Am. J. Hum. Genet. 48:351–361, 1991

The Glycine Cleavage System: Structure of a cDNA Encoding Human H-Protein, and Partial Characterization of Its Gene in Patients with Hyperglycinemias

Hirohisa Koyata and Koichi Hiraga

Department of Biochemistry, Toyama Medical and Pharmaceutical University School of Medicine, Toyama, Japan

Summary

We have isolated a 1,192-base-long cDNA which encodes the entire structure of a precursor form of human H-protein. The tentatively calculated number of copies for this cDNA appeared to be about four times as many as that of the antithrombin III gene specified by a single locus in the human haploid genome. Southern analysis using H-protein cDNA probe demonstrates a deletion of the 5.0-kb *Sac*I fragment in the genome of a patient with an atypical nonketotic hyperglycinemia in whom there was an inactive H-protein. This *Sac*I fragment was also deleted from the genome of one of seven patients with nonketotic hyperglycinemia resulting from the lesion of glycine decarboxylase. The remaining six patients had common aberrations identified with the 5.2-kb *Eco*RI and 5.5-kb *Sac*I fragments. Although implication of these defective fragments in pathogenesis is unclear at present, it is suggested that rearrangements occur in multiple genomic loci of patients with nonketotic hyperglycinemia and that this H-protein cDNA can be used for carrier screening.

Introduction

Hydrogen carrier protein (H-protein) is one of the four constituents of the glycine cleavage system which catalyzes reversibly the catabolic degradation of glycine (Kikuchi 1973). This protein could form an enzyme complex with another constituent, glycine decarboxylase. Since glycine decarboxylase is inactive by itself, the observed interaction has seemed to be essential for expression of the catalytic reaction of glycine decarboxylase (Hiraga and Kikuchi 1980*a*, 1980*b*). H-protein also has appeared to carry an intermediary product through lipoyl moiety in the overall reaction (Motokawa et al. 1970; Kikuchi 1973). Therefore, an aberrant H-protein causes a complete deficit of the glycine cleavage activity as reported in a previously published study (Hiraga et al. 1981). Because the glycine cleavage reaction plays a major role in the glycine catabolism in vertebrates (Yoshida et al. 1969; Yoshida and Kikuchi 1970), a defective glycine cleavage system results in pathologic conditions called hyperglycinemia in man (Yoshida et al. 1969; Tada and Hayasaka 1987; Nyhan 1989).

There have been two differently classified types of abnormal H-protein so far. One is an H-protein devoid of the prosthetic lipoic acid which is covalently bound with apo-H-protein. This was found in the liver of a patient with an atypical nonketotic hyperglycinemia in whom there was an additional lesion, a progressive degeneration in the central nervous system (Hiraga et al. 1981; Trauner et al. 1981). Although an impaired H-protein is rarely involved in the molecular lesion of nonketotic hyperglycinemia, this important case can be used to determine whether the altered structure resides in the genome. The present analysis surveys what causes the two different abnormalities in one patient.

Nonketotic hyperglycinemia has differed from ketotic hyperglycinemia, in that, in the latter, patients have had organic acidemia such as propionic acidemia and methylmalonic acidemia (Childs et al. 1961; Nyhan et al. 1961; Rosenberg et al. 1968; Baumgartner et al. 1969). An inactive H-protein, which has not

Received March 22, 1990; revision received August 9, 1990.

Address for correspondence and reprints: Dr. Koichi Hiraga, Department of Biochemistry, Toyama Medical and Pharmaceutical University School of Medicine, 2630 Sugitani, Toyama 930–01, Japan.

0002-9297/91/4802-0020$02.00

been well characterized yet, has been immunochemically detected in the liver of a patient with ketotic hyperglycinemia (Kume et al. 1988). The genome of this patient would be a control for the above study.

Among three essential constituents of this enzyme system, glycine decarboxylase has most frequently been assigned to a molecular lesion of nonketotic hyperglycinemia (Tada and Hayasaka 1987). It is also interesting to determine whether patients with this molecular lesion have the genomic alterations detectable by the H-protein cDNA probe, because the lesion in the central nervous system is universally involved in this disease (Tada and Hayasaka 1987; Nyhan 1989).

We isolated some human H-protein cDNA clones by using a specific antibody in a previous study, but the cloned cDNA appeared to have no initiator methionine codon and poly(A) region (Hiraga et al. 1988). Through a couple of further trials for selection by using an immunoreactive H-protein cDNA as a probe, we have obtained a longer H-protein cDNA encoding the entire structure of the precursor H-protein. We report here several features involved in the human H-protein cDNA and its genes, in both normal individuals and patients with hyperglycinemias.

Material and Methods

cDNA Probe and Other Materials

We cloned previously an H-protein cDNA from a human liver cDNA expression library. The immunoreactive cDNA clone designated as pHH15B9 appeared to encode the partial primary structure of the purified H-protein (Hiraga et al. 1988). However, because no initiator methionine codon and poly(A) tail were found in the sequence, the structures of the cDNA and deduced protein have not been reported yet. In the present study, we used the nick-translated pHH15B9 cDNA (744 bp) for the selection of a longer cDNA. Restriction endonucleases, DNA- and RNA-modifying enzymes, and other reagents were commercially obtained. Radioactive nucleotides were products of DuPont—New England Nuclear.

