

# Serine Hydroxymethyltransferase Is Maternally Essential in *Caenorhabditis elegans*\*

(Received for publication, November 3, 1997, and in revised form, December 23, 1997)

Greg P. Vatcher‡, Colin M. Thacker§, Titus Kaletta¶, Heinke Schnabel¶, Ralf Schnabel¶, and David L. Baillie‡\*\*

From the ‡Institute of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada, the §Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada, the ¶Max-Planck-Institut für Biochemie, Martinsried D-82152, Federal Republic of Germany, and the ¶Institut für Genetik, Braunschweig D-38023, Federal Republic of Germany

The *mel-32* gene in the free living soil nematode *Caenorhabditis elegans* encodes a serine hydroxymethyltransferase (SHMT) isoform. Seventeen ethylmethanesulfonate (EMS)-induced mutant alleles of *mel-32*(SHMT) have been generated, each of which causes a recessive maternal effect lethal phenotype. Animals homozygous for the SHMT mutations have no observable mutant phenotype, but their offspring display an embryonic lethal phenotype. The Mel-32 phenotype has been rescued with a transgenic array containing only *mel-32*(SHMT) genomic DNA. Heteroduplex analysis of the 17 alleles allowed 14 of the mutations to be positioned to small regions. Subsequent sequence analysis has shown that 16 of the alleles alter highly conserved amino acids, while one allele introduces a stop codon that truncates two thirds of the predicted protein. *mel-32*(SHMT) has a 55–60% identity at the amino acid level with both isoforms of SHMT found in yeast and humans and a 50% identity with the *Escherichia coli* isoform. The *C. elegans mel-32* mutation represents the first case where SHMT has been shown to be an essential gene.

Serine hydroxymethyltransferase (SHMT)<sup>1</sup> is a highly conserved, ubiquitous, pyridoxal 5'-phosphate (PLP)-containing enzyme that catalyzes the reversible conversion of serine and tetrahydrofolate to glycine and 5,10-methylenetetrahydrofolate (1). The  $\beta$ -carbon of serine is the major source of one-carbon units in the one-carbon metabolic pool (2). The methyl group passed on through the folate cofactor of SHMT is used in thymidylate, methionine, lipid, and purine biosynthesis. SHMT also catalyzes many secondary reactions, such as amino acid transaminations (for a full review of the reactions catalyzed by SHMT, see Ref. 1).

Eukaryotic organisms have two nuclear genes that encode two distinct isoforms of this enzyme, a cytosolic form and a mitochondrial form. Recent evidence suggests that there is a subcellular partitioning of the reactions catalyzed by SHMT, with the serine to glycine conversion occurring in the mitochondria, and the glycine to serine conversion occurring in the cytoplasm (3). A comparison of the determined and predicted

amino acid sequences of many SHMTs reveals a striking degree of homology between all the known forms of this enzyme (2, 4, 5). These alignments show that the amino and carboxyl termini of SHMT are less conserved but that the middle two thirds of the sequence have long stretches of very high identity.

Serine hydroxymethyltransferase levels are elevated in rapidly proliferating cell lines and tumors (6, 7). When lymphocytes are treated with a mitogenic stimulus, the enzymatic activity of SHMT is increased and the incorporation of the  $\beta$ -carbon of serine into DNA is increased (6). SHMT activity is also increased during the S phase of the cell cycle, suggesting that a product of SHMT activity is utilized during cell division. Conversely, when cells stop proliferating, SHMT levels decrease. Retinoic acid treatment of P19 embryonal carcinoma cells stops proliferation and stimulates differentiation. Furthermore, it has been shown recently that retinoic acid causes a 50% decrease in SHMT transcript levels (7). Based on these findings, serine hydroxymethyltransferase has been proposed as a potential chemotherapy target (8). A chemical or drug that decreases SHMT activity may cause rapidly proliferating tumor cells to quiesce.

Prior to this study, the only observable phenotype of SHMT deficiency was glycine auxotrophy. In *Escherichia coli*, when the single copy of SHMT (*glyA*) is mutated, glycine auxotrophy results (9). In the yeast *Saccharomyces cerevisiae*, glycine auxotrophy is observed only when both forms of SHMT and a third gene, *glyA*, are mutated (4) suggesting a functional redundancy in the glycine synthesis pathway. In addition, a line of Chinese hamster ovary cells that lack mitochondrial SHMT activity are glycine auxotrophs (3, 10). This deficiency is rescued when the cells are transfected with a cDNA encoding the human mitochondrial SHMT (11).

This report presents the cloning and characterization of a *Caenorhabditis elegans* SHMT homolog called *mel-32*. Mutations in *mel-32* result in a maternal effect lethal (Mel) phenotype. The *C. elegans mel-32* mutation is the first reported case of a serine hydroxymethyltransferase deficiency causing lethality.

## EXPERIMENTAL PROCEDURES

**Growth and Handling of *C. elegans***—All strains of *C. elegans* were grown at 20 °C on nematode growth medium plates streaked with *E. coli* OP50 as a food source. Standard genetic manipulation followed previously described protocols (12).

**Computer Analysis**—The analysis of sequence data, sequence comparisons, and data base searches were performed with ACeDB (a *C. elegans* data base) (13),<sup>2</sup> the BLAST<sup>3</sup> (provided by the NCBI server) and FASTA programs (14, 15), CLUSTALW (16), and MacDNASIS Pro

\* This work was supported by a Natural Sciences and Engineering Research Council of Canada grant (to D. L. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

\*\* To whom correspondence should be addressed. Tel./Fax: 604-291-4597; E-mail: dbaillie@darwin.mbb.sfu.ca.

<sup>1</sup> The abbreviations used are: SHMT, serine hydroxymethyltransferase; PLP, pyridoxal 5'-phosphate; EMS, ethylmethanesulfonate; *mel*, maternal effect lethal; bp, base pair(s); PCR, polymerase chain reaction.

<sup>2</sup> ACeDB documentation, code, and data are available from anonymous FTP servers at lirmm.lirmm.fr, cele.mrc-lmb.cam.ac.uk, and ncbi.nlm.nih.gov.

<sup>3</sup> The BLAST program is available via the World Wide Web (<http://www.ncbi.nlm.nih.gov/blast>).

(Hitachi Software Engineering Co., Ltd.). PCR primers were designed with the aid of Oligo (17).

**Cloning the C. elegans SHMT Gene**—The cosmid clone of C05D11 (GenBank accession number U00048) was kindly provided by Dr. Alan Coulson (Medical Research Council, Hinxton, United Kingdom). A computer restriction analysis of C05D11 (MacDNASIS) revealed that coding elements 11 (SHMT) and 13 could be isolated as a 5052-base pair (bp) *PvuII-ScaI* (Life Technologies, Inc., Pharmacia Biotech Inc.) fragment (residues 36311–41363 of the GenBank entry). The plasmid pGV9 (Fig. 2.) was constructed by ligating the *PvuII-ScaI* fragment into *EcoRV* (Life Technologies, Inc.)-cut pBluescript II KS<sup>+</sup> (Stratagene). The plasmid pC05.11 (Fig. 2.) was constructed by cutting pGV9 with *PstI* (New England Biolabs) and religating, leaving a 3196-bp fragment of C05D11 (residues 36311–39507), containing only open reading frame 11 (SHMT), in pBluescript.

