only in the mammalian and \textit{C.elegans} cytosolic isozymes. The
other cSHMT proteins, the mitochondrial isozymes or the prokaryotic SHMT proteins
contain an alanine residue at this position further suggesting that the C204 is not a
critical residue for catalysis.

W111A and W111F mutations produced significantly less active cSHMT proteins
at the 95% confidence level and approximately 2 and 3 times more cSHMT protein
respectively, was required to get 50% of the maximum wild-type activity in the assay
using serine and THF compared to wild-type. In the assay using dl-allothreonine as a
substrate, preliminary results suggested that the affinity of W111A mutated cSHMT
protein for dl-allothreonine was reduced. A \textit{Km} for dl-allothreonine for W111F
cSHMT could not be determined, as at each concentration of dl-allothreonine
measured, the activity was too low.

In the assay using serine and THF as substrates, the activities of W15A, W15F and
W15Y proteins were reduced at a statistically significant level compared to wild-type
protein activity. The activity W15F cSHMT was higher than the other W15 mutants
suggesting that phenylalanine can partially perform the role of tryptophan at amino
acid position 15 in cSHMT. The assay using dl-allothreonine as the substrate
suggested that for the mutation W15Y, the difference in \textit{Km} between this mutant and
wild-type was not significant at the 95% level but the \textit{Vmax} value was significantly
different from the wild-type. Preliminary indications were that these trends may be
mirrored in the other W15 mutants. However, the data suggests that affinity of the
W15 mutants for dl-allotheonine is higher than for H256QK257A mutant cSHMT and activity is not completely lost in either assay.

Vijayalakshmi and Rao, (1989) showed using chemical modification with N-bromosuccinimide (NBS) that at least one tryptophan residue was necessary for the activity of human cSHMT but the enzyme was protected against inhibition by serine and THF. There are only two tryptophan residues in human cSHMT and they have both been mutated in this study. As mentioned previously the reason that two assays were performed is that the dl-allotheonine assay does not use THF so if the mutant Km values were not affected by the mutation but the EC\textsubscript{50} was, then this would indicate that THF binding was affected. The results of the assays of the tryptophan mutants discussed above are inconclusive in determining which of the tryptophan residues is involved in THF binding as both of the assays showed a reduction in activity when either of the tryptophan residues was mutated. The changes in activity seen could be due to a change in the secondary, tertiary or quaternary structures of the protein caused by the change in the primary structure and, therefore, it cannot be ruled out that THF binding is not affected. Further work needs to be performed to determine whether the structures of the mutant cSHMT proteins are different to the wild-type. Possible methods to determine whether the structures are different are circular dichroism and differential scanning calorimetry. Differential scanning calorimetry has previously been used on SHMT to determine the enthalpy of denaturation (Schirch et al., 1991). Mutations in SHMT can either increase or decrease the enthalpy, corresponding to a more stable or less stable protein respectively.

If a tryptophan residue is required by human cSHMT for THF binding, there is conflicting evidence as to which of the two tryptophan residues is necessary.
Alignments of eSHMT to dialkylglycine decarboxylase (DGD) and aspartate aminotransferase (AAT) have been performed and it has been proposed that eSHMT shares the same protein folds (Pascarella et al., 1993). The X-ray structures of DGD and AAT are known (Toney et al., 1993, McPhalen et al., 1992). In both DGD and AAT the N-terminus of one subunit interacts with the other subunit of the dimer. Cai and Schirch, (1996b) demonstrated that on refolding denatured E.coli SHMT, the first 55 amino acids remain in a disordered form until dimerisation occurs and so, therefore, the N-terminus of the protein may have a similar role in cSHMT. It has also been demonstrated that by removing the first 25 amino acids of rabbit cSHMT, the catalytic activity was increased when serine and THF were used as substrates but was not changed when dl-allothreonine was used (Schirch et al., 1984) implying that W15 is not essential for THF binding.

In AAT, tryptophan residues from the N-terminus interact with hydrophobic pockets in the second subunit (McPhalen et al., 1992). If this homology is extended to human cSHMT then mutation of W15 to alanine, phenylalanine or tryptophan may prevent interaction of the subunits disrupting the quaternary structure and distorting the tertiary structure.

Therefore, it could be hypothesised that the first 25 amino acids, including W15 are near to the THF binding site but play a structural role rather than a catalytic role and by removing them THF can access the site more easily. However, by changing amino acid W15, the tertiary structure of the first 25 amino acids is altered, perhaps because they cannot associate correctly with a second subunit, and the THF binding site is obscured so THF has more difficulty accessing the site. This hypothesis could be tested by calculating the affinity of THF for the cSHMT with the first 25 amino acids removed and the W15 mutants and comparing to the affinity for the wild-type. The
same "partial blocking" effect would be seen by the addition of a chemical group to a tryptophan close to the THF binding site. Therefore, it cannot be ruled out that inactivation of human cSHMT by NBS is also caused by steric hindrance.

Alignments of all the amino acid sequences for SHMT (appendix 2) show that the tryptophan at amino acid 15 is conserved only in mammalian cSHMTs. It is, therefore, unlikely that the W15 plays a major role in SHMT catalytic activity.

The data for the mutants generated does not rule out W111 as being the tryptophan which binds THF but alignment of the amino acid sequences of all SHMTs shows that W111 is conserved in all eukaryotic SHMTs. This residue corresponds to a gap in the alignment in all prokaryotic sequences. If this tryptophan residue was involved in THF binding it is expected that this residue would be conserved in all species. One difference between eSHMT and eukaryotic SHMTs is that eSHMT forms dimers whereas, eukaryotic SHMTs from tetramers. Therefore, W111 may play a role in the formation of cSHMT quaternary structure and mutation of this residue affects the interaction of the sub units and hence the activity of the protein.

Although the activities of the mutant cSHMT proteins have been calculated using serine and THF or dl-allothreonine as substrates further work is required to establish the precise roles of these mutations. The Km for serine and THF can be calculated by coupling the SHMT reaction to that of C1-THF synthase which reduces NADP⁺, hence the reaction can be followed spectrophotometrically. However, C1-THF synthase is not commercially available and purification from cells is expensive and laborious. Dissociation constants for serine and THF can also be measured (Schirch and Ropp, 1967). The mutants could also be incubated with glycine and serine and the appearance of the absorption maxima mentioned in section 1.2 determined. As each absorption maxima corresponds to a particular intermediate in the catalytic mechanism
(figure 1.1), any changes affecting the mechanism will be manifested by the changes in these maxima.

Only elucidation of a tertiary and quaternary structure will really determine the accuracy of theories on SHMT structure and catalytic mechanism derived from all the chemical modification and site directed mutagenesis studies performed here and by other researchers.
CHAPTER 5

CHARACTERISATION OF A HUMAN CYTOSOLIC

SHMT PSEUDOGENE
5.1 Summary

A number of cDNA's encoding human and rabbit cytosolic and mitochondrial SHMT isozymes have been isolated (Byrne, 1992; Xu, 1992; Garrow et al., 1993; Whitehouse, 1996). It has been demonstrated (chapters 3 and 4) that the isolation of cDNAs allows human SHMT isozymes to be expressed in large quantities. The sequencing of cDNAs may also provide evidence of translational regulation. For example, an orf overlapping the cSHMT orf in rabbit cSHMT cDNA was identified and experiments indicated a possible role for the overlapping orf in reducing the translation of the cSHMT orf (Byrne et al., 1995). However, cDNA sequence gives no information on the regulation of transcription of the genomic sequences from which the cDNAs were derived. Sequences within introns normally lost in cDNA cloning may also contain sequences which affect transcription, as shown in the cSHMT gene in *N. crassa* (McClung et al., 1992). In the thymidylate synthase gene which is expressed in a cell cycle dependent manner, loss of several introns leads to reduced transcription (Deng et al., 1989) and in constitutive expression (Takayanagi et al., 1992).

Human cells have two isozymes of SHMT which are differentially regulated (section 1.1.1) and SHMT expression is increased in the build up to cell division. It can, therefore, be hypothesised that the genomic sequences encoding the cytosolic and mitochondrial isozymes of SHMT contain sequences involved in differential regulation.

Previous work to isolate human genomic sequences encoding SHMT isozymes and identify potential regulatory elements in the promoter and introns resulted in the isolation of putative SHMT genomic clones from a human genomic DNA library constructed in lambda phage EMBL3A (Byrne, personal communication). Two of these clones (3 and 4) were digested with *EcoRI* and *BamHI* and fragments were subcloned into pUC18 (Byrne, personal communication).
The work presented in this section describes the further characterisation of the subcloned fragments and the two genomic clones. The genomic DNA hybridising to cSHMT was sequenced and found to encode a human cSHMT processed pseudogene.