Selection of the H-Protein cDNA

About 5×10^5 phages of the above library were again tested, and a longer cDNA was selected. Phage DNA fixed on a nitrocellulose filter was treated with a prehybridization solution composed of 50 mM Tris-HCl buffer (pH 8.0), 1 M NaCl, 5 mM EDTA, 0.1% SDS, 50% formamide, 100 μg/ml of single-stranded salmon testes DNA, and 5 × Denhardt's solution at 42°C for 6–14 h and hybridized with the radioactive probe in this solution. The filter was washed at 42°C–50°C with several changes of the solution for low or high stringency. For low stringency, 100 mM Tris-HCl buffer (pH 8.0) containing 250 mM NaCl, 1 mM each of EDTA and disodium phosphate, 0.1% SDS, and 1 × Denhardt's solution was used. Concentrations of Tris-HCl buffer (pH 8.0) and NaCl were adjusted to 10 and 25 mM, respectively, for high stringency. A plasmid vector, Bluescript SK (M13 –) (Stratagene Cloning Systems, San Diego), was used.

DNA Sequencing

Nucleotide sequence was determined by the method of Sanger et al. (1977) by using 7-deaza-dGTP (Mizusawa et al. 1986). The promoter sequences for T7 and T3 RNA polymerases on the plasmid were used for the priming sites. The oligonucleotides identical to both strands at the particular positions of the H-protein cDNA sequence were prepared using a DNA synthesizer, model 381A (Applied Biosystems, Inc., Japan, Tokyo) and were designated as DNA 1 (nucleotides 321–338 in the cDNA sequence), DNA 2 (nucleotides 475–492), and their complementary counterparts. These were also used for primers.

Northern Analysis and Southern Analysis

Total RNA was isolated, from macroscopically normal portions within a surgically resected and cancer-bearing liver, by the method of Fyrberg et al. (1980) and was subjected to northern analysis (Lehrach et al. 1977). High-molecular-weight DNA was obtained from the nucleated blood cells of 13 normal Japanese individuals by the method of DiLella and Woo (1987).

Portions of the frozen liver specimens from a patient with an atypical nonketotic hyperglycinemia (patient LS) (Hiraga et al. 1981; Trauner et al. 1981) and from a patient with a propionic acidemia–associated ketotic hyperglycinemia (patient TI) (Kume et al. 1988) were ground into a fine powder in liquefied nitrogen, and genomic DNA was isolated by a method similar to the one described above. The frozen liver specimens from seven patients with nonketotic hyperglycinemia resulting from the deficient glycine decarboxylase activity were transported to our laboratory through Dr. Tada, Department of Pediatrics, Tohoku University School of Medicine. In the present study, the livers from patients FS, VM, and DT (Kume et al. 1988), from patient 6 in the previously published report (pa-

tient MR) (Hayasaka et al. 1987), and from patient JL (Hayasaka et al. 1983) were subjected to analysis. There is no report on patients AH and YH. The small amount of the protein reactive to an anti-chicken glycine decarboxylase antibody was in the liver extract of patient DT (Kume et al. 1988). Using this antibody in the liver extracts from patients FS and VM (Kume et al. 1988) and in the remaining patients (data not shown), we confirmed that no immunoreactive material was detected. These livers showed both H-protein and T-protein activities at levels comparable to those found in the control livers. A 1.0% agarose gel was used for northern analysis, and a 0.8% gel was used for Southern analysis. Probing was conducted as described in the preceding section. DNA fragments with known sizes were used as size markers.

The Number of Copies for the H-Protein cDNA in the Human Genome

We isolated a human antithrombin III (AT III) cDNA by using a specific antibody (Hoechst Japan, Tokyo). This insert (1,470 bp) encompassed nucleotides 13–1,482 of the known sequence reported by Prochownik et al. (1983). Since it has been demonstrated that AT III is specified by a single genomic locus (Kao et al. 1984; Bock et al. 1985), we employed the AT III gene as a standard to estimate the number of copies of the H-protein cDNA in the human genome by Southern analysis. Two different cDNA probes with similar sizes and specific radioactivities would reveal similar integrated intensities of their hybrids on an autoradiogram. Southern analysis detected the difference in the number of copies for certain pseudogenes (Piechaczyk et al. 1984).

To avoid hybridization with the sequences rich in A and T bases, the poly(A) region was removed by cutting the pHH24S cDNA with *Alu*I. The resultant size of the probe was 1,120 bp. Genomic DNA from a control individual was thoroughly digested with *Eco*RI. An equal amount (20 μg for each lane) of the *Eco*RI-treated genomic DNA was separated on an agarose gel, was transferred onto a single nitrocellulose filter, and was baked at 80°C. The pHH24L and pHH24S inserts (1 μg and 0.3 μg, respectively) were mixed to an equimolar concentration and were nick-translated. The AT III cDNA (1.7 μg) was also nick-translated in parallel. Thus, in terms of molar amount of DNA molecule, equal amounts of the two probes were prepared. Nevertheless, on the basis of the 1,120-bp size, the H-protein cDNA probe had 1.47 times more radioactivity than did the AT III cDNA probe. Southern analysis was carried out using these probes under comparable conditions. Signals were located by autoradiography, and their intensities were measured using a Shimadzu TLC Scanner CS-910 equipped with an integrator, Chromatopac C-R1A. Integrated density of signals revealed by the AT III probe was multiplied by 1.47 and then was compared with that of all the signals revealed by the H-protein cDNA probe.

Results

Structure of the Human H-Protein cDNA

By rescreening, we isolated an H-protein cDNA longer than the pHH15B9 insert and designated λHH24. This cDNA contained one internal *Eco*RI site at which two fragments (about 800 and 300 bp) were formed (fig. 1*A*). The fragments thus formed with *Eco*RI were separately subcloned and designated as pHH24L (for the longer insert) and pHH24S (for the shorter insert).