**Mutant Rescue**—The subclones containing the SHMT gene were injected into the syncytial gonad of adult wild-type (N2) *C. elegans* hermaphrodites together with the dominant marker *rol-6(su1006)* (18, 19) contained on plasmid pCes1943.<sup>4</sup> Stable transgenic strains, those expressing the roller phenotype in successive generations, were used in the rescue experiments.

**Heteroduplex Analysis**—DNA from individual *C. elegans* heterozygous for each of the SHMT mutations was isolated as described by Barstead *et al.* (20) with the modifications of Williams *et al.* (21). PCR using *Taq* polymerase (BioCan Scientific) was performed as follows; 5 pmol of each of five sets of primers (synthesized by DNAgency, Malvern, PA) were used (Table I). The samples were incubated at 94 °C for 1 min on an Idaho Technology 1605 Air Thermo-Cycler before commencement of 30 cycles of amplification (94 °C for 10 s, 59 °C (62 °C for the D primer set) for 20 s, and 72 °C for 40 s). Following a 2-min incubation at 72 °C, 1.25 µl of EDTA (0.1 M, pH 7.5) was added to terminate the reaction. Heteroduplex analysis was performed using a mutation detection enhancement gel matrix (J.T. Baker) following the method of Nijbroek *et al.* (22) with the following changes; a heteroduplex denaturation/reannealing profile of 95 °C for 3 min, 85 °C for 3 min, 75 °C for 5 min, 65 °C for 5 min, 55 °C for 5 min, and 37 °C for 5 min was used. The heteroduplex DNA was then resolved on a 0.5× mutation detection enhancement gel at 400 V for 16–20 h.

**DNA Sequence Analysis**—DNA from individual nematodes homozygous for the *mel-32*(SHMT) mutations was isolated and amplified with the same primers used in the heteroduplex analysis. A total of 0.5 µl of this template DNA was incubated at 95 °C for 3 min on a Precision Scientific GTC-2 Genetic Thermo Cycler before commencement of 35 cycles of amplification (94 °C for 45 s, 59 °C (62 °C for the D primer set) for 30 s, and 72 °C for 1 min) followed by a polishing step of 72 °C for 7 min. The PCR products from two separate reactions were pooled, purified by agarose gel electrophoresis, and collected with a Qiagen QIAquick gel extraction kit. A total of 100–200 ng of each PCR product was sequenced on both strands using FS *Taq* terminator chemistry (Applied Biosciences) on a Perkin-Elmer GeneAmp PCR System. The reactions were run on an Applied Biosciences model 373A automated DNA sequence analyzer located at the Nucleic Acid-Protein Service Unit, University of British Columbia.

## RESULTS AND DISCUSSION

**Identification of a C. elegans Homolog of SHMT**—A search of the ACeDB program revealed one SHMT homolog, C05D11.11, located on chromosome III within the area defined by the cosmid C05D11. The position of the CeSHMT gene was based on both a Genefinder<sup>5</sup> prediction and on five partial cDNAs isolated as part of the *C. elegans* genome sequencing project. The predicted gene contains four exons coded in 1599 bp of genomic sequence. The predicted 484-amino acid protein sequence was used to search the GenBank data bank with the FASTA algorithm (14, 15), and the closest homologs were rabbit and human cytosolic SHMTs (61.7% and 61.3% identity, respectively, in a 470-amino acid overlap).

A sequence comparison of SHMTs shows the highly conserved nature of this protein and reveals many conserved domains (Fig. 1). The amino- and carboxyl-terminal 50 amino acids are the least conserved, and the central three-quarters of

the protein contain large stretches of completely conserved amino acids. The PLP cofactor binding site (residues 301–305 in the *C. elegans* protein) and the active site lysine (residue 306) are conserved in every case.

**Rescue of *mel-32* with C. elegans SHMT**—Our laboratory is in the process of constructing a transgenic library of sequenced cosmids that can be used for high resolution genetic mapping (23). Stable transgenic arrays are generated, which can act as cosmid sized duplications in rescue experiments. If the genomic DNA present in the extrachromosomal array rescues the recessive lethality of an essential gene then the wild-type copy of the mutation must be present in the DNA defined by the transgenic. Previous rescue experiments in our laboratory (23) have placed five essential genes in the genomic region defined by the cosmid C05D11. The genes *let-713*, *let-721*, *let-725*, *let-756*, and *mel-32* are all rescued by C05D11, making these genes candidates for potential mutations in CeSHMT.

The *mel-32* mutant phenotype was rescued with CeSHMT genomic DNA, indicating that *mel-32* encodes SHMT. All rescue experiments were performed with the canonical allele of *mel-32*, *s2518*, which was isolated in an EMS mutagenesis screen for maternal effect lethals.<sup>6</sup> Subsequently, *mel-32* was found to be allelic with 16 EMS-induced Mel alleles from a collection of maternal effect embryonic lethal mutations on chromosome III isolated by H. Schnabel and R. Schnabel.<sup>7</sup> Hermaphrodites homozygous for the *mel-32* mutations have no observable mutant phenotype, but their self-fertilized offspring display an embryonic lethal phenotype and arrest at about the 100 cell stage. The Mel phenotype of *mel-32* was rescued with pGV9, which contains coding element 11 (SHMT) and 13 (Fig. 2). To determine in which of these two genes the mutations reside, a smaller subclone, pC05.11 (Fig. 2), containing only gene 11 (SHMT) was constructed. This subclone, pC05.11, gave a partial rescue of the Mel-32 phenotype. From *mel-32*;pC05.11 transgenic hermaphrodites, a small number of progeny hatch and grow to adulthood but are not themselves fertile. pGV9 contains the entire 1254 bp of intergenic sequence between gene 11 (CeSHMT) and gene 13, while pC05.11 contains only 416 bp upstream of the CeSHMT ATG start codon. The partial rescue by pC05.11 suggests that some important regulatory sequences are missing from the smaller subclone.

