Volume 234, number 2, 395–399 FEB 06087 July 1988

Isolation of cDNA clones of human argininosuccinate lyase and corrected amino acid sequence

Seiji Matuo, Michiko Tatsuno, Keiko Kobayashi, Takeyori Saheki, Toshiyuki Miyata°, Sadaaki Iwanaga°, Yoshihiro Amaya+ and Masataka Mori+

Department of Biochemistry, Faculty of Medicine, Kagoshima University, Kagoshima 890,
°Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812 and
+Institute for Medical Genetics, Kumamoto University Medical School, Kumamoto 862, Japan

Received 20 May 1988

In the present study, we isolated clones of human argininosuccinate lyase (ASL) cDNA from a liver cDNA library using a clone of rat ASL cDNA and analyzed human ASL cDNA nucleotide sequence. The results reveal that the sequence of human ASL cDNA published by O'Brien et al. in 1986 [Proc. Natl. Acad. Sci USA 83, 7211–7215] had one-base deletions at three independent positions in the coding regions near the COOH-terminus, which caused frame-shift variations in the amino acid sequence. Amino acid sequencing of peptides prepared from purified human liver ASL showed our predicted amino acid sequence to be correct.

Argininosuccinate lyase; Cloning; Nucleotide sequence; Amino acid sequence; Argininosuccinic aciduria

1. INTRODUCTION

Argininosuccinate lyase (ASL) is a urea cycle enzyme which catalyzes the cleavage of argininosuccinate into arginine and fumarate. Deficiency of ASL results in argininosuccinic aciduria, the second most frequently found enzyme deficiency in the urea cycle [2]. In Japan, more than 10 cases have been found and described. We have shown that most Japanese patients with argininosuccinic aciduria are completely deficient in ASL activity and immune cross-reacting material [3,4], but Simard et al. [5] reported that, in all of the 28 patients in the USA, cross-reacting material was detected by Western blotting analysis.

Correspondence address: T. Saheki, Department of Biochemistry, Faculty of Medicine, Kagoshima University, Kagoshima 890, Japan

Abbreviations: ASL, argininosuccinate lyase; HPLC, high-performance liquid chromatography

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession number Y00753

In 1986, O'Brien et al. [1] isolated human ASL cDNA clones, predicted the amino acid sequence and showed partial homology in the deduced amino acid sequences between yeast and human enzymes. Amaya et al. [6] described the nucleotide sequence of rat ASL cDNA clones. They found a high homology between the human and the rat enzymes along the entire sequence except for the 66 amino acid residues of the human sequence near the COOH-terminus, and suggested that the reported amino acid sequence of the corresponding region of the human enzyme might arise from the different reading frame of the cDNA. We isolated cDNA clones for ASL from a human liver cDNA library in order to reexamine the human cDNA sequence and to use the cDNA for future analysis of ASL deficiency at the molecular level.

2. MATERIALS AND METHODS

2.1. *Screening of cDNA library and DNA sequence analysis*

Complementary DNA clones for human ASL were isolated from a human liver cDNA library constructed in λgt11 expression vector essentially according to the plaque hybridization method of Maniatis et al. [7]. Previously cloned rat ASL cDNA [6], labeled with [^{32}P]dCTP, was used as a probe for screening.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