As summarized in figure 1*B*, DNA sequencing has demonstrated that the pHH24L insert is 853 bp and that its sequence near one end coincides with the 3′ region of the pHH15B9 cDNA (see fig. 1 legend). The first methionine codon begins 93 bp downstream from the other end and precedes an open reading frame encoding 173 amino acid residues until the translation termination codon at nucleotides 612–614. The deduced peptide contains both the dodecapeptide assigned to the amino-terminal region of the purified human H-protein (underlined in fig. 1*B*) and the carboxy-terminal glutamate residue, which are confirmatory of the previous observations (Hiraga et al. 1988). Therefore, we concluded that the pHH24L insert encodes part of human H-protein mRNA and that the precursor H-protein comprises a mitochondrial presequence of 48 amino acid residues and the mature protein of 125 amino acid residues.

The pHH24S insert is 339 bp and has a 39-bp poly(A) tail at one end. The structure of the *Eco*RI site at the other end, however, differs from that of the *Eco*RI linker, indicating that the 1,192-bp human H-protein cDNA sequence is reconstituted by direct linkage of the two cDNA fragments through the internal *Eco*RI site.

Two genuine consensus sequences for the poly(A) site are found in the 3′ untranslated region (nucleotides 629–634 and 1,135–1,140). Northern analysis has also demonstrated that the H-protein cDNA probe hybridized to two distinctly sized RNAs (1,400 and

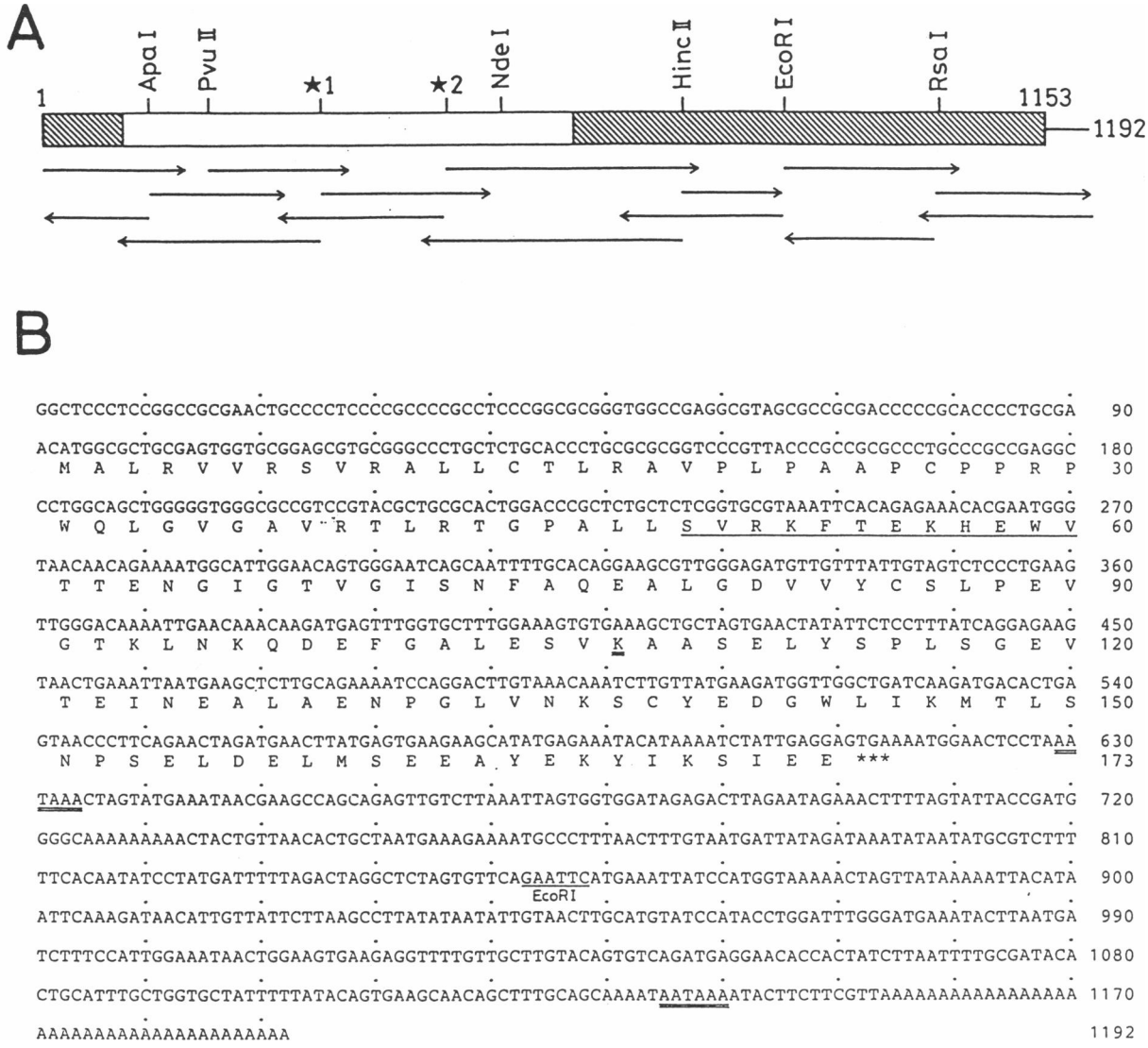

Figure 1 Structure of human H-protein cDNA. *A*, Restriction map of λHH24 insert. The open and hatched bars indicate the protein coding and untranslated regions, respectively. The line at the 3′ region stands for the poly(A) tail. The sites for restriction endonucleases were determined from the nucleotide sequence and were confirmed by digestion with the enzymes. The numerals with asterisks (*1 and *2) indicate the positions of the sequence from which the two 17-mer oligodeoxynucleotides were synthesized. Arrows indicate both the strands and the regions of the nucleotide sequences determined. *B*, Nucleotide sequence of human H-protein cDNA. The deduced primary structure of H-protein is displayed under the sequence. The amino-terminal region of mature H-protein is underlined. The translation termination codon is denoted by the triple asterisks (***), and the two genuine consensus sequences for the poly(A) sites are double underlined. The pHH15B9 cDNA spans nucleotides 114–857.