**Heteroduplex Analysis of *mel-32*(SHMT) Mutations**—The 17 alleles of *mel-32* were analyzed by heteroduplex analysis to define more precisely the regions where the mutations occurred. Five sets of overlapping PCR products were generated from animals heterozygous for each allele. The primers used for PCR are listed in Table I, the exact location of each is shown in Fig. 3, and the overlap of each product is shown in Fig. 4. Each amplified DNA product was heat-denatured and allowed to cool slowly. Any PCR product that contains a mutation will have a mixture of homoduplexes and heteroduplexes, paired wild-type and mutant strands that contain base pair mismatches. These duplexes were run on a special mutation detection enhancement gel (see “Experimental Procedures”), which can reveal a single base pair mismatch in a short strand of DNA (24). The presence of multiple bands indicates a mutation in that particular overlapping fragment. Using this procedure, 13 out of the 17 mutations were placed into specific regions of the gene.

These mutations were distributed as follows. Region A (Fig. 4) contained one mutation: *t1473* (which was also detected in region B; see below). Region B contained six mutations: *t1473* (see below), *t1555*, *t1597*, *t1665*, *t1679*, and *s2518*. Region C contained three mutations: *t1552*, *t1616*, and *t1631*. Region D

<sup>4</sup> S. J. M. Jones and W. B. Barbazuk, personal communication.

<sup>5</sup> P. Green and L. Hillier, manuscript in preparation.

<sup>6</sup> G. Vatcher, unpublished results.

<sup>7</sup> H. Schnabel and R. Schnabel, unpublished results.

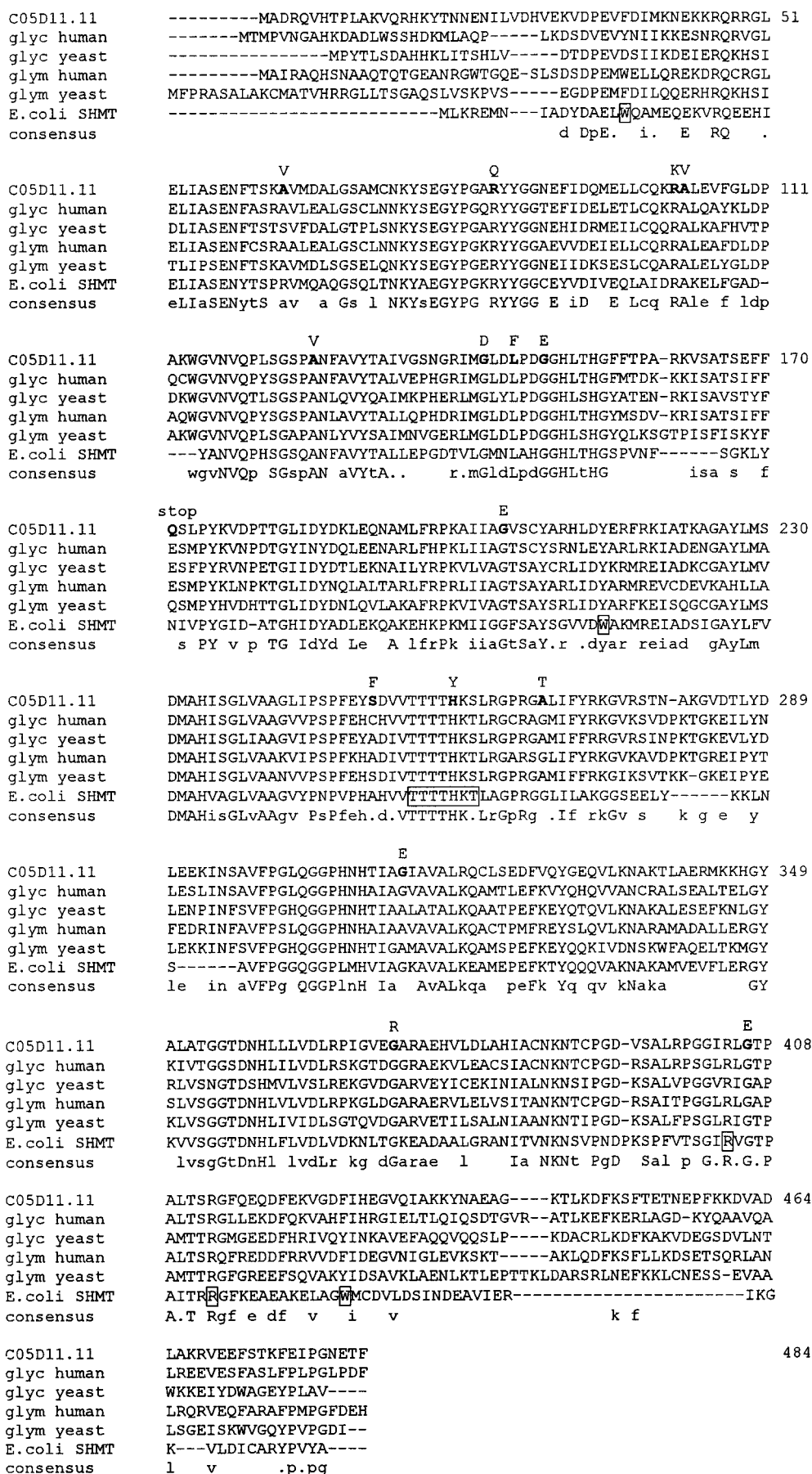


FIG. 1. Comparison of the amino acid sequences of selected SHMTs. The aligned sequences are *C. elegans mel-32* (C05D11.11), human cytosolic (glyc human) and mitochondrial (glym human), *S. cerevisiae* cytosolic (glyc yeast) and mitochondrial (glym yeast), and *E. coli GlyA* (*E. coli*



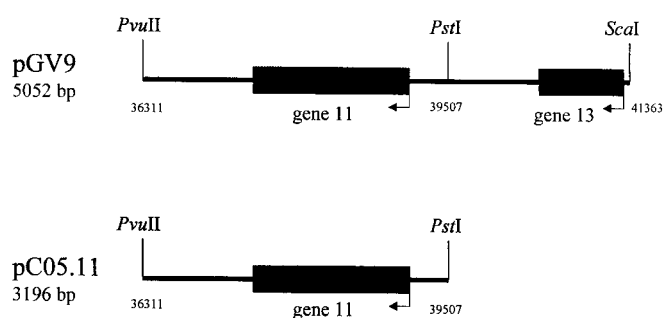


FIG. 2. SHMT subclones. The genomic subclones used to rescue *mel-32* are shown. Nucleotide numbers correspond to the GenBank entry for C05D11 (U00048). Restriction sites used for subcloning are indicated.

TABLE I  
PCR primers

Primer	Sequence	Product length
		bp
A1	ATGCATATACGTGTTGCC	399
A2	TAATACCGAGCTCCAGGG	
B1	GAAGCGGCAGCGTCGTGG	408
B2	GAGACTTTGCGAGCTGGG	
C1	TATGGGCTCGATCTTCC	483
C2	GTGTCAACTCCCTTTGCG	
D1	CAACGACGCACAAGTCGC	449
D2	GAGACGAATGCCTCCTGG	
E1	CGGAGTTGAAGGAGCTCG	474
E2	AAGTGATAAAATGGCGGG	

contained three mutations: *t1456*, *t1474*, and *t1576*. Region E contained a single mutation: *t1632*. For allele *t1666*, no mutations were detected in regions A, C, D, and E, but a single PCR product could not be obtained for region B. It was therefore assumed that the lesion in this allele was contained in region B, which was later confirmed by direct sequence analysis. The position of the molecular lesions in alleles *t1520*, *t1607*, and *t1671* could not be detected using this procedure.

