Vol. 185, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 29, 1992 Pages 277-283

Alternative Splicing of the Human Isoaspartyl Protein Carboxyl Methyltransferase RNA Leads to the Generation of a C-terminal -RDEL Sequence in Isozyme II

Duncan C. MacLaren, Ron M. Kagan, and Steven Clarke

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

Received April 14, 1992

We have isolated two cDNA clones that correspond to the mRNAs for two isozymes of the human L-isoaspartyl/D-aspartyl protein carboxyl methyltransferase (EC 2.1.1.77). The DNA sequence of one of these encodes the amino acid sequence of the C-terminal half of the human erythrocyte isozyme I. The other cDNA clone includes the complete coding region of the more acidic isozyme II. With the exception of potential polymorphic sites at amino acid residues 119 and 205, the deduced amino acid sequences differ only at the C-terminus, where the -RWK sequence of isozyme I is replaced by a -RDEL sequence in isozyme II. The latter sequence is identical to a mammalian endoplasmic reticulum retention signal. With the previous evidence for only a single gene for the L-isoaspartyl/D-aspartyl methyltransferase in humans, and with evidence for consensus sites for alternative splicing in corresponding mouse genomic clones, we suggest that alternative splicing reactions can generate the major isozymes previously identified in human erythrocytes. The presence of alternative splicing leads us to predict the existence of a third isozyme with a -R C-terminus. The calculated isoelectric point of this third form is similar to that of a previously detected but uncharacterized minor methyltransferase activity.

Three classes of protein carboxyl methyltransferases have been found to date. They are distinguished by their methyl–acceptor substrate specificity (1-3). The type II methyltransferase catalyzes the transfer of a methyl group from S–adenosyl–L–methionine to the free carboxyl groups of D–aspartyl and L–isoaspartyl residues. These substrates are spontaneous isomerization and racemization products of L–aspartyl and L–asparginyl residues. Upon methylation, the methyl esters undergo spontaneous hydrolysis and the original residue configuration may be restored. The action of this class of enzymes has been proposed as a repair process to counteract non-enzymatic damage to aging proteins (4).

0006-291X/92 $4.00

The type II methyltransferase has been extensively studied in mammalian tissues. Two similar activities have been isolated from bovine brain (5) and human erythrocytes (6, 7). These isozymes are monomeric polypeptides of about 25,000 Da and have similar catalytic properties, but differ by about 1 pH unit in isoelectric point (7). A third activity has been noted, with an isoelectric point between isozymes I and II, but the isozyme has not been isolated or characterized (8). The complete amino acid sequences for isozyme I from bovine brain (9) and human erythrocytes (10) as well as most of that of isozyme II from human erythrocytes (11) have been presented. The sequence of the human isozyme I is 96% identical with the bovine enzyme. The amino acid sequences of the two human isozymes appear to be identical with the exception of two amino acids at the C–terminus that can account for the difference in isoelectric point (11). Southern blot analysis suggested that both isozymes were products of a single gene (11). Nucleotide sequences that code for enzymes similar to isozyme I in rat brain and mouse testis have recently been determined (12, 13). We now present a cDNA sequence encoding human isozyme I and the first eukaryotic sequence encoding isozyme II. We suggest that these isozymes, as well as a potentially new species, can arise by alternative splicing of a single gene transcript.

MATERIALS AND METHODS

cDNA Library Synthesis and Clone Screening— a cDNA library constructed from the temporal cortex of the brain of a 2 year–old female human was purchased from Stratagene (#935205). The cDNA was synthesized from oligo–dT isolated mRNA, and packaged into the *Eco*RI sites of their lambda ZAP bacteriophage vector. The library was propagated in *E. coli* BB4 and 22 plates containing 5 x 10^5 plaques each (1.1 x 10^7 plaques total) were screened using a radio–labeled 769 bp *HaeIII* fragment from the coding region of a 1580 bp murine methyltransferase cDNA (13). The fragment was labelled with [α–^{32}P]–dCTP to a specific activity of 10^9 cpm/μg with the PRIME–IT random priming kit (Stratagene). Standard plaque lift and Southern blot procedures (14) produced three positive signals. The clones for these plaques were isolated by subsequent screenings. The clones were repackaged into plasmids in XL1–Blue cells via *in vivo* excision according to λZAP protocol. Successful excision was denoted by ampicillin resistance. The cells containing the insert carrying plasmids of interest were grown in LB/Ampicillin medium, and their plasmids isolated and purified using Qiagen plasmid isolation columns.

