Vol. 170, No. 1, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 16, 1990 Pages 231-237

MOLECULAR CLONING OF TWO SPECIES OF cDNAs FOR HUMAN α-N-ACETYLGALACTOSAMINIDASE AND EXPRESSION IN MAMMALIAN CELLS

Toyoaki Yamauchi[1], Masao Hiraiwa[2], Hisashi Kobayashi[1], Yutaka Uda[2], Tadashi Miyatake[1] and Shoji Tsuji[1]

[1]Depatment of Neurology, Brain Research Institute, Niigata University, 1 Asahimachi, Niigata 951, JAPAN

[2]Department of Health Chemistry, Niigata Pharmaceutical College, Niigata 951, JAPAN

Received June 1, 1990

Summary: Two species of cDNAs for human α-N-acetylgalactosaminidase were isolated from a human fibroblast cDNA library. The two species differ each other by a 70 bp insertion in the coding region. Transient expression study in COS cells demonstrated that only the cDNA without the 70 bp insertion expressed α-N-acetylgalactosaminidase activity. Analysis of mRNA species utilizing polymerase chain reaction revealed that the majority of the mRNA does not contain the 70 bp insertion, and the mRNA containing the 70 bp insertion is present only in a minor amount in human brain.

α-N-acetylgalactosaminidase (EC 3.2.1.49) is a lysosomal hydrolase whose deficiency in human has been recently described (1,2). The molecular analysis of mutations causing the disease, however, has not been reported. Recently our laboratory has succeeded in molecular cloning of a full-length cDNA for human α-N-acetylgalactosaminidase (3). The cDNA, pcD2-HS1204, codes for 358 amino acids with the first 17 amino acids representing a putative signal peptide. The identity of the pcD2-HS1204 as coding for human α-N-acetylgalactosaminidase was confirmed by colinearlity of the deduced amino acid sequences with those obtained by microsequencing of the purified human α-N-acetylgalactosaminidase (3). Here we describe identification of two species of mRNA for human α-N-acetylgalactosaminidase, expression of two species of mRNA in various tissues and cDNA-mediated expression of α-N-acetylgalactosaminidase activity in COS cells.

MATERIALS AND METHODS

Molecular cloning of cDNAs for human α-N-acetylgalactosaminidase: Using a 5' fragment of pcD2-HS1204 (3), we screened a human fibroblast cDNA library (kindly provided by Dr. H. Okayama) and isolated 3 additional cDNA clones, pcD2-HS1207, pcD2-HS1225 and pcD2-HS1237.

0006-291X/90 $1.50

Nucleotide sequence analysis: The nucleotide sequence was analyzed by dideoxynucleotide chain terminator method using double stranded plasmid DNA as a template (4-6).

Transient expression of α-N-acetylgalactosaminidase in COS cells: Plasmid DNA used for transfection was prepared by standard protocols (7). COS cells were plated 24 h before transfection at 6 X 10^5 cells per 25 cm^2 in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. The COS cells were transfected with DNA-calcium phophate coprecipitate according to the standard protocol (7).

Measurement of enzyme activities: α-Galactosidase activity was measured using 4-methylumbelliferyl-α-galactopyranoside as a substrate. α-N-acetylgalactosaminidase activity was measured using p-nitrophenyl-N-acetyl-α-galactosaminide as a substrate (8,9).