**Sequence Analysis of *mel-32*(SHMT)**—The same oligonucleotide primers used to generate PCR products for the heteroduplex analysis (Table I) were used to amplify DNA from homozygous mutant animals from the region in which the mutation for each allele was detected. For alleles *t1520*, *t1607*, and *t1671*, in which heteroduplexes were not observed, DNA from all five regions was amplified from mutant strains. The PCR products were then directly sequenced in both orientations (see “Experimental Procedures”).

The sequence analysis results are summarized in Table II and Fig. 4. Allele *t1631* contains the only nonsense mutation identified for *mel-32*. The codon for glutamine 171 is mutated into an amber stop codon. Any truncated protein produced in this mutant would lack the terminal two thirds of the enzyme, including the active site. This allele is assumed to be a null. The phenotype of this mutant is indistinguishable from the other 16 alleles, suggesting that all of the alleles may be null, or that all reduce the SHMT activity below the threshold required for embryonic survival.

Alleles *t1456* and *t1474* both contain the same base pair mutation, which changes glycine 313 to glutamic acid. In allele *t1616*, alanine 268 is changed to threonine. The amino acids at positions 313 and 268 of the SHMT consensus sequence (Fig. 1) are always either a glycine or an alanine. It is assumed that

replacing a small, neutral amino acid with a larger acidic or polar one warps the secondary structure of SHMT enough to disrupt its enzymatic function.

Eukaryotic forms of the SHMT holoenzyme exist as tetramers of identical subunits, and the predicted secondary structure contains alternating helices and strands, placing SHMT in the  $\alpha/\beta$  class of proteins (5). Fig. 1 reveals that there are 22 completely conserved glycines in the SHMT consensus sequence. Most of the conserved glycines are located at predicted turns, indicating a probable tertiary structural conservation (5). This suggests that these glycines are structurally important, being located at turns, at positions where helices cross, or at other constrained locations. The importance of having a glycine at five of these positions is proven by this work. In allele *t1576*, glycine 372 is changed to arginine; in allele *t1679*, glycine 143 is changed to aspartic acid; in alleles *t1552*, *t1632*, and *t1665*, conserved glycines 204, 406, and 149, respectively, are all changed to glutamic acids. Replacing these constrained glycines with large charged groups probably has a deleterious effect on the structure of SHMT.

Two of the mutations cause alanine to valine alterations, *t1666* (alanine 103) and *s2518* (alanine 126). These residues are conserved alanines in the SHMT consensus sequence. We assume there are very strong structural constraints at these positions, as the conservative change of an alanine to a valine disrupts the structure enough to affect SHMT activity. A single PCR product could not be obtained for allele *t1666* with primer set B in the heteroduplex analysis, but the point mutation does not interfere with the primer binding site (Fig. 3.). To eliminate the possibility that *t1666* contained a second mutation in one of the B primer binding sites, the overlapping regions, A and C, were sequenced. Neither region contained a mutation, so the failure of this allele to yield a single PCR product with the B primer set remains a mystery.

Allele *t1473* contains a double mutation, as suggested by the heteroduplex analysis where region A and B both indicated the position of a mutation. Alanine 63 is changed to valine, and leucine 146 is changed to phenylalanine. Residue 63 is an alanine four out of six times in the SHMT consensus sequence, but the *E. coli* protein has an arginine at this position, so we assume that changing this alanine to valine would have a minimal steric effect. However, residue 146 is a conserved leucine in the SHMT consensus sequence, suggesting that the leucine to phenylalanine missense mutation is the primary cause of the mutant phenotype in allele *t1473*.

In alleles *t1671* and *t1520*, serine 251 is changed to phenylalanine. Residue 251 is only three amino acids away from the active site and is always small in the SHMT consensus sequence. The replacement of this small amino acid with a large aromatic one probably distorts the conformation of the active site and may also interfere with the PLP aromatic ring.

In allele *t1607*, histidine 259 is changed to tyrosine. The mutation in allele *t1607*, residue 259 in the SHMT consensus sequence, is the only one from our collection that has been mutated previously. This histidine is in the conserved active site VTTTTHK(S/T) motif found in all SHMTs and is adjacent to the active site lysine. Every residue in this motif, except the valine, has been mutated in the *E. coli* isoform of SHMT to determine the effects on catalysis (Fig. 1.). When the active site histidine is changed to asparagine, there is no structural change in the enzyme, but the catalytic activity is greatly

SHMT). The alignment was made using the ClustalW program. In the SHMT consensus sequence (*consensus*), uppercase letters indicate amino acids identical in all six proteins, lowercase letters indicate residues identical in at least four out of six proteins, and a dot indicates conserved amino acids. The numbers indicate the positions in the *C. elegans* protein. The amino acids mutated in the *mel-32* alleles are in bold, and the mutant amino acid for each is listed above the sequence. Residues that have been previously mutated in *E. coli* are boxed (see text for details).