5.2 Results

5.2.1 Sequencing of restriction digest fragments hybridising to cSHMT cDNA

Restriction mapping of 3kb fragments produced by digestion of clones 3 and 4 by EcoRI showed that these fragments had identical restriction sites. Preliminary sequencing of the ends of these fragments produced identical sequence, indicating that they were identical fragments. The 3kb EcoRI fragment from clone 4 and a 1.7kb BamHI fragment from clone 3, both subcloned into pUC18, were digested with KpnI, BamHI, PstI, HindII, SacI, EcoRI, HindIII and XbaI and the released fragments were subcloned into pUC18. After digestion with the above enzymes, parts of the original fragments were not liberated from pUC18 and were, therefore, religated to form smaller vectors. The resulting subcloned plasmids and a 0.6kb BamHI fragment from clone 3 were sequenced using universal and reverse primers. After sequencing, two gaps remained in the sequence, one in the 3kb EcoRI fragment and the other in the 1.7kb BamHI fragment. Oligonucleotides kjc1-4 were designed and sequence spanning these gaps was obtained. Restriction maps and sequencing strategy for the 3kb EcoRI and 1.7kb BamHI fragments are shown in figures 5.1 and 5.2.

A further 0.6kb SalI/EcoRI fragment from clone 4 hybridising to cSHMT cDNA (pUS1213) was identified by Southern blotting. This fragment was subcloned into pUC18 and sequenced using universal and reverse primers. The SalI terminus of this 0.6kb fragment contained sequence homologous to one arm of the EMBL3A vector.

Comparison of the sequenced fragments revealed that the 3kb EcoRI fragment overlapped the 1.7kb and 0.6kb BamHI fragments. The sequence of the 0.6kb SalI/EcoRI fragment from
Figure 5.1 Diagram of the 3kb fragment isolated from clone 4 by digestion with EcoRI. Indicated by arrows are the sequencing reactions performed. Also indicated is the area of homology of the sequence to human cSHMT cDNA (pUS1213).
Figure 5.2 Diagram of the 1.7kb fragment from clone 3 isolated by digestion of the clone with BamHI. Indicated by arrows are the sequencing reactions. Also indicated is the area of homology of the fragment to human cSHMT cDNA.
Figure 5.3 Relationships between the sequenced fragments of human genomic DNA. Indicated is the area of homology with human cSHMT cDNA. Also shown is the homology between the sequenced fragments. The Sau3A site indicates the end of the sequence in clone 4.
clone 4 was matched to sequence in the 1.7kb BamHI fragment from clone 3. Figure 5.3 shows the relationship of the fragments sequenced to each other. Sequences from the overlapping regions of the fragments from clones 3 and 4 were identical. The sequences of the fragments were merged to form one continuous sequence 4.3kb in length (figure 5.4).

By alignment with human cSHMT cDNAs it was demonstrated that the entire cSHMT coding sequence with some 5' and 3' non coding sequence was present in the 4.3kb sequence between nt 1157-3491 (figure 5.4). There were, however, many base changes, insertions and deletions in the genomic DNA fragment sequenced compared to the cSHMT cDNA sequence. These base changes include the mutation of the ATG start codon to ATA (nt 1177-9). No promoter sequences for RNA polymerase II were found in the sequence upstream of the mutated start codon. The point mutations and frame shifts had also introduced two premature stop codons in frame with the mutated start codon. The 0.6kb SalI/EcoRI fragment ends at the Sau3A site at nt 3680. The sequence in figure 5.4 was submitted to the Genbank DNA database and was given the accession number X85980.

5.2.2 Mapping of genomic clones

As mentioned in the previous section lambda clones 3 and 4 contained related DNA containing the same cSHMT sequence. Some differences between the two clones was, however, indicated by restriction mapping whole lambda clones 3 and 4.

Two mapping strategies were employed to further define the genomic DNA inserts in clones 3 and 4. A short range map was built up by hybridisation with DIG labeled probes from either the cSHMT cDNA or from the 5' end of the sequenced section. Secondly, restriction sites were mapped using partial digests followed by hybridisation to the cos ends (Rackwitz et al., 1984).
Figure 5.4  Sequence of the human genomic DNA fragments, merged to produce a 4.3kb fragment. Sequence in red italics indicates the direct repeats found at either end of the cSHMT cDNA homologous sequence. Sequence in blue is sequence with no homology to cSHMT cDNA within the direct repeat boundary. Sequence in green is the mutated start codon. Sequences in magenta are mutations compared to the cSHMT cDNA. Insertions in the sequence compared to cSHMT cDNA are underlined and sequences in bold border deletions in the human genomic DNA compared to the cDNA.
Using cSHMT cDNA (pUS1213) as a probe, fragments from EcoRI, XbaI, HindIII, and NdeI digests were identical in clones 3 and 4 and could be identified as spanning the coding region. Hybridisation of this probe to SalI, BamHI, KpnI and SmaI digests gave different products in the two clones which hybridised to the cDNA. Hybridisation of cDNA probe to PstI digested DNA gave too many bands to be useful in mapping the fragments (figures 5.5, 5.6). The total inserts for lambda clones 3 and 4 are approximately 12.5kb and 17.5kb respectively as digestion with SalI releases the DNA inserted between the lambda arms and there were no SalI sites inside the inserts.

A 1.4kb EcoRI/KpnI fragment derived from the 3kb EcoRI fragment from clone 4 was used as a probe to detect sequences which hybridised to the 5' end of the deduced sequence. The use of this probe meant that the BamHI, KpnI and SmaI restriction fragments could be orientated with respect to the sequenced section and additional SmaI and KpnI sites were located (Figure 5.7).

Further evidence was provided that the same piece of human genomic DNA had been inserted into the two clones as many of the fragments were identical for both clones. It was already known that the Sau3A site at nt 3680 in the sequence was the end of the clone 4 insert and was, therefore, ligated to one of the arms. The restriction maps suggested that this was the right arm. For clone 3, a 1.2kb SalI/EcoRI fragment corresponded to the 3' end of the 1.7kb BamHI fragment suggesting that the end of the sequence derived in 5.4 was one end of the insert in clone 3. The KpnI, HindIII and SmaI digests suggested that this end was ligated to the left arm of EMBL3A.

To confirm information obtained from probing digested DNA with sequence from the inserts of the lambda clones, partial digestion of the full lambda clones and mapping with DIG labeled oligonucleotides complementary to the left and right cos ends of the lambda
<table>
<thead>
<tr>
<th>Lane</th>
<th>Enzyme</th>
<th>Clone 3 (kb)</th>
<th>Clone 4 (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SalI</td>
<td>12.5</td>
<td>17.5</td>
</tr>
<tr>
<td>2</td>
<td>SalI/EcoRI</td>
<td>3.0, 1.4</td>
<td>3.0, 0.6</td>
</tr>
<tr>
<td>3</td>
<td>EcoRI</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>4</td>
<td>Xbal/EcoRI</td>
<td>2.7</td>
<td>2.7</td>
</tr>
<tr>
<td>5</td>
<td>Xbal</td>
<td>3.8</td>
<td>3.8</td>
</tr>
<tr>
<td>6</td>
<td>PstI</td>
<td>nd</td>
<td>1.8, 0.4, 0.3</td>
</tr>
<tr>
<td>7</td>
<td>Ndel</td>
<td>4.8</td>
<td>4.8, 3.7</td>
</tr>
<tr>
<td>8</td>
<td>HindIII</td>
<td>4.0</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Figure 5.5  Hybridisation of human cSHMT cDNA (pUS1213) to digests of lambda clones.

The table indicates the sizes of the fragments hybridising to the probe.  nd not determined

Lane M  \( \lambda \text{HindIII} \) DNA size markers
Figure 5.6 Hybridisation of human cSHMT cDNA (pUS1213) to digested lambda clones.

Indicated in the table are fragments hybridising to the probe. nd not determined

<table>
<thead>
<tr>
<th>Lane</th>
<th>Enzyme</th>
<th>Clone 3 (kb)</th>
<th>Clone 4 (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>SalI</em></td>
<td>12.5</td>
<td>17.5</td>
</tr>
<tr>
<td>2</td>
<td><em>EcoRI</em></td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>3</td>
<td><em>BamHI</em></td>
<td>10.0, 1.7, 0.6</td>
<td>10.0, 0.6</td>
</tr>
<tr>
<td>4</td>
<td><em>PstI</em></td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>5</td>
<td><em>KpnI</em></td>
<td>4.4</td>
<td>11.0</td>
</tr>
<tr>
<td>6</td>
<td><em>HindIII</em></td>
<td>22.0, 4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>7</td>
<td><em>SmaI</em></td>
<td>3.5</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Lane M  \( \lambda HindIII \) DNA size markers
<table>
<thead>
<tr>
<th>Lane</th>
<th>Enzyme</th>
<th>clone 3 (kb)</th>
<th>clone 4 (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SalI</td>
<td>12.5</td>
<td>17.5</td>
</tr>
<tr>
<td>2</td>
<td>EcoRI</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>3</td>
<td>BamHI</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>4</td>
<td>PstI</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>5</td>
<td>KpnI</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>6</td>
<td>HindIII</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>7</td>
<td>SmaI</td>
<td>12.5</td>
<td>14.5</td>
</tr>
</tbody>
</table>

Figure 5.7 Hybridisation of a 1.4kb EcoRI/KpnI fragment from clone 4 to the digested lambda clones. Fragments hybridising to the probe are shown in the table.