900 bp) in liver total RNA (fig. 2). The distance (500 bp) of the two consensus sequences appears to be equal to the size difference of the RNAs. Densitometric estimation clearly indicated that the longer RNA is dominant in abundance. There was about 50 times more of the longer RNA than of the shorter RNA.

RFLP Revealed by Using H-Protein cDNA in the Normal Human Genome

Prior to examination of the H-protein gene structure in the patients with hyperglycinemia, we elucidated RFLP detectable by the H-protein cDNA in the normal genome. For this purpose, we chose *Eco*RI, *Pst*I, and

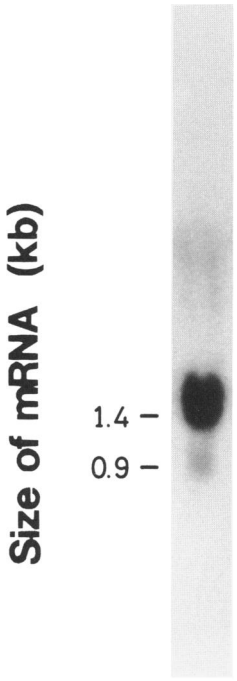

Figure 2 Northern analysis using nick-translated pHH24L cDNA probe. About 40 μg of human liver total RNA was used. *Escherichia coli* ribosomal RNA was used as a size marker.

*Sac*I. The genomic DNA digested with one of them was probed with either pHH24L cDNA or pHH24S cDNA.

Polymorphic *Pst*I and *Sac*I sites were found. One of 13 DNA preparations had polymorphic 3.3-kb *Pst*I fragments revealed by both probes (figs. 3*A* and 3*B*, lanes 5). A 1.3-kb *Sac*I fragment which also hybridized with both probes has been lost in the other genomic DNA. This DNA preparation did not reveal a 6.4-kb *Sac*I fragment and, instead, appeared to possess a 7-kb fragment hybridized with both probes (figs. 3*A* and 3*B*, lanes 9). The two 3.3-kb *Pst*I fragments appear to exist together, and the polymorphic disappearance of the 1.3- and 6.4-kb *Sac*I fragments is found to be associated with the appearance of a novel 7.0-kb fragment. However, the precise linkage of these individual polymorphisms is unclear at present.

In the *Eco*RI-treated DNA, the pHH24L probe (0.8 kb) detects nine fragments (about 45 kb in integrated length), and the alternative 0.3-kb probe hybridizes to six fragments of about 15 kb in total length (figs. 3*A* and 3*B*, lanes 1, 4, and 7). However, no polymorphic *Eco*RI site was found in the normal genome. These results demonstrate that the fragments detected by the

