Vol. 203, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 30, 1994 Pages 491-497

AMINO ACID POLYMORPHISMS OF THE HUMAN L-ISOASPARTYL/D-ASPARTYL METHYLTRANSFERASE INVOLVED IN PROTEIN REPAIR

William Tsai and Steven Clarke

Department of Chemistry and Biochemistry and the Molecular Biology Institute, University of California, Los Angeles, California 90024-1569

Received July 11, 1994

SUMMARY: We have analyzed DNA from three exons of the human protein-L-isoaspartate(D-aspartate) *O*-methyltransferase gene in 30 individuals. We present evidence for two polymorphisms in these regions that result in amino acid changes. At a site corresponding to amino acid position 119, we find the ATA codon for Ile at a frequency of 0.77 and the GTA codon for Val at a frequency of 0.23. At the site corresponding to amino acid position 205, we find the AAG codon for Lys at a frequency of 0.98 and the AGG codon for Arg at a frequency of 0.02. These amino acid changes may affect the ability of this enzyme to recognize and catalyze the first step in the repair of proteins spontaneously damaged in the aging process. © 1994 Academic Press, Inc.

L-Aspartyl and L-asparaginyl residues in polypeptides are especially labile to non-enzymatic degradation reactions. Their spontaneous deamidation, isomerization, and racemization proceed via a common succinimidyl intermediate and result in the formation of L-isoaspartyl, L-aspartyl, D-isoaspartyl, and D-aspartyl residues (1). Proteins containing such abnormal residues can be functionally impaired and it may be essential for cells to metabolize them in some way. Once such mechanism is the conversion of protein-bound L-isoaspartyl and D-aspartyl residues to L-aspartyl residues (and D-isoaspartyl residues) initiated by the widely dispersed *S*-adenosylmethionine-dependent protein L-isoaspartate(D-aspartate) *O*-methyltransferase (E.C. 2.1.1.77) (2-4). Recently, this repair hypothesis has been strengthened by the demonstration of the importance of the L-isoaspartyl methyltransferase activity in *Escherichia coli* stationary phase survival and heat shock resistance (5). We are now interested in more fully understanding the biological role of this enzymatic activity in repairing protein damage in higher organisms, particularly humans.

In this study, we address the question of variant forms of the human methyltransferase. Initial amino acid sequence studies of the human erythrocyte enzyme revealed a clear site of polymorphism at position 119, where isoleucine was found in two preparations and valine in another (6,7). Additionally, evidence was found for the presence of both isoleucine and leucine at position 22 (7). Since there appears to be only a single gene encoding this enzyme on human chromosome 6q22.3-6q24 (8), these amino acid changes represent allelic variations. Subsequent

0006-291X/94 $5.00
*Copyright © 1994 by Academic Press, Inc.
All rights of reproduction in any form reserved.*

analysis of a human brain cDNA library revealed two cDNA species that confirmed the polymorphism at position 119 - one of these encoded a valine residue at position 119 (GTA) and one encoded an isoleucine residue (ATA), while both had an identical ATC codon for isoleucine at position 22 (9). Furthermore, one had a AAG encoding a lysine residue at position 205 (the same as found in the protein sequence), while one had an AGG encoding an arginine residue at this same position (9). The latter nucleotide change may reflect an additional amino acid polymorphism or an artifact of the cDNA preparation.

The high frequency of amino acid differences seen in this limited number of samples suggests that the methyltransferase may be highly polymorphic on an amino acid level. Because this enzyme must recognize a wide variety of damaged proteins for repair, we were intrigued with the idea that the amino acid polymorphisms may enlarge or restrict the substrate range of this enzyme, in the same way that the abundant polymorphisms of the MHC complex allow it to recognize distinct subsets of peptides for T cell recognition (10). Significantly, alternative splicing reactions have been shown to result in the formation of two isozymes of the human methyltransferase that differ at the C-terminus (-RK versus -DEL) (9), further enlarging the number of distinct chemical forms of this enzyme. We were thus interested in looking at a larger population of individuals in order to determine the frequency of various potential alleles of the human methyltransferase gene.

