Volume 222, number 1, 95–98 FEB 05094 September 1987

Mitochondrial aldehyde dehydrogenase

Homology of putative targeting sequence to that of carbamyl phosphate synthetase I revealed by correlation of cDNA and protein data

John Hempel*, Jan-Olov Höög and Hans Jörnvall

*Department of Biochemistry, University of Pittsburgh, Pittsburgh, PA 15261, USA and Department of Chemistry I, Karolinska Institutet, S-104 01 Stockholm, Sweden

Received 21 July 1987

Comparison of existing protein and cDNA data for human liver mitochondrial aldehyde dehydrogenase reveals deviations in two segments. They are shown to correspond exactly to localized frameshifts in the cDNA data, which are likely to have three reading errors. After correction of the cDNA frameshifts, a deduced amino acid sequence corresponds exactly to the data established at the protein level. In addition, extension of the shifted frame into the cDNA corresponding to the mitochondrial leader sequence allows reinterpretation of that sequence. The new leader sequence is consistent with characteristics of such segments of other mitochondrial protein pro-forms. Furthermore, the sequence displays a homology, when centered around the cleavage site, with the leader sequence of rat liver carbamyl phosphate synthetase I, suggesting a novel similarity between mitochondrial targeting sequences of two different enzymes.

Aldehyde dehydrogenase; Carbamyl phosphate synthetase; cDNA; Frameshift; Leader sequence; Mitochondrial targeting

1. INTRODUCTION

NAD-linked aldehyde dehydrogenases (EC 1.2.1.3) effect oxidation of aliphatic and aromatic aldehydes to carboxylic acids. Substrates include acetaldehyde, thus involving the enzyme in ethanol metabolism. Two forms of the enzyme, one cytosolic and one mitochondrial, occur in human and other mammalian livers [1,2]. The complete primary structures of both these human enzymes have been characterized directly at the protein level [3,4]. The same applies to both forms of the horse liver enzyme ([5] and Johansson et al., unpublished). In addition, substitution of lysine [6] for glutamic acid [7,8] at position 487 has been linked to the defect in the low-activity form of the human mitochondrial enzyme, which is frequent in Oriental individuals. Comparisons of the structures have revealed similarities and differences which are taken to reflect various functional aspects [4,9,10].

Most of the structures of the cytosolic and mitochondrial human aldehyde dehydrogenases (except the N-terminal 160 and 101 residues, respectively) have also been characterized at the cDNA level [8]. These indirectly deduced amino acid sequences are in agreement with those determined directly at the protein level [3,4] except at one position in the cytosolic enzyme (Val/Ile) which likely reflects a neutral polymorphism, and one position in the mitochondrial enzyme (Glu/Val) which potentially could also derive from a single base substitution, although such a polymorphism which would result in differing electrophoretic mobilities has not been noted. Recently, two reports have appeared [11,12], covering

Correspondence address: H. Jörnvall, Department of Chemistry I, Karolinska Institutet, S-104 01 Stockholm, Sweden

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

also the N-terminal part of the human mitochondrial enzyme as indirectly deduced from a cloned cDNA.

Those reports are of interest because of the possibility of correlation with the known protein data [4] and because of the possibility of judging the targeting signals for mitochondrial transport in the pro-form of mitochondrial aldehyde dehydrogenase. The latter is transported to the mitochondria from its site of synthesis in the cytoplasm, and is then proteolytically processed to lose an N-terminal targeting signal (frequently quite long and often with basic residues [13,14]). Consequently, this segment is lacking in the mature protein and has therefore not been studied by protein analysis. However, based on the observation of ragged N-terminal ends in mitochondrial aldehyde dehydrogenase [4] and other mature mitochondrial enzymes [15,16], it was suggested that the proteolytic removal of the targeting sequence is not always quantitative at a single site.

The protein sequence indirectly deduced from the recently reported cDNA structure [11,12] is puzzling, since it does not fit the one directly obtained from the protein [4]. Similarly, knowing the start of the mature protein [4], the signal sequence deduced appears somewhat shorter than that typical in other cases [13,14]. Consequently, the reports on the cDNA sequence [11,12] have now been re-examined. As a result, the protein analyses are reinforced and a novel homology to the pro-forms of another enzyme is suggested regarding the targeting signals.

2. MATERIALS

All data for mitochondrial human liver aldehyde dehydrogenase are taken from the protein analysis [4] and from the cDNA analyses [11,12]. The numbering system from the protein analyses is used to designate all amino acid residues, while nucleotides are numbered according to the cDNA reported.