Lane M $\lambda$HindIII DNA size markers
arms was performed. (this technique is described in section 1.11). Results of the hybridisation of the partially digested fragments to the left and right arms of EMBL3A are shown in figures 5.8, 5.9 and 5.10. The sizes of the partially digested fragments, derived from agarose gel electrophoresis, were often inaccurate due to the high sizes of most fragments meaning that a 1mm difference on the gel was equivalent to several kilobases. However, the relationship of the different restriction sites to each other can be seen by comparing the locations of the signals detected on the X-ray film for each enzyme to each other. This method confirmed that in clone 4 the sequenced fragment was located adjacent to the right arm. The sequence of the fragment was running from 5' to 3' in clone 4. For clone 3, evidence from the mapping gave credence to the theory that the sequenced part of the insert was ligated to the left arm. In this case the BamHI site which occurs at the end of the sequenced region is ligated to the BamHI site of the left arm. Therefore, the sequence was running 3' to 5' in clone 3. By putting all the information together maps were produced for lambda clones 3 and 4 (figure 5.11).

5.3 Discussion

5.3.1 Sequencing of genomic DNA homologous to cSHMT cDNA

The sequencing of the human genomic library clones hybridising to human cSHMT revealed a sequence similar to that of human cSHMT cDNA. Using the BESTFIT program (Genetics Computer Group) the genomic sequence was aligned to the human cSHMT cDNA with 91% identity over the region nt 1157-3491 (figure 5.4). BESTFIT alignment of the genomic sequence to mSHMT cDNA yielded only a 65% identity. It can, therefore, be assumed that the human genomic sequence presented here was derived from the cSHMT gene.

The sequence of the DNA hybridising to the human cDNA probe has been compared to all the human SHMT cDNA’s that have been isolated. The region corresponding to the coding
Figure 5.8 Mapping of clone 3 by partial digestion and hybridisation to oligonucleotides complementary to the cos ends.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Enzyme</th>
<th>Left arm oligo</th>
<th>Right arm oligo</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>SalI</em></td>
<td>16.4, 21.2</td>
<td>9.6, 19, 28</td>
</tr>
<tr>
<td>2</td>
<td><em>EcoRI</em></td>
<td>23</td>
<td>16.3, 19, 28</td>
</tr>
<tr>
<td>3</td>
<td><em>BamHI</em></td>
<td>16.3</td>
<td>10.6, 23</td>
</tr>
<tr>
<td>4</td>
<td><em>KpnI</em></td>
<td>15.4, 23</td>
<td>13.7, 16.4, 18, 20.5, 25</td>
</tr>
<tr>
<td>5</td>
<td><em>HindIII</em></td>
<td>23</td>
<td>4.6, 13.8, 16.3, 20, 25</td>
</tr>
<tr>
<td>6</td>
<td><em>XmnI</em></td>
<td>1</td>
<td>2.4</td>
</tr>
<tr>
<td>7</td>
<td><em>ClaI</em></td>
<td>4.2, 15.4, 23</td>
<td>2.0, 4.3, 9.3</td>
</tr>
</tbody>
</table>

Lane M  \*HindIII* DNA size markers

Lane 8  control lambda DNA partial digest
<table>
<thead>
<tr>
<th>Lane</th>
<th>Enzyme</th>
<th>Left arm oligo (kb)</th>
<th>Right arm oligo (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>SalI</em></td>
<td>16.5, 23</td>
<td>9.4, 23</td>
</tr>
<tr>
<td>2</td>
<td><em>EcoRI</em></td>
<td>24</td>
<td>9.6, 12, 23</td>
</tr>
<tr>
<td>3</td>
<td><em>BamHI</em></td>
<td>17.8</td>
<td>9.7, 23</td>
</tr>
<tr>
<td>4</td>
<td><em>HindIII</em></td>
<td>25</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td><em>KpnI</em></td>
<td>15.4, 23</td>
<td>10, 23</td>
</tr>
<tr>
<td>6</td>
<td><em>SmaI</em></td>
<td>17.8, 23</td>
<td>5, 22, 23</td>
</tr>
<tr>
<td>7</td>
<td><em>ClaI</em></td>
<td>4.3, 6.5</td>
<td>2, 5.5</td>
</tr>
</tbody>
</table>

Figure 5.9 Mapping of clone 4 by partial digestion and hybridisation to oligonucleotides complementary to the cohesive ends (cos) of the lambda arms.

Lane M  \(\lambda HindIII\) DNA size markers

Lane 8  control lambda DNA partial digest
<table>
<thead>
<tr>
<th>Lane</th>
<th>Enzyme</th>
<th>Left arm oligo (kb)</th>
<th>Right arm oligo (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BamHI (10U/µl)</td>
<td>20</td>
<td>9.6</td>
</tr>
<tr>
<td>2</td>
<td>BamHI (1U/µl)</td>
<td>20</td>
<td>9.6</td>
</tr>
<tr>
<td>3</td>
<td>EcoRI (10U/µl)</td>
<td>23</td>
<td>9.6, 13, 25</td>
</tr>
<tr>
<td>4</td>
<td>EcoRI (1U/µl)</td>
<td>23</td>
<td>9.6</td>
</tr>
<tr>
<td>5</td>
<td>HindIII (10U/µl)</td>
<td>22</td>
<td>4.5, 9.6</td>
</tr>
<tr>
<td>6</td>
<td>HindIII (1U/µl)</td>
<td>23</td>
<td>4.5, 9.6, 13.5, 25</td>
</tr>
<tr>
<td>7</td>
<td>SalI (10U/µl)</td>
<td>20</td>
<td>9.8</td>
</tr>
</tbody>
</table>

Figure 5.10 Mapping of clone 4 by partial digestion and hybridisation to oligonucleotides complementary to the cohesive ends (cos) of the lambda arms. Two different concentrations of enzyme were used for BamHI, HindIII and EcoRI.

Lane M  λHindIII DNA size markers

Lane H  High size DNA markers

Lane 8  Lambda DNA control
Figure 5.11 Diagram of lambda clones 3 and 4. In black are the lambda DNA arms of EMBL3A and in red are the human genomic DNA inserts. Information used to generate this map was derived from hybridisation of cos probes to partial digests of the clones and from probing restriction digests with cSHMT cDNA and a 1.4kb EcoRI/KpnI fragment from clone 4.

Key to restriction enzymes
B BamHI
E EcoRI
H HindIII
K KpnI
L Sall
S SmaI
sequence for the protein spans nt 1177-2609 and the 3' untranslated region from nt 2610-3491. However, as mentioned in the results the start codon has been mutated from ATG to ATA and there were many point mutations, deletions and insertions in the genomic sequence compared to the cSHMT cDNA.

Kozak, (1987a) studied 699 vertebrate sequences and deduced a consensus sequence for initiation of translation. The translation initiation consensus was compared to the human genomic sequence isolated here and to cSHMT cDNA below with the start codon underlined.

<table>
<thead>
<tr>
<th>consensus</th>
<th>(GCC)GCC A/G CCATGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>genomic sequence</td>
<td>(ACC)CAT G CAATAG</td>
</tr>
<tr>
<td>cSHMT cDNA</td>
<td>(ACC)AGT G CAATGA</td>
</tr>
</tbody>
</table>

Although the human cSHMT cDNA sequence is not a perfect match for Kozak's consensus, the most highly conserved element of a purine at -3 is conserved. The cytosine residues at -1, -2, -4 and -5 are less conserved than the purine at -3 and translation will occur when these positions are different. (Kozak, 1987b). The human genomic sequence does not match the Kozak consensus either, although it is interesting that there is the more favoured G residue at +4. However, translation would not be initiated as the start codon is mutated. If translation of the genomic SHMT isolated were possible, the resulting protein would be totally dysfunctional and would be truncated due to two in frame early stop codons produced by the mutations.

Introns are a common feature of human genes, they are present in the mSHMT gene (Stover et al., 1985) and would have been expected in the cSHMT gene as well. However, there are no introns present within the region homologous to cSHMT.

Upstream of the mutated start codon there are 20 nucleotides (nt 1157-1176) of sequence homologous to the 5' untranslated region of previously isolated human cSHMT cDNA's
(Byrne, personal communication) some of which may be incompletely spliced and still contain intron sequences. Two of these human cDNA's contain a sequence consistent with an intron having been spliced from the mRNA at the point where pseudogene and cDNA homology ends (pUS1216 and pUS1217) and one sequence (pUS1214) contains an intron at this point. Nucleotides 1-1156 have no homology to SHMT cDNA's isolated or to the intron in pUS1214. The region spanning nt 1140-1158 in the genomic DNA does not correspond to the consensus for an intron splice acceptor site (Ohshima and Gotoh, 1987) and it is, therefore, unlikely that the sequence upstream of nt 1156 is an intron. There are also no sequences corresponding to a RNA polymerase II promoter suggesting that the cSHMT sequence is not transcribed by RNA polymerase II. Searching of the Genbank database reveals no homology to known sequences between nt 1 and 1156.