MATERIALS AND METHODS

Human DNA samples:

Our studies were approved by the UCLA Human Subject Protection Committee and written informed consent was obtained from all participants. Approximately 200-300 μL of whole blood, obtained by a finger prick from healthy individuals, were mixed with 600 μL of a lysis buffer composed of 8% dodecyltrimethylammonium bromide, 1.5 M NaCl, 50 mM EDTA and 100 mM Tris-HCL (pH 8.6) (11). After incubation of the sample at 68°C for 5 min, 900 μL of chloroform:isoamyl alcohol (24:1) was added. The sample was mixed by gentle inversion, and centrifuged at 16,000 g for 2 min. The aqueous layer, containing the genomic DNA, was aliquoted into a new tube. 900 μL of H_2O and 100 μL of 5% cetyltrimethylammonium bromide made up in 0.4 M NaCl was added to the sample. Again, the sample was mixed by gentle inversion, followed by centrifugation at 16,000 g for 2 min. After decanting the supernatant, the DNA pellet was resuspended in 300 μL of 1.2 M NaCl. 750 μL of ethanol was added, and the sample was centrifuged at 16,000 g for 2 min. Finally, the DNA pellet was rinsed with 70% ethanol, and resuspended in 30 μL of H_2O. The yield of DNA from 300 μL of whole blood is approximately 5 μg (11). DNA samples were also obtained from Dr. Jake Lusis (University of California, Los Angeles).

Polymerase chain reaction (PCR) amplification:

Amplifications of exon 2, exon 5, and exon 7a were done in a final volume of 50 μL. Each reaction sample contained 200-300 ng of genomic DNA from a single individual (estimated by the fluorescence intensity of ethidium bromide staining), 40 pmol of each of the flanking primers, 0.2 mM of each of the four deoxynucleotide triphosphates, 2.0 mM of $MgCl_2$, 1x buffer from the supplier, and 2.5 U of Taq DNA Polymerase (Promega). The primers were synthesized using ß-cyanoethyl N,N-diisopropylphosphoramidite chemistry in a Gene Assembler Plus DNA synthesizer (Pharmacia LKB Biotechnology). PCR cycling conditions were 95°C for 2 min (without enzyme), then 35-40 cycles of 95°C for 30 sec and 58°C for 30 sec. To ethanol precipitate the PCR-amplified product, 3 volumes of ethanol and 1/10 volume of 3 M sodium acetate (pH 4.8) were added and centrifuged for 30 min at 16,000 g. The amplified DNA template was rinsed with 70% ethanol and resuspended in 20 μL of H_2O.

DNA sequencing:

Direct sequencing of the respective DNA fragments with both the forward and reverse primers were subsequently performed using the Δ Taq Sequencing Kit (2.0, United States Biochemicals).

Restriction endonuclease digestion:

*Hae*III restriction enzyme digestion was performed on the PCR amplified product from exon 7a (see above). Amplified DNA was resuspended in 16 μL of H_2O and 20 units of *Hae*III and 1x buffer from the supplier (Gibco BRL) was added and the sample incubated at 37°C for 2 h. Products were analyzed on a 2% agarose gel.

RESULTS

From a sample of 30 individuals, we amplified genomic DNA encoding three regions of the L-isoaspartyl/D-aspartyl methyltransferase polypeptide (Fig. 1). One fragment (272 bp) provided information on DNA encoding amino acids 19-52 in exon 2 where evidence for a leucine/isoleucine polymorphism at position 22 had been reported (7). A second fragment (118 bp) provided information on DNA encoding amino acids 108-129 in exon 5, where a