Nucleotide Sequence Determination and Analysis— The nucleotide sequences of the clones were determined on both strands by the dideoxy chain–terminating method (15) using the Sequenase 2.0 kit (USB), M13 and T7 universal primers, and synthesized 22mer primers. The sequence data were analyzed with DNAStar programs on a Macintosh computer.

Murine methyltransferase oligonucleotide primers and the 1580 bp murine methyltransferase cDNA clone were generous gifts of Dr. Clare O'Connor (Worcester Foundation for Experimental Biology, Shrewsbury, MA).

RESULTS AND DISCUSSION

Isolation of Human Brain cDNA Clones for the Isoaspartyl Methyltransferases— A cDNA library of the brain of a 2 year–old female human was screened with a radio–labeled probe against the coding region of a murine methyltransferase cDNA (13).

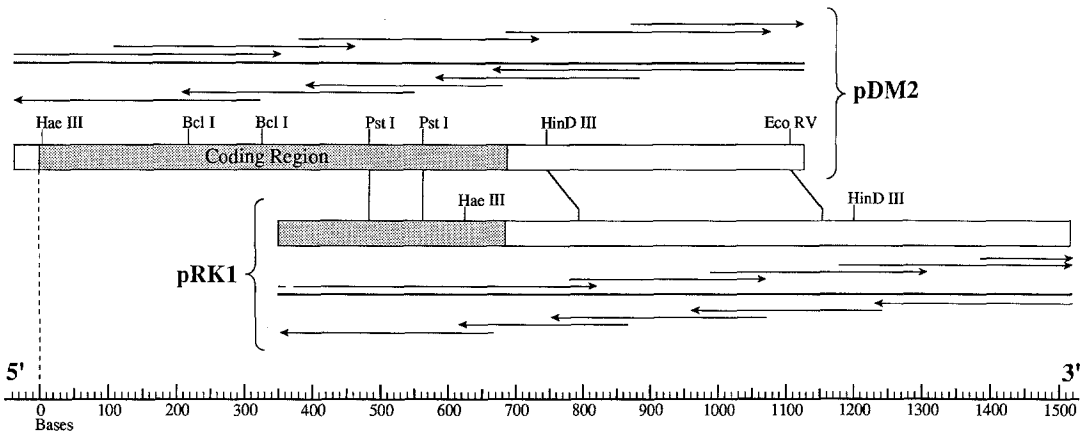

FIG. 1. **Restriction map and sequencing strategy of methyltransferase cDNA inserts.** Coding and untranslated regions are represented by the *shaded* and *open* boxes respectively. Sequence data were obtained as indicated by length and direction of the arrows. The inserts are bounded by *Eco*RI linker sites.

Three clones out of 1.1 x 10^7 plaques gave a positive signal and were isolated. The sequences of two of the clones were determined from both strands (Figs. 1 & 2). Sequence analysis of the insert of pDM2 identified 37 bases of 5'–untranslated, 684 bases of coding region, and 447 bases of 3'–untranslated nucleotides including the stop codon and a vestigial 8 base poly(A) tail for a total length of 1167 bases. The insert of pRK1 was missing the 5' half of the coding region but had a much longer 3'–untranslated end. Its sequence started at the codon for amino acid 119. The insert of pRK1 included a coding region of 324 bases and a 3'–untranslated region of 844 bases for a total length of 1168 nucleotides. The sequencing strategy, common and unique regions, and restriction endonuclease sites of the cDNA clones are shown in Fig. 1. Preliminary sequence analysis identified one clone, pDM1, as two 5' halves of a methyltransferase coding region with one half inverted and concatenated onto the other, i.e., abutted at the 3' ends of both halves. This structure was present in the original lambda ZAP vector and may have been an artifact of the library construction.

