

Research

Use of molecular beacons to verify that the serine hydroxymethyltransferase pseudogene *SHMT-ps1* is unique to the order Primates

Eric J Devor

Address: Molecular Genetics and Bioinformatics, Integrated DNA Technologies, Inc., 1710 Commercial Park, Coralville, IA 52241, USA.
E-mail: rdevor@idtdna.com

Published: 29 January 2001

Genome Biology 2001, **2**(2):research0006.1-0006.5

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2001/2/2/research/0006>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

Received: 22 November 2000

Revised: 20 December 2000

Accepted: 2 January 2001

Abstract

Background: The serine hydroxymethyltransferase processed pseudogene *SHMT-ps1* has been suggested to be unique to the order Primates because of the failure to amplify this sequence by PCR from genomic DNAs of any non-primate mammal species. Here, 'molecular beacon' probes specific to *SHMT-ps1* were used in an attempt to verify this suggestion.

Results: In a search for *SHMT-ps1*-specific sequences using molecular beacons across a range of mammalian species, *SHMT-ps1* was only found in primates. The molecular beacon assays also showed that *SHMT-ps1* is present in both Old World and New World species but not among prosimians.

Conclusions: These results suggest that *SHMT-ps1* originated close to the origin of the Anthropoidea, some 40 to 50 million years ago.

Background

Serine hydroxymethyltransferase (EC 2.1.2.1, SHMT) is a pyridoxal-phosphate-binding enzyme that catalyzes the reversible reaction of serine and tetrahydrofolate to glycine and 5,10-methylene tetrahydrofolate [1]. SHMT-catalyzed reactions are important as a primary source of one-carbon units used in folate-dependent metabolism, including purine synthesis [2]. Two isoforms of SHMT are found in eukaryotic cells: one expressed in the cytoplasm (SHMT-cyt) and one in mitochondria (SHMT-mit). The gene encoding the former is located on human chromosome 17p11.2 and the gene encoding the latter is on human chromosome 12q13.2 [3,4]. A third member of this small gene family is a processed pseudogene, *SHMT-ps1*, located on human chromosome 1p32.3-p33 [5,6]. Analyses of known SHMT amino acid and nucleotide sequences indicate that eukaryotic *SHMT-cyt* is descended from a prokaryotic *glyA*-like ancestor and that both *SHMT-mit* and *SHMT-ps1* have been derived from *SHMT-cyt* [3-5,7].

PCR-based analyses of evolutionary conservation of the *SHMT-ps1* nucleotide sequence suggest that the reverse transcription event giving rise to *SHMT-ps1* occurred after the divergence of the order Primates from the rest of the Mammalia, thus making this locus a genetic marker unique to the primates [7]. Here, this suggestion is verified using two end-point molecular beacon assays. In addition, results presented here further refine the time frame in which the reverse transcription and retrotransposition events occurred.

Results and discussion

End-point molecular beacon assays (see Materials and methods) were carried out on 19 primate and 12 non-primate mammalian genomic DNAs. The results of those assays confirmed the previous conclusion that *SHMT-ps1* is not found outside members of the order Primates. There was, however, one notable departure from expectation among the primate DNA samples. In previous studies,

genomic DNA from *Lemur catta* had given a 292 base pair (bp) *SHMT-ps1* amplicon from which sequence had been obtained [7]. In the present study, the genomic DNA sample from the ruffed lemur, *Varecia variegata*, gave no evidence of the pseudogene with either molecular beacon probe, E8/E9 or E9/E10 (Figure 1). To establish which of these results was reliable, genomic DNAs from an additional 17 lemurs covering nine living species were examined by PCR using the C1/B6 primer pair under the amplification conditions given in the Materials and methods. In no case did the lemur samples yield the expected *SHMT-ps1* amplicon, whereas the other primate controls succeeded each time they were used. On the basis of these new data, it now appears that the original *L. catta* sample did not contain pure lemur genomic DNA, and that the *SHMT-ps1* locus is not present in the prosimians as previously thought.