Analysis of mRNA species: Total RNA was extracted from HeLa cells, human fibroblasts, a human liver and human brains, and poly A(+) RNA was isoltated (10). Complementary DNA was synthesized by MuLV (murine leukemia virus) reverse transcriptase (Bethesda Research Laboratory, Gaithersburg, MD, U.S.A.) in 50 mM TrisHCl, pH 8.3 containing 75 mM KCl, 10 mM DTT and 5 mM $MgCl_2$ using 5 μg of poly A(+) RNA as a template at 37 ^oC for 1 hr. The aliquot of the reaction mixture was used directly for polymerase chain reaction (PCR) (11,12). Two oligonucleotides (primer 1; 5'GCTGCTCATTGGGAACTTTG, primer 2; 5'GCTAGCCTTGTTGGACAGAG) were synthesized by the phosphoramidite method using an automated DNA synthesizer (model 391, Applied Biosystems, Foster City, CA, U.S.A.). The condition for PCR consists of 30 cycles of denaturation (94^oC, 30 sec.), annealing (55^oC, 30 sec.) and extension (72^oC, 3 min.). PCR products were electrophoresed through 1.8 % agarose gel and transferred to a nitrocellulose membrane. The filter was hybridized with a ^{32}P-labelled KpnI fragment of pcD2-HS1204 spanning the region to be amplified (nucleotides 850-1608), washed in 0.1X SSC (1X SSC = 150 mM NaCl, 15 mM sodium citrate) at 55^oC and autoradiographed with an intensifying screen (13).

RESULTS

Using a 5' fragment of pcD2-HS1204 (3) as a probe, we isolated three additional cDNA clones, pcD2-HS1207, pcD2-HS1225 and pcD2-HS1237. Comparison of physical maps of these cDNAs and pcD2-HS1204 was shown in Fig. 1 A. pcD2-HS1207, pcD2-HS1225 and pcD2-HS1237 showed identical patterns on restriction mapping. The pcD2-HS1204 differs from pcD2-HS1207, pcD2-HS1225 and pcD2-HS1237 by an insertion of 70 bp in the coding region near the carboxy terminus, which leads to a frame shift of the open reading frame of pcD2-HS1207, pcD2-HS1225 and pcD2-HS1237. As the result, the deduced amino acid sequence downstream to the site is different between the two species of cDNAs. The nucleotide sequence of the 70 bp insertion shows highly homologous sequence to the portion immediately upstream to the insertion. Furthermore, there is a repeated motif of 5'CTGGATGCCTAAGGGATCCTG within the 70 bp insertion (Fig. 1 B, C).

To determine the functions of the gene products of the two species of cDNAs, we have investigated the expression of α-N-acetylgalactosaminidase activities in COS cells transfected with these cDNAs. As the pcD2 vector is an expression vector containing the SV40 early promoter (14), we transfected COS cells with pcD2-HS1204, pcD2-HS1207, pcD2-HS1225 and pcD2-HS1237 by