Analysis of Nucleotide and Deduced Amino Acid Sequences— The analysis of the pDM2 cDNA identified an open reading frame encoding 228 residues including the initiator methionine (Fig. 2). This sequence is identical to that of the experimentally derived amino acid sequence of human isozyme II (11) except at the C–terminus. Here, protein sequencing gave a –RDD end (11) whereas the nucleotide sequence codes for a –RDEL C–terminus. Either of these ends would give a similar isoelectric point. It is likely that the actual C–terminal sequence of isozyme II is –RDEL as suggested by the cDNA isolated here. The previous chemical assignment of –RDD was based on the sequence of a *Staphylococcus aureus* protease V8 fragment. The presence of the E residue would result in a cleavage releasing the C-terminal L residue. The apparent presence of a second D residue rather than a E residue in the Edman sequencing may have resulted from carryover of the first D residue combined with a low yield of the terminal E residue

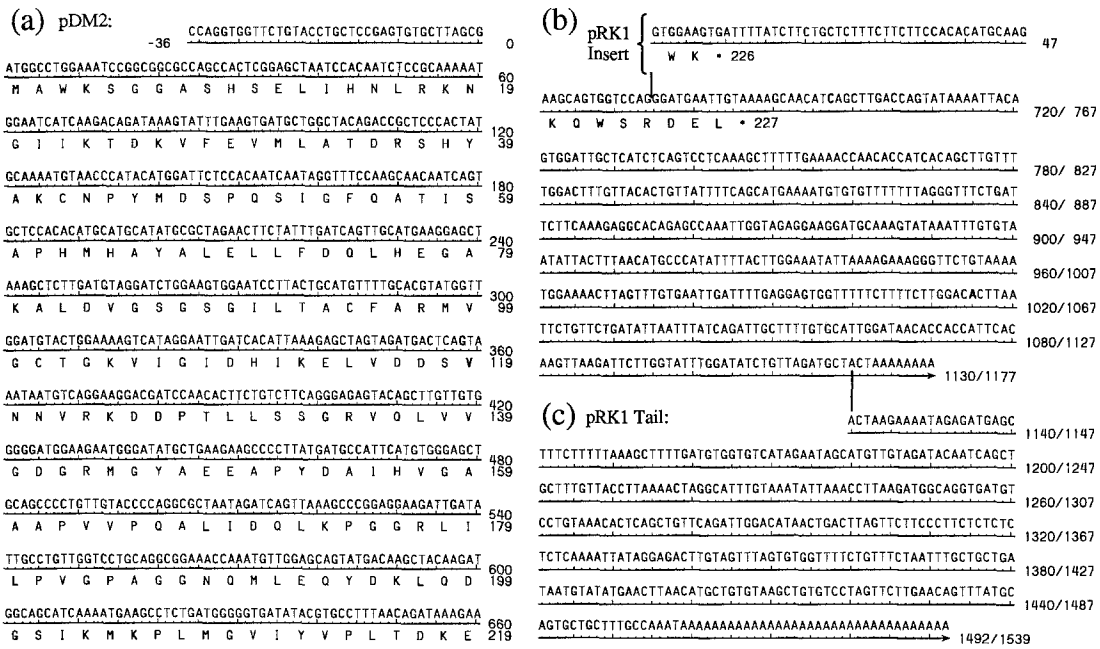

FIG. 2. **Nucleotide and encoded amino acid sequences of L-isoaspartyl/D-aspartyl methyltransferase cDNAs pDM2 and pRK1.** The A of the initiator ATG in pDM2 is numbered one and the encoded methionine is numbered zero. This is done to match the numbering scheme of the final protein due to the excision of the initiator methionine. Numbering for both clones begins at these positions. Clone pRK1 begins at position 358. Numbers to the left of the divisor represent pDM2. Those to the right of the divisor are for pRK1. The nucleotide and encoded amino acid sequences of pDM2 is shown under (a) and continues to (c). (b) and (c) represent the 47 base insert and the extended tail found in clone pRK1.

of the peptide (11). The N-terminal sequence suggests that the initiator methionine residue will be excised and the penultimate alanine will be acetylated (16). This assignment is consistent with the previous chemical analysis of isozyme II (6, 10). The calculated molecular weight of 24,592 is in good agreement with the empirically determined weight of 25,000 (7). Taken together, these results suggest that the pDM2 insert is derived from the mRNA encoding isozyme II.