-416	TGCAGTATCGTTTGCACTTGTTTTAA	-389
-390	ATGGTTTCTTGCATAGTTTCTGTAGCAATGTTTTGTGTTGCATTTGAAAGTTATCTG	-331
-330	CATTATATTTTGTATTAGAGACCCTTTTTGTTTTTATTACTAAGTTAGTGTACACT	-271
-270	TGTTAAAACCACGCTCGAATACCTCTTAATTACAATTTCTAATGAGTAACTTTTGACTT	-211
-210	ATCAGTTTTGATCTACGATCCGGGATCGGCCAATCTCGAAAACCTCCCGGTTTCAGATAT	-151
-150	CATCATAGCCTACCGCAATCCAGTTTTATTCTTATTTTCTCCACACTTTCCCATGCATA	-91
-90	<u>TACGTGTTGCCAACGCGTCAGCTTTTGTGTTGAATATGTTCCCGATTTTTTCGCTTCTTT</u>	-31
1	<u>AI</u>	
1	M A D R Q V H T P L	10
-30	GCTTAATTTAATTTGGTTTTTACAGTGTAAAATGGCTGATCGTCAAGTGCACACACCATTG	30
11	A K V Q R H K Y T N N E N I L V D H V E	30
31	GCTAAAGTTCAGCGCCACAAGTACCAACAACGAGAACATTTTGGTGGACCATGTTGAG	90
31	K V D P E V F D I M K N	42
91	AAAGTTGATCCAGAAGTTTTTCGATATCATGAAAAATgtaagacttaatgtttctcaaatc	150
43	E K K R Q R R G L E L I A	55
151	ttagtcatttgTTTTTTTcagGAGAAGAAGCGGCAGCGTCGTGGACTTGAGCTCATCGCT	210
56	S E N F T S K <sup>2</sup> A V M D A L G S A M C N K	75
211	TCCGAGAATTTCAAGCAAGGCTGTATGGATGCTCTTGGCTCGGCAATGTGCAACAAA	270
76	Y S E G Y P G A <sup>8</sup> R Y Y G G N E F I D Q M	95
271	TACAGTGAAGGATACCCCTGGAGCTCGGTATTACGGAGGAATGAGTTCATCGATCAGATG	330
96	E L L C Q K <sup>6</sup> R <sup>14</sup> A L E V F G L D P A K W G	115
331	GAGCTCCTTTGCCAGAAGAGAGCTCTTGAGGTATTCCGACTTGATCCAGCCAAGTGGGGA	390
116	V N V Q P L S G S P <sup>17</sup> A N F A V Y T A I V	135
391	GTCAATGTGCAGCCATTGTCCGGATCACCAGCCAATTTCCGAGTCTACACTGCCATCGTT	450
136	G S N G R I M <sup>16</sup> G <sup>2</sup> L D <sup>13</sup> L P D G G H L T H G	155
451	GGATCCAATGGACGCATTATGGGCCTCGATCTCCAGATGGAGGTCATTTGACTCATGGg	510
156	F F T P	159
511	taaggttatctagattagactcgaatgaattatacatTTAATTTcagATTCTTCCACCCA	570
160	A R K V S A T S E F F <sup>11</sup> Q S L P Y K V D P	179
571	<u>GCTCGCAAAGTCTCTGCGACTTCTGAATTCCTCCAGTCTTCCATACAAGGTTGATCCA</u>	630
180	T T G L I D Y D K L E Q N A M L F R P K	199
631	ACAACTGGATTGATCGACTATGACAAGCTTGAGCAGAATGCAATGCTTTTCCGTCCAAAA	690
200	A I I A <sup>5</sup> G V S C Y A R H L D Y E R F R K	219
691	GCCATCATTGCCGGAGTTTCTGTACGCTCGTCATCTCGATTATGAACGTTTCCGTAAG	750
220	I A T K A G A Y L M S D M A H I S G L V	239
751	ATTGCCACAAAGGCTGGAGCCTATTTGATGTCTGATATGGCTCACATCTCCGACTTGT	810
240	A A G L I P S P F E Y <sup>4,15</sup> S D V V T T T T H <sup>9</sup>	259
811	GCCGCTGGACTTATCCCATCACCATTTCGAGTATTCTGATGTTGTAACCACAACGAGCAC	870
260	K S L R G P R G <sup>10</sup> A L I F Y R K G V R S T	279
871	<u>AAGTCGCTCAGAGGACCACGTGGAGCTTTGATCTTACAGAAAGGTTGTCGATCTACC</u>	930
280	N A K G V D T L Y D L E E K I N S A V F	299
931	<u>AACGCAAAGGGAGTTGACACTTTGTATGATTTGGAGGAGAAGATCAACTCGGCCGTGTT</u>	990
300	P G L Q G G P H N H T I A <sup>1,3</sup> G I A V A L R	319
991	CCAGGACTTCAAGGTGGACCACACAATCACACTATTGCTGGAATTGCCGTTGCTTTGAGA	1050
320	Q C L S E D F V Q Y G E Q V L K N A K T	339
1051	CAATGCCTTTCTGAAGATTTTGTTCAGTACGGAGAGCAAGTGTGAAGAATGCCAAAACC	1110

FIG. 3. Nucleotide sequence of the *C. elegans* SHMT genomic region and derived amino acid sequence. Intronic regions are in *lowercase*, the primers used in the heteroduplex and sequence analysis are *underlined*, a putative SL1 leader sequence is denoted by a *dashed line*, and the poly(A) signal is *double underlined*. The nucleotides and amino acids mutated in the *mel-32* alleles are in *bold*, and the allele number for each is listed. This sequence corresponds to positions 39507–37321 (in reverse orientation) of the C05D11 GenBank entry (accession number U00048).

340	L A E R M K K H G Y A L A T G G T D N H	359
1111	TTGGCCGAGAGAATGAAGAAGCATGGATATGCTTTGGCAACCCGGAGGAACCCACAATCAT	1170
360	L L L V D L R P I G V E <sup>7</sup> G A R A E H V L	379
1171	TTGTTGCTTGTGACTTGCGCCCAATCGGAGTTGAAGGAGCTCGTGCTGAGCATGTGCTT	1230
	<sub>E1</sub>	
380	D L A H I A C N K N T C P G D V S A L R	399
1231	GATCTCGCTCATATTGCTTGAACAAGAATACGTGCCCAGGAGACGTTTCTGCTTTGAGA	1290
400	P G G I R L <sup>12</sup> G T P A L T S R G F Q E Q D	419
1291	CCAGGAGGCATTTCGTCTCGGA <sup>12</sup> ACTCCAGCTCTCACCTCCCGTGGATTCCAGGAGCAAGAT	1350
	<sub>B2</sub>	
420	F E K V G D F I H E	429
1351	TTCGAGAAAGTCGGAGATTTTCATTCATGAGGgtatgttttttcaactgtaatcttcaaca	1410
430	G V Q I A K K Y N A E A G	442
1411	ttaatctaaaatattgatttcagGTTC <sup>12</sup> CAATCGCAAAGAAGTACAATGCTGAGGCTGG	1470
443	K T L K D F K S F T E T N E P F K K D V	462
1471	AAAGACATTGAAAGACTTCA <sup>12</sup> AATCATTCACTGAGACCAATGAACCATTCAAGAAAGACGT	1530
463	A D L A K R V E E F S T K F E I P G N E	482
1531	CGCTGATCTCGCCAAACGTGTGAGGAGTCTCCACGAAGTTCGAGATTCCAGGAAACGA	1590
483	T F *	484
1591	AACTTTCTAATTTTAATCATCATCGTCATTATCAACATTGCATTGTATTTATAAAAAACA	1650
1651	TTCCCGCCATTTTATCACTTT <sup>12</sup> CATGTTTGAATTTTAAACAGGAGAAATAAATATTGATCT	1710
	<sub>E2</sub>	
1711	TTTCAAATTAATTTGTTTTTTTATATTTAGAAAGCAATTGAGATTATCTACGGCATTGG	1770

FIG. 3—continued

reduced (25). When the active site histidine is changed to aspartic acid, the activity of the physiological reactions is reduced, but the activity of some alternate reactions is increased (26). A series of kinetic and spectral studies on these mutant enzymes revealed that the active site histidine is not catalytically essential and is not the base that accepts the  $\alpha$  proton. These studies have also shown that the active site histidine interacts with the amino acid substrate or PLP. This histidine is believed to have a critical role in determining reaction specificity by determining the structure of the one-carbon binding site and controlling the orientation of the substrate and PLP ring (25, 26). Because *E. coli* SHMT is catalytically active without an active site histidine, it is unlikely that the imidazole ring of histidine 259 participates as an electron donor or acceptor in the physiological reactions of CeSHMT. In allele *t1607*, inserting the aromatic ring of tyrosine at the active site may change the environment of the PLP aromatic ring such that it is no longer in line with critical residues required for catalysis. The PLP and substrate binding pocket would also be distorted.