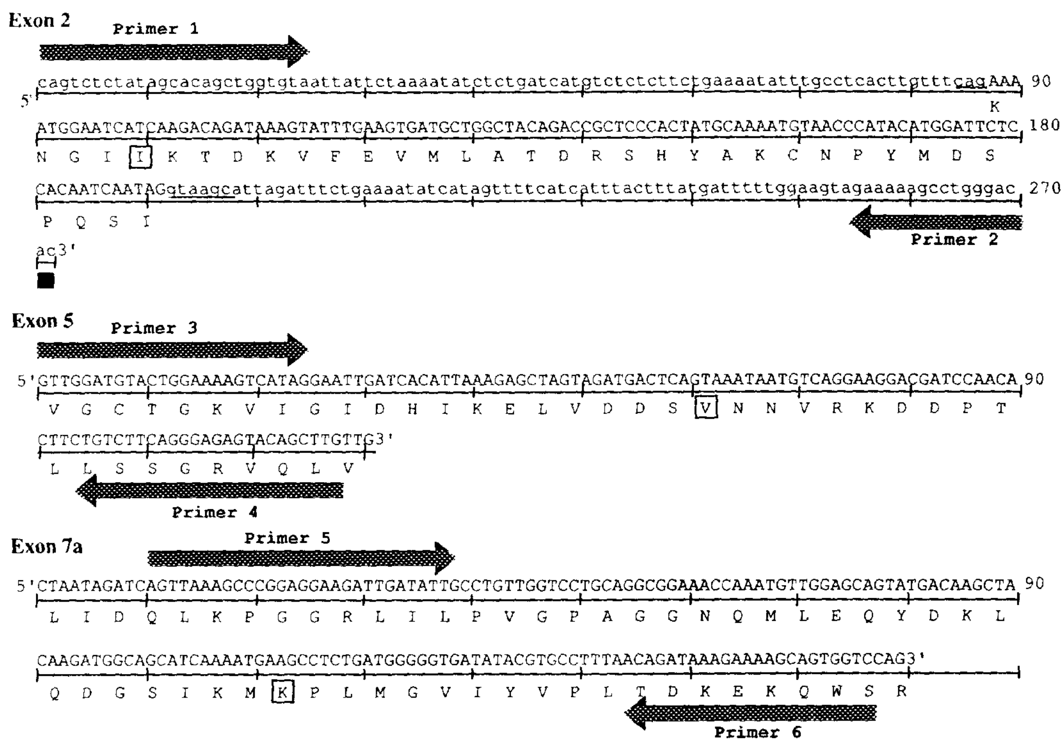

Figure 1. Sequence strategy for exons 2, 5 and 7a of the human L-isoaspartyl (D-aspartyl) methyltransferase. Primers 1-6 were synthesized based on the sequence of human cDNA (9) or genomic DNA (W. Tsai and S. Clarke, unpublished data). Boxed amino acids represent the positions of the possible polymorphisms. Upper case letters represent exon sequence and lower case letters represent intron sequence. Underlined nucleotides represent consensus splicing sequences.

valine/isoleucine polymorphism at position 119 had been found (7). Finally, a fragment (157 bp) gave information on DNA encoding amino acids 181-215 in exon 7a, where a possible lysine/arginine polymorphism at position 205 was reported (9). The amplified fragments were subsequently analyzed on a 2% agarose gel in order to verify both the presence of the expected size products and the absence of DNA contamination in controls performed without the addition of DNA template. DNA sequencing and/or restriction endonuclease analysis of each amplified product was then performed. In sequencing exon 2, we determined approximately 110 bp and 95 bp from primers 1 and 2, respectively. In sequencing exon 5, we determined approximately 80 bp and 85 bp from primers 3 and 4, respectively. Finally, in sequencing exon 7a, we determined approximately 115 bp and 120 bp from primers 5 and 6, respectively.

We find that the most prevalent polymorphism is at position 119, corresponding to exon 5 (Table I). A first position nucleotide change from ATA, which codes for Ile, to GTA changes the amino acid in this position to Val. Out of the 30 individuals screened, 2 individuals were Val/Val homozygotes, 10 individuals were Ile/Val heterozygotes, and 18 individuals were Ile/Ile

Table I
Amino Acid Polymorphisms of the Human L-Isoaspartyl/D-Aspartyl Methyltransferase

