BIOCHEMICAL LETTERS

Human aldehyde dehydrogenase E3: the N-terminal primary structure

It is not unusual to find errors in DNA sequence databases [1]. There are two types of sequencing errors: the substitution of one base for another without changing the reading frame, and the insertion or deletion of one or more bases. The first type of error is of less consequence when predicting a protein sequence, owing to the nature of the genetic code, in which a triplet of nucleotides encodes one amino acid residue. The second type of error will result in a reading frameshift that most likely will lead to complete loss of homology in the alternatively translated amino acid sequence. This may well perplex researchers in working out the correct amino acid sequence of the protein. The human aldehyde dehydrogenase (EC 1.2.1.3) E3 isoenzyme, which has a broad substrate specificity, is active with short-chain aliphatic aldehydes, but also metabolizes aminoaldehydes, such as γ -aminobutyraldehyde [2], metabolites of polyamines [3] and betaine aldehyde [4]. The partial primary structure (462 amino acid residues) obtained by sequencing human cDNA [5] was verified by comparison with the amino acid sequence of about one third of the E3 isoenzyme. The primary structure of the E3 isoenzyme resembled *Escherichia coli* betaine aldehyde dehydrogenase (52.7% positional identity) more closely than that of any other aldehyde dehydrogenase.

More recently, a complete coding nucleotide sequence (under a new name, ALDH9) and the deduced amino acid sequence (493 amino acid residues) have been reported by Lin et al. [6]. The C-terminal end (462 amino acid residues) agreed with that reported by Kurys et al. [5], except for two polymorphic nucleotide positions. The deduced new sequence of 31 amino acid residues at the N-terminal end has never been confirmed by any peptide sequence derived from the enzyme protein.

Recently, we have purified an E3 isoenzyme equivalent from rat liver mitochondria (M.-K. Chern and R. Pietruszko, unpublished work). Four peptides derived from tryptic digest of the protein were sequenced. Three of these sequences either matched exactly or aligned well with the deduced amino acid sequence from Kurys et al. [5] or Lin et al. [6], indicating homology between the rat enzyme and the human E3 isoenzyme. Three matching segments (corresponding to peptides 2, 3 and 4) were from amino acids 89–97, 227–238 and 326–337 respectively (numbering according to Lin et al. [6]). Only one peptide (peptide 1) did not match the E3 sequence with a good score on the BLAST search program [7]. This peptide contained 11 amino acid residues; five residues at the C-terminal end were found to exactly match the amino acids 25–29 of Lin et al. [6], whereas the other six residues at the N-terminus did not match any amino acid in the Lin et al. sequence [6] (Figure 1). However, when the nucleotide sequence was read in an alternate frame (**Frame 2**, Figure 1) in which one base (**N**) was added between 72C and 73T and the nucleotide sequence upstream to 72C was shifted by one base toward the 3' end with the sequence downstream to 73T staying unchanged, these six N-terminal residues became well aligned. Four residues were positionally identical and two were substituted with similar residues (Leu¹⁹ → Val and Ala²² → Val (Figure 1).

An attempt was then made to identify the open reading frame in **Frame 2**: two stop codons were created at the triplet of nucleotides from 2–4 and from –170 to –168, with the first methionine codon found at the triplet of nucleotides from –245 to –243. This would prevent the translation downstream from nucleotides –170 or 2, indicating that **Frame 2** was incorrect in the N-terminal region and more correction might be introduced. Assuming the human and rat isoenzymes had the same tryptic cleavage site in this region, lysine or arginine may be the

Figure 1 Sequence alignment of peptide 1 from a human E3 isoenzyme equivalent of rat with human γ -aminobutyraldehyde dehydrogenase (ALDH9) [6] in two different reading frames

The human nucleotide sequence is from –15 to 96. **Frame 1** (triplet codons separated by a space) is the translation originally proposed by Lin et al. [6]. **Frame 2** (triplet codons referred to by double arrows) is an alternative translation resulting from addition of a nucleotide (**N**) between 72C and 73T. The nucleotide sequence at the 5' side to 72C is shifted toward the 3' end by one base, while that at the 3' side to 72C is unaltered. Positionally identical amino acid residues are in **bold** type.