The *E. coli* active site is VTTTHKT. Each of the threonine residues in this motif was changed to an alanine to determine the kinetic and spectral properties of the mutant enzymes (27). When the first or fourth threonine is converted to alanine, the enzyme is essentially wild-type. When the second or fifth threonine is changed to alanine, the mutants are structurally unchanged but have shifted kinetic properties. These results indicated that these four threonines do not play a critical role in the mechanism of SHMT. However, when the third threonine is mutated to an alanine, the enzyme loses 97% of its catalytic activity. When this residue is mutated to a serine, the activity is essentially wild-type, so the presence of a hydroxyl-containing side chain is very important. These studies also revealed that none of these threonines bind PLP (27).

The active site lysine in *E. coli* has also been mutated. When this lysine is changed to a glutamine, the enzyme catalyzes one turnover of product at wild-type levels, but cannot release the product (28). When the lysine is changed to arginine or histidine, the PLP cofactor cannot readily form the external aldi-

mine. These results suggest that the active site lysine expels the product by converting the external aldimine to an internal aldimine and that lysine is not the base that removes the  $\alpha$  proton of the substrate.

It has been proposed that there are two bases at the active site of SHMT on opposite sides of the PLP ring (29). These bases are not the active site histidine (25, 26, 29) or lysine (28). A study on sheep liver SHMT has shown that there are arginine residues present at the active site. Arginine 269 and arginine 462 from sheep liver were protected from chemical modification by tetrahydrofolate binding (33). Arginine 462 is not conserved but arginine 269 (residue 273 in CeSHMT) is conserved as an arginine in all eukaryotic SHMTs and as a lysine in *E. coli* SHMT. Arginine 363 and arginine 372 in *E. coli* SHMT (conserved residues 404 and 413 in CeSHMT, Fig. 1.) were changed to both alanine and lysine (30). Both of the arginine 372 mutations had wild-type activity, suggesting that this residue, although conserved, is not critical for catalytic activity. The R363A mutant enzyme had no activity with serine as a substrate and could not bind serine or glycine. The R363K mutant enzyme had only 0.03% of wild-type activity and a 15-fold decreased affinity for serine and glycine. The conserved arginine at this position is catalytically essential and believed to be the binding site of the amino acid substrate carboxyl group (30). It has been proposed that an arginine-carboxyl interaction might be preferred over simple charge interactions because the guanidinium group presents charged hydrogen bonds rather than the single bond formed by lysine or histidine (33). This may explain why replacing an arginine with a lysine has such a dramatic effect on enzyme activity.

The mutant phenotype of allele *t1555* is caused by an arginine (Arg-102) to lysine mutation. The arginine at this position is 100% conserved in all sequenced SHMTs. This conservation suggests that the presence of the guanidinium group of arginine at this position is essential to the enzymatic function of SHMT. In allele *t1597*, arginine 84 is changed to glutamine. This arginine is also 100% conserved, so the presence of a basic group at this position is probably also essential to the enzymatic function of

FIG. 4. Exon/intron structure of *mel-32*(SHMT) and summary of sequencing results. The overlapping PCR products used for heteroduplex analysis are shown above the gene. The wild-type and mutant amino acid(s) are given for each allele. The active site is indicated by a shaded box.

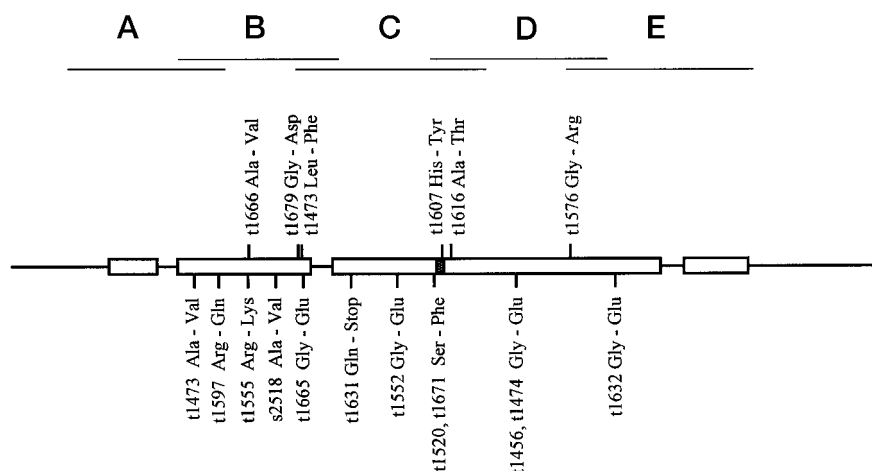


TABLE II  
*mel-32* mutant alleles

	Allele	Codon mutation	Amino acid mutation
1	<i>t1456</i>	GGA → GAA	Gly-313 → Glu
2	<i>t1473</i>	GCT → GTT	Ala-63 → Val
		CTT → TTT	Leu-146 → Phe
3	<i>t1474</i>	GGA → GAA	Gly-313 → Glu
4	<i>t1520</i>	TCT → TTT	Ser-251 → Phe
5	<i>t1552</i>	GGA → GAA	Gly-204 → Glu
6	<i>t1555</i>	AGA → AAA	Arg-102 → Lys
7	<i>t1576</i>	GGA → AGA	Gly-372 → Arg
8	<i>t1597</i>	CGG → CAG	Arg-84 → Gln
9	<i>t1607</i>	CAC → TAC	His-259 → Tyr
10	<i>t1616</i>	GCT → ACT	Ala-268 → Thr
11	<i>t1631</i>	CAG → TAG	Gln-171 → Stop
12	<i>t1632</i>	GGA → GAA	Gly-406 → Glu
13	<i>t1665</i>	GGA → GAA	Gly-149 → Glu
14	<i>t1666</i>	GCT → GTT	Ala-103 → Val
15	<i>t1671</i>	TCT → TTT	Ser-251 → Phe
16	<i>t1679</i>	GGC → GAC	Gly-143 → Asp
17	<i>s2518</i>	GCC → GTC	Ala-126 → Val

SHMT. Either or both of these arginine residues could be at the active site of the enzyme.