Thermodynamic studies of molecular beacons suggest that conformationally constrained probes display enhanced hybridization specificity as a general feature [8]. The states in which a molecular beacon can exist are a random coil, the closed hairpin/loop structure, and a linearized structure bound to the target PCR product. When two or more competing molecular beacons are present, the enhanced hybridization specificity results in the ability to discriminate between target sequences that differ by as little as a single nucleotide [8-11]. In the present study, in which molecular beacon probes were used singly as end-point reporters, they were able to tolerate multiple mismatches in the target sequences. Comparative sequences homologous to the E8/E9 and E9/E10 probes (Figure 2) from several species

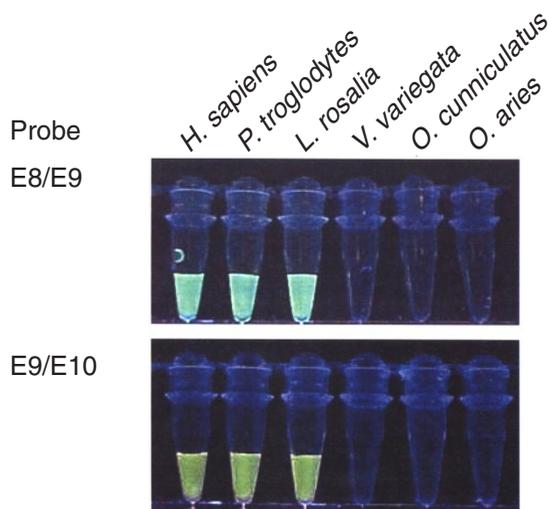


Figure 1
Results of molecular beacon end-point assays using the E8/E9 and E9/E10 probes on a representative genomic DNA panel. There is no difficulty in distinguishing a positive from a negative result in this assay. Note that the ruffed lemur sample (*V. variegata*) is clearly negative for both probes.

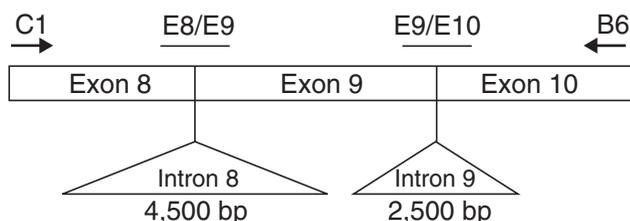


Figure 2
Representation of the *SHMT-ps1* region amplified by the C1/B6 primers. The positions of the two molecular beacon sequences E8/E9 and E9/E10 are shown in relation to the exon-exon boundaries in *SHMT-ps1*.

are shown in Table 1. As can be seen, the species-specific target sequences differed from the probe sequences by as many as nine bases, with the average difference being 2.5 bases. Confidence in the ability of these highly sensitive probes to reliably detect the presence of target sequences that differ from the probe sequence by numerous bases is supported by the observation that the most divergent targets, those belonging to the Old World monkey species for the E8/E9 probe, still gave positive, albeit weakened, results. On the other hand, potential targets that were perfect matches or that differed by only one or two bases did not give any positive indications in the lemur and non-primate mammal samples (Figure 1).

The appearance of a processed pseudogene such as *SHMT-ps1* is due to the reverse transcription of a processed mRNA followed by the random insertion of the derived cDNA into the genome via retrotransposition [12,13]. On the basis of the results presented here, refinement of the point at which those events giving rise to *SHMT-ps1* occurred must now exclude the prosimians. Kay *et al.* [14] suggest that the divergence of the line leading to the monkeys and apes (Haplorhini) from that leading to the prosimians (Strepsirrhini) occurred around the beginning of the Eocene some 55 million years ago. They also note that the divergence of the Old World monkeys and apes (Catarrhini) and the New World monkeys (Platyrrhini) probably took place before the end of the Eocene, or before 34 million years ago [14]. Given that *SHMT-ps1* is found in both New World and Old World monkeys, the window in which the reverse transcription event could have taken place is further bracketed within a 12 to 15 million year period between the lower and middle Eocene. One species that has not so far been examined that would shed light on this question is, of course, the tarsier. Should the tarsier not have *SHMT-ps1* in its genome, the window would be further constrained to a lower bound corresponding to the origin of the Anthropoidea, which is placed somewhat nearer the middle Eocene [14].