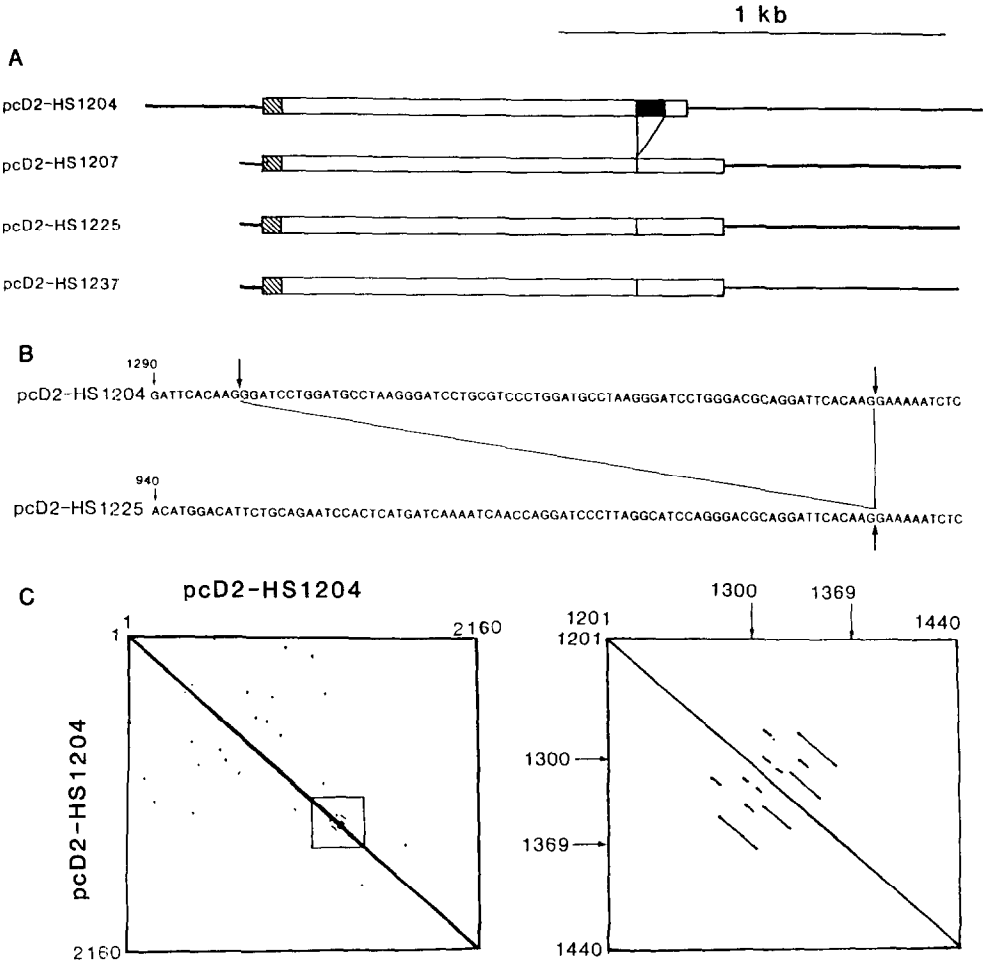

Fig. 1. (A) Physical maps of pcD2-HS1204, pcD2-HS1207, pcD2-HS1225 and pcD2-HS1237. Coding regions for mature human α-N-acetylgalactosaminidase are shown by open boxes. Hatched boxes show putative signal peptides. A 70 bp insertion in pcD2-HS1204 is indicated by a solid box. (B) Comparison of nucleotide sequences of pcD2-HS1204 and pcD2-HS1225. The 70 bp insertion in pcD2-HS1204 is indicated by arrows. (C) Dot-matrix analysis. The complete nucleotide sequence of pcD2-HS1204 are represented on both horizontal and vertical axes (left panel). The dot-matrix analysis of the part of the pcD2-HS1204 (nucleotides 1201-1440) is shown on the right panel. The dot-matrix plot was prepared using a commercially available software, Genetyx (Tokyo, Japan). Segment of 20 nucleotides were compared and a dot was placed when 15 nucleotides are matched.

calcium phosphate method (7). As shown in Table 1, COS cells transfected with pcD2-HS1207, pcD2-HS1225 and pcD2-HS1237 showed 9-11 times higher α-N-acetylgalactosaminidase activities compared to that of mock-transfected COS cells. On the other hand, COS cells transfected with pcD2-HS1204 containing the 70 bp insertion did not show significant increase of α-N-acetyl-galactosaminidase activity.

Fig. 2 shows the nucleotide sequence and the deduced amino acid sequence of pcD2-HS1225. The pcD2-HS1225 is 1821 base pairs in length and

Table 1. Transient expression of α-N-acetylgalactosaminidase in COS cells

DNA	α-N-acetylgalactosaminidase activity
None	1.2 ± 0.6 n moles/min./mg protein
pcD2-HS1204	1.4 ± 0.3
pcD2-HS1207	10.2 ± 3.4
pcD2-HS1225	16.2 ± 1.4
pcD2-HS1237	12.8 ± 2.3

Values are expressed as mean + S.E.M.

contains a poly A tail as well as a polyadenylation signal. The cDNA codes for 411 amino acids with the first 17 amino acids representing a putative signal peptide, as the amino acid sequence of the amino termius was determined by microsequencing of the purified human α-N-acetylgalactosaminidase (3). Within the mature enzyme, six potential N-linked glycosylation sites were identified.

