Volume 215, number 2, 233-236 FEB 04668 May 1987

Evidence for a signal peptide at the amino-terminal end of human mitochondrial aldehyde dehydrogenase

Thomas Braun, Eva Bober, Surjit Singh, Dharam P. Agarwal and H. Werner Goedde

Institut für Humangenetik, Universität Hamburg, Butenfeld 32, D-2000 Hamburg 54, FRG

Received 26 February 1987

A full-length cDNA clone coding for human mitochondrial aldehyde dehydrogenase (ALDH I) was isolated from a human fetal muscle cDNA library. Sequence analysis revealed structural similarities between the amino-terminal end of ALDH I and other known targeting sequences responsible for protein uptake into the mitochondria.

Leader peptide; cDNA sequence; Aldehyde dehydrogenase I; (Human)

1. INTRODUCTION

Human liver aldehyde dehydrogenase (ALDH, EC 1.2.1.3) consists of at least four isozymes. The multiple forms of ALDH show a considerable heterogeneity in their tissue and organ distribution as well as subcellular localization [1,2]. ALDH I is also referred to in the literature as E_2 or ALDH2 isozyme [2]. Localization of this isoenzyme in the mitochondria as well as its assignment to chromosome 12 [3,4], suggests the existence of a leader peptide directing the protein uptake into the mitochondria. Most nuclear coded mitochondrial proteins are translated on cytoplasmic polysomes at larger precursors containing leader peptides of 20–60 residues in length at the amino-terminal end of the protein [5]. Usually these peptides are characterized by the absence of acidic residues and an average to above average content of basic residues periodically separated by uncharged residues [6,7]. Beside these general structural similarities, no significant sequence homologies were found between mitochondrial presequences.

Correspondence address: H.W. Goedde, Institute of Human Genetics, Butenfeld 32, D-2000 Hamburg 54, FRG

In a recent study, it has been shown that even a duplication of a part of the amino-terminal end ('cryptic mitochondrial targeting signal') of a cytosolic protein is sufficient to direct its incorporation into the mitochondria [8].

Previously, a partial cDNA clone pEXAL21 coding for the carboxy-terminal end of ALDH I was isolated and characterized from a human liver cDNA library [9,10]. Until now, no sequence data were available for the amino-terminal end of the ALDH I protein. The cDNA clone pEXAL21 was chosen to screen a human fetal muscle cDNA library in order to isolate a full-length cDNA clone and especially to obtain information on the extreme amino-terminal end.

2. MATERIALS AND METHODS

Human fetal muscle RNA was isolated from abortion material in the 22nd week of gestation. The cDNA library was established in the expression vector λ gt11. RNA isolation and cDNA synthesis were performed as in [11]. 2×10^6 pfu of the human fetal muscle cDNA library were screened using a nick-translated cDNA fragment derived from the ALDH I clone pEXAL21. DNA was isolated from positive λ clones by a small-scale

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

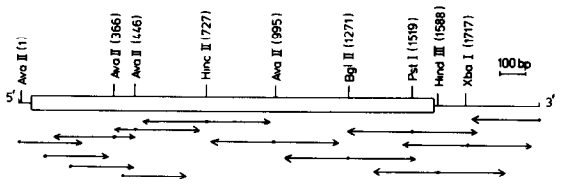

Fig.1. Partial restriction maps and sequencing strategy of the cDNA insert of λ cALI 23. Arrows mark the restriction sites sequenced from and the extent of the individual sequence determinations.

procedure and analyzed further by Southern blot hybridization and restriction mapping [12].

The inserts chosen for sequence analysis were subcloned in M13 mp18 and mp19 and sequenced according to Sanger and Coulson [13]. Full-length ALDH I cDNA was subcloned downstream of a T_3-RNA-polymerase promotor [14] and transcribed in vitro with and without capping reagent (0.1 mM GpppG). The transcription reaction mixture was placed on ice and 10 μl aliquots were directly translated without further purification in a rabbit reticulocyte lysate in the presence of [^{35}S]methionine. Proteins synthesized were separated on an SDS-polyacrylamide gel and detected by fluorography.