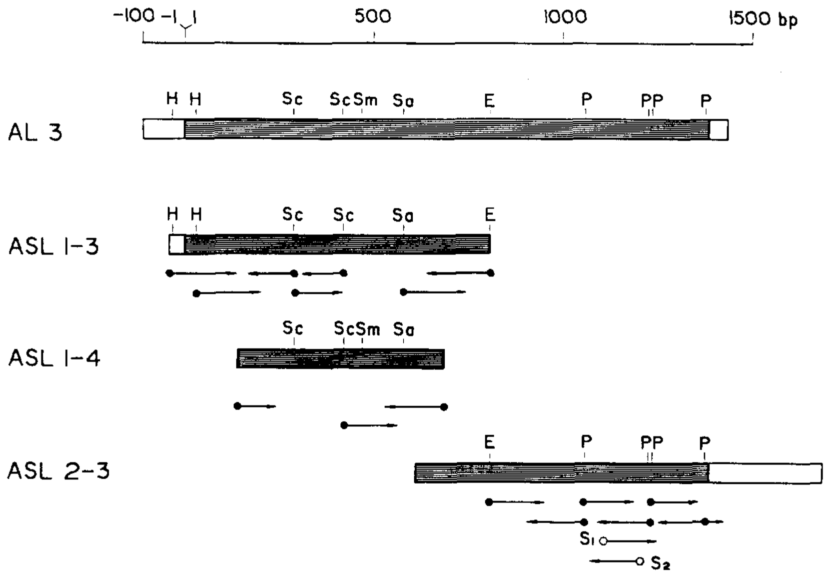
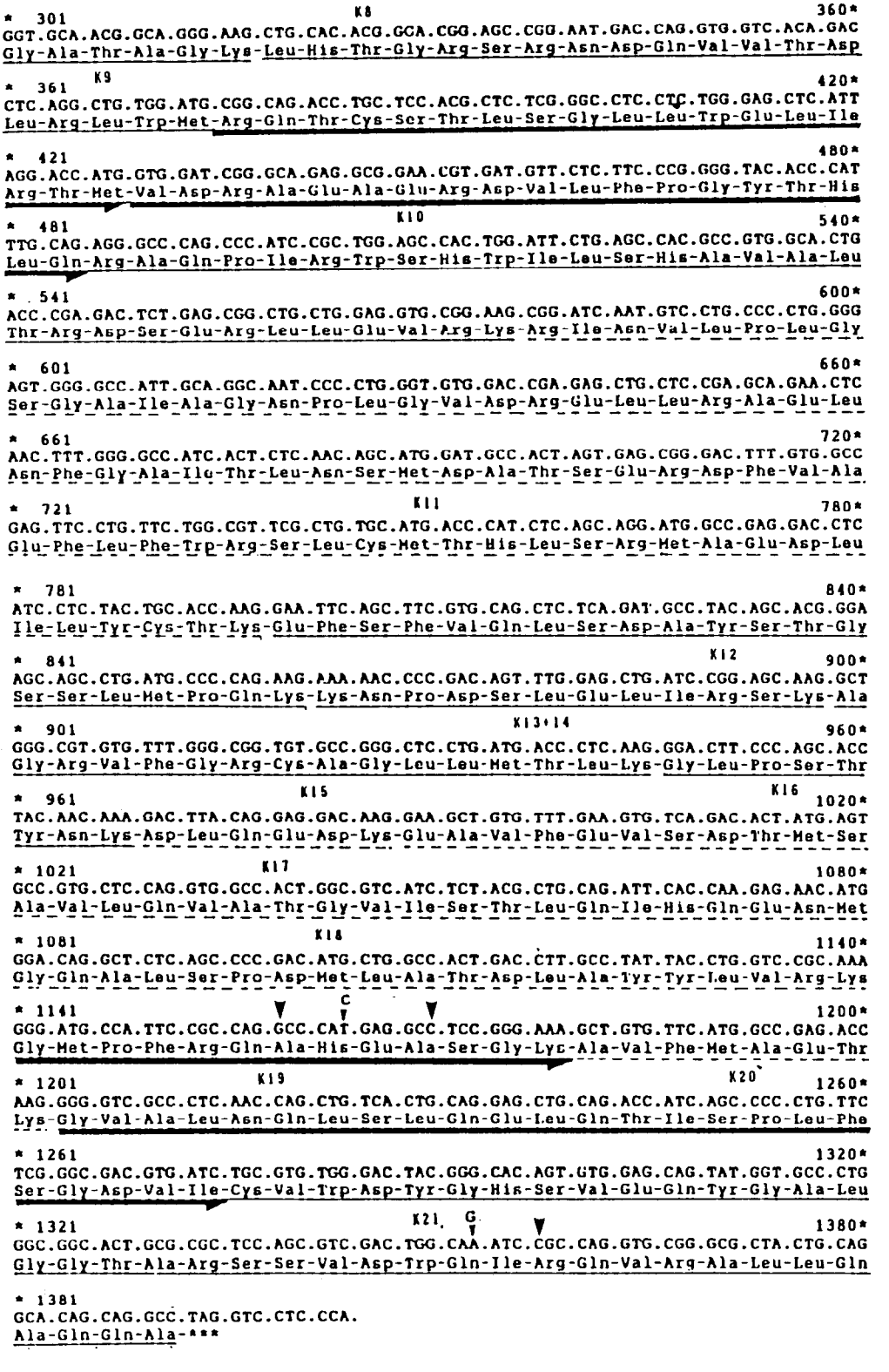