To investigate if the two species of mRNA are expressed in various tissues, we devised a strategy utilizing polymerase chain reaction (PCR).

```
CAGAGCCCAACACATACAGCTGATACACGGCAGACCAGATCTGGTCAGGTCCTCGGAAGCTGAGTCCAGAGCG ATG CTG CTG AAG ACA GTG CTC TTG CTG GGA CAT   105
                                                                            Met Leu Leu Lys Thr Val Leu Leu Leu Gly His    11

GTG GCC CAG GTG CTG ATG CTG GAC AAT GGG CTC CTG CAG ACA CCA CCC ATG GGC TGG CTG GCC TGG GAA CGC TTC CGC TGC AAC ATT   192
Val Ala Gln Val Leu Met↑Leu Asp Asn Gly Leu Leu Gln Thr Pro Pro Met Gly Trp Leu Ala Trp Glu Arg Phe Arg Cys Asn Ile    40

AAC TGT GAT GAG GAC CCA AAG AAC TGC ATA AGT GAA CAG CTC TTC ATG GAG ATG GCT GAC CGG ATG GCA CAG GAT GGA TGG CGG GAC   279
Asn Cys Asp Glu Asp Pro Lys Asn Cys Ile Ser Glu Gln Leu Phe Met Glu Met Ala Asp Arg Met Ala Gln Asp Gly Trp Arg Asp    69

ATG GGC TAC ACA TAC CTC AAC ATT GAT GAC TGC TGG ATC GGT GGT CGC GAT GCC AGT GGC CGC CTG ATG CCG GAT CCC AAG CGC TTC   366
Met Gly Tyr Thr Tyr Leu Asn Ile Asp Asp Cys Trp Ile Gly Gly Arg Asp Ala Ser Gly Arg Leu Met Pro Asp Pro Lys Arg Phe    98

CCT CAT GGC ATT CCT TTC CTG GCT GAC TAC GTT CAC TCC CTG GGC CTG AAG TTG GGT ATC TAC GCG GAC ATG GGC AAC TTC ACC TGC   453
Pro His Gly Ile Pro Phe Leu Ala Asp Tyr Val His Ser Leu Gly Leu Lys Leu Gly Ile Tyr Ala Asp Met Gly Asn◆Phe Thr Cys   127

ATG GGT TAC CCA GGC ACC ACA CTG GAC AAG GTG GTC CAG GAT GCT CAG ACC TTC GCC GAG TGG AAG GTA GAC ATG CTC AAG CTG GAT   540
Met Gly Tyr Pro Gly Thr Thr Leu Asp Lys Val Val Gln Asp Ala Gln Thr Phe Ala Glu Trp Lys Val Asp Met Leu Lys Leu Asp   156

GGC TGC TTC TCC ACC CCC GAG GAG CGG GCC CAG GGG TAC CCC AAG ATG GCT GCT GCC CTG AAT GCC ACA GGC CGC CCC ATC GCC TTC   627
Gly Cys Phe Ser Thr Pro Glu Glu Arg Ala Gln Gly Tyr Pro Lys Met Ala Ala Ala Leu Asn◆Ala Thr Gly Arg Pro Ile Ala Phe   185

TCC TGC AGC TGG CCA GCC TAT GAA GGC GGC CTC CCC CCA AGG GTG AAC TAC AGT CTG CTG GCG GAC ATC TGC AAC CTC TGG CGT AAC   714
Ser Cys Ser Trp Pro Ala Tyr Glu Gly Gly Leu Pro Pro Arg Val Asn◆Tyr Ser Leu Leu Ala Asp Ile Cys Asn Leu Trp Arg Asn   214