3. RESULTS AND DISCUSSION

3.1. *Comparison of protein structures directly deduced from peptide studies and those indirectly deduced from cDNA data*

Inspection of the two amino acid sequences

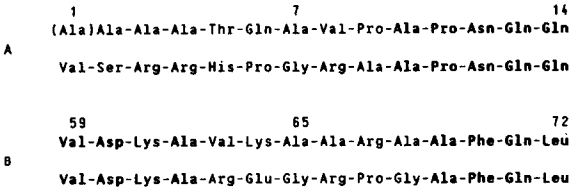

Fig.1. Multi-residue segments of human liver mitochondrial aldehyde dehydrogenase which differ as determined directly (upper line) at the protein level from overlapping peptides [4] and as deduced (lower line) from cDNA [11,12]. (A) Segment with the N-terminus of the mature protein, (B) internal segment. Residue numbers given are those from the mature protein ([4], with Ala-1 shown in parentheses as in the original report to indicate its uncertain assignment). Thin type-face indicates the differing segments, which are flanked by complete agreements (bold-face).

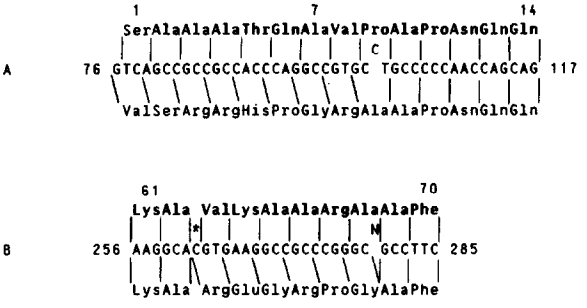

Fig.2. The differing segments from fig.1A and B can be transformed to agree exactly between protein and cDNA data by introduction in the cDNA of single frameshifts. The cDNA from [11,12] is shown in between the segments from fig.1 of human mitochondrial aldehyde dehydrogenase as determined directly (upper line) and as deduced (lower line). Vertical lines show nucleotide groupings corresponding to codons for the respective amino acids. In order to effect the necessary frameshift to fit the cDNA data to the protein data, only two nucleotide insertions (C and N above the line of the nucleotide sequence) and one deletion (C with an asterisk) are required, as indicated. All codons then fit exactly from the cDNA to the top-line bold-face amino acid sequence determined at the protein level.

reported for the mature protein of human mitochondrial aldehyde dehydrogenase reveals two segments (residues 1–9 and 63–68) and a single amino acid residue (Glu-320) that differ between the protein and cDNA data. It is apparent that the

two differing regions do not correspond at all in amino acid sequence and that the deviating residues are contiguous (fig.1), suggesting the possibility of some frameshifts at the cDNA level.

Indeed, such a situation appears to have occurred in both of the present deviating regions (fig.2). Hence, insertion of a cytidine after either nucleotide 100 or 101 (numbering system of [11,12]) alters the reading frame in the 5′-direction and results in deduction of the same sequence as that already found at the protein level for the N-terminal segment (disregarding residue 1, which was previously not firmly established). Similarly, at amino acid positions 63–68, frameshifts in the corresponding cDNA sequence can explain the deviations. Thus, removal of cytidine-262 in the cDNA and insertion of any nucleotide after position 279 in the cDNA yields a deduced amino acid sequence corresponding exactly to that from the protein directly analyzed.

The coincidences support the protein data and strongly suggest that the initially reported deviations are derived from single frameshifts (fig.2). This situation could most easily derive from insertions/failures to detect nucleotides in interpretations of sequencing gels. In summary, comparison of the cDNA data recently reported for human mitochondrial aldehyde dehydrogenase [11,12] with the previous data obtained directly at the protein level [4] suggests that three frameshift errors occurred in the reading of the cDNA data. This underscores the utility of making maximal use of all available protein data when assembling protein sequences by deduction from cDNAs.

3.2. *Comparison of pro-form segments in mitochondrial aldehyde dehydrogenase and other mitochondrial enzymes*

Using the frameshift at nucleotide 100 or 101 defined above, and assuming that the region of the cDNA towards the 5′-direction is correct without further introduction of additional frameshifts, a novel amino acid sequence may be deduced from the existing cDNA sequence [11,12]. Several facets of the structure of the leader sequence may then be noted after translation in the revised frame.