Fig.1. Restriction map of human argininosuccinate lyase cDNAs (ASL 1–3, ASL 1–4 and ASL 2–3) and the strategy for nucleotide sequence determination. The scale at the top indicates the nucleotide position relative to the protein initiator codon beginning at position 1. The restriction map of human ASL cDNA (AL 3) obtained by O'Brien et al. [1] is also shown. H, Sc, Sm, Sa, P, and E denote restriction sites cut by restriction enzymes, *Hin*dIII, *Sac*I, *Sma*I, *Sal*I, *Pst*I and *Eco*RI, respectively. The protein coding region as predicted from the cDNA is indicated by a closed box, and 5′- and 3′-untranslated regions are indicated by open boxes. The arrows beneath the cDNAs indicate the direction and extent of sequence determination, starting from the universal primer or reverse primer (●), or the specific primer (S_1, CTCAGCCCCGACATGCT; S_2, AGTGACAGCTGGTTGAG) (○).

```
* -45                                                                  -1
CAG.ACC.CGG.AGA.CCG.AAG.CTT.CCG.GAC.GAC.GAG.GAA.CCG.CCC.AAC.

*   1                                                                 60*
ATG.GCC.TCG.GAG.AGT.GGG.AAG.CTT.TGG.GGT.GGC.CGG.TTT.GTG.GGT.GCA.GTG.GAC.CCC.ATC
Met-Ala-Ser-Glu-Ser-Gly-Lys-Leu-Trp-Gly-Gly-Arg-Phe-Val-Gly-Ala-Val-Asp-Pro-Ile

*  61       K1                                         K2            120*
ATG.GAG.AAG.TTC.AAC.GCG.TCC.ATT.GCC.TAC.GAC.CGG.CAC.CTT.TGG.GAG.GTG.GAT.GTT.CAA
Met-Glu-Lys-Phe-Asn-Ala-Ser-Ile-Ala-Tyr-Asp-Arg-His-Leu-Trp-Glu-Val-Asp-Val-Gln

* 121                                        K3                      180*
GGC.AGC.AAA.GCC.TAC.AGC.AGG.GGC.CTG.GAG.AAG.GCA.GGG.CTC.CTC.ACC.AAG.GCC.GAG.ATG
Gly-Ser-Lys-Ala-Tyr-Ser-Arg-Gly-Leu-Glu-Lys-Ala-Gly-Leu-Leu-Thr-Lys-Ala-Glu-Met

* 181                    K4                          K5              240*
GAC.CAG.ATA.CTC.CAT.GGC.CTA.GAC.AAG.GTG.GCT.GAG.GAG.TGG.GCC.CAG.GGC.ACC.TTC.AAA
Asp-Gln-Ile-Leu-His-Gly-Leu-Asp-Lys-Val-Ala-Glu-Glu-Trp-Ala-Gln-Gly-Thr-Phe-Lys

* 241   K6                                           K7              300*
CTG.AAC.TCC.AAT.GAT.GAG.GAC.ATC.CAC.ACA.GCC.AAT.GAG.CGC.CGC.CTG.AAG.GAG.CTC.ATT
Leu-Asn-Ser-Asn-Asp-Glu-Asp-Ile-His-Thr-Ala-Asn-Glu-Arg-Arg-Leu-Lys-Glu-Leu-Ile
```