Sample	Race	Sex	Age	Amino Acid 22	Amino Acid 119	Amino Acid 205
1	Caucasian	F	47	Ile	Ile	Lys
2	Caucasian	F	46	Ile	Ile	Lys
3	Asian	M	23	Ile	Ile	Lys
4	Asian	F	22	Ile	Ile	Lys
5	Caucasian	M	23	Ile	Ile	Lys*
6	Asian	F	26	Ile	Ile	Lys
7	Asian	M	23	Ile	Ile	Lys
8	Caucasian	M	34	Ile	Ile	Lys*
9	Caucasian	M	31	Ile	Ile	Lys*
10	Caucasian	M	24	Ile	Ile	Lys*
11	Caucasian	M	24	Ile	Ile	Lys*
12	Caucasian	M	31	Ile	Ile	Lys*
13	Caucasian	F	27	Ile	Ile	Lys*
14	Caucasian	F	26	Ile	Ile	Lys*
15	Asian	F	23	Ile	Ile	Lys*
16	African	M	44	Ile	Ile	Lys*
17	Hispanic	F	33	Ile	Ile	Lys*
18	Asian	M	27	Ile	Ile	Lys/Arg
19	Caucasian	F	37	Ile	Ile/Val	Lys*
20	Asian	M	23	Ile	Ile/Val	Lys*
21	Hispanic	F	23	Ile	Ile/Val	Lys*
22	Caucasian	M	22	Ile	Ile/Val	Lys*
23	Asian	M	23	Ile	Ile/Val	Lys*
24	Caucasian	M	68	Ile	Ile/Val	Lys
25	Caucasian	M	38	Ile	Ile/Val	Lys
26	Caucasian	F	25	Ile	Ile/Val	Lys
27	Caucasian	F	47	Ile	Ile/Val	Lys
28	Caucasian	M	44	Ile	Ile/Val	Lys
29	Caucasian	?	?	N/A	Val	Lys
30	Caucasian	F	54	Ile	Val	Lys

*Lys detected by lack of *Hae*III polymorphism.

homozygotes (Table I). The calculated frequency for the Val and Ile alleles at this position are 0.23 and 0.77, respectively (Table II). Applying the Hardy-Weinberg equilibrium relationship, the expected frequency of homozygote Val is 0.054 (1.6 individuals), the expected frequency of homozygote Ile is 0.59 (17.6 individuals), and the expected frequency of heterozygote Val/Ile is 0.36 (10.6 individuals). These expected frequencies match closely the experimentally determined frequencies.

Of the 29 individuals screened for the possible polymorphism at position 22, corresponding to exon 2, we did not confirm the presence of the polymorphism; all of the individuals screened were homozygote Ile (Table I). Because the Leu allele previously identified in this position was initially found through amino acid sequencing (7), our data thus suggest that the Leu variant is a rare polymorphism.

However, of the 30 individuals screened for the possible polymorphism in position 205, corresponding to exon 7a, we identified one sample with both the lysine and arginine alleles (Table I). This heterozygous variant was confirmed both by direct DNA sequencing and *Hae*III restriction digest analysis in multiple replicates. DNA sequencing initially detected a second position nucleotide change from AAG, which codes for Lys to AGG, which codes for Arg. Because the two nucleotides immediately downstream of position 205 are CC (Fig. 1), an AAG to AGG change introduces a *Hae*III restriction site (GGCC). *Hae*III digestion of this variant sample resulted in the partial cleavage of the exon 7a 157 bp fragment to give 53 and 104 bp fragments that was visualized by ethidium bromide staining of the agarose gel.

Finally, in these analyses of individual genomic DNA sequences, we did not detect any clear examples of additional nucleotide polymorphisms.

DISCUSSION

In this study, we found that approximately one out of three individuals would be expected to be heterozygous at position 119 of the gene encoding the human methyltransferase, and a

Table II
Frequency of Polymorphisms of the Human L-Isoaspartyl/D-Aspartyl Methyltransferase

	Amino Acid Residue	Nucleotide Encoding Sequence	Allele Frequency
Exon 2	Ile_{22}	ATC	>0.98
	Leu_{22}	?	<0.02
Exon 5	Ile_{119}	ATA	0.77
	Val_{119}	GTA	0.23
Exon 7a	Lys_{205}	AAG	0.98
	Arg_{205}	AGG	0.02

smaller fraction of individuals would be expected to be heterozygous at position 22 and 205. The three variants are located outside the conserved motifs associated with *S*-adenosylmethionine binding (12). Although the amino acids at each these sites are similar (Ile/Val, Ile/Leu, and Lys/Arg), the difference in the length of the side chains may have a significant effect on the binding of different L-isoaspartyl-containing peptide substrates by the methyltransferase, and can thus affect the catalytic efficiency of the repair activity.