Finally, of interest here is the report of another primate pseudogene, a processed immunoglobulin J chain pseudogene,

Table 1

Aligned sequences homologous to molecular beacons E8/E9 and E9/E10

(a) Beacon E8/E9 sequence	
1 Human <i>SHMT-ps1</i>	CCGCACGATCTTCTACAG/GAAAGGAGGGC
2 Chimpanzee <i>SHMT-ps1</i>	CCGCACGATCTTCTACAG/GAAAGGAGTGC
3 Gorilla <i>SHMT-ps1</i>	CCGCACGATCTTCTACAG/GAAAGGAGGGT
4 Orangutan <i>SHMT-ps1</i>	CCGCACGATCTTCTACAG/GAAAGGAGGGC
5 Gibbon <i>SHMT-ps1</i>	CCGCACGATCTTCTACAG/GAAAGGAGGGT
6 Macaque <i>SHMT-ps1</i>	AGCCGGCATGATCTACGG/GAAAGGAGTGC
7 Baboon <i>SHMT-ps1</i>	AGCCGGCATGATCTACGG/GAAAGGAGTGC
8 Capuchin <i>SHMT-ps1</i>	CCGCACGATCTTCTACAG/GAAAGGAGGGC
9 Spider monkey <i>SHMT-ps1</i>	CCGCACGATCTTCTACAG/GAAAGGAGGGC
10 Tamarin <i>SHMT-ps1</i>	CCGCACGATCTTCTACAG/GAAAGGAGGGC
11 Human <i>SHMT-cyt</i> cDNA	TGGCATGATCTTCTACAG/GAAAGGAGBGA
12 Rabbit <i>SHMT-cyt</i> cDNA	CGGCATGATCTTCTACAG/GAGAGGAGTGC
13 Sheep <i>SHMT-cyt</i> cDNA	CGGCATGATCTTCTACAG/GAAGGGAGTGC
(b) Beacon E9/E10 sequence	
1 Human <i>SHMT-ps1</i>	CACAACCACACCATTGCTG/GGGTTGCTG
2 Chimpanzee <i>SHMT-ps1</i>	CACAACCACGCCACTGCTG/GGGTTGCTG
3 Gorilla <i>SHMT-ps1</i>	CACAACCACACCATTGTTG/GGGTTGCTG
4 Orangutan <i>SHMT-ps1</i>	CACAACCACACCATTGCTG/GGGTTGCTG
5 Gibbon <i>SHMT-ps1</i>	CACAACCACACCATTGTTG/GGGTAGCTG
6 Macaque <i>SHMT-ps1</i>	CACAACCATGCCATTGCTG/GGGTTGCTG
7 Baboon <i>SHMT-ps1</i>	CACAACCATGCTATTGCTG/GGGTTGCTG
8 Capuchin <i>SHMT-ps1</i>	CACAACCACACCATTGTTG/GGGTTGCTG
9 Spider monkey <i>SHMT-ps1</i>	CACAACCACACCATTGTTG/GGGTAGCTG
10 Tamarin <i>SHMT-ps1</i>	CACAACCAAACCAATGCTG/GGGTTTCTG
11 Human <i>SHMT-cyt</i> cDNA	CACAACCACGCCATTGCTG/GGGTTGCTG
12 Rabbit <i>SHMT-cyt</i> cDNA	CACAACCACGCCATCGCAG/GGGTGGCCG
13 Sheep <i>SHMT-cyt</i> cDNA	CACAACCACGCCATTGCTG/GGGTCGCGG

Exon/exon boundaries are indicated with a slash, and nucleotide differences from sequence 1 are in bold. Sequence sources are as follows: 1-10, all from [7]; 11, GenBank L11931; 12, GenBank Z11846; 13, GenBank X80024.