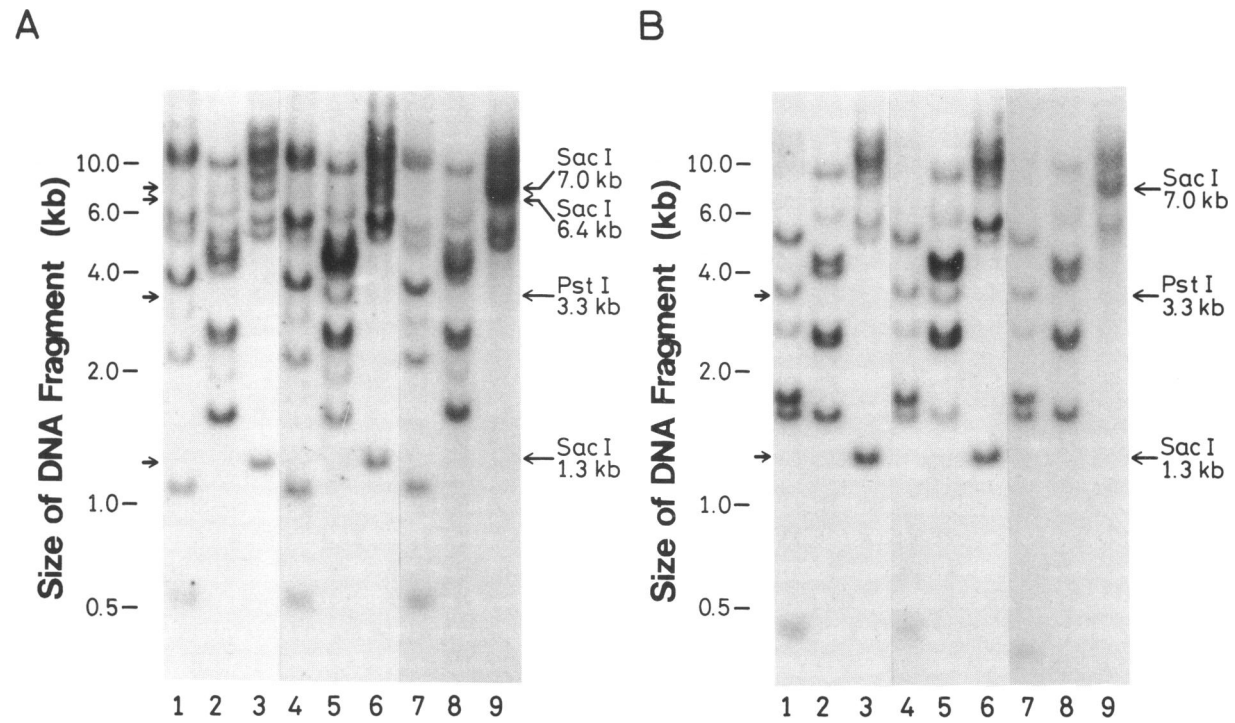

Figure 3 RFLP in normal human genome. The genomic DNAs from normal individuals (20 μg) were treated with either *Eco*RI (lanes 1, 4, and 7 in both *A* and *B*), *Pst*I (lanes 2, 5, and 8 in both *A* and *B*), and *Sac*I (lanes 3, 6, and 9 in both *A* and *B*). The genomic DNAs containing no polymorphic site identified (lanes 1–3 in both *A* and *B*) and the polymorphic *Pst*I (lanes 4–6 in both *A* and *B*) or *Sac*I (lanes 7–9 in both *A* and *B*) site are probed with the pHH24L and pHH24S cDNAs in *A* and *B*, respectively.

1-kb-long cDNA sequence is coded in the extremely long stretch of about 60 kb in the genome. This size is similar to the integrated length of the fragments displayed in the *Pst*I- and *Sac*I-treated DNAs. When all of these considerations are taken into account, the human genome is expected to contain multiple copies of the H-protein cDNA sequence.

The Number of Copies for the H-Protein cDNA in the Human Genome

To further examine this issue, the *Eco*RI-treated genomic DNA was subjected to the estimation of the number of copies that is described in Material and Methods. An equimolar mixture of the pHH24L and pHH24S probes reveals 11 fragments (fig. 4, lane 1). In contrast, identical DNA preparation has two fragments hybridized with the AT III cDNA probe (fig. 4, lane 2). Their sizes are 11 and 4 kb and are consistent with the finding reported by Prochownik et al. (1983). These results also suggest that no fragment formed by partial digestion with *Eco*RI is included in this DNA preparation and that the washing conditions used were stringent enough to remove nonspecific hybrids.

Next, the integrated density of all the hybridization signals (fig. 4, lane 1) was calculated to be 17.8 in arbitrary units, whereas that revealed with the AT III cDNA probe was 3.33 (fig. 4, lane 2). Therefore, the latter was corrected to be 4.9 units on the basis of the 1.1-kb probe. In this context, contents of G + C bases in the H-protein cDNA and in the AT III cDNA were 44% and 50%, respectively. The calculated T_m values of the hybrids that corresponded exactly in size to these cDNA probes were about 66°C for the hybrid with the H-protein cDNA and 69°C for that with the AT III cDNA under the condition for hybridization, and they were 75°C and 78°C, respectively, for the high-stringency washing. These values suggest that the average efficiencies in hybridization during Southern analysis are not largely different in either experimental system, although, when the theoretical consideration regarding the nick-translation is taken into account, the actual sizes of the radioactive probes are thought to be shorter than 1,100 bp. From these considerations, we judged that the number of copies for the H-protein cDNA probe is at least four times greater than that of the AT III gene in the genome.

Occurrence of Genomic Rearrangement in a Patient with an Atypical Nonketotic Hyperglycinemia

The occurrence of multiple copies in the genome made it likely that there would be a great deal of difficulty in interpreting the results that would be shown by Southern analysis. We, however, attempted to determine whether a particularly altered structure is included in the genomes of patients LS and TI. For this purpose, the *Eco*RI-, *Pst*I-, or *Sac*I-treated genomic DNAs from these patients were probed with the pHH24L and pHH24S cDNAs.

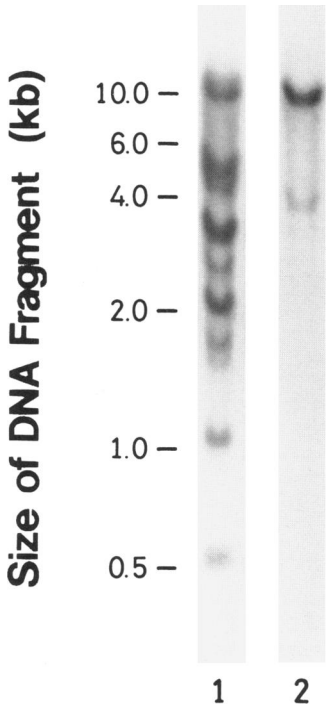

Figure 4 Number of copies for H-protein cDNA in human genome. The locations of the fragments revealed by using the equimolar mixture of the pHH24L and *Alu*I-treated pHH24S inserts (lane 1) or the antithrombin III cDNA probe (lane 2) are shown. An aliquot (20 μg) of identical *Eco*RI-treated DNA was used. The autoradiogram was made on a single sheet of film. The 0.4-kb fragment falls into the background under the condition for printing this photograph.

There was no aberrant distribution of fragments revealed by the pHH24L and pHH24S probes in the *Eco*RI-treated DNA of patient LS (figs. 5*A* and 5*B*, lanes 1, 3, and 5). Similarly, no significant aberration was found in the *Pst*I-treated DNA of this patient, although both the 4.7-kb fragment from the longer probe and the 6-kb fragment from the shorter probe looked obscure (data not shown).