Chemical modification studies on the cysteine residues in rabbit liver cytosolic (31) and mitochondrial (32) SHMT, sheep liver cytosolic SHMT (5), and *E. coli* SHMT (34) have shown that there are no disulfide bonds in eukaryotic homotetramers or prokaryotic homodimers, and that PLP and substrates protect a catalytically essential active site cysteine. There are no cysteine mutations in our collection, but residue 99 in the SHMT consensus sequence is a conserved cysteine in all the eukaryotic isoforms. This residue could be the important active site cysteine.

In a series of experiments to study folding intermediates of *E. coli* SHMT, a set of mutant proteins was constructed in which the three tryptophan residues were replaced with phenylalanine residues (35). The three double mutants and a triple mutant were essentially wild-type enzymes, with only minor structural differences. Tryptophans 16, 183, and 385 in *E. coli* correspond to positions 37, 214, and 426 in the SHMT consensus sequence (Fig. 1). These tryptophans are not conserved, but, intriguingly, position 37 in the SHMT consensus sequence contains five out of six aromatic residues; position 214 has six out of six aromatic residues, with all five eukaryotic forms having a tyrosine; position 426 also has six out of six aromatic residues, tryptophan in *E. coli* and a mixture of phenylalanine and tyrosine in the eukaryotic forms. Clearly, the presence of an aromatic side chain at these positions is important.

It is assumed that each of the *mel-32* mutations abolishes, or

greatly reduces, the SHMT activity, resulting in the observed maternal effect lethal phenotype. We hypothesize that a product of SHMT, such as glycine or some byproduct of the one-carbon metabolic pool, is required for embryonic development in the egg. The essential metabolite is normally supplied by the diet but cannot diffuse through the eggshell. *mel-32* homozygotes have enough maternally provided SHMT activity to develop and hatch into feeding larvae, where the diet can supply enough of the required nutrient. These *mel-32* homozygotes, however, have no SHMT activity to pass on to their offspring, which quickly use up the required metabolite pool in the egg and arrest as embryos. This hypothesis can also explain why *C. elegans* is unique in its requirement for SHMT activity for survival. The SHMT mutations in *E. coli*, *S. cerevisiae*, and Chinese hamster ovary cells all result in glycine auxotrophy (3, 4, 9, 10). In all three cases, the deficiency is rescued by addition of glycine. The egg shell of *C. elegans*, which is impervious to most chemicals (36), would imprison the embryo in a forced starvation. This hypothesis can be tested, and the missing metabolite identified, by supplying the developing larvae with metabolic precursors, for example glycine and thymidine.

The 17 point mutations found in *C. elegans mel-32*(SHMT) are clustered within the middle two thirds of the protein, the most conserved region. This suggests that the ends of the protein are not required for enzymatic function and that any point mutations occurring there would not be detected in a screen for maternal effect lethals, the likely null phenotype. This observation is supported by the fact that if the amino-terminal 25–30 amino acids of rabbit liver cytosolic SHMT are removed with proteases the enzyme remains catalytically active and structurally stable (37). Sixteen of the *mel-32*(SHMT) alleles contain a single base pair substitution, and one, *t1473*, contains two base pair substitutions. All of the mutations are G/C to A/T transitions, as expected in EMS-induced alleles.

*mel-32* has 17 alleles, making it a high frequency hit gene, as most Mel genes only have one or two alleles. Several large scale screens for maternal effect lethals have been carried out which have produced hundreds of Mel alleles (38–40).<sup>6,7</sup> Most of these mutations have only been identified genetically, but a few have been cloned and sequenced. Most of the identified Mel genes are involved in polarizing the embryo or determining cell fate. For example, *par* (for partitioning defective) mutant embryos arrest as amorphous masses of differentiated cells (39). *par-1* encodes a conserved Ser/Thr kinase, and the two alleles that have been sequenced reveal mutations in invariant kinase domain residues, suggesting that PAR-1 kinase activity has an essential function (41). *par-2* encodes a 628-amino acid protein with a putative ATP binding site and zinc ring finger domain. Two sequenced alleles of *par-2* introduce stop codons that trun-



cate the protein to a form lacking the ATP binding domain (42). The *pie-1* gene encodes a zinc finger protein, which interferes with transcription (43). The only sequenced allele of this gene contains a 217-bp deletion in the 5' end of the gene (44). The *mex-1* gene contains two copies of the zinc finger domain found in *pie-1* and mutations in *mex-1* alter the fates of some somatic blastomeres. Two deletion mutants have been sequenced; one deletes the NH<sub>2</sub>-terminal 36 amino acids, and the other deletes the COOH-terminal 80 amino acids (45). The *mex-3* gene contains two RNA binding KH domains and regulates blastomere identity in embryos. One allele of *mex-3* is deleted for the first 92 bp of coding sequence, whereas three other alleles have point mutations that change conserved glycines in the KH domain (46). The transcription factor *skn-1* is required for correct specification of cell fates. One mutant form of *skn-1* has a stop codon introduced in the DNA binding domain (47, 48). Finally, the *apx-1* gene of *C. elegans* controls early cell fates and is a homolog of the *Drosophila Delta* and *Serrate* genes. One mutant allele of *apx-1* has a 900-bp deletion in the center of the gene (49).

The extreme degree of conservation in SHMT across millions of years of evolution suggests that the conserved residues must play some catalytically or structurally essential roles. Several of these important residues have previously been identified. From our collection of 17 mutations, only one, the active site histidine, has been mutated previously. Therefore, this work defines 13 new residues (not including the nonsense and two repeated mutations) that are potentially essential to SHMT's catalytic activity or structural integrity.

SHMT is believed to have a folding pattern similar to aspartate aminotransferase and dialkylglycine decarboxylate, both of which have crystallographic structure data available (50). Three out of four of the residues conserved in all aminotransferases are conserved in SHMT. In the CeSHMT numbering scheme, these residues are: aspartic acid 231, which hydrogen bonds to N1 of PLP; lysine 260, which forms a Schiff base with PLP; and arginine 404, which hydrogen bonds with the  $\alpha$ -carboxyl group of the substrate. Future crystallographic analysis of SHMT may allow characterization of more essential residues. The residues corresponding to arginine 404 and lysine 260 have been mutated in *E. coli* (Fig. 1) and shown to be essential (28, 30), but there are no known mutations affecting aspartic acid 231.

Injecting SHMT subclones containing site-directed mutations for aspartic acid 231, and other conserved residues, into *mel-32*(SHMT) nulls may allow a quick assay of enzymatic function. If the mutated subclone rescues, then the mutant SHMT is functional; if it does not rescue, then we can assume the mutation has abolished or decreased SHMT function. *C. elegans* could also be used to test chemotherapy drugs directed against SHMT. If shells of wild-type eggs are solubilized, and the embryos treated with drugs, any drug that reduces or eliminates SHMT function should cause a Mel phenotype.

**Acknowledgments**—We thank Diana Janke, Jacquie Schein, and The Ha for expert technical assistance and Raja Rosenbluth for advice.