At the 3' end of the gene, homology to cSHMT cDNAs ends at nt 3491. Directly following this nucleotide is a 16 base sequence AAAAGGTAAGATTTA which is also present between nucleotides 938-953. A comparison of the 3' untranslated region downstream of the second direct repeat to sequences within the Genbank database shows only homology to an Alu repeat element starting at 4313 and running to the end of the known sequence.

Alu repeats are the most abundant family of repeated sequence in the human genome. The name was derived from the fact that the repeat is digested by the restriction endonuclease AluI (Houck et al., 1979). The Alu repeat sequences are closely related 300nt sequences, with an average homology to the consensus of 87%. Each Alu repeat consists of a tandem duplication of a 130bp sequence with an additional 31bp inserted into the right half of the sequence (Deninger et al., 1981). Alu repeats also have direct repeats at each end due to duplication of the target site at the point of insertion into the chromosome. As the homology
to the Alu repeat occurs at the end of the isolated human genomic sequence, it cannot be confirmed that this is a genuine Alu repeat sequence.

The presence of such an adenine rich direct repeat at each end of the coding sequence combined with the sequence mutations, lack of introns and lack of promoter sequence indicates that this sequence is a processed pseudogene (for review see Vanin, 1985). It is proposed that processed pseudogenes arise from the insertion of either mature mRNA or cDNA into a staggered break in a chromosome. (Vanin, 1985). Recently it has been demonstrated that human cells have an endogenous reverse transcriptase activity which can generate cDNA copies of mRNA. Using reporter gene technology and following the insertion of processed pseudogenes into the HeLa cell genome it has been shown that processed pseudogene formation is still continuing and is not an evolutionary relic (Maestre et al., 1995).

In processed pseudogenes studied by Vanin, (1985) sequence divergence between the original and the pseudogene between the two direct repeats was uncommon. As mentioned above the genomic sequence between nt 954 (the end of the 5' direct repeat sequence) and 1156 does not have any homology to cSHMT cDNA's isolated so far but is within the direct repeats bordering the pseudogene. Vanin, (1985) proposed the formation of some processed pseudogenes was due to transcription by RNA polymerase III rather than RNA polymerase II. It has previously been shown that the human β-globin gene can be transcribed by RNA polymerase III and that this transcript is processed in the same way as transcripts generated by RNA polymerase II (Carlson and Ross, 1983). Sharp et al., (1981) and Geiduschek et al., (1988) defined the consensus sequences for boxes A and B where transcription factors responsible for the transcription of tRNA genes by RNA polymerase III bind to the DNA.
These consensus sequences were compared to the sequence between 954 and 1156 of the human genomic DNA isolated.

Box A  
TRGCNNAGY.GG  T in bold is generally at +8 of transcript sequence  
TAGCAGAGC.CT  corresponding T is at nt 987

Box B  
GGTTCGANTCC  T in bold is generally at 55 of transcript sequence  
GCTTCGGACCC  corresponding T is at nt 1131

It can be seen from these comparisons that there is almost an exact match for box A in the human genomic sequence. The alignment with box B is less exact. The distance between the two putative boxes in the pseudogene is higher than in the consensus. However, in yeast tRNA genes the distance between the two boxes varies between 31 and 93 bp (Geiduschek et al., 1988). The distance of box A from the 5’ end also varies. tRNA genes without the second control region (box B) can be transcribed but with reduced efficiency (Sharp et al., 1981). Therefore, it may be possible for RNA polymerase III to have generated a transcript from the cSHMT gene which was processed in the same way as transcripts generated by RNA polymerase II and subsequently incorporated into the genome as a processed pseudogene.

An estimation of the time since the pseudogene was created from the cSHMT gene can be performed based on the number of substitutions made in the pseudogene compared to the cDNA. As the pseudogene has no RNA polymerase II promoter it is assumed that the gene was inactivated as soon as the duplication event occurred and therefore, any mutations in the pseudogenes are neutral. Using the figure of 0.7% per million years as the rate of fixation of neutral mutations in the genome (Perler et al., 1980; Ochman and Wilson, 1987), and there is 7.2% nucleotide substitution (excluding insertions and deletions) over the region that used to code for cSHMT protein compared to cSHMT cDNA, the duplication event can be estimated as being 11.5 million years (MY) ago.
This method of calculation assumes that the rate of fixation is the same in pseudogenes as in functional genes. In functional genes there are two types of mutations, ones that are silent and do not affect the coding sequence and those that do affect the coding sequence, known as replacement mutations. The rate of fixation of replacement sites in functional genes is slower than for silent site as those mutations changing the amino acid often lead to a dysfunctional protein and are eliminated. The other assumption in the calculation is that there has been no change in the sequence in the functional gene since the duplication event. Others authors have modified Perler’s assumption and attributed half the nucleotide substitution rate at the third codon position to evolutionary drift, i.e. the functional gene has accumulated silent site mutations since the time of divergence, and these mutations are likely to be at third position of the codon (Lee et al., 1983). Using this assumption the time since the cSHMT pseudogene was created is 7.7 MY, a reduction of about 4 MY from the previous calculation.

There is evidence that the mutation rate in inactivated pseudogenes is faster than in functional genes as there is no selective pressure on the mutation. Using the method of Perler et al., (1980), Proudfoot and Maniatis, (1980) estimated the time since divergence of the mouse α globin and its pseudogene was 30 MY, between human α globin and its pseudogene was 60 MY and between rabbit β globin and its pseudogene was 55 MY. However, Li et al., (1981) calculated the time since duplication for the same samples as above as 27MY, 49MY and 44MY. This method was achieved by calculating substitutions since duplication between the pseudogene and two functional counterparts. The substitutions between the two functional genes were also calculated. It was found that the rates of change in all three codon positions were equal in the pseudogene as opposed to biased towards the third codon in the functional genes and were higher in all three positions in the pseudogene than in the functional gene from which it was derived.
Attempts to produce a figure for human cSHMT pseudogene divergence were made using the method in Li et al., (1981) similar results were obtained in that the rate of nucleotide substitution was equally spread over the three codon positions and was faster than for the human cSHMT gene. However, the final calculation using rabbit cSHMT as the second functional gene sequence, gave a time since divergence of 24.25MY, a result higher than estimated using the other methods. It is noteworthy here that the ratio of third codon substitutions was higher when comparisons were made between the pseudogene and the rabbit cSHMT than between the human cSHMT and the rabbit cSHMT whereas in the example given by Li et al., (1981) this ratio was lower and this change would affect the outcome. However, the authors used two mammalian comparisons and averaged them out. This could not be done for cSHMT as rabbit and human cSHMTs are the only two mammalian DNA sequences known which span the entire coding sequence for cSHMT. Once other mammalian sequences are available the calculations of age can be repeated to establish whether this method of estimation of pseudogene age is useful for SHMT.

Some processed pseudogenes are still functional after creation and time since inactivation can be calculated in using the results from the nucleotide substitutions calculated above (Li et al., 1981). Function may be retained if, for instance, RNA polymerase III is responsible for the creation of some processed pseudogenes, the internal promoters are still present and mRNA transcripts can still be made. Under some circumstances where RNA polymerase III transcription was started before the RNA polymerase II promoter, the signals for transcription by RNA polymerase II are still present. It is possible, therefore, that if the human cSHMT pseudogene was generated by RNA polymerase III it was still transcribable by RNA polymerase III after its creation and remained functional. This transcription may have been at a low level and maybe the mutation occurred in the translation start consensus at +4 to
increase the efficiency of translation as a result. It was only later that other mutations including the mutated start codon accumulated which made the cSHMT pseudogene unfunctional.

The cSHMT pseudogene has been mapped to chromosome 1p33.2-33 by in situ hybridisation. (Byrne et al., 1996). cDNA encoding human cSHMT was also mapped using the same technique. In 12.5% of the spreads there was a signal at 1p32.2-33 and in 11.5% there was a signal at 17p11-12 when cSHMT cDNA was used as a probe. The cSHMT gene has previously been mapped to 17p11.2 using cSHMT cDNA (Garrow et al., 1993). However, these author did not detect a signal on chromosome 1.
CHAPTER 6
ISOLATION OF HUMAN GENOMIC SEQUENCES
ENCODING CYTOSOLIC SHMT
6.1 Summary

Several cDNAs encoding cSHMT had previously been isolated (Byrne, 1992; personal communication), several of which had cSHMT sequence adjacent to sequences with no homology to SHMT. At the junctions between the cSHMT and non cSHMT sequences there were motifs with good homology to intron splice junctions suggesting that the non cSHMT sequences were introns which had not yet been spliced from the mRNA used to synthesise the cDNA library. The locations of these potential introns are shown relative to cSHMT cDNA in figure 6.1. Three other potential intron splice junctions were identified by Xu, (1992) who isolated human cSHMT cDNAs encoding proteins shorter than the full length sequence by 39 and 80 amino acids, due to deletions near the C terminus (figure 6.1).