As to the *Sac*I-treated DNA, because of the disappearance not only of the 1.3- and 6.4-kb fragments from the pHH24L probe but also of the 1.3-kb fragment from the pHH24S probe, and because of the

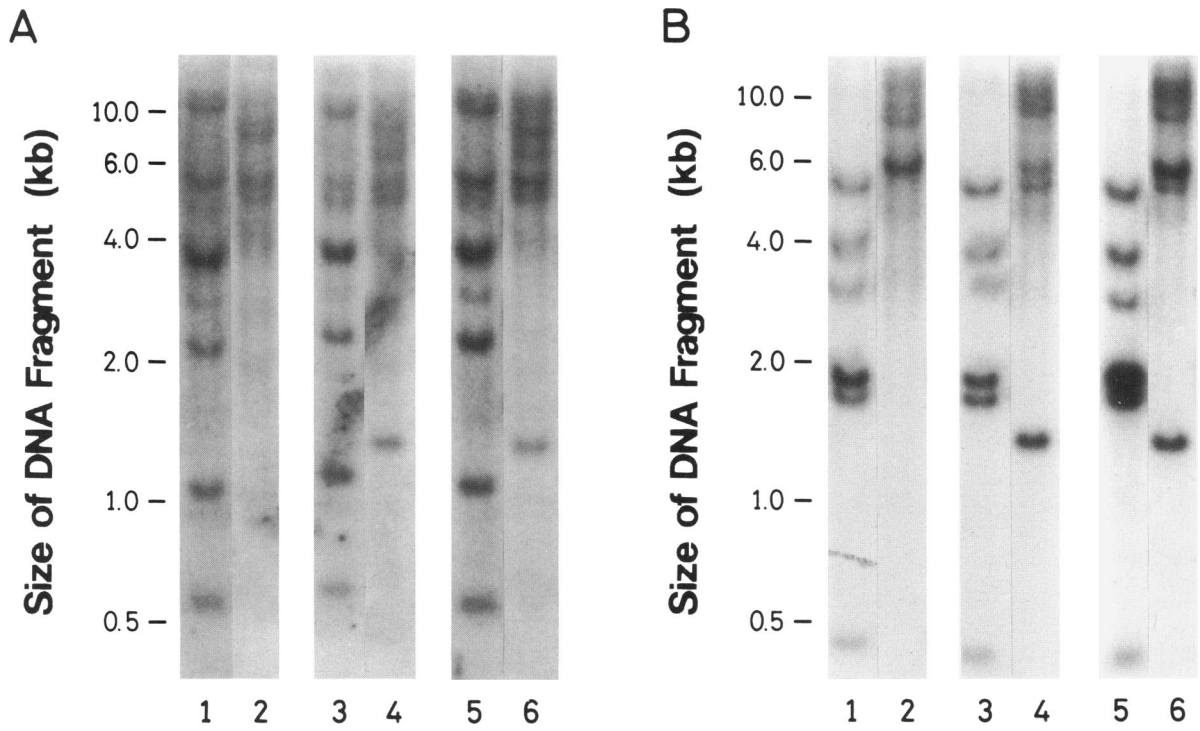

Figure 5 RFLP in patients with hyperglycinemia resulting from inactive H-protein. Genomic DNA preparations (20 μg) from patient LS (lanes 1 and 2 in both *A* and *B*), patient TI (lanes 3 and 4 in both *A* and *B*), and a control (lanes 5 and 6 in both *A* and *B*) were treated with either *Eco*RI (lanes 1, 3, and 5 in both *A* and *B*) or *Sac*I (lanes 2, 4, and 6 in both *A* and *B*). The nick-translated pHH24L or pHH24S probes were used in *A* and *B*, respectively.

occurrence of a 7-kb fragment from each of the two probes (figs. 5*A* and 5*B*, lanes 2), this patient was found to carry the *Sac*I polymorphism shown in figure 3. However, it is also clear that the 5.0-kb fragment, which is deficient in this patient (fig. 5*B*, lane 2), is not included in this polymorphism. These results led us to the conclusion that patient LS had the significantly rearranged genome. On the other hand, the genome of patient TI showed no abnormal distribution of hybridization signals detectable by this procedure (figs. 5*A* and 5*B*, lanes 3 and 4).

Some Characteristics of Genomic Rearrangements in Patients with Molecular Lesion in Glycine Decarboxylase

We also examined the genomic structures of seven patients in whom there may be a lesion involved in the glycine decarboxylase gene. The *Eco*RI-treated DNA was first examined (fig. 6). Patient YH showed no abnormal distribution of the hybridization signals revealed by both the pHH24L and the pHH24S probes. Patient AH seemed to carry a novel 4.0-kb fragment. Patients VM and DT have lost the 5.2-kb fragment. This 5.2-kb fragment is obscure in the genome from patients MR, FS, and JL. Instead, signals at the regions corresponding to 4.4–4.7 kb in size appear to be broad in the lanes for these three patients. Because the 5.2-kb fragment has disappeared, it is unclear whether the genuine 4.7-kb fragment exists at its intrinsic position in lanes 2–4. The faint 10–12-kb fragments are observed in lanes 1–5. All the patients tested had no defective and aberrant fragment detectable by the pHH24S probe (fig. 6*B*).