Fig.2. Nucleotide and deduced amino acid sequences of human argininosuccinate lyase constructed from those of three cDNA clones, ASL 1–3, ASL 1–4 and ASL 2–3. Nucleotides are numbered in the 5′- to 3′-direction, beginning with the first residue of the ATG triplet encoding the initiator methionine. Arrowheads above nucleotide sequence at bases 1159, 1170 and 1357 show bases deleted in the sequence published by O'Brien et al. [1]. T at base 1164 and A at 1354 are substituted by C and G, respectively, in the sequence of O'Brien et al., as indicated. K-peptides predicted to be produced by digestion with lysylendopeptidase are shown by lines beneath the amino acid sequence. Dotted lines show that peptides with the amino acid compositions predicted from the nucleotide sequence were not isolated by the present HPLC analysis, and solid lines indicate that peptides with the same amino acid compositions were isolated. Bold lines with arrows indicate sequenced peptides and the extent of sequencing. After cyanogen-bromide digestion, two middle fragments of K-10 were sequenced.


* 301 K8 360*
GGT.GCA.ACG.GCA.GGG.AAG.CTG.CAC.ACG.GGA.CGG.AGC.CGG.AAT.GAC.CAG.GTG.GTC.ACA.GAC
Gly-Ala-Thr-Ala-Gly-Lys-Leu-His-Thr-Gly-Arg-Ser-Arg-Asn-Asp-Gln-Val-Val-Thr-Asp
* 361 K9 420*
CTC.AGG.CTG.TGG.ATG.CGG.CAG.ACC.TGC.TCC.ACG.CTC.TCG.GGC.CTC.CTC.TGG.GAG.CTC.ATT
Leu-Arg-Leu-Trp-Met-Arg-Gln-Thr-Cys-Ser-Thr-Leu-Ser-Gly-Leu-Leu-Trp-Glu-Leu-Ile
* 421 480*
AGG.ACC.ATG.GTG.GAT.CGG.GCA.GAG.GCG.GAA.CGT.GAT.GTT.CTC.TTC.CCG.GGG.TAC.ACC.CAT
Arg-Thr-Met-Val-Asp-Arg-Ala-Glu-Ala-Glu-Arg-Asp-Val-Leu-Phe-Pro-Gly-Tyr-Thr-His
* 481 K10 540*
TTG.CAG.AGG.GCC.CAG.CCC.ATC.CGC.TGG.AGC.CAC.TGG.ATT.CTG.AGC.CAC.GCC.GTG.GCA.CTG
Leu-Gln-Arg-Ala-Gln-Pro-Ile-Arg-Trp-Ser-His-Trp-Ile-Leu-Ser-His-Ala-Val-Ala-Leu
* 541 600*
ACC.CGA.GAC.TCT.GAG.CGG.CTG.CTG.GAG.GTG.CGG.AAG.CGG.ATC.AAT.GTC.CTG.CCC.CTG.GGG
Thr-Arg-Asp-Ser-Glu-Arg-Leu-Leu-Glu-Val-Arg-Lys-Arg-Ile-Asn-Val-Leu-Pro-Leu-Gly
* 601 660*
AGT.GGG.GCC.ATT.GCA.GGC.AAT.CCC.CTG.GGT.GTG.GAC.CGA.GAG.CTG.CTC.CGA.GCA.GAA.CTC
Ser-Gly-Ala-Ile-Ala-Gly-Asn-Pro-Leu-Gly-Val-Asp-Arg-Glu-Leu-Leu-Arg-Ala-Glu-Leu
* 661 720*