TAT GAT GAC ATC CAG GAC TCC TGG TGG AGC GTG CTC TCC ATC CTG AAT◆TGG TTC GTG GAG CAC CAG GAC ATA CTG CAG CCA GTG GCC   801
Tyr Asp Asp Ile Gln Asp Ser Trp Trp Ser Val Leu Ser Ile Leu Asn Trp Phe Val Glu His Gln Asp Ile Leu Gln Pro Val Ala   243

GGC CCT GGG CAC TGG AAT GAC CCT GAC ATG CTG CTC ATT GGG AAC TTT GGT CTC AGC TTA GAG CAA TCC CGG GCC CAG ATG GCC CTG   888
Gly Pro Gly His Trp Asn Asp Pro Asp Met Leu Leu Ile Gly Asn Phe Gly Leu Ser Leu Glu Gln Ser Arg Ala Gln Met Ala Leu   272

TGG ACG GTG CTG GCA GCC CCC CTC TTG ATG TCC ACA GAC CTG CGT ACC ATC TCC GCC CAG AAC ATG GAC ATT CTG CAG AAT CCA CTC   975
Trp Thr Val Leu Ala Ala Pro Leu Leu Met Ser Thr Asp Leu Arg Thr Ile Ser Ala Gln Asn Met Asp Ile Leu Gln Asn Pro Leu   301

ATG ATC AAA ATC AAC CAG GAT CCC TTA GGC ATC CAG GGA CGC AGG ATT CAC AAG GAA AAA TCT CTC ATC GAA GTG TAC ATG CGG CCT  1062
Met Ile Lys Ile Asn Gln Asp Pro Leu Gly Ile Gln Gly Arg Arg Ile His Lys Glu Lys Ser Leu Ile Glu Val Tyr Met Arg Pro   330

CTG TCC AAC AAG GCT AGC GCC TTA GTC TTC TTC AGC TGC AGG ACC GAT ATG CCT TAT CGC TAC CAC TCC TCC CTT GGC CAG CTG AAC  1149
Leu Ser Asn Lys Ala Ser Ala Leu Val Phe Phe Ser Cys Arg Thr Asp Met Pro Tyr Arg Tyr His Ser Ser Leu Gly Gln Leu Asn   379

TTC ACC GGG TCT GTG ATA TAT GAG GCC CAG GAC GTC TAC TCA GGT GAC ATC ATC AGT GGC CTC CGA GAT GAA ACC AAC TTC ACA GTG  1236
Phe Thr Gly Ser Val Ile Tyr Glu Ala Gln Asp Val Tyr Ser Gly Asp Ile Ile Ser Gly Leu Arg Asp Glu Thr Asn◆Phe Thr Val   388

ATC ATC AAC CCT TCA GGG GTA GTG ATG TGG TAC CTG TAT CCC ATC AAG AAC CTG GAG ATG TCC CAG CAG TGAGGAGCTGGGACATGTGACAG  1328
Ile Ile Asn◆Pro Ser Gly Val Val Met Trp Tyr Leu Tyr Pro Ile Lys Asn Leu Glu Met Ser Gln Gln

GCTGTGGTGGCACCACTGAGCCTAGACCATGGAGCCTTGGCATGCCCAGGGCAAGTGGGGAGGTTCTCTGCTCCCCAGGCCTGCTCGGTGACTGACCCCATCATACCCAAAGTGCA  1444

ATCTCACGGCCAGGTTCTATGCCCTGTCCAAGCGTAAACCCTCTTGGAAACTTCTTTTGGGGCAATTTTCCTGTGGCCTTCCTGGCCTCTACTTCCATGTGCGCAGCCCCACAGAC  1560

GTTGCTGAGCAACTCGCCAGCCTCCTGAGCTCCATGCCCATCAGGACTCTAGCCTCTGACCTTGCTGTTGACTCTGAAATCAGGATTTGGAAGTTTTCGAATTAGGAGTAGAGAGA  1676

TCTGACCTCTTGCCAGGAATGCCCATGGATCATGTGATTGGCTTTTCTACCCATAGAGGGCCTTGCAGCCTGATACCACCTGGGAGTGAGGGTCACAAAGGAGACCTTGGCTCCCT  1792

CAGGTCACCAATAAACCTGTTCTTTAATC(A)n
```