#### REFERENCES

- Schirch, L. (1982) *Adv. Enzymol.* **53**, 83–112
- Garrow, T. A., Brenner, A. A., Whitehead, V. M., Chen, X. N., Duncan, R. G., Korenberg, J. R., and Shane, B. (1993) *J. Biol. Chem.* **268**, 11910–11916
- Narkewicz, M. R., Sauls, S. D., Tjoa, S. S., Teng, C., and Fennessey, P. V. (1996) *Biochem. J.* **313**, 991–996
- McNeil, J. B., McIntosh, E. M., Taylor, B. V., Zhang, R., Tang, S., and Bognar, A. L. (1994) *J. Biol. Chem.* **269**, 9155–9165
- Usha, R., Savithri, H. S., and Rao, N. A. (1994) *Biochim. Biophys. Acta* **1204**, 75–83
- Eichler, H. G., Hubbard, R., and Snell, K. (1981) *Biosci. Rep.* **1**, 101–106
- Nakshatri, H., Bouillet, P., Bhat-Nakshatri, P., and Chambon, P. (1996) *Gene (Amst.)* **174**, 79–84
- Trakatellis, A., Dimitriadou, A., Exindari, M., Christodoulou, D., Malissiovas, N., Antoniadis, A., and Haitoglou, K. (1994) *Postgrad. Med. J.* **70**, (Suppl. 1), S89–S92
- Stauffer, G. V., Plamann, M. D., and Stauffer, L. T. (1981) *Gene (Amst.)* **14**, 63–72
- Pfender, W., and Pizer, L. I. (1980) *Arch. Biochem. Biophys.* **200**, 503–512
- Stover, P. J., Chen, L. H., Suh, J. R., Stover, D. M., Keyomarsi, K., and Shane, B. (1997) *J. Biol. Chem.* **272**, 1842–1848
- Brenner, S. (1974) *Genetics* **77**, 71–94
- Durbin, R., and Thierry-Mieg, J. (1997) *C. elegans II*, p. 17, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2444–2448
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
- Rychlik, W., and Rhoads, R. E. (1989) *Nucleic Acids Res.* **17**, 8543–8551
- Kramer, J. M., French, R. P., Park, E.-C., and Johnson, J. J. (1990) *Mol. Cell. Biol.* **10**, 2081–2089
- Mello, C. C., Kramer, J. M., Stinchcomb, D., and Ambros, V. (1991) *EMBO J.* **10**, 3959–3970
- Barstead, R. J., Kleiman, L., and Waterston, R. H. (1991) *Cell Motil. Cytoskel.* **20**, 69–78
- Williams, B. D., Schrank, B., Huynh, C., Shownkeen, R., and Waterston, R. H. (1992) *Genetics* **131**, 609–624
- Nijbroek, G., Sood, S., McIntosh, I., Francomano, C. A., Bull, E., Pereira, L., Ramirez, F., Pyeritz, R. E., and Dietz, H. C. (1995) *Am. J. Hum. Genet.* **57**, 8–21
- Janke, D. L., Schein, J. E., Ha, T., Franz, N. W., O'Neil, N. J., Vatcher, G. P., Stewart, H., I., Kuervers, L. K., Baillie, D. L., and Rose, A. M. (1997) *Genome Res.* **7**, 974–985
- Keen, J., Lester, D., Inglehearn, C., Curtis, A., and Bhattacharya, S. (1991) *Trends Genet.* **7**, 5
- Hopkins, S., and Schirch, V. (1986) *J. Biol. Chem.* **261**, 3363–3369
- Stover, P., Zamora, M., Shostak, K., Gautam-Basak, M., and Schirch, V. (1992) *J. Biol. Chem.* **267**, 17679–17687
- Angelaccio, S., Pascarella, S., Fattori, E., Bossa, F., Strong, W., and Schirch, V. (1992) *Biochemistry* **31**, 155–162
- Schirch, D., Delle Fratte, S., Iurescia, S., Angelaccio, S., Contestabile, R., Bossa, F., and Schirch, V. (1993) *J. Biol. Chem.* **268**, 23132–23138
- Shostak, K., and Schirch, V. (1988) *Biochemistry* **27**, 8007–8014
- Delle Fratte, S., Iurescia, S., Angelaccio, S., Bossa, F., and Schirch, V. (1994) *Eur. J. Biochem.* **225**, 395–401
- Gavilanes, F., Peterson, D., and Schirch, L. (1982) *J. Biol. Chem.* **257**, 11431–11436
- Gavilanes, F., Peterson, D., Bullis, B., and Schirch, L. (1983) *J. Biol. Chem.* **258**, 13155–13159
- Usha, R., Savithri, H. S., and Rao, N. A. (1992) *J. Biol. Chem.* **267**, 9289–9293
- Schirch, V., Hopkins, S., Villar, E., and Angelaccio, S. (1985) *J. Bacteriol.* **163**, 1–7
- Cai, K., and Schirch, V. (1996) *J. Biol. Chem.* **271**, 2987–2994
- Wood, W. B. (1988) *The Nematode Caenorhabditis elegans*, p. 217, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Schirch, V., Schirch, D., Martini, F., and Bossa, F. (1986) *Eur. J. Biochem.* **161**, 45–49
- Kemphues, K. J., Kusch, M., and Wolf, N. (1988) *Genetics* **120**, 977–986
- Kemphues, K. J., Priess, J., Morton, D., and Cheng, N. (1988) *Cell* **52**, 311–320
- Mains, P. E., Sulston, I. A., and Wood, W. B. (1990) *Genetics* **125**, 351–369
- Guo, S., and Kemphues, K. J. (1995) *Cell* **81**, 611–620
- Levitani, D. J., Boyd, L., Mello, C. C., Kemphues, K. J., and Stinchcomb, D. T. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 6108–6112
- Seydoux, G., Mello, C. C., Pettitt, J., Wood, W. B., Priess, J. R., and Fire, A. (1996) *Nature* **382**, 713–716
- Mello, C. C., Schubert, C., Draper, B., Zhang, W., Lobel, R., and Priess, J. R. (1996) *Nature* **382**, 710–712
- Guedes, S., and Priess, J. R. (1997) *Development* **124**, 731–739
- Draper, B. W., Mello, C. C., Bowerman, B., Hardin, J., and Priess, J. R. (1996) *Cell* **87**, 205–216
- Bowerman, B., Eaton, B. A., and Priess, J. R. (1992) *Cell* **68**, 1061–1075
- Blackwell, T. K., Bowerman, B., Priess, J. R., and Weintraub, H. (1994) *Science* **266**, 621–628
- Mello, C. C., Draper, B. W., and Priess, J. R. (1994) *Cell* **77**, 95–106
- Pascarella, S., Schirch, V., and Bossa, F. (1993) *FEBS Lett.* **331**, 145–149