First, the cDNA after the ATG initiation codon was previously interpreted to encode 35 amino acid residues N-terminal to the start of the mature enzyme subunit [11,12]. However, based on the known start points of the mature enzyme [4], that cDNA translation actually corresponds only to an additional 15 residues N-terminal to position 1 of the mature protein. Furthermore, the cDNA reported [11,12], as presently frameshifted, will not cover an ATG initiation of translation. Consequently, it now appears that more residues were originally present between an ATG and the start of the mature enzyme. This proves that the targeting segment should also be longer, and resolves the question of the otherwise unexpectedly short pro-form.

```
                    ↓
CPS I:  Pro-Gly-Ile-Arg-Leu-Leu-Ser-Val-Lys-Ala-Gln-Thr-Ala
AlDH :  Pro-Gly-Arg-Arg-Leu-Leu-Ser-Ala-Ala-Ala-Thr-Gln-Ala
                          ↑
```

Fig.3. Alignment of portions of the pre-sequences of human liver mitochondrial aldehyde dehydrogenase (AlDH) with carbamyl phosphate synthetase I (CPS I), to reveal maximal identities (bold-face) after centering the alignment on leader sequence cleavage sites indicated by arrows. Other cleavage sites have been identified for both proteins [4,16]. The CPS I sequence is from [16] and the AlDH sequence is based on the frameshift (fig.2A) introduced in the cDNA data of [11,12], allowing the cDNA reading frame to correspond to the known protein sequence [4].

Second, the leader now deduced is highly basic, encompassing three Arg residues, which is consistent with previously observed features of mitochondrial targeting segments in other proteins [13,14].

The third observation which arises from the frameshift is a short but intriguing homology between the 'new' putative leader sequence and that of carbamyl phosphate synthetase I [16]. As shown in fig.3, an alignment centered on one of the cleavage sites of each enzyme (two have been reported for the synthetase) reveals five matches over the six positions of the leader closest to the cleavage site, and three matches in seven positions extending into the sequence of the mature protein, for an overall positional identity of 62%. Although the similarities do not extend further into the leader sequences, this is to our knowledge the most convincing degree of homology which has been noted in connection with mitochondrial targeting segments.

ACKNOWLEDGEMENTS

This work was supported by NIAAA (AA 06985-01), the Alcoholic Beverage Medical Research Foundation, and the Swedish Medical Research Council (03X-3532). Special help from Melanie Popa and Kristina Idner is gratefully acknowledged.

REFERENCES

[1] Eckfeldt, J.H. and Yonetani, T. (1976) Arch. Biochem. Biophys. 175, 717–722.
[2] Greenfield, N.J. and Pietruszko, R. (1977) Biochim. Biophys. Acta 483, 35–45.
[3] Hempel, J., Von Bahr-Lindström, H. and Jörnvall, H. (1984) Eur. J. Biochem. 141, 21–35.
[4] Hempel, J., Kaiser, R. and Jörnvall, H. (1985) Eur. J. Biochem. 153, 13–28.
[5] Von Bahr-Lindström, H., Hempel, J. and Jörnvall, H. (1984) Eur. J. Biochem. 141, 37–42.
[6] Yoshida, A., Huang, I.-Y. and Ikawa, M. (1984) Proc. Natl. Acad. Sci. USA 81, 258–261.
[7] Hempel, J., Kaiser, R. and Jörnvall, H. (1984) FEBS Lett. 173, 367–373.
[8] Hsu, L.C., Tani, K., Fujiyoshi, T., Kurachi, K. and Yoshida, A. (1985) Proc. Natl. Acad. Sci. USA 82, 3771–3775.
[9] Jörnvall, H., Hempel, J., Von Bahr-Lindström, H., Höög, J.-O. and Vallee, B.L. (1987) in: Alcohol and Alcoholism, published as suppl.1, Advances in Biomedical Alcohol Research (Lindros, K.O. et al. eds) pp.13–23, Pergamon, Oxford.
[10] Hempel, J. and Jörnvall, H. (1987) in: Enzymology and Molecular Biology of Carbonyl Metabolism (Weiner, H. and Flynn, T.G. eds) pp.1–14, Liss, New York.
[11] Braun, T., Bober, E., Singh, S., Agarwal, D.P. and Goedde, H.W. (1987) FEBS Lett. 215, 233–236.
[12] Braun, T., Bober, E., Singh, S., Agarwal, D.P. and Goedde, H.W. (1987) Nucleic Acids Res. 15, 3179.
[13] Hay, R., Böhni, P. and Gasser, S. (1984) Biochim. Biophys. Acta 779, 65–87.
[14] Schatz, G. (1987) Eur. J. Biochem. 165, 1–6.
[15] Wieland, B., Tomasselli, A.G., Noda, L.H., Frank, R. and Schulz, G.E. (1984) Eur. J. Biochem. 143, 331–339.
[16] Nyunoya, H., Broglie, K.E., Widgren, E.E. and Lusty, C.J. (1985) J. Biol. Chem. 260, 9346–9356.