A previously undetected processed pseudogene encoding human cSHMT was characterised in chapter 5. However, there are no introns or RNA polymerase II promoter sequences within the pseudogene so no information on the potential regulation of cSHMT expression has been generated. The work in this section describes attempts to characterise the cSHMT gene. Several approaches were taken to isolate DNA fragments encoding cSHMT. The human PromoterFinder kit (Clontech, section 1.13) was used in conjunction with primers complementary to the 3' end of the human cSHMT to determine how much of the whole sequence could be isolated. Evidence from the isolation of the human mSHMT gene (Stover et al., 1995) indicated that the coding sequence spanned approximately 4.5kb of genomic DNA. Therefore, if the same was true for the cSHMT gene, much of the gene could be amplified in one PCR reaction if the reported sizes of products produced by the human PromoterFinder kit were accurate.

Once the products of this reaction were characterised and it was found that only sequence 3' of nt 1349 (appendix 1) had been obtained two other reactions were performed, one at the 5' end of the gene to try to isolate the promoter and one to determine the sequence of the
introns present in the alternatively spliced region of the cSHMT gene (introns 6, 7 and 8, figure 6.1). In order to determine the size and sequence of the intron in pUS1206 (intron 5, figure 6.1) identified by Byrne, (1992), this region of human genomic DNA was amplified by PCR using primers either side of the intron junction.

6.2 Results

6.2.1 Amplification of human cSHMT genomic DNA using sequences from the 3' end of cSHMT cDNA sequence

Oligonucleotides were designed which were complementary to nt 2300-2326 (gsp1) and nt 2041-2069 (gsp2) of human cSHMT cDNA (appendix 1). The oligonucleotides were used with the human PromoterFinder kit to amplify human genomic DNA. The results from the amplification of the five libraries are shown in figure 6.2.

Fragments from libraries 1, 2 and 5 were subcloned into pT7blue(R) (Novagen). After transforming DH5α cells to ampicillin resistance, white colonies were grown overnight and the plasmids were extracted by alkaline lysis. The 0.8kb product from library 1, the 1.5kb and 3.5kb products from library 2 and the 1.4kb product from library 5 were all successfully subcloned into pT7blue(R).

The primers gsp1 and gsp2 also have homology to the cSHMT pseudogene and, therefore, amplification of this DNA is also possible. The subcloned PCR products were sequenced with gsp2 to check whether the PCR products obtained contained the cSHMT gene. All except the 3.5kb product contained sequence identical to human cSHMT cDNA. Alignments were done with the mSHMT cDNA and the cSHMT pseudogene sequence to rule out the amplification of these sequences. The clones containing cSHMT sequences were then sequenced to establish the 5' end of the amplified products with primer ap2 from the human
Figure 6.1 Putative intron locations in the cSHMT gene derived from alternatively spliced and incompletely spliced cSHMT cDNAs.

1-5 were obtained from incompletely spliced cDNAs (Byrne, personal communication).

6-8 were obtained from the alternatively spliced cDNAs isolated by Xu, (1992).
Figure 6.2 Agarose gel showing results of PCR amplification of the human genomic libraries in the human PromoterFinder kit using nested primers gsp1 and gsp2 with the adaptor primers supplied.

Lane M  | λHindIII markers  
Lane 1-5 | libraries 1-5  
Lane 6 | negative control  
Lane 7 | positive control (3.9kb)

A 0.8kb band is present in lane 1 but only just visible on this figure.

A 1.5kb band is present in lane 3 but only just visible on this figure.
Figure 6.3 Sequence of the products obtained using primers gspl and gsp2 in conjunction with the human PromoterFinder kit. The whole sequence spans from the start of the genomic DNA from library 2 to the end of the gsp2 primer. The products from libraries 5 and 1 start at nt 87 and 670 respectively. Splice acceptor sites are highlighted in red, splice donor sites in blue and motifs matching the branch point consensus in green. Underlined are the bases between which the introns are removed from the exons. The sequence of the gsp2 primer is italicised.
PromoterFinder kit, and with internal primers INT1 and INT2. The consensus sequence of the PCR fragments is shown in figure 6.3. Two sections of the sequence (nt 1-373, and nt 485-849, figure 6.3) had no homology to cSHMT cDNA. The sequence nt 1-373 had a splice acceptor site at the 3' end and the sequence between nt 485-849 had splice donor and acceptor sites (Ohshima and Gotoh, 1987). It follows that these sequences are introns. The 5' end of the first intron was not identified but is at least 373bp in length. The second intron was 364bp in length. Also included in the sequence were cSHMT exons between nt 375-484 and from nt 850 to the end of the sequence which corresponded to cSHMT cDNA sequence (appendix 1) from nt 1350-1459 and from nt 1460 to the end of the gsp2 primer. The sequence of the exons isolated was 100% homologous to the cSHMT cDNA sequence. The intron-exon boundaries fell between nt 1349-1350 and nt 1459-1460 of cSHMT cDNA sequence (appendix 1). A search of the Genbank database with the sequence obtained from the 3.5kb product from library 2 with gsp2 as a primer revealed no homologies and no further work was performed with this clone.

6.2.2 Amplification of human cSHMT genomic DNA from the 5' end of cSHMT cDNA towards promoter region

Amplification of the cSHMT gene indicated that the whole genomic sequence could not be amplified from the 3' end. In addition evidence from Elsea et al., (1995) indicated that the cSHMT gene might cover up to 25kb of DNA. Work, therefore, focused on amplifying specific regions of the gene. In order to amplify upstream sequences from known cSHMT sequence, primers complementary to human cSHMT cDNA from nt 237-260 (pfk1) and nt 45-70 (pfk2) were used in conjunction with the human PromoterFinder kit. After the first round of amplification, the expected smears were seen in all libraries. However, after the second round of amplification, there were no discreet bands present. An additional primer complementary to nt 138-62 (pfk3) was synthesised and used on the first round amplification
products of pfk1. Again no discreet bands were produced after the second round of amplification.

In another attempt two further primers complementary to human cSHMT cDNA were synthesised namely kjc22 (nt 341-363) and kjc23 (nt 313-332) (appendix 1). The primary amplification with kjc22 yielded smears when the products from the libraries were visualised on an agarose gel. Secondary amplification with kjc23 yielded two products in each of libraries 1, 2 and 4 (figure 6.4). The largest product was approximately 1kb in library 4. This 1kb band was subcloned into pT7blue(R) and sequenced with kjc23, ap2 and internal primers 309int- and 22/234B6P. The sequence of the 1kb fragment is shown in figure 6.5.

The sequence obtained was compared to cSHMT DNA. Nucleotides 1-1091 (figure 6.5) showed no homology to cSHMT cDNA but had a splice acceptor site. Sequence from nt 1092-1159 was identical to human cSHMT cDNA (appendix 1) from nt 274-332. The intron identified is at least 1091bp in length and the 5' end of the intron has not been found.

6.2.3 Amplification of human cSHMT genomic DNA to amplify across the intron junction boundary at nt 1231

The variant cSHMT cDNAs identified by Xu, (1992) coded for alternatively spliced cSHMT proteins, identifying 3 putative intron splice sites between nt 991-992, nt 1108-1109 and nt 1231-1232 in cSHMT cDNA sequence (appendix 1). The sequences of any introns sequences adjacent to these potential splice sites had not been determined.

One of the primers previously synthesised for sequencing cSHMT cDNA (kjc8) was complementary to nt 1262-1281 cSHMT cDNA (appendix 1). Another primer, namely kjc21 complementary to nt 1238-1256 was synthesised and kjc8 and kjc21 were used with the human PromoterFinder kit to amplify genomic DNA upstream of these two primers.
Figure 6.4  Agarose gel of the PCR amplification of human genomic libraries 1 to 5 of the human PromoterFinder kit using primers kjc22 as the first primer and kjc23 as the nested primer.

Lane M  \( \lambda \text{HindIII} \) markers

Lanes 1-5 libraries 1-5

Lane 6 negative control

Lane 7 positive control (1.8kb)
Figure 6.5 Sequence of the Lkb product produced by amplification of library 4 of the human PromoterFinder kit with primers kjc22 and kjc23. In italics is the primer ap2 (at 5’ end) and kjc23 (at 3’ end). The sequence of the genomic DNA starts at nt 47. Splice acceptor motif is shown in red. Underlined bases indicates the intron-exon boundary.
Figure 6.6  Agarose gel of amplification of the human genomic libraries in the human PromoterFinder kit libraries with primers kjc8 and kjc21.

Lane M  \(\lambda HindIII\) markers

Lanes 1-5  libraries 1-5

Lane 6  negative control

Lane 7  positive control (0.9kb)
putative branch point consensus in green. represents gaps of unknown length in the primers kjc8 and kjc21. The whole sequence corresponds to the 2.6kb product from library 3, Figure 6.7 Sequence of the products isolated from the using human PromoterFinder kit with (2609-2627) is indicated.