AAC.TTT.GGG.GCC.ATC.ACT.CTC.AAC.AGC.ATG.GAT.GCC.ACT.AGT.GAG.CGG.GAC.TTT.GTG.GCC
Asn-Phe-Gly-Ala-Ile-Thr-Leu-Asn-Ser-Met-Asp-Ala-Thr-Ser-Glu-Arg-Asp-Phe-Val-Ala
* 721 K11 780*
GAG.TTC.CTG.TTC.TGG.CGT.TCG.CTG.TGC.ATG.ACC.CAT.CTC.AGC.AGG.ATG.GCC.GAG.GAC.CTC
Glu-Phe-Leu-Phe-Trp-Arg-Ser-Leu-Cys-Met-Thr-His-Leu-Ser-Arg-Met-Ala-Glu-Asp-Leu
* 781 840*
ATC.CTC.TAC.TGC.ACC.AAG.GAA.TTC.AGC.TTC.GTG.CAG.CTC.TCA.GAT.GCC.TAC.AGC.ACG.GGA
Ile-Leu-Tyr-Cys-Thr-Lys-Glu-Phe-Ser-Phe-Val-Gln-Leu-Ser-Asp-Ala-Tyr-Ser-Thr-Gly
* 841 K12 900*
AGC.AGC.CTG.ATG.CCC.CAG.AAG.AAA.AAC.CCC.GAC.AGT.TTG.GAG.CTG.ATC.CGG.AGC.AAG.GCT
Ser-Ser-Leu-Met-Pro-Gln-Lys-Lys-Asn-Pro-Asp-Ser-Leu-Glu-Leu-Ile-Arg-Ser-Lys-Ala
* 901 K13+14 960*
GGG.CGT.GTG.TTT.GGG.CGG.TGT.GCC.GGG.CTC.CTG.ATG.ACC.CTC.AAG.GGA.CTT.CCC.AGC.ACC
Gly-Arg-Val-Phe-Gly-Arg-Cys-Ala-Gly-Leu-Leu-Met-Thr-Leu-Lys-Gly-Leu-Pro-Ser-Thr
* 961 K15 K16 1020*
TAC.AAC.AAA.GAC.TTA.CAG.GAG.GAC.AAG.GAA.GCT.GTG.TTT.GAA.GTG.TCA.GAC.ACT.ATG.AGT
Tyr-Asn-Lys-Asp-Leu-Gln-Glu-Asp-Lys-Glu-Ala-Val-Phe-Glu-Val-Ser-Asp-Thr-Met-Ser
* 1021 K17 1080*
GCC.GTG.CTC.CAG.GTG.GCC.ACT.GGC.GTC.ATC.TCT.ACG.CTG.CAG.ATT.CAC.CAA.GAG.AAC.ATG
Ala-Val-Leu-Gln-Val-Ala-Thr-Gly-Val-Ile-Ser-Thr-Leu-Gln-Ile-His-Gln-Glu-Asn-Met
* 1081 K18 1140*
GGA.CAG.GCT.CTC.AGC.CCC.GAC.ATG.CTG.GCC.ACT.GAC.CTT.GCC.TAT.TAC.CTG.GTC.CGC.AAA
Gly-Gln-Ala-Leu-Ser-Pro-Asp-Met-Leu-Ala-Thr-Asp-Leu-Ala-Tyr-Tyr-Leu-Val-Arg-Lys
* 1141 C 1200*
GGG.ATG.CCA.TTC.CGC.CAG.GCC.CAT.GAG.GCC.TCC.GGG.AAA.GCT.GTG.TTC.ATG.GCC.GAG.ACC
Gly-Met-Pro-Phe-Arg-Gln-Ala-His-Glu-Ala-Ser-Gly-Lys-Ala-Val-Phe-Met-Ala-Glu-Thr
* 1201 K19 K20 1260*
AAG.GGG.GTC.GCC.CTC.AAC.CAG.CTG.TCA.CTG.CAG.GAG.CTG.CAG.ACC.ATC.AGC.CCC.CTG.TTC
Lys-Gly-Val-Ala-Leu-Asn-Gln-Leu-Ser-Leu-Gln-Glu-Leu-Gln-Thr-Ile-Ser-Pro-Leu-Phe
* 1261 1320*
TCG.GGC.GAC.GTG.ATC.TGC.GTG.TGG.GAC.TAC.GGG.CAC.AGT.GTG.GAG.CAG.TAT.GGT.GCC.CTG
Ser-Gly-Asp-Val-Ile-Cys-Val-Trp-Asp-Tyr-Gly-His-Ser-Val-Glu-Gln-Tyr-Gly-Ala-Leu
* 1321 K21 G 1380*
GGC.GGC.ACT.GCG.CGC.TCC.AGC.GTC.GAC.TGG.CAA.ATC.CGC.CAG.GTG.CGG.GCG.CTA.CTG.CAG
Gly-Gly-Thr-Ala-Arg-Ser-Ser-Val-Asp-Trp-Gln-Ile-Arg-Gln-Val-Arg-Ala-Leu-Leu-Gln
* 1381
GCA.CAG.CAG.GCC.TAG.GTC.CTC.CCA.
Ala-Gln-Gln-Ala-***