Comparison of the cDNA clones showed that pRK1 was missing a 5'–untranslated region and the nucleotides coding for the first 118 residues. The nucleotides in the coding region translate exactly to those determined experimentally for the C–terminal half of isozyme I (10). Assuming the missing portion of the coding region for pRK1 matches that of pDM2, as protein sequence analysis would indicate, the coded protein would be a monomer 226 residues long, with a molecular weight of 24,549, again in agreement with the experimentally determined value (7, 8). These results suggest that pRK1 is derived from the mRNA for isozyme I.

Sequence Comparisons— Fig. 3 summarizes the differences between the nucleotide and deduced amino acid sequences of the two human cDNAs. Using the numbering system for pDM2 we see that the most obvious difference occurs with a 47 base insert in pRK1 after position 674—changing the –RDEL C–terminus in pDM2 to –RWK for pRK1. These changes make isozyme I

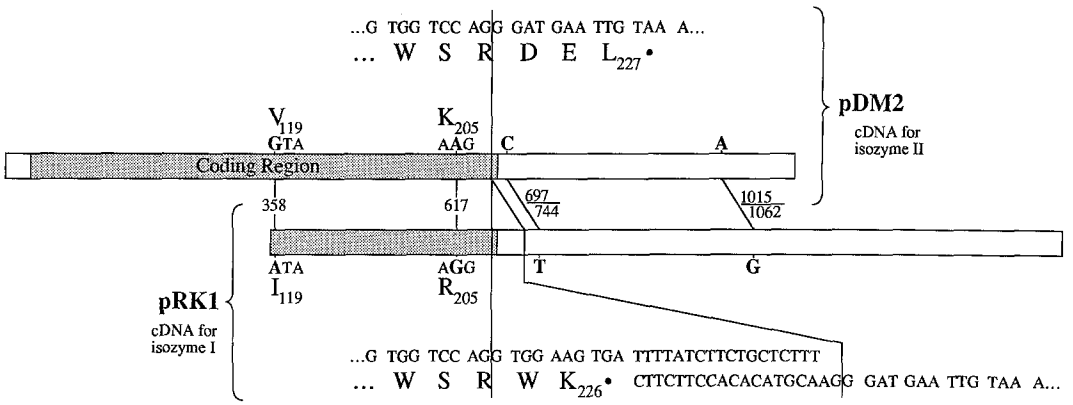

FIG. 3. **Nucleotide and amino acid comparison between the cDNAs pDM2 and pRK1.** Coding and untranslated regions are represented by the *shaded* and *open* boxes respectively. Differences in nucleotide and encoded amino acids are shown and numbered with respect to the start site of pDM2. A, of the initiator ATG, is numbered one and the encoded methionine is numbered zero. Splice site determination was aided by comparison with mouse genomic DNA.

one residue shorter and considerably more basic— the calculated pI for isozyme I is 6.72 versus a pI of 6.13 for isozyme II. These values can be compared with the experimentally determined isoelectric points of 6.6 and 5.5 (7). Besides the major difference in the coding for the C-terminus, there are a total of four other nucleotide changes. Since it appears that there is only a single gene for this methyltransferase (11), these results suggest the presence of genetic polymorphism. Two of the base changes occur in the coding region and cause amino acid changes. A change of G to A at nucleotide 358 changes V_{119} to I in pRK1, and a change of A to G at nucleotide 617 changes K_{205} to R. Protein sequence analysis of isozyme I in human erythrocytes shows both valine and isoleucine at position 119. There may be a similar polymorphism at position 205, but arginine has not been detected at this site. The other two nucleotide changes occur in the 3'–untranslated region.

Neither sequence has any obvious polyadenylation signal, e.g., AATAAA, though the sequences near the 3' end are rich in thymidine bases, which are necessary for the pre–mRNA cleavage reaction and are traditionally thought to be close to the polyadenylation site (17).