The *Sac*I-treated genomic DNAs show the more obvious alterations in structure of the genome (fig. 7). Patient YH showed the normal distribution of the fragments in lane 6 of figure 7*A* but showed one deleted fragment (5.0 kb) in lane 6 of figure 7*B*. The 5.5-kb fragments revealed by both probes in the control DNA are undetectable in the genomes of patients VM and DT (figs. 7*A* and 7*B*, lanes 5 and 7). From the findings in this figure, patient DT appears to belong to the *Sac*I polymorphism (fig. 7*A* and 7*B*, lanes 7). The 5.5-kb fragments are not located at the intrinsic position by both probes in lanes 1–4 (patients AH, MR,

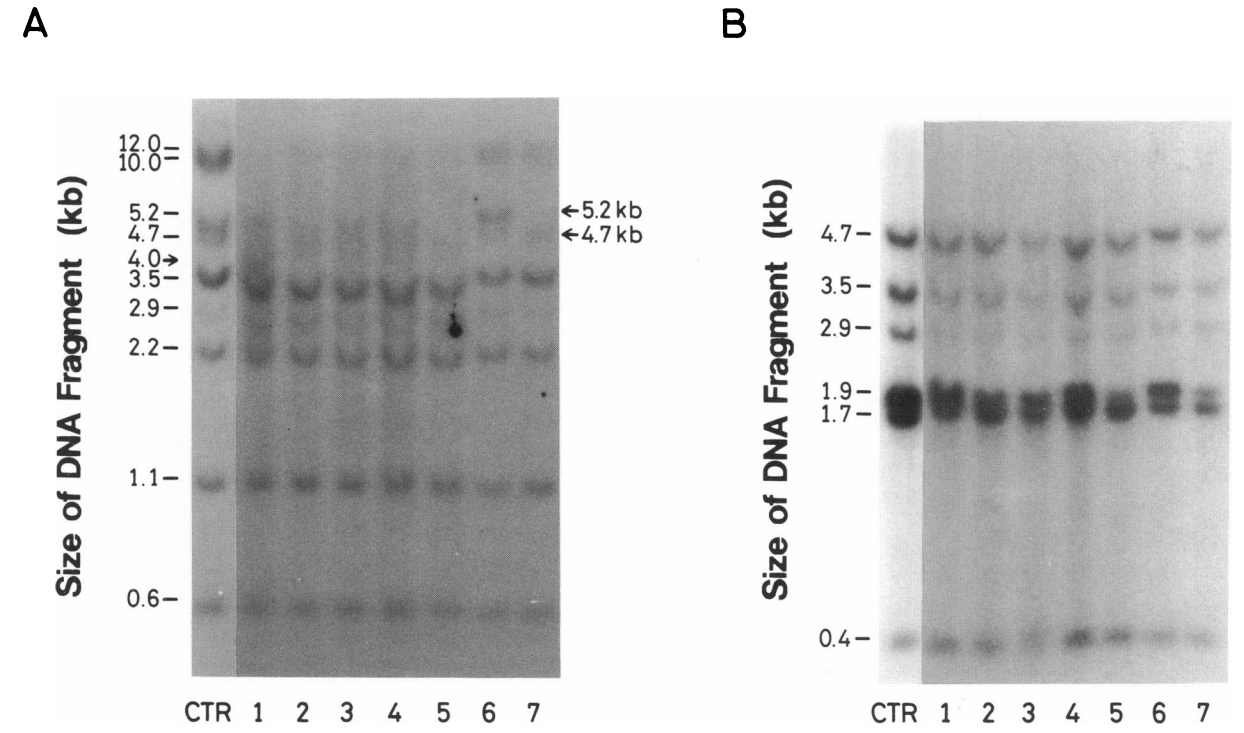

Figure 6 Rearrangements of *Eco*RI fragments from genomes of patients with nonketotic hyperglycinemia due to lesion in glycine decarboxylase. Genomic DNAs (20 μg) from a control individual (lane CTR) and from patients AH, MR, FS, JL, VM, YH, and DT (lanes 1–7) were treated with *Eco*RI. The pHH24L and pHH24S cDNAs were used as probes in *A* and *B*, respectively.