All nucleotide sequencing was performed by the chain-termination method of Sanger [8], using an M13 sequencing kit. M4 and RV primers from Takara or synthetic 17-mer oligonucleotides were used as primers. Fuji Gensor gels were used for electrophoresis.

2.2. *Purification of human liver ASL*

Human liver ASL was purified according to the method of Palekar and Mantagos [9] as modified by Tatsuno et al. [4]. The purified enzyme had a specific activity of 20 μmol/min per mg of protein at 37°C.

2.3. *Sequence analysis of peptides*

The dialyzed enzyme preparation was reduced and treated with vinylpyridine. After lyophilization, the *S*-pyridylethylated enzyme (400 μg) was digested with lysyl-endopeptidase (2 μg) (Wako Pure Chemical Industries, Ltd, Tokyo) in 50 mM Tris-HCl, pH 9.0, containing 4 M urea for 8 h at 37°C. The peptides obtained were then separated with high-performance liquid chromatography (HPLC) on a Vydac protein C4 column (0.46 × 15 cm). Amino acid composition of the separated peptides was determined by the phenylthiocarbamyl method [10] using a Waters Pico-Tag amino acid analysis system after hydrolysis for 20 h in an evacuated, sealed tube at 110°C in 5.7 M HCl. Automated sequence analyses of the peptides were performed on an Applied Biosystems model 477 protein sequencer [11] equipped with a model 120A PTH analyzer.

3. RESULTS AND DISCUSSION

From 150000 plaques, 10 positive clones were separated using a rat ASL cDNA probe. The length of the cDNA inserts cut from the vector DNA by *Eco*RI digestion was 550 to 1100 bp. Fig.1 shows the restriction map of the inserts. These inserts were further digested by restriction enzymes shown in fig.1, subcloned into pUC18 and sequenced as indicated. Fig.2 shows the nucleotide sequence of human ASL worked out from the sequences of three clones, ASL 1–3, ASL 1–4 and ASL 2–3. There was no difference in the human ASL cDNA nucleotide sequence between our data and O'Brien's data from postulated initiator codon ATG to base number 1158 (counted from ATG), but O'Brien's data had one-base omissions, G, C and C at base numbers 1159, 1170 and 1357, respectively. The first and second one-base omissions cause frame-shift variations in the predicted amino acid sequence of ASL and the third deletion causes the reading frame to return to the initial frame. We also found two base-substitutions which, however, cause no amino acid substitutions in the present reading frame (see fig.2).

In order to confirm our nucleotide sequence, we sequenced peptides (K-peptides) which had been derived from purified human liver ASL by digestion with lysylendopeptidase. We isolated 15 different peptides, whose amino acid compositions were consistent with those of the peptides predicted from the cDNA sequence. We sequenced three of these peptides.