Underlined bases indicate the intron-exon cleavage site. Sequence of primer kjc21 and 4 begin at nt 1154 and 1915 respectively. The splice acceptor site is shown in red and the starting from the end of the human genomic DNA fragment. The products from libraries 2

```
1  AAAAGAGCACA TACGAAAAAG A TACA TAGAA TA TGAGTC A TTTA TAGA TC
51  A TAAAGA GTGT AAGCAAAAT TTA AAAAA T TTTAAGA AGAATCGT A TCTCTAGA
101  A TACCTGAGCT TGAGTAC A TAA TTAGA AGTTGTCAGCT GGCACAGTT
151  A AACAAATAC TGAAGAC TCTTTA TTTAACAGA TAGAGTGGG AAGTGAATAA
201  A CAAATTGCC TTTCTCCTAT TAGAAT TTAGGAC CACCTTACAG CATEGTAGA
251  A TCCAGCTTCT CTTAACATGA ACAATTGCA ACATTAAATG TCTACCTTCT
301  A TTAGTGTAAT GTGCTCCTAG TAGAAGGAA TTAGTCACAT TTTCAATTT
351  A TTTAGTTACT GTGCTCCTAG TAAAGAGAAT TTAGACTCAT TTTCTTGTGT
401  A TCTATACACT CTCTACATA TGA AAACCCG A CACACCTGG TAGGGGACAC
451  A CAGAGCTCTG TGCTGAGTCC T GAGTACTGT CAGGGGAGAT GCTCCTCTCC
501  A ACCCGGCTGA CGGGGGGCTTT CCCCTCCCAG TTTCTCCTGA
551  A ATTCCTGGGG AGTTGGGTCTT CTTGTGCTAG TAC TCTCTGCAC
601  A TCTGAGCTGG CAGAGCTCTCC GAGGGGGGAT CAGGGGACCTT
651  A GCAAATTTTTG GAGGAGGGAC ATTACTAGTGC TCTCCAGGAG AGGGGCACTT
701  A TATCCTGTAG CCCAAAGGAA TGGCGAGTTGT GCTCAGATTG ACTCCTCCCT
751  A GAA ATAGGAG TGGCTGCTTG TGAAGACTG CTGGCCAGAG AGGGGCGATT
801  A TGCTCTTTGC CTTTATCGAG GTATTACGAC CCTCTGACTC AGAATTGCTG
851  A GAGGGGTATAC AGGCTCCTCTGG TTTAAGAGGA TGGGGGATCC ACCAACTCTC
901  A GCTGAATGGGG CGAGTCCCTGG CAGGGGAGTT CAGGGGACAC
951  A GACATCCAGCT GCTTCTCCCT CATGTTGTGA ATGCTTGACT TTTTTTTTCC
1001  A GTTGAGGATTT GACCTCAAAA CACTCCCCCG ATGACCCTCG TGAAAACACT
1051  A GTCTCCTACA TAAAGGCTTC GTGTTGCCC A AAGGGTGGGG GACCACTGCT
1101  A GCCACCTCTTT GGTGTTGTAGT GGTTTAGCCC ATGCTTGACT TTTTTTTTCC
1151  A ACCTGTTGTGG TGGCAGCTTT GAGGACACTG GAGGGACAGT
1201  A CTGCGGCTAG AGCTTGGGCA CTTGAGTCTG GGGGCGCTTC AGATGCTGGT
1251  A CAGCTTTAGA TTCTCAGAGG AGCTGAGAAC GGTGTTGAC TGCAAATGTT
1301  A AGGGAATCTAG GTGGGCTTCTT GACAATTTCA GACAAACTCT
1351  A TTTTCTACAT GTGAGGCTTG CTTGTGCTAG TAC TCTCTGCAC
1401  A GAGGGGCTAA TGCTTTCTTG GAGGGGGTTT CAC TCTCTCTCTC
1451  A GCCAAGCTCT GGCACTCGGG C GCACTCGGG C AATGGACAAT
1501  A TGGAATGCTT GGTTTTTTTT GCCCTTCCCC ACCAGCAGCT
1551  A CTAGACCTGG AGCAGTTCTT TTTTATTTTT TTTTTTTTTTT
1601  A GCTTTTTTTT CTCTCTCTCT TCACATCTCT GTGCTAGGAG GACCACTGCT
1651  A GAGGAGGAGG C TGAAGGCTTC GGGGCGCTTC AGATGCTGGT
1701  A TTGAGGATTT GACCTCAAAA CACTCCCCCG ATGACCCTCG TGAAAACACT
1751  A AAAGCACTCTG TCGAGTGCTG GTGCCAGACT GAGGGCAGCT
1801  A TAAACCTTCTCTG TCAAGAGGAA TGGGGGATCC ACCAACTCTC
1851  A TGCGAGAGCT ATGGGCAAGG GCAGAGGAGG GGGGCGCTTC AGATGCTGGT
1901  A CAGTCTGGGG GCAGGCGCTG GGGGCGCTTC AGATGCTGGT
1951  A TCCACCCGCA TCTCTGGAAAG ATCCGCAGCC ACCAGCAGCT
2001  A CCCCAGCTCT GCCAGCAGCT GCCCTCCTCTG TAAAGAGAGA
2051  A AGAAGGATTT AGTCTTCTGT ATGACCTGGA GCACTCGGG C AATGGACAAT
2101  A TTAAAGGCTTA TGCTGCTAGC TAAACTATCA TTAATACACA TCGCCCTCAC
2151  A AGGATCCGGG AAAAAACCTAC GACAGGGCTC AGTGTCTGCG ACCAAATGGT
2201  A CTTGATGGAAC ATGCGCTTTA AACCCCAAGA TAAATAGTAC TCGCTCCAA
2251  A ATCCATGCTG CTTCAAGGG GAGCGTCTTC CTAAGAGGTT AGGGGATTCG
2301  A CCAAACACAG TGGCAGAAG AACAACGAGG GAGAGGCTTC AGAAGGCTGT
2351  A CCCTCCCTCTC TGGAGGAGTC AGGCGCTTCT CTTGTAAGAC ACCAACTCTC
2401  A TTCCAGAATGT CTTCTGCTCTT CTTTTCACTAC TTTCTGAGGC GGCACTTTTT
2451  A TGGGAGGTTT AGAAGAGCTG GCACTCTGCT GGGGGCTCTC ACCAACTCTC
2501  A CCTGAGGTTC CTGAGGCTCTC TTAAGACGCA TCTGAGGATC
2551  A CTTGACACA TTAGAGTGG CTGAGATTCT A CTAAGCAAT TTGCTGCTAG
2601  A AGTGTCTTCT GCAAGACCTTT TCTGAAGAC

kjc 21
```

Figure 6.7 Sequence of the products isolated from the using human PromoterFinder kit with primers kjc8 and kjc21. The whole sequence corresponds to the 2.6kb product from library 3, starting from the end of the human genomic DNA fragment. The products from libraries 2 and 4 begin at nt 1154 and 1915 respectively. The splice acceptor site is shown in red and the putative branch point consensus in green. "." represents gaps of unknown length in the sequence. Underlined bases indicate the intron-exon cleavage site. Sequence of primer kjc21 (2609-2627) is indicated.
The products from the 5 libraries were visualised by agarose gel electrophoresis and staining with ethidium bromide. As figure 6.6 shows, there are products present in libraries 2, 3, 4 and 5 after the second round of amplification. The DNA in libraries 2, 3 and 4 was purified from the mineral oil used in the PCR reaction and ligated into pT7blue(R) (Novagen). After using the ligation mixed to transform DH5α cells to ampicillin resistance, white colonies were grown in broth and plasmids extracted by alkaline lysis. A 700bp product from library 4 was isolated along with a 1.5kb product from library 2 and an approximately 2.5kb product from library 3. The products were sequenced initially with kjc21 to check for homology to cSHMT cDNA and to rule out presence of the cSHMT pseudogene. Further sequencing was performed with primer ap2 and internal primers, INT3, 8/2124N, and 8/2124P. Sequencing with these primers left two small gaps in the sequence estimated to be no more than 100bp. The sequence is shown in figure 6.7 was compared to cSHMT cDNA. Sequence from nt 1-2603 of the clone showed no homology to cSHMT DNA. Sequence from nt 2603-2627 (figure 6.7) was identical to cSHMT sequence between nt 1232-1256 (appendix 1). At the junction between non-homologous and homologous sequence there is a motif matching the consensus for a splice acceptor (Ohshima and Gotoh, 1987). There 5' end of the intron is unknown but the intron is at least 2062bp long.

6.2.4 Amplification of the intron present between nt 696-697 of cSHMT cDNA

A human cDNA (pUS1206) had previously been isolated containing 850bp of sequence with no homology to cSHMT cDNA followed by sequence from nt 697 of cSHMT (appendix 1) (Byrne, 1992). At the junction between the two sequences was a consensus sequence for a splice acceptor site. Amplification of human genomic DNA (Promega) was performed using Expand Long Template DNA polymerase (Boehringer Mannheim) using primer pairs kjc9 and kjc16 or kjc20 and kjc16 which had previously been synthesised for sequencing cSHMT.
cDNA. These primer pairs were situated in the cDNA either side of the site where the intron was inserted between nt 696-697 (appendix 1). Using PCR reaction buffer 2 several bands were produced for each PCR reaction (figure 6.8). Amplification with kje9 and kje16 yielded a strong band at 267bp corresponding to amplification of the pseudogene. Other discreet bands ranged from 1.8kb to approximately 9kb. Amplification of human genomic DNA with kje20 and kje16 yielded bands between 2kb and approximately 9kb (figure 6.8). A fragment corresponding to the pseudogene was not observed which was not unexpected as this product would have only been 75bp. DNA from the amplification reactions was transferred to a nylon membrane by Southern blotting and hybridised with the DIG labeled intron sequence from pUS1206. It was found that the DNA hybridising to the intron probe corresponded to 8.0kb for kje9/kje16 and 6.8kb for kje20/kje16 primer pairs respectively (figure 6.8). Further PCR amplification of human genomic DNA using primers kje9 and kje16 and reaction buffer 3 yielded a single 8kb band. Initial attempts to subclone the 8kb band into pT7blue(R) failed and no sequence data is available.