Figure 2 A third reading frame that recovers arginine as a tryptic cleavage site

Frame 3 is generated by an additional frame shift due to the addition of one nucleotide (**N**) between 55G and 56C (numbering according to Lin et al. [6]). The nucleotide sequence at the 5' side of 55G is again shifted toward the 3' end by one base. The recovered arginine residue is in **italic bold** type.

next residue to the N-terminal end of peptide 1. This residue is proline in **Frame 2** or alanine in **Frame 1** (Figure 1); however, an additional frameshift (**Frame 3** in Figure 2) by adding one nucleotide between 55G and 56C can recover an arginine residue immediately next to the N-terminal end of the peptide. In **Frame 3**, the segment of 13 amino acids upstream to the recovered arginine was found to exhibit some features of mitochondrial targeting sequences, being rich in basic and hydrophobic residues as well as lacking acidic residues (Figure 2). This suggests that the E3 isoenzyme might be processed in the mitochondria and then released into cytoplasm in a way analogous to yeast fumarase [8]. **Frame 3**, however, failed to regain a proper open reading frame, owing to the presence of a stop codon at the triplet of nucleotides from -160 to -158 with the first methionine codon from -265 to -263. Whether the human E3 isoenzyme indeed has an arginine residue next to the N-terminal end of peptide 1 and how further correction could be made should await the resequencing of human DNA encoding the E3 isoenzyme or the direct sequencing of the N-terminal of this enzyme. It may be noted that the sequences around both shifted sites are rich in G,C bases and prone to form secondary structures that might interfere with the processing of polymerases when the nucleic acid acts as a template. This situation also occurred with the cDNA encoding human mitochondrial E2 isoenzyme [9], the correct sequence of which was eventually obtained using the chemical method of Maxam and Gilbert [9].

In view of the growing number of aldehyde dehydrogenase

genes being sequenced, more attempts may be made to establish the structure-function relationship of the enzyme based on its sequence data. Therefore, the accuracy of the amino acid sequence is essential not only for blotting experiments but also for functional expression and mutagenic experiments with the enzyme.

Ming-Kai CHERN and Regina PIETRUSZKO¹

Center of Alcohol Studies and Department of Molecular Biology and Biochemistry, Rutgers University, The State University of New Jersey, 607 Allison Road, Piscataway, NJ 08854-8001, U.S.A.

¹ To whom correspondence should be addressed.

- 1 Roberts, L. (1991) *Science* **252**, 1255-1256
- 2 Kurys, G., Ambroziak, W. and Pietruszko, R. (1989) *J. Biol. Chem.* **264**, 4715-4721
- 3 Ambroziak, W. and Pietruszko, R. (1991) *J. Biol. Chem.* **266**, 13011-13018
- 4 Chern, M.-K. and Pietruszko, R. (1995) *Biochem. Biophys. Res. Commun.* **213**, 561-568
- 5 Kurys, G., Shah, P. C., Kikonyogo, A., Reed, D., Ambroziak, W. and Pietruszko, R. (1993) *Eur. J. Biochem.* **218**, 311-320
- 6 Lin, S. W., Chen, J. C., Hsu, L. C., Hsieh, C.-L. and Yoshida, A. (1996) *Genomics* **34**, 376-380
- 7 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403-410
- 8 Stein, I., Peleg, Y., Even-Ram, S. and Pines, O. (1994) **14**, 4770-4778
- 9 Braun, T., Bober, E., Singh, S., Agarwal, D. P. and Goedde, H. W. (1988) *FEBS Lett.* **233**, 440

Received 29 May 1998