The presence of at least three amino acid polymorphisms within the methyltransferase gene suggests that the heterogeneity has a useful function. Previous studies had suggested an average variation in only about 3/10,000 non-degenerate coding sequences in human genes (13). In this study, we sequenced about 270 bp of the 678/684 bp encoding the L-isoaspartyl/D-aspartyl methyltransferase, and thus should expect about 4.9 variants in the 30 samples examined but actually found 15. Since the rate of variation in degenerate positions was two to four times higher than that in non-degenerate positions (13), we would expect a correspondingly higher frequency of DNA changes that do not result in amino acid changes. Surprisingly, however, we found no evidence in this study for any DNA variation that did not result in amino acid changes.

It is interesting to note that evidence exists for even more extensive polymorphisms in the wheat L-isoaspartyl methyltransferase (14). Here, there appear to be at least 12 amino acid variants. One of these polymorphic sites occurs at a position corresponding to the Lys/Arg residue at position 205 in the human enzyme, where both Arg and Val are detected in the wheat enzyme (14).

Polymorphisms identified within protein coding regions are especially important because of their possible phenotypic and clinical consequences. There are numerous examples in which a polymorphism identified at a specific site within a genetic locus can affect susceptibility to human disease. For instance, a polymorphism at codon 129 of the prion protein is associated with the age of onset and duration of the prion disease (15). Likewise, polymorphic variants of the ApoE protein have been shown to contribute to the development of familial combined hyperlipidaemia (16). The presence of polymorphisms can also confer resistance to disease. For example, heterozygosity of the gene encoding the human enzyme glucose-6 phosphate dehydrogenase can confer a selective advantage on individuals living where malaria is endemic (17).

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institute of Health (GM 26020). We thank Duncan MacLaren for his many contributions to the early stages of this work, and Dr. Jake Lusis for assistance in obtaining some of the DNA samples used in this work.

REFERENCES

1. Clarke, S., Stephenson, R. C., and Lowenson, J. L. (1992) in *Stability of Protein Pharmaceuticals, Part A*. Chemical and Physical Pathways of Protein Degradation (Ahern, T.J. and Manning, M.C., eds.) pp. 1-29, Plenum Press, New York.

2. McFadden, P. N. and Clarke, S. (1987) *Proc. Natl. Acad. Sci. USA.* **84**, 2595-2599.
3. Johnson, B. A., Murray, E. D. Jr., Clarke, S., Glass, D.B., and Aswad, D. W. (1987) *J. Biol. Chem.* **262**, 5622-5629.
4. Galletti, P., Ciardiello, A., Ingrosso, D., Di Donato, A., and D'Alessio, G. (1988) *Biochemistry.* **27**, 1752-1757.
5. Li, C. and Clarke, S. (1992) *Proc. Natl. Acad. Sci. USA. 89* 9885-9889.
6. Gilbert, J. M., Fowler, A., Bleibaum, J., and Clarke, S. (1988) *Biochemistry* **27**, 5227-5233.
7. Ingrosso, D., Fowler, A.V., Bleibaum, J. and Clarke, S. (1989) *J. Biol. Chem.* **264**, 20131-20139.
8. MacLaren, D.C., O'Connor, C.M., Xia, Y., Mehrabian, M., Klisak, I., Sparkes, R. S., Clarke, S., and Lusis, A.J. (1992) *Genomics* **14**, 852-856.
9. MacLaren, D. C., Kagan, R. M., and Clarke, S. (1992) *Biochem. Biophys. Res. Commun.* **185**, 277-283.
10. Abastado, J.P., Casrouge, A., and Kourilsky, P. (1993) *J. Immunol.* **151**, 3569-3575.
11. Gustincich, S., Manfioletti, G., Del Sal, G., Schneider, C., and Carninci, P. (1991) *Biotechniques*, **11**, 298-301.
12. Kagan, R. M., and Clarke, S. (1994) *Arch. Biochem. Biophys.* **310**, 417-427.
13. Li, W.-H., and Sadler, L.A. (1991) *Genetics* **129**, 513-523.
14. Mudgett, M.B. and Clarke, S. (1993) *Biochemistry* **32**, 11100-11111.
15. Ridley, R. M. and Baker, H.F. (1993) *Develop. Biol. Stand.* **80**, 15-23.
16. Davignon, J., Gregg, R.E., Sing, C.F. (1988) *Arteriosclerosis*, **8**, 1-21.
17. Vulliamy, T., Mason, P., Luzzatto, L. (1992) *Trends Genetics*, **8**, 138-43.