3. RESULTS AND DISCUSSION

40 hybridization signals were obtained after screening of 2×10^6 plaques of the human cDNA library with a nick-translated ALDH I cDNA fragment derived from pEXAL21 [3,9]. 24 signals were processed through three additional rounds of screening until distinct plaques could be isolated.

```
                  15                  30                  45                  60
GGT CCC CTC CCT GTC CGC TAG CCC GCT GCG ATG TTG CGC GCT GCC GCC GCT TGG CCC GCC
                                        MET Leu Arg Ala Ala Ala Trp Pro Ala

                  75                  90                 105                 120
TGG GCG CCG CCT CTT GTC AGC CGC CGC CAC CCA GGC CGT GCT GCC CCC AAC CAG CAG CCC
Trp Ala Pro Pro Leu Val Ser Arg Arg His Pro Gly Arg Ala Ala Pro Asn Gln Gln Pro

                 135                 150                 165                 180
GAG GTC TTC TGC AAC CAG ATT TTC ATA AAC AAT GAA TGG CAC GAT GCC GTC AGC AGG AAA
Glu Val Phe Cys Asn Gln Ile Phe Ile Asn Asn Glu Trp His Asp Ala Val Ser Arg Lys

                 195                 210                 225                 240
ACA TTC CCC ACC GTC AAT CCG TCC ACT GGA GAG GTC ATC TGT CAG GTA GCT GAA GGG GAC
Thr Phe Pro Thr Val Asn Pro Ser Thr Gly Glu Val Ile Cys Gln Val Ala Glu Gly Asp

                 255                 270                 285                 300
AAG GAA GAT GTG GAC AAG GCA CGT GAA GGC CGC CCG GGC GCC TTC CAG CTG GGC TCA CCT
Lys Glu Asp Val Asp Lys Ala Arg Glu Gly Arg Pro Gly Ala Phe Gln Leu Gly Ser Pro

                 315                 330                 345                 360
TGG CGC CGC ATG GAC GCA TCA CAC AGC GGC CGG CTG CTG AAC CGC CTG GCC GAT CTG ATC
Trp Arg Arg MET Asp Ala Ser His Arg Gly Arg Leu Leu Asn Arg Leu Ala Asp Leu Ile

                 375                 390                 405                 420
GAG CGG GAC CGG ACC TAC CTG GCG GCC TTG GAG ACC CTG GAC AAT GGC AAG CCC TAT GTC
Glu Arg Asp Arg Thr Tyr Leu Ala Ala Leu Glu Thr Leu Asp Asn Gly Lys Pro Tyr Val

                 435                 450                 465                 480
ATC TCC TAC CTG GTG GAT TTG GAC ATG GTC CTC AAA TGT CTC CGG TAT TAT GCC GGC TGG
Ile Ser Tyr Leu Val Asp Leu Asp MET Val Leu Lys Cys Leu Arg Tyr Tyr Ala Gly Trp

                 495                 510                 525                 540
GCT GAT AAG TAC CAC GGG AAA ACC ATC CCC ATT GAC GGA GAC TTC TTC AGC TAC ACA CGC
Ala Asp Lys Tyr His Gly Lys Thr Ile Pro Ile Asp Gly Asp Phe Phe Ser Tyr Thr Arg

                 555                 570                 585                 600
CAT GAA CCT GTG GGG GTG TGC GGG CAG ATC ATT CCG TGG AAT TTC CCG CTC CTG ATG CAA
His Glu Pro Val Gly Val Cys Gly Gln Ile Ile Pro Trp Asn Phe Pro Leu Leu MET Gln

                 615                 630                 645                 660
GCA TGG AAG CTG GGC CCA GCC TTG GCA ACT GGA AAC GTG GTT GTG ATG AAG GTA GCT GAG
Ala Trp Lys Leu Gly Pro Ala Leu Ala Thr Gly Asn Val Val Val MET Lys Val Ala Glu

                 675                 690                 705                 720
CAG ACA CCC CTC ACC GCC CTC TAT GTG GCC AAC CTG ATC AAG GAG GCT GGC TTT CCC CCT
Gln Thr Pro Leu Thr Ala Leu Tyr Val Ala Asn Leu Ile Lys Glu Ala Gly Phe Pro Pro

                 735                 750                 765                 780
GGT GTG GTC AAC ATT GTG CCT GGA TTT GGC CCC ACG GCT GGG GCC GCC ATT GCC TCC CAT
Gly Val Val Asn Ile Val Pro Gly Phe Gly Pro Thr Ala Gly Ala Ala Ile Ala Ser His

                 795                 810                 825                 840
GAG GAT GTG GAC AAA GTG GCA TTC ACA GGC TCC ACT GAG ATT GGC CGC GTA ATC CAG GTT
Glu Asp Val Asp Lys Val Ala Phe Thr Gly Ser Thr Glu Ile Gly Arg Val Ile Gln Val

                 855                 870                 885                 900
GCT GCT GGG AGC AGC AAC CTC AAG AGA GTG ACC TTG GAG CTG GGG GGG AAG AGC CCC AAC
Ala Ala Gly Ser Ser Asn Leu Lys Arg Val Thr Leu Glu Leu Gly Gly Lys Ser Pro Asn