J Ψ , whose origin is via the same process at about the same time. Max *et al.* [15] examined several primate species, including Old World and New World monkey species, and concluded that this locus, too, arose in the Primates some 40 to 50 million years ago. We have recently been able to confirm this conclusion by a PCR study of the same genomic DNA panel used in this paper (E.J.D. and H. Trill, unpublished observations).

Is the fact that two retroprocessed pseudogenes were produced in the order Primates at nearly the same time unusual?

The answer to this is probably no. Goncalves *et al.* [16] suggest that retropseudogenes are a common phenomenon in mammalian genomes and estimate that the human genome contains some 23,000 to 33,000 such loci. They also note several common features among the genes giving rise to retropseudogenes: they are widely expressed; they are highly conserved; they are usually short; and they are relatively GC poor [16]. *SHMT-cyt*, the antecedent locus of *SHMT-ps1*, does, in fact, meet most of these criteria, as does the antecedent locus of J Ψ . Thus, although *SHMT-cyt* is quite large [3,4] and neither gene is particularly GC poor [3,4,15], the appearance of these two retroprocessed pseudogenes at about the same time in the evolutionary history of the primates is probably not unusual. Both may, however, prove to be valuable molecular tools in primate phylogenetic research.

Materials and methods

Genomic DNA samples

Genomic DNA samples used in this study were obtained both commercially and through the generosity of a number of colleagues. DNA samples from humans (*Homo sapiens*) as well as from chimpanzee (*Pan troglodytes*), orangutan (*Pongo pygmaeus*), tamarin (*Leontopithecus rosalia*), horse (*Equus equus*), cow (*Bovis domesticus*), dog (*Canis familiaris*), cat (*Felis domesticus*), sheep (*Ovis aries*), pig (*Sus scrofa*), rabbit (*Oryctolagus cuniculatus*), rat (*Rattus norvegicus*), hamster (*Mesocricetus auratus*), and the northern short-tailed shrew (*Blarina brevicauda*) were obtained by the author. DNA samples from gibbon (*Hyllobates lar*), macaque species (*Macaca mulatta*, *M. assamensis*, *M. nemestrina*, *M. arctoides*, and *M. fuscata*), baboon species (*Papio hamadryas* and *P. anubis*), the African green monkey (*Chlorocebus aethiops*), New World capuchins (*Cebus capucinus*, *C. apella*, and *C. albifrons*), squirrel monkey (*Saimiri boliviensis*) and spider monkey (*Ateles paniscus*) were provided by Moses Schanfield. DNA from the ruffed lemur (*Varecia variegata*) was one of a number of lemur DNAs provided by Edward Lewis (Henry Doorly Zoo, Omaha, Nebraska). DNA from the short-tailed opossum (*Monodelphis domestica*) was provided by John Vandenberg (Southwest Foundation for Biomedical Research, San Antonio, Texas) and DNA from the blind mole rat super-species *Spalax ehrenbergi* by Eviatar Nevo (University of Haifa, Israel).

PCR amplification and molecular beacon design

To date, analyses of *SHMT-ps1* in primates and other mammals have focused on a region of the pseudogene that flanks a His-Lys dimer at positions 256-257 [4,6]. This dimer is the pyridoxal-phosphate-binding site and is thus a critical functional region of the protein [17,18]. PCR amplifications of this small region of the *SHMT-ps1* sequence are carried out using the primers SHMT-C1 (5'-GGCGTGGTGC-CCTCCCATTT-3') and SHMT-B6 (5'-TGCAGTTGGCCAC-CACCTGGT-3'). Usual amplification conditions are 94°C for

3 min followed by 30 cycles of 94°C for 30 sec; 62°C for 30 sec; 72°C for 45 sec with a final 72°C extension for 5 min. PCR amplicons are visualized directly on 1.3% low melting temperature agarose gels (1x TBE) after ethidium bromide staining. Amplicons from PCR with the C1/B6 primer pair are three different sizes in primates. The shortest, 281 bp, has been seen in humans, chimpanzees, and gorillas as a result of the 11 bp deletion in those species. The longest, 317 bp, is only seen in *C. aethiops*, due to the presence of the 25 bp insertion. All other primate species so far examined (18 species in all) display a 292 bp amplicon [7].