6.3 Discussion

All of the intron location data obtained in these experiments was added to figure 6.1 to form figure 6.9.

Sequences which had homology to human cSHMT cDNA were amplified by PCR using the human PromoterFinder kit. The two introns isolated by amplification with gsp1 and gsp2 (introns 9 and 10, figure 6.9) had not been previously identified by isolation of alternatively or incompletely spliced cDNAs. However, a PCR reaction performed using primers derived from the cSHMT cDNA sequence yielded a 660bp product instead of the expected 270bp (Elsea et al., 1995). Analysis of the PCR primers used by these researchers indicates that the primers were located either side of the second of the introns identified (intron 10).
Figure 6.8 Agarose gel of PCR products from human genomic DNA amplified using kjc9/kjc16 and kjc20/kjc16 primer pairs. A second gel was run with these samples and DIG labelled DNA size markers, blotted and hybridised with a DIG labeled probe containing human cSHMT intron sequence from pUS1206.

Lane M \( \lambda \text{HindIII markers} \)

Lane L \( \phi X174 \text{HaeIII markers} \)

Lane 1 amplification with primers kjc9 and kjc16

Lane 2 amplification with primers kjc20 and kjc16
Figure 6.9 Illustration of the sequences isolated in this chapter in relation to human cSHMT cDNA and the putative intron boundaries illustrated in figure 6.1. Highlighted in red are the intron sites confirmed in this study.

Intron 2a was isolated using primers kj c22 and kj c23 with the human PromotrFinder kit.

Intron 8 was isolated using primers kj c8 and kj c21 with the human PromoterFinder kit.

Introns 9 and 10 were isolated using primers gsp1 and gsp2 with the human PromoterFinder kit.

Intron 5 was identified by amplification of human genomic DNA with kj c20 and kj c16.
Figure 6.10 Illustration of the human genomic DNA sequences isolated compared to human cSHMT cDNA. Indicated in black are the introns identified and sequenced. Marked on the cDNA sequence are the positions of the intron boundaries confirmed in this study.
Approximately 170bp of cDNA sequence at the 3' end of the cSHMT cDNA downstream of gsp2 remains uncloned.

The amplification of the libraries in the human PromoterFinder kit using primers kjc22 and kjc23 did not result in the isolation of a promoter sequence. Another intron was identified (intron 2a, figure 6.9) which was not predicted by the cSHMT cDNAs isolated to date. In order to isolate the promoter, further PCR reactions need to be performed, perhaps using sequence from the 5' end of the intron identified here as a template.

Other attempts to isolate the promoter region using primers pfk1 to pfk3 indicated the limitations of the nested PCR method used. From the estimation of the length of the human cSHMT gene spanning 25kb (Elsea et al., 1995) and evidence from the introns isolated here, it is possible that some of the introns within the human cSHMT gene are several kilobases long. It is also possible, therefore that there are restriction sites in these introns that match those used to generate the libraries provided in the human PromoterFinder kit. If there is a long intron between the sites of the nested gene specific primers then the first PCR reaction product may not be of sufficient length to include the binding site for the primer used in the second round of amplification and no products will be seen. It is proposed that this is the reason that the reactions using primers pfk1 to pfk3 failed.

Using primers kjc8 and kjc21 yielded products in the libraries of the human PromoterFinder kit. Sequencing of these products showed that the intron site (intron 8) predicted by the alternatively spliced cSHMT cDNAs isolated by Xu, (1992) was genuine. The length of the intron meant analysis of further splice sites predicted from the alternatively spliced cSHMT cDNAs was not possible in this particular experiment. Only one strand of DNA has been sequenced in all the human PromoterFinder kit isolates. In order to confirm the sequence is genuine, the other strand should be sequenced.
PCR amplification using primers kjc20 and kjc16 followed by Southern blotting produced a band of 6.8kb which hybridised to the intron sequence isolated by Byrne, (1992). If this primer pair was used on the cSHMT cDNA, a product of 75bp would be produced. It follows that the intron previously identified is approximately 6.7kb in length. The primer kjc9 is a further 193bp upstream of kjc20 but the differences in size between the products produced in PCR reactions using kjc9/kjc16 and kjc20/kjc16 is nearer to 1.2kb in size. Therefore, it is likely there is at least one more intron present relative to the cSHMT cDNA sequence between nt 483 (the start of primer kjc9) and nt 675 (appendix 1). The potential location of one intron (intron 4) were identified from incompletely spliced cSHMT cDNAs (Byrne, personal communication) in this region of cDNA (figure 6.9). Further work needs to be performed to characterise the PCR products obtained and to determine whether intron 4 has also been isolated.

The subcloning of PCR products into the pT7blue(R) vector relies on the fact that a single overhanging adenine nucleotide is added to each end of the PCR product by the DNA polymerase. In fact the overhanging nucleotide is not always an adenine and the nucleotide added depends on the preceding nucleotide (Hu, 1993). Sanchez et al., (1996) demonstrated with primers beginning with one of the four nucleotides that the success of cloning into vectors with T overhangs depended on the base at the 5' end of the PCR primers. The descending order of efficiency was A,T,G,C. Primer kjc9 starts with a G and kjc20 starts with a T. Therefore, the overhanging nucleotides added to this PCR product may not be adenine and pT7blue(R) may not be suitable for the insertion of this particular PCR product. An alternative strategy for cloning the PCR product would be to remove the overhanging nucleotides and ligate into SmaI digested pUC18.

All sequences determined here were used to search the Genbank and EMBL databases against other human DNA sequences. No significant homologies were found other than to
human cSHMT. The sequences were also searched for direct and inverted repeats. A direct repeat where 29 out of 34 basepairs matched was found in the product isolated using kjc8 and kjc21 (figure 6.7) between nt 868-901 and 902-935. This direct repeat is present in the intron (intron 8) which is present 3' to two exons which may be alternatively spliced. An intron at the same site has been identified in the human mSHMT gene (Stover et al., 1995). The intron in the mSHMT gene is only 87bp and there is no evidence to date for alternative splicing in mSHMT at this point. Therefore, it is possible that the length of this intron in the cSHMT gene and repeat sequences within it may play a role in the selection of alternatively spliced exons. No repeat sequences were found in the other introns isolated.

Holland and Blake, (1987) proposed that introns are ancient pieces of DNA which enabled the shuffling of exons to form functional proteins. The introns were then lost in prokaryotes and in eukaryotes can be retained, lost, or the splice sites may drift, due to mutation. Craik et al., (1982) postulated that the drifting of intron boundaries may have evolutionary advantages in that it may generate functional differences in the same protein family. The gene encoding human mSHMT has been published (Stover et al., 1995). Comparing the mSHMT gene to the known and potential intron splice sites in the cSHMT gene, many of the splice sites either match exactly or are within 25bp of each other. As yet some of the mSHMT splice sites have no known equivalents in the cSHMT gene as some areas of the latter have still to be characterised. By comparing the introns found in this study with the mSHMT gene it is found that the introns at nt 1231-2 of cSHMT cDNA (intron 8) is matched in the mSHMT gene but the intron found at nt 273-4, 1348-1349 and 1459-60 (appendix 1) (introns 2a, 9 and 10) do not have equivalents in the mSHMT gene. Conversely, there is an intron boundary in the 3' end of the coding sequence of the mSHMT gene which is not present in the cSHMT gene. The intron boundaries in the N.crassa gene (McClung et al., 1992) are not conserved in the human mSHMT gene. There is not enough information to say whether this is also the case.
for the cSHMT gene. The drifting of some intron junctions when comparing cSHMT and mSHMT genes may account for the functional differences between the two proteins which lead to the mSHMT protein having a higher affinity for glycine than the cSHMT isozyme (Schirch and Peterson, 1980). Only determination of the tertiary and quaternary protein structures of the SHMT isozymes and the full characterisation of the cSHMT gene will determine whether this hypothesis is correct.

In conclusion, the work described here has gone someway to determining the structure of the SHMT gene. Some intron boundaries suggested by incompletely spliced and alternatively spliced cSHMT cDNAs have been confirmed as genuine. The complete sequence of the cSHMT gene still has to be determined.
CHAPTER 7

CONCLUDING REMARKS AND FURTHER STUDIES
7.1 Concluding remarks and further study

The expression of human and rabbit cSHMT in *E.coli* described here has made it possible to rapidly generate large quantities of highly pure cSHMT. Work is in progress at the Institute of Cancer Research to determine the structure of human cSHMT by X-ray crystallography and to screen folate analogues for cSHMT inhibition. An alternative to using protein crystals to generate a tertiary and quaternary structure is to use NMR. The addition of label to the protein for NMR analysis is much simpler and cheaper when the protein is expressed in *E.coli* as only one labeled nutrient has to be added to the medium instead of many labeled amino acids that need to be added to animal cell cultures. Therefore, the expression system described here could be used for NMR studies as well.