                 915                 930                 945                 960
ATC ATC ATG TCA GAC GCC GAT ATG GAT TGG GCC GTG GAA CAG GCC CAC TTC GCC CTG TTC
Ile Ile MET Ser Asp Ala Asp MET Asp Trp Ala Val Glu Gln Ala His Phe Ala Leu Phe

                 975                 990                1005                1020
TTC AAC CAG GGC CAG TGC TGC TGT GCC GGC TCC CGG ACC TTC GTG CAG GAG GAC ATC TAT
Phe Asn Gln Gly Gln Cys Cys Cys Ala Gly Ser Arg Thr Phe Val Gln Glu Asp Ile Tyr

                1035                1050                1065                1080
GAT GAG TTT GTG GTG CGG AGC GTT GCC CGG GCC AAG TCT CGG GTG GTC GGG AAC CCC TTT
Asp Glu Phe Val Val Arg Ser Val Ala Arg Ala Lys Ser Arg Val Val Gly Asn Pro Phe

                1095                1110                1125                1140
GAT AGC AAG ACC GAG CAG GGG CCG CAG GTG GAT GAA ACT CAG TTT AAG AAG ATC CTC GGC
Asp Ser Lys Thr Glu Gln Gly Pro Gln Val Asp Glu Thr Gln Phe Lys Lys Ile Leu Gly

                1155                1170                1185                1200
TAC ATC AAC ACG GGG AAG CAA GAG GGG GCG AAG CTG CTG TGT GGT GGG GGC ATT GCT GCT
Tyr Ile Asn Thr Gly Lys Gln Glu Gly Ala Lys Leu Leu Cys Gly Gly Gly Ile Ala Ala

                1215                1230                1245                1260
GAC CGT GGT TAC TTC ATC CAG CCC ACT GTG TTT GGA GAT GTG CAG GAT GGC ATG ACC ATC
Asp Arg Gly Tyr Phe Ile Gln Pro Thr Val Phe Gly Asp Val Gln Asp Gly MET Thr Ile

                1275                1290                1305                1320
GCC AAG GAG GAG ATC TTC GGG CCA GTG ATC CAG ATC CTG AAG TTC AAG ACC ATA GAG GAG
Ala Lys Glu Glu Ile Phe Gly Pro Val MET Gln Ile Leu Lys Phe Lys Thr Ile Glu Glu

                1335                1350                1365                1380
GTT GTT GGG AGA GCC AAC AAT TCC ACG TAC GGG CTG GCC GCA GCT GTC TTC ACA AAG GAT
Val Val Gly Arg Ala Asn Asn Ser Thr Tyr Gly Leu Ala Ala Ala Val Phe Thr Lys Asp

                1395                1410                1425                1440
TTG GAC AAG GCC AAT TAC CTG TCC CAG GCC CTC CAG GCG GGC ACT GTG TGG GTC AAC TGC
Leu Asp Lys Ala Asn Tyr Leu Ser Gln Ala Leu Gln Ala Gly Thr Val Trp Val Asn Cys

                1455                1470                1485                1500
TAT GAT GTG TTT GGA GCC CAG TCA CCC TTT GGT GGC TAC AAG ATG TCG GGG AGT GGC CGG
Tyr Asp Val Phe Gly Ala Gln Ser Pro Phe Gly Gly Tyr Lys MET Ser Gly Ser Gly Arg

                1515                1530                1545                1560
GAG TTG GGC GAG TAC GGG CTG CAG GCA TAC ACT GAA GTG AAA ACT GTC ACA GTC AAA GTG
Glu Leu Gly Glu Tyr Gly Leu Gln Ala Tyr Thr Glu Val Lys Thr Val Thr Val Lys Val

                1575                1590                1605                1620
CCT CAG AAG AAC TCA TAA GAA TCA TGC AAG CTT CCT CCC TCA GCC ATT GAT GGA AAG TTC
Pro Gln Lys Asn Ser

                1635                1650                1665                1680
AGC AAG ATC AGC AAC AAA ACC AAG AAA AAT GAT CCT TGC GTG CTG AAT ATC TGA AAA GAG

                1695                1710                1725                1740
AAA TTT TTC CTA CAA AAT CTC TTG GGT CAA GAA AGT TCT AGA ATT TGA ATT GAT AAA CAT

                1755                1770                1785                1800
GGT GGG TTG GCT GAG GGT AAG AGT ATA TGA GGA ACC TTT TAA ACG ACA ACA ATA CTG CTA

                1815                1830                1845                1860
GCT TTC AGG ATG ATT TTT AAA AAA TAG ATT CAA ATG TGT TAT CCT CTC TCT GAA ACG CTT

                1875                1890                1905                1920
CCT ATA ACT CGA GTT TAT AGG GGA AGA AAA AGC TAT TGT TTA CAA TTA TAT CAC CAT TAA

                1935                1950                1965                1980
GGC AAC TGC TAC ACC CTG CTT TGT ATT CTG GGC TAA GAT TCA TTA AAA ACT AGC TGC TCT

                1995
TAA AAA AAA AAA AAA AAA AAA A
```