Fig. 2. Nucleotide sequence analysis of human α-N-acetylgalactosaminidase cDNA (pcD2-HS1225) and the deduced amino acid sequence. The signal peptidase cleavage site is shown by an arrow. Potential N-linked glycosylation sites are shwon by diamonds. Putative polyadenylation signal is boxed. Amino acid sequences determined by direct sequencing of the amino terminus and a tryptic peptide are underlined.

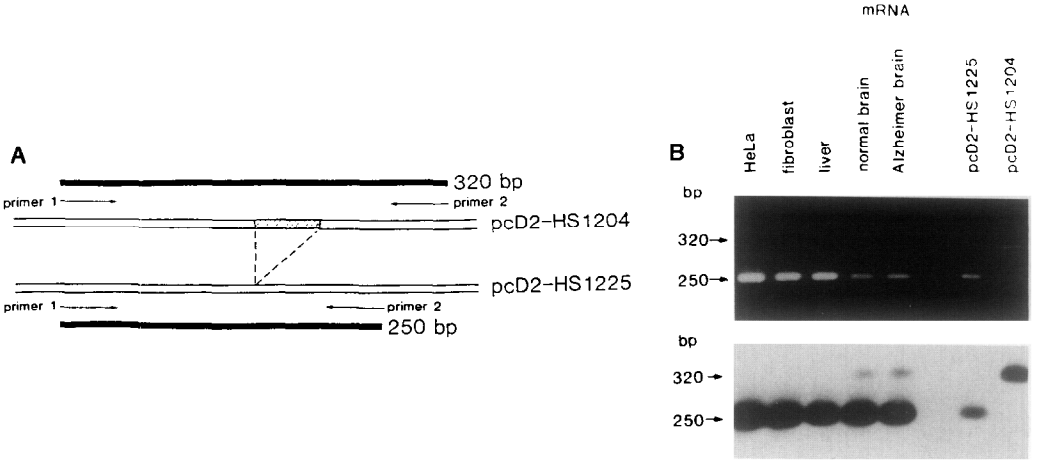

Fig. 3. (A) The strategy for the analysis of mRNA species. (B) Poly A(+) RNA (5 μg) was converted to complementary DNA by reverse transcriptase and the aliquot of the reaction mixture was used directly for polymerase chain reaction (PCR). The PCR products were electrophoresed through 1.8 % agarose gel (upper panel), transferred to a nitrocellulose membrane and hybridized to ^{32}P-labelled KpnI fragment of pcD2-HS1204 spanning the region of the 70 bp insertion.

Under the strategy shown in Fig. 3 A, it was demonstrated that there are two species of cDNAs in the pcD2 fibroblast cDNA library and the majority was the cDNA containing the 70 bp insertion (data not shown). To analyze mRNA species, poly A(+) RNA was isolated from various tissues and analyzed by PCR after the poly A(+) RNA was converted to cDNA by reverse transcriptase. As shown in Fig. 3B, in HeLa cells, human fibroblasts and a human liver, only the species of mRNA lacking the 70 bp insertion was detected. In human brains, however, it was demonstrated that an mRNA containing the 70 bp insertion is present only in a minor amount.

DISCUSSION

Of the two species of cDNAs we have isolated, only the cDNA lacking the 70 bp insertion showed increase of α-N-acetylgalactosaminidase activity in COS cells. Although the pcD2-HS1204 did not show expression of the enzyme activity, we have shown that the mRNA containing the 70 bp insertion is present in human brains in a minor amount. The analysis of mRNA species by PCR has proved extremely useful for distinguishing mRNAs with such a small difference of the 70 bp insertion. The result excludes the possibility that the cDNA containing the 70 bp insertion was generated by a cloning artifact. In the previous paper, we showed that the nucleotide sequence of pcD-2-HS1204 has highly homologous sequence to human α-galactosidase A (3). Compared to the nucleotide sequence of human α-galactosidase A genomic DNA

(15), it was discovered that the 70 bp insertion found in the human α-N-acetylgalactosaminidase cDNA, pcD2-HS1204, corresponds to the boundary of exon 6 and exon 7 of human α-galactosidase A gene. Although we have not analyzed the human α-N-acetylgalactosaminidase genomic DNA, it seems likely that the two species of mRNA for the α-N-acetylgalactosaminidase are generated by an alternative splicing from the single gene for α-N-acetylgalactosaminidase. This point should be clarified by detailed analysis of the human α-N-acetylgalactosaminidase gene.