Alignment of the mouse 1580 bp cDNA (13) with human clone pRK1 gives a 78.2% identity at the nucleotide level (Martinez/Needleman-Wunsch method). Alignment of the nucleotides in the coding regions gives an 89.3% identity, reflecting 95.2% identity of the encoded amino acids. These results indicate a high degree of conservation of the mouse and human DNA sequences.

Alternative Splicings and a Third Isozyme— From the sequences of mouse methyltransferase genomic DNA and cDNAs (13), the sequences of the human methyltransferase cDNAs reported here, and the consensus sequences for alternative splicing (18), it is possible to suggest the origin of the mRNAs for human isozyme I and II. Figure 4 summarizes how alternative splicing could lead to the formation of two mRNAs that code for a –RWK C-terminus (Mouse 825, and Mouse 1580 and Human pRK1) and a third mRNA that encodes the –RDEL terminus (Human

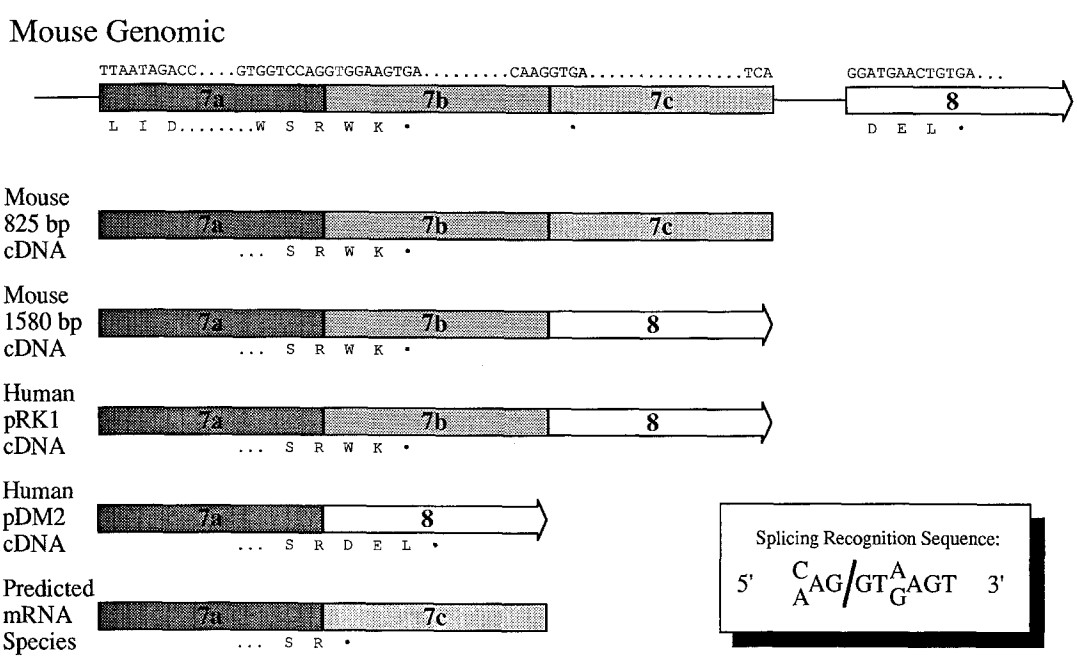

FIG. 4. **Alternate mRNA splicing patterns found in mouse and human isoaspartyl methyltransferase cDNA clones.** A schematic of the exon pattern and splicing boundaries of mouse exons 7 and 8 are shown at the top of the figure. The four drawings below the genomic layout show the known splicing patterns found in mouse (13) and human cDNAs. Mouse 1580 and human pRK1 have the same pattern. The existence of cDNA pDM2 and the combination of these possible splicing reactions suggests the existence of a third mRNA species – 7a/7c. The box at the bottom right shows the splicing boundary recognition sequence.

pDM2). From the known mouse splicing reactions and the mammalian splicing recognition sequence (18) we propose the formation of a third methyltransferase isozyme, 224 residues long, with a –R C-terminus as shown at the bottom of Fig. 4. The calculated isoelectric point of this third isozyme would be about 6.48 and is equivalent to that of a minor form with a measured pI value intermediate between the two major species – 6.0 (8). The ratio of the pI differences, (III–II)/(I–III), for the experimental and theoretical values are 1.5 and 1.46 respectively, and is consistent with our prediction of this third methyltransferase isozyme.