Fig.2. Nucleotide sequence of the cDNA insert of λ cALI 23 as determined by the strategy outlined in fig.1.

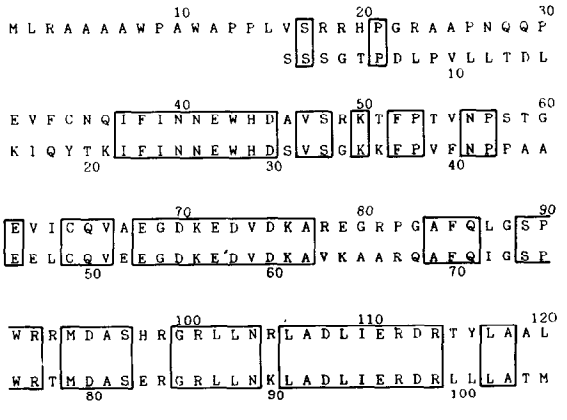

Fig.3. Comparison of the first 100 amino acids at the amino-terminal ends of ALDH I (upper sequence) and ALDH II (lower sequence).

Two clones with an approximate length of 2 kb were sequenced and found to be identical. Fig.1 shows a limited restriction map of the clone λ cALI 23 and the sequence strategy. The nucleotide sequence of the cDNA insert of λ cALI 23, which corresponds to the entire coding region of ALDH I mRNA and the adjacent non-coding regions, is shown in fig.2. The nucleotide sequence is in good agreement with a previously published partial sequence [15]. The deduced ALDH I amino acid sequence was subsequently compared to the known sequence of cytoplasmic isozyme ALDH II [2] in order to define the regions forming the mitochondrial signal peptide. As expected considerable homology (70%) was observed in the major part of the enzyme with exception of the amino-terminal end of both proteins (fig.3). The first 20 amino acids of ALDH II showed no sequence homology to the corresponding ALDH I sequences. Furthermore, an additional peptide of 15 residues without a corresponding counterpart in ALDH II was found at the amino-terminal end of ALDH I. Based on these observations the whole of the 35 amino acid amino-terminal end of ALDH I seems to represent a mitochondrial leader peptide. Similar to other mitochondrial leader sequences (fig.4) this peptide is characterized by lack of acidic residues and an above average content of basic residues typically separated from each other by uncharged residues [6,7]. Whether all 35 amino acids belong to the leader sequence remains to be established. Recently, Hurt et al. [16] demonstrated that the first 12 amino acids of an imported mitochondrial protein are sufficient to direct a cytoplasmic protein into the mitochondrial matrix, although most leader peptides sequenced so far comprise 20–60 residues. In order to assess the actual length of ALDH I signal peptide, in vitro studies using recombinant ALDH I protein are essential. In pursuance of this goal, the whole cDNA insert of λ cALI 23 was subcloned into the vector pSP6/T3. This construct was transcribed in vitro by T_3-RNA-polymerase, and the resulting capped and uncapped mRNAs were translated in a cell-free reticulocyte lysate system. The synthesized polypeptide migrated similarly to authentic ALDH I on SDS-PAGE indicating the absence of major posttranslational modifications (fig.5). This protein product has an apparent molecular mass of 50 kDa which is in good agreement with the molecular mass of ALDH I monomers [1]. To define the critical regions in the putative leader peptide, in vitro targeting studies with modified ALDH I precursor are in progress.