In previous work, evidence for the absence of the *SHMT-ps1* locus in non-primate mammals was taken from a consistent failure to amplify the pseudogene sequence from genomic DNAs via PCR using a variety of primers in addition to the C1/B6 primer pair. Among these other primer pairs were sequences specific to published non-primate mammal *SHMT* cDNAs. Here, this assay is taken further by employing the more sensitive and specific method of molecular beacons [9]. Molecular beacons contain a sequence-specific probe of up to 25 bases flanked by a self-complementary stem of 5-6 bases. On the 5' end of the probe is a fluorescent moiety and on the 3' end is a dark quencher. In the ground state, the molecule is folded such that the fluorophore and the quencher are in direct contact and the transfer of fluorescent energy from the former to the latter is subsequently emitted as heat rather than as light. Thus, in the ground state, molecular beacons are dark. In the presence of a target sequence, the molecular beacon undergoes a conformational change, with the internal probe sequence hybridizing to its target and the fluorophore and quencher becoming separated by a distance sufficient to prevent quenching. In this state, the fluorophore is free to emit light in its specific wavelength [19,20].

Girgis *et al.* [21] presented the genomic organization of *SHMT-cyt*. Their data show that the 281-317 bp *SHMT-ps1* amplicon generated by the PCR primer pair C1/B6 contains parts of exons 8 and 10 as well as all of exon 9. *SHMT-cyt* genomic sequences also show that exons 8 and 9 are separated by a 4,500 bp intron and that exons 9 and 10 are separated by a 2,500 bp intron. To eliminate the possibility of false hybridization to *SHMT-cyt* genomic sequences instead of the intronless *SHMT-ps1* sequence, two molecular beacons were designed to span exon-exon boundaries (Figure 2). Following the general design rules set out by Kramer and colleagues [19,20], the sequences of these two molecular beacons are SHMT-E8/E9: 5'-FAM-**CCCGGCAC-GATCTTCTACAG/GAAAGGAGGGCCGGG**-DABCYL-3' and SHMT-E9/E10: 5'-HEX-**GCCGCACAACCACACCATTGCTG/GGGTTGCTGCGGC**-DABCYL-3'. The self-complementary hairpin sequence in each molecular beacon is in bold and the exon-exon boundaries are indicated by the slashes. Finally, the fluorophores FAM (5-carboxyfluorescein) and HEX (5-hexachlorofluorescein) are paired with the universal

fluorescence quencher DABCYL (4-(4'-dimethylaminophenylazo)benzoic acid).

As the analyses carried out in this study were end-point assays having a simple yes/no answer, the PCR amplifications were all performed in the presence of 100 ng genomic DNA. In addition to the target DNAs, 50 pmol of the PCR primers C1 and B6, 10 mM each dNTP, 2.5 U Taq polymerase, and standard Taq PCR buffer (Roche) were used. Amplification conditions were 94°C for 3 min followed by 40 cycles of 94°C for 30 sec, 62°C for 30 sec, 72°C for 45 sec and a final extension step for 7 min at 72°C. Following amplification, 250 pmol of molecular beacon probe was added to the reaction, the mixture was heated to 94°C for 3 min, and then held at 60°C before visualization of the reaction tubes on a UV transilluminator. Each set of reactions was photographed on the transilluminator in the same manner as ethidium bromide-stained agarose gels (Figure 1).

Acknowledgements

I thank Moses Schanfield, Edward Lewis, John Vandenberg and Eviatar Nevo for their generosity in providing samples of mammalian DNAs.