There are two potential phosphorylation sites (at amino acids 38-41 and 158-161, appendix 1). Now that large quantities of pure human cSHMT are available, studies can be performed to see whether these sites can be phosphorylated and the role that any phosphorylation may play on the SHMT activity. Site directed mutagenesis of the potential phosphorylation sites may also aid these studies.

Some residues have been identified by chemical modification as being important for SHMT activity, although in some cases a more likely cause of inhibition was steric hindrance preventing substrates binding. Work performed here using site directed mutagenesis shows that this is the case for iodoacetate inactivation of cSHMT activity as performing the mutation C204A did not significantly alter SHMT activity.

Now that site directed mutagenesis has been shown to be a valuable tool for determining whether amino acids implicated in SHMT activity by chemical modification really are important, other potential sites for site directed mutagenesis would be other amino acids implicated in SHMT activity by chemical modification. For example, R270 is implicated in
THF binding (Usha et al., 1992). Another potential area of research would be the site directed mutagenesis of amino acids conserved in all SHMT protein sequences (appendix 2). The structures of mutants created by site directed mutagenesis could also be ascertained to help determine the residues required for substrate binding and catalysis.

A pseudogene, previously undetected by in situ hybridisation (Garrow et al., 1993) has been characterised and found to be located on chromosome 1p32-33. Several other genes have been identified within this region of chromosome 1 which may contribute to breast cancer (J. Shipley, personal communication). Further study of this region of chromosome 1 is required to isolate the genes within this area of the genome.

The lowest estimated age of the cSHMT pseudogene is 7.7MY so it would be interesting to see whether the pseudogene is also present in monkeys as the divergence of monkeys and humans occurred after this date.

Human genomic DNA has been isolated which encodes cSHMT gene. Further work needs to be performed to further characterise the gene and its promoter region. Probing of a human genomic DNA library with human cSHMT cDNA led to the isolation of only the pseudogene. As the pseudogene does not contain any introns, then a combination of cSHMT intron and exon sequences could be used to isolate the cSHMT gene from a genomic library in bacteriophage lambda or in cosmids. As the cSHMT gene fragments isolated here contain both intron and exon sequence, these sequences could be used as the probes.

Alternatively, as the cSHMT gene is predicted to be 25kb (Elsea et al., 1995), the promoter alone could be isolated using the human PromoterFinder kit and used in conjunction with reporter genes such as Green Fluorescent Protein, which allows the real time study of expression in cells, to determine how the cSHMT gene is regulated. It has recently been suggested that cSHMT may be regulated by retinoic acid (Nakshatri et al., 1996). It would be
interesting to determine whether there are any potential binding sites for retinoic acid dependent transcription factors present in the cSHMT gene promoter.

Isolation of the complete SHMT sequence would also allow the construction of a SHMT minigene. This minigene would consist of all the exons and the 5’ and 3’ sequences. Introns can then be added to determine whether any particular introns affect transcription. This technique was used on the human thymidylate synthase gene by Deng et al., (1989) and several introns were found to affect transcription. It was also found that intron 1 of the thymidylate synthase gene was required for cell cycle dependent expression (Takayanagi et al., 1992).

Mitochondrial SHMT genomic and cDNA sequences have been identified (Garrow et al., 1993; Stover et al., 1995, Whitehouse, 1996). These sequences could be used in conjunction with cSHMT sequences to probe for SHMT RNA in tissues by in situ hybridisation. This would be of particular use in brain tissue to determine the potential role of SHMT in schizophrenia. Further information will arise from the development and characterisation of mice genetically engineered for transgenic expression or null expression of different SHMT isozymes (Thomas and Capecchi, 1987, Gu et al., 1994).

It is critical for understanding the diseases in which SHMT has been implicated to determine the contribution to human physiology of the individual SHMT isozymes and their regulation. The work presented in this thesis has increased knowledge of the gene and protein structures of cSHMT and the foundations have been laid to enable characterisation of the precise role that SHMT plays in mammalian metabolism.
CHAPTER 8

REFERENCES


Du, Z. EMBL/Genbank/DDBJ Accession number P50432.


Hopkins, S. and Schirch, V. (1986). Properties of a serine hydroxymethyltransferase in which an active site histidine has been changed to an asparagine by site-directed mutagenesis. Journal of Biological Chemistry 261, 3363-3369.


189


Additional references


Appendix 1  DNA and Protein sequence of human cytosolic SHMT cDNA. Highlighted in red is the asparagine residue changed to aspartic acid by the mutation generated in the PCR of the cSHMT orf. Highlighted in green are the amino acids changed by site directed mutagenesis. Underlined are the nucleotides between which introns have been isolated in the human cSHMT gene.
<table>
<thead>
<tr>
<th>Organism</th>
<th>SwissProt accession number</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurospora crassa cytosolic</td>
<td>P34898</td>
<td>McClung et al., (1992)</td>
</tr>
<tr>
<td>Saccharomyces cerevisae cytosolic</td>
<td>P37291</td>
<td>McNeil et al., (1994)</td>
</tr>
<tr>
<td>Schizosaccharomyces pombe cytosolic</td>
<td>Q10104</td>
<td>Connor et al., (1995)</td>
</tr>
<tr>
<td>Flaveria pringlei mitochondrial</td>
<td>P49357</td>
<td>Kopriva and Bauwe, (1993)</td>
</tr>
<tr>
<td>Flaveria pringlei mitochondrial</td>
<td>P49358</td>
<td>Kopriva and Bauwe, (1993)</td>
</tr>
<tr>
<td>Pisum sativum mitochondrial</td>
<td>P34899</td>
<td>Turner et al., (1992)</td>
</tr>
<tr>
<td>Solanum tuberosum mitochondrial</td>
<td>P50433</td>
<td>Kopriva and Bauwe, (1995)</td>
</tr>
<tr>
<td>Homo sapiens mitochondrial</td>
<td>P34897</td>
<td>Garrow et al., (1993)</td>
</tr>
<tr>
<td>Oryctolagus cuniculus mitochondrial</td>
<td>P14519</td>
<td>Martini et al., (1989)</td>
</tr>
<tr>
<td>Homo sapiens cytosolic</td>
<td>P34896</td>
<td>Garrow et al., (1993)</td>
</tr>
<tr>
<td>Mus musculus cytosolic</td>
<td>P50431</td>
<td>Nackshatri et al., (1996)</td>
</tr>
<tr>
<td>Oryctolagus cuniculus cytosolic</td>
<td>P07511</td>
<td>Martini et al., (1987)</td>
</tr>
<tr>
<td>Ovis aries cytosolic</td>
<td>P35623</td>
<td>Usha et al., (1994)</td>
</tr>
<tr>
<td>Saccharomyces cerevisae mitochondrial</td>
<td>P37292</td>
<td>McNeil et al., (1994)</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>P00477</td>
<td>Plamann et al., (1983)</td>
</tr>
<tr>
<td>Salmonella typhimurium</td>
<td>P06192</td>
<td>Steiert et al., (1990b)</td>
</tr>
<tr>
<td>Actinobacillus actinomycetemcomitans</td>
<td>P34894</td>
<td>Brogan and Demuth, (1993)</td>
</tr>
<tr>
<td>Bradyrhizobium japonicum</td>
<td>P24060</td>
<td>Rossbach and Hennecke, (1991)</td>
</tr>
<tr>
<td>Methylbacterium extorquens</td>
<td>P50435</td>
<td>Chistoserdova and Lidstrom, (1994)</td>
</tr>
<tr>
<td>Hyphomicrobium methyllovorum</td>
<td>P34895</td>
<td>Miyata et al., (1993)</td>
</tr>
<tr>
<td>Bacillus subtilis</td>
<td>P39148</td>
<td>Glaser and Danchin, (1994)</td>
</tr>
<tr>
<td>Campylobacter jejuni</td>
<td>P24531</td>
<td>Chan and Bingham, (1988)</td>
</tr>
<tr>
<td>Mycoplasma genitalium</td>
<td>P47634</td>
<td>Fraser et al., (1995)</td>
</tr>
<tr>
<td>Corynebacterium sp</td>
<td>P50434</td>
<td>Chlumsky et al., (1995)</td>
</tr>
<tr>
<td>Methanobacterium thermoautotrophicum</td>
<td>P50436</td>
<td>Vaupel et al., (1996)</td>
</tr>
</tbody>
</table>

Appendix 2 Protein sequences for SHMT proteins in the SWISSPROT database. These protein sequences were used to produce the alignment overleaf using PILEUP and PRETTY from the GCG suite of programs (Genetics Computer Group). In the alignment residues marked with a * are conserved in all SHMT protein sequences. † are W15, W111, N182 and C204 residues mentioned in the text.
MATERIAL REDACTED AT REQUEST OF UNIVERSITY