Does the C-terminal -RDEL sequence of isozyme II direct its localization in the cell?— Secreted proteins with -KDEL or –RDEL C-termini are marked for retention in the endoplasmic reticulum (19). Is it possible that the presence of such a sequence in isozyme II of the human L-isoaspartyl/D-aspartyl protein methyltransferase fulfills a similar role? In intact red cells, both isozymes are cytosolic (6, 7). However, the process of red cell maturation involves the loss of the endoplasmic reticulum in the reticulocyte stage and proteins located in the lumen would be released in the cytosol. In other cells types, including brain, sub-cellular fractionation studies have suggested that some protein methyltransferase activity is localized in the membrane fraction (20). Although there is no evidence for N-terminal signal sequences in these enzymes, additional splicing reactions at the 5'–end of the mRNA might generate a species that could be either secreted or localized to the lumen of the endoplasmic reticulum.

ACKNOWLEDGMENTS

This work was supported by USHHS National Institutional Research Service Awards T32 GM08375 to D.C.M. and T32-GM07185 to R.M.K., and by grant GM26020 from the National Institutes of Health. Thanks are due to Tina Marie-Glatt, Joan and Ken for experimental insights, and a very special thank you to Dr. Clare O'Connor (Worcester Foundation for Experimental Biology, Shrewsbury, MA) for helpful discussions, providing essential materials, and sharing results before publication.

REFERENCES

1. Clarke, S. (1985) *Annu. Rev. Biochem.* **54**, 479-506.
2. Barten, D. M. and O'Dea, R. F. (1990) *Life Sciences* **47**, 181-194.
3. Ota, I. M. and Clarke, S. (1990) in *Protein Methylation* (W. K. Paik and S. Kim, Eds.), pp. 179-194. CRC Press, Florida, Boca Raton, Florida.
4. McFadden, P. N. and Clarke, S. (1982) *Proc. Natl. Acad. Sci. USA.* **79**, 2460-2464.
5. Aswad, D. W. and Deight, E. A. (1983) *J. Neurochem.* **40**, 1718-1726.
6. Gilbert, J. M., Fowler, A., Bleibaum, J. and Clarke, S. (1988) *Biochemistry* **27**, 5227-5233.
7. Ota, I. M., Gilbert, J. M. and Clarke, S. (1988) *Biochem. Biophys. Res. Commun.* **151**, 1136-1143.
8. Ingrosso, D. and Clarke, S. (1991) in *Adv. Exper. Med. Biol.* **307**, 263-276.
9. Henzel, W. J., Stults, J. T., Hsu, C.-A. and Aswad, D. W. (1989) *J. Biol. Chem.* **264**, 15905-15911.
10. Ingrosso, D., Fowler, A. V., Bleibaum, J. and Clarke, S. (1989) *J. Biol. Chem.* **264**, 20131-20139.
11. Ingrosso, D., Kagan, R. M. and Clarke, S. (1991) *Biochem. Biophys. Res. Commun.* **175**, 351-358.
12. Sato, M., Yoshida, T. and Tuboi, S. (1989) *Biochem. Biophys. Res. Commun.* **161**, 342-347.
13. Romanik, E. A., Killoy, L. C., Ladino, C. A., Desrosiers, R. R. and O'Connor, C. M. (1992) *Gene, in press* .
14. Sambrook, J., Fritch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual, 2nd Ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
15. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
16. Persson, B., Flinta, C., von Heijne, G. and Jornvall, H. (1985) *Eur. J. Biochem.* **152**, 523-527.
17. Wickens, M. (1990) *Trends Biochem. Sci.* **15**, 277-281.
18. Smith, C. W. J., Patton, J. G. and Nadal-Ginard, B. (1989) *Annu. Rev. Gen.* **23**, 527-77.
19. Pelham, H. R. B. (1990) *Trends Biochem. Sci.* **15**, 483-486.
20. Sellinger, O. Z., Kramer, C. M., Fischer-Bovenkerk, C. and Adams, C. M. (1986) *Neurochem. Int.* **10**, 155-166.