References

- Schirch L: **Serine hydroxymethyltransferase**. *Adv Enzymol* 1982, **53**:83-112.
- Shane B: **Folypolyglutamate synthesis and role in the regulation of one-carbon metabolism**. *Vitam Horm* 1989, **45**:263-335.
- Byrne PC, Sanders PG, Snell K: **Nucleotide sequence and expression of a cDNA encoding rabbit liver serine hydroxymethyltransferase**. *Biochem J* 1992, **286**:117-123.
- Garrow TA, Brenner AA, Whitehead VM, Chen XN, Duncan RG, Korenberg JR, Shane B: **Cloning of human cDNAs encoding mitochondrial and cytosolic serine hydroxymethyltransferases and chromosomal localization**. *J Biol Chem* 1993, **268**:11910-11916.
- Byrne PC, Shipley JM, Chave KJ, Sanders PG, Snell K: **Characterization of a human serine hydroxymethyltransferase pseudogene and its localization to 1p32.3-33**. *Hum Genet* 1996, **97**:340-344.
- Devor EJ, Dill-Devor RM: **Nucleotide sequence, chromosome localization, and evolutionary conservation of a serine hydroxymethyltransferase processed pseudogene**. *Hum Hered* 1997, **47**:125-130.
- Devor EJ, Dill-Devor RM, Magee HJ, Waziri R: **Serine hydroxymethyltransferase pseudogene, SHMT-ps1: A unique genetic marker of the Order Primates**. *J Exp Zool* 1998, **282**:150-156.
- Bonnet G, Tyagi S, Libchaber A, Kramer FR: **Thermodynamic basis of the chemical specificity of structured DNA probes**. *Proc Natl Acad Sci USA* 1999, **96**:6171-6176.
- Kostrikis LG, Tyagi S, Mhlanga MM, Ho DD, Kramer FR: **Spectral genotyping of human alleles**. *Science* 1998, **279**:1228-1229.
- Vet JAM, Majithia AR, Marras SAE, Tyagi S, Dube S, Poesz BJ, Kramer FR: **Multiplex detection of four pathogenic retroviruses using molecular beacons**. *Proc Natl Acad Sci USA* 1996, **96**:6394-6399.
- Täpp I, Malmberg L, Rennel E, Wik M, Syvänen A-C: **Homogeneous scoring of single-nucleotide polymorphisms: comparison of the 5'-nuclease TaqMan assay and molecular beacon probes**. *BioTechniques* 2000, **28**:732-738.
- Vanin EF: **Processed pseudogenes: characteristics and evolution**. *Annu Rev Genet* 1985, **19**:253-272.
- Tchenio T, Segal BE, Heidmann T: **Generation of processed pseudogenes in murine cells**. *EMBO J* 1993, **12**:1487-1497.
- Kay RF, Ross C, Williams BA: **Anthropoid origins**. *Science* 1997, **275**:797-804.

15. Max EE, Jahan N, Yi H, McBride WO: **A processed J chain pseudogene on human chromosome 8 that is shared by several primate species.** *Mol Immunol* 1994, **31**:1029-1036.
16. Goncalves I, Duret L, Mouchiroud D: **Nature and structure of human genes that generate retropseudogenes.** *Genome Res* 2000, **10**:672-678.
17. Hopkins S, Schirch V: **Properties of a serine hydroxymethyltransferase in which the active site histidine has been changed to an asparagine by site-directed mutagenesis.** *J Biol Chem* 1986, **261**:3363-3369.
18. Schirch D, Delle Fratte S, Iurescia S, Angelaccio S, Contestabile R, Bossa F, Schirch V: **Function of the active-site lysine in *Escherichia coli* serine hydroxymethyltransferase.** *J Biol Chem* 1993, **268**:23132-23138.
19. Tyagi S, Kramer FR: **Molecular beacons: probes that fluoresce upon hybridization.** *Nat Biotechnol* 1996, **14**:303-308.
20. Tyagi S, Bratu DP, Kramer FR: **Multicolor molecular beacons for allele discrimination.** *Nat Biotechnol* 1998, **16**:49-53.
21. Girgis S, Nasrallah IM, Suh JR, Oppenheim E, Zanetti KA, Mastri MG, Stover PJ: **Molecular cloning, characterization, and alternative splicing of the human cytoplasmic serine hydroxymethyltransferase gene.** *Gene* 1998, **210**:315-324.