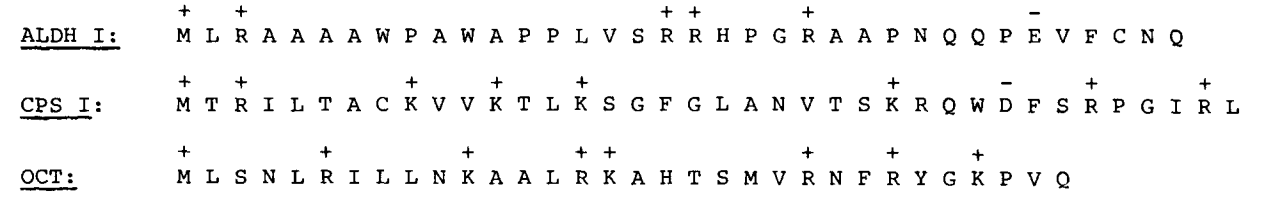

Fig.4. Comparison of ALDH I leader peptide and presequences found in other mitochondrial precursor proteins. (+) Basic amino acid, (−) acidic residue. OCT, ornithine carbamyltransferase, CPS I, carbamyl-phosphate synthetase I, ALDH I, aldehyde dehydrogenase I.

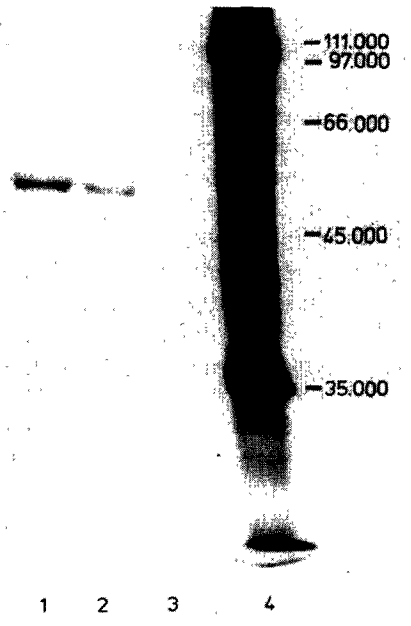

Fig.5. In vitro transcription-translation of the precursor of ALDH I. A fluorogram of the gel is shown. Molecular mass markers are indicated (in kDa). Lanes: 1, translation products of uncapped ALDH I mRNA; 2, translation products of capped ALDH I mRNA; 3, without exogenous RNA; 4, translation products of 5 μg brome mosaic virus RNA.

ACKNOWLEDGEMENTS

We wish to thank the Deutsche Forschungsgemeinschaft, the Bundesministerium für Forschung und Technologie, Bonn-Bad Godesberg, and the Boehringer Ingelheim Fonds, Stuttgart, for financial support; and the Frauenklinik, Universitäts-Krankenhaus Eppendorf, Hamburg, for the supply of the abortion material.

REFERENCES

[1] Goedde, H.W., Agarwal, D.P. and Harada, S. (1983) in: Isozymes 8, pp.175–193, A.R. Liss, New York.
[2] Hempel, J., Von Bahr-Lindström, H. and Jörnvall, H. (1984) Eur. J. Biochem. 141, 21–35.
[3] Braun, T., Grzeschik, K.H., Bober, E., Singh, S., Agarwal, D.P. and Goedde, H.W. (1986) Hum. Genet. 73, 365–367.
[4] Hsu, L.C., Tani, K., Fujiyoshi, T., Kurachi, K. and Yoshida, A. (1986) Am. J. Hum. Genet. 38, 641–648.
[5] Hay, R., Böhni, P. and Gasser, S. (1984) Biochim. Biophys. Acta 779, 65–87.
[6] Douglas, M. and Takeda, M. (1985) Trends Biochem. Sci. 10, 192–194.
[7] Nguyen, M., Argan, C., Lusty, C.J. and Shore, G.C. (1986) J. Biol. Chem. 261, 800–805.
[8] Hurt, E.C. and Schatz, B. (1987) Nature 325, 499–503.
[9] Goedde, H.W. and Agarwal, D.P. (1986) in: Ethnic Differences to Reactions to Drugs and Xenobiotics, pp.113–138, A.R. Liss, New York.
[10] Braun, T., Bober, E., Schaper, J., Agarwal, D.P., Singh, S. and Goedde, H.W. (1987) Proceedings 3rd ISBRA-Congress Helsinki, in press.
[11] Bober, E., Singh, S., Agarwal, D.P. and Goedde, H.W. (1986) Biosci. Rep. 6, 633–638.
[12] Maniatis, T., Fritsch, E.F. and Sambrock, J. (1982) in: Molecular Cloning: a New Laboratory Manual, Cold Spring Harbour Laboratory Press, NY.
[13] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[14] Krieg, P.A. and Melton, D.A. (1984) Nucleic Acids Res. 12, 7057–7070.
[15] Hsu, L.C., Tani, K., Fujiyoshi, T., Kurachi, K. and Yoshida, A. (1985) Proc. Natl. Acad. Sci. USA 82, 3771–3775.
[16] Hurt, E.C., Pesold-Hurt, B. and Schatz, G. (1984) EMBO J. 3, 3149–3256.