In the present study we demonstrated that mRNA which lacks the 70 bp insertion codes for active human α-N-acetylgalactosaminidase. At the present, functional implication of the mRNA containing the 70 bp insertion is still unclear. As shown in Fig. 1, the 70 bp insertion shows highly homologous sequence to the portion immediately upstream to the insertion. Similar structures have been reported in erythrocyte ankyrin cDNA (16).

It has been shown that human α-N-acetylgalactosaminidase catalyzes hydrolysis of various substrates containing not only α-linked N-acetylgalactosamine but also α-linked galactose (17) at non-reducing ends. Furthermore very recently structures of amino acid glycosides excreted in urine of a patient with the α-N-acetylgalactosaminidase deficiency have been determined (18). The availability of expression system of human α-N-acetylgalactosaminidase cDNA should enable us detailed analyses of the α-N-acetylgalactosaminidase activities toward various kinds of natural substrates, which are under way in our laboratory.

ACKNOWLEDGMENTS

This study was supported in part by grants from the Ministry of Education of Japan, and Grant (11-B) from National Center of Neurology and Psychiatry (NCNP) of the Ministry of Health and Welfare Japan.

REFERENCES

1. van Diggelen, O.P., Schindler, D., Kleijer, W.J., Huijmans, J.M.G., Galjaard, H., Linden, H.U., Peter-Katalinic, J., Egge, H., Dabrowski, U. and Cants, M. (1987) Lancet 1, 804
2. Kanzaki, T., Yokota, M., Mizuno, N., Matsumoto, Y. and Hirabayashi, Y. (1989) Lancet i,875-877
3. Tsuji, S., Yamauchi, T., Hiraiwa, M., Isobe, T., Okuyama, T., Sakimura, K., Takahashi, Y., Nishizawa, M., Uda, Y. and Miyatake, T. (1989) Biochem. Biophys. Res. Commun. 163,1498-1504
4. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74,5463-5467
5. Chen, E.Y. and Seeburg, P.H. (1985) DNA 4,165-170
6. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. U.S.A. 80,3963-3965
7. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) In Molecular Cloning. A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
8. Beutler, E. and Kuhl, W. (1972) J. Biol. Chem. 247,7195-7200

9. Schram, A.W., Hamers, M.N. and Tager, J.M. (1977) Biochim. Biophys. Acta 482,138-144
10. Kobayashi, H., Sakimura, K., Kuwano, R., Sato, S., Ikuta, F., Takahashi, Y., Miyatake, T. and Tsuji, S. J. Mol. Neurosci. (in press)
11. Saiki, R., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985) Science 230,1350-1354
12. Chelly, J., Kapaln, J.-C., Marie, P., Gautron, S. and Kahn, A. (1988) Nature 333,858-860
13. Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc. Natl. Acad. Sci. U.S.A. 76,3683-3687
14. Chen, C. and Okayama, H. (1987) Mol. Cell. Biol. 7,2745-2752
15. Bishop, D.F., Kornreich, R. and Desnick, R.J. (1988) Proc. Natl. Acad. Sci. U.S.A. 85,3903-3907
16. Lux, S.E., John, K.M. and Bennet, V. (1990) Nature 344,36-42
17. Desnick, R.J. and Bishop, D.F. (1989) In The Metabolic Basis of Inherited Disease. (Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D.V., Ed.) 6th ed. pp1751-1796, McGraw-Hill, New york, N.Y.
18. Hirabayashi, Y., Matsumoto, Y., Matsumoto, M., Toida, T., Iida, N., Matsubara, T., Kanzaki, T., Yokota, M. and Ishizuka, I. (1990) J. Biol. Chem. 265,1693-1701