These were assumed to be K-10, K-19 whose nucleotide sequence involves two bases omitted in O'Brien's sequence, and K-21, the C-terminal peptide. The results revealed that the sequences of the middle portion of K-10, the whole of K-19 and the N-terminal 24 residues of K-21 are exactly as predicted (fig.2). It therefore seems that the amino acid sequence deduced by our nucleotide analysis is almost entirely correct (we have no information about the N-terminal amino acid) and that O'Brien's nucleotide sequence had two one-base omissions at the position shown in the present study. We were unable to find direct evidence for the third omission, but as Amaya et al. [7] pointed out, rat and human ASL amino acid sequences become homologous again just after the one-base omission we posit, strongly supporting the presence of a third one-base omission in O'Brien's sequence. As the results show, the predicted amino acid sequence of human ASL in the present study is highly homologous with the rat sequence along the entire sequence (90% identical).

Wistow and Piatigorsky [12] reported that δ crystallin, the major protein of the embryonic lens in birds and reptiles, belongs to the same protein superfamily as ASL. They showed that chicken δ1 and δ2 crystallins are, respectively, 58 and 62% identical to the predicted amino acid sequence published by O'Brien et al. [1]. They further postulated a frame shift in the human ASL sequence by an insertion of G after base 1272 and an insertion of C after 1281 of O'Brien's cDNA sequence (these correspond to bases 1159 and 1170 in the present sequence), deletions of G at positions 1457 and 1461 (bases 1345 and 1349 in the present sequence), and gaps in the amino acid sequence. These postulations yield predicted protein sequences that are 64 and 69% identical to chicken δ1 and δ2 crystallins, respectively. As we show in fig.2, the one-base insertion at the two positions they postulated are at the correct positions. The amino acid sequence predicted from the present cDNA study is 68% identical to chicken δ2

crystallin without any rearrangement of nucleotide and amino acid sequences. We are now analyzing ASL deficiency using the probes prepared from the cDNA clones isolated in the present study.

Acknowledgements: This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan, and the Kodama Memorial Fund for Medical Science Research.

REFERENCES

[1] O'Brien, W.E., McInnes, R., Kalumuck, K. and Adcock, M. (1986) Proc. Natl. Acad. Sci. USA 83, 7211–7215.
[2] Walser, M. (1983) in: The Metabolic Basis of Inherited Disease (Stanbury, J.P. et al. eds) pp.402–438, McGraw-Hill, New York.
[3] Kobayashi, K., Itakura, Y., Saheki, T., Nakano, K., Sase, M., Oyanagi, K., Okamoto, R. and Mino, M. (1986) Clin. Chim. Acta 159, 59–67.
[4] Tatsuno, M., Kobayashi, K., Ichiki, H., Tachikura, T., Nakamura, M. and Saheki, T. (1988) Clin. Chem. Enzymol. Commun., in press.
[5] Simard, L., O'Brien, W.E. and McInnes, R.R. (1986) Am. J. Hum. Genet. 39, 38–51.
[6] Amaya, Y., Matsubasa, T., Takiguchi, M., Kobayashi, K., Saheki, T., Kawamoto, S. and Mori, M. (1988) J. Biochem. (Tokyo) 103, 177–181.
[7] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[8] Sanger, F. (1981) Science 214, 1205–1210.
[9] Palekar, A.G. and Mantagos, S. (1981) J. Biol. Chem. 258, 9192–9194.
[10] Bidlingmeyer, B.A., Cohen, S.A. and Tarvin, T.L. (1984) J. Chromatogr. 336, 93–104.
[11] Hewick, R.M., Hunkapiller, M.W., Hood, L.E. and Dreyer, W.J. (1981) J. Biol. Chem. 256, 7990–7997.
[12] Wistow, G. and Piatigorsky, J. (1987) Science 236, 1554–1556.