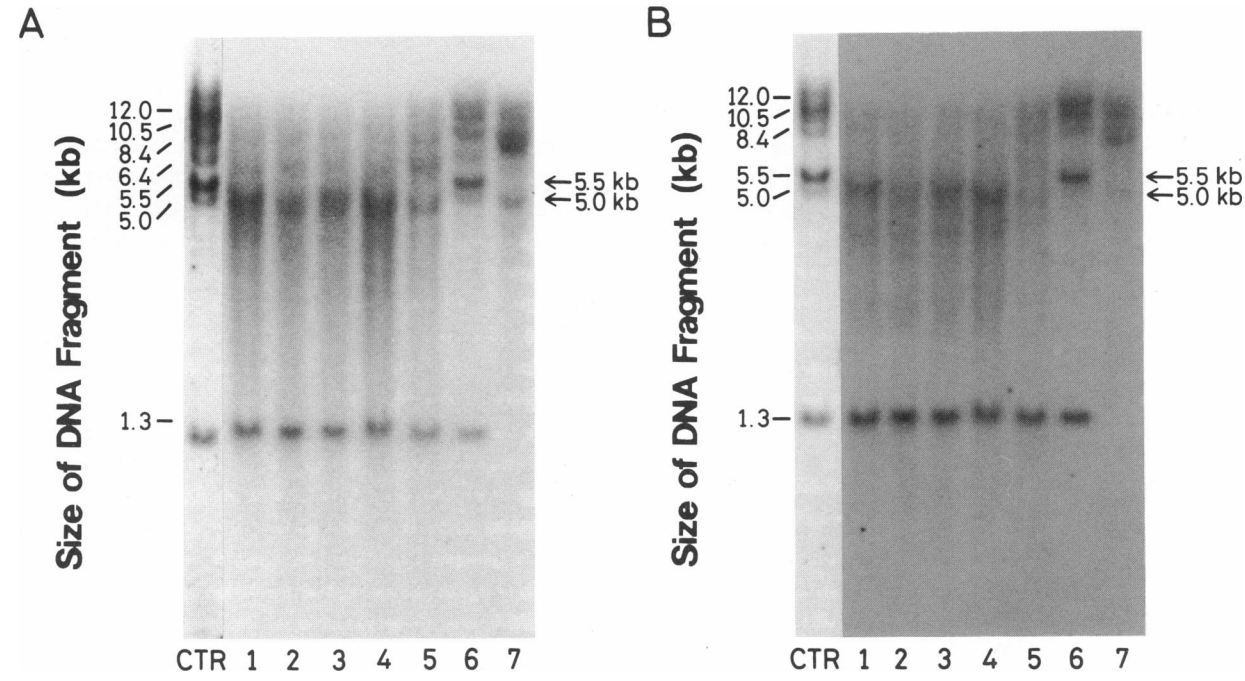

Figure 7 Rearrangements of *Sac*I fragments from genomes of patients with nonketotic hyperglycinemia due to lesion in glycine decarboxylase. The same genomic DNAs as used in fig. 6 were treated with *Sac*I and were similarly probed by using the pHH24L and pHH24S cDNAs as probes in *A* and *B*, respectively. Lanes CTR and 1–7 represent the control and the patients, in the same order as given in the legend to fig. 6.

FS, and JL) (figs. 7*A* and 7*B*). However, these patients again revealed broad signals showing sizes shorter than 5 kb in figure 7*A*. The 12- and 10.5-kb fragments are also obscure in lanes 1–5 in both panels of figure 7.

Discussion

We have clarified the human H-protein cDNA structure—and thereby the deduced primary structure of human H-protein. The precursor and mature forms of human H-protein are deduced to be composed of 173 and 125 amino acids, respectively, and their calculated molecular weights are 18,912 and 13,815, respectively. Human H-protein appears to be highly homologous to chicken H-protein of 125 amino acid residues which was chemically determined and reported by Fujiwara et al. (1986).

By comparison, there are 18 amino acid substitutions within the mature protein (Ser49-Glu173). Of these, 11 substitutions occur between Asp and Glu, between Thr and Ser, or within branched-chain amino acids at both the amino- and the carboxy-terminal regions (Ser49-Thr62 and Leu149-Glu173). The amino acid sequence of 54 residues between these regions (Glu63-Leu116; fig. 1*B*) therefore shows 96% homology to the comparable region in the chicken enzyme. As a consequence, the lipoic acid binding site was assigned to Lys107 in the human enzyme (double-underlined in fig. 1*B*). Moreover, as suggested in a previously published study (Hiraga et al. 1981), the two conserved cysteine residues, one of which appeared to participate in an intermolecular disulfide bonding, were assigned to Cys85 and Cys138. The highly conserved structure of the H-proteins from different vertebrates can be substantially implicated in their common properties, such as immunochemical and enzymological similarities.

As would be expected from the fact that the integrated length of the fragments (which were hybridized with the separate but consecutive sequences of this H-protein cDNA [approximately 1 kb] in Southern analysis) appears to encompass more than 60 kb, the number of copies for the H-protein cDNA is suggested to be at least four in the human haploid genome. This is supported by our preliminary findings that several cloned fragments of the human genome carry processed sequences similar to the H-protein cDNA sequence and that the two clones having an obvious overlap show an organized exon/intron-like structure which codes for H-protein (authors' unpublished data; the completed results will appear elsewhere).

Therefore, we can postulate at present that the processed genes (Hollis et al. 1982) or the retropseudogenes (Weiner et al. 1986) which hybridize to the H-protein cDNA probe exist in the human genome in addition to a single genuine H-protein gene. We are unable to exclude, at present, a possibility that the common sequence of some genes with a related structure, if any, might be included in the fragments detected by the procedures used in the present study, although the washing condition was highly stringent.

Northern analysis has demonstrated that two differently sized H-protein mRNAs exist in the human liver. The two types of H-protein mRNAs may result from the alternative use of the two poly(A) sites found in the cDNA sequence. The occurrence of the genomic regions encoding the different structures for the 3′ untranslated region might also be predicted.

To gain more insight into the mechanism which gave rise to an abnormal structure in H-protein of patient LS, we examined the RFLP of the H-protein gene. We could first elucidate in the normal genome the occurrence of the polymorphic *Pst*I and *Sac*I sites by which the 3.3-kb *Pst*I and 7.0-kb *Sac*I fragments are formed and by which the 1.3- and 6.4-kb *Sac*I fragments disappear. The genome of this patient revealed the additional, defective 5.0-kb *Sac*I fragment, suggesting that this patient carried the significantly rearranged genome. Unfortunately, because the multiple copies of the H-protein cDNA probe prevented us from a more precise analysis, it is unclear, at present, which of the genomic fragments really encodes the genuine H-protein gene. Therefore, what the observed rearrangements caused in the patient is still unknown. In this context, it is noteworthy that patient LS had an apo-H-protein (Hiraga et al. 1981). This reported fact complicates the understanding of the genetic lesion in this patient, because an additional enzyme which produces the covalent bonding of lipoic acid to the Lys107 is indispensable for forming holo-H-protein. This idea suggests that the observed rearrangement in the genome of this patient may not necessarily imply the production of an abnormal H-protein. As a comparable example, patient TI had an inactive H-protein in the liver (Kume et al. 1988) and contained no abnormal genomic structure detectable by Southern analysis in the present study.

As reviewed elsewhere, most patients with nonketotic hyperglycinemia are associated with some neurologically distinguished symptoms as part of the clinical features of their disease (Tada and Hayasaka 1987; Nyhan 1989). The reports indicate that the neuropa-

thology of nonketotic hyperglycinemia consists of diffuse alteration of myelination and gliosis (Shuman et al. 1978; Brun et al. 1979; Trauner et al. 1981). These alterations may occur in patients in whom this disease is caused by a molecular lesion involved in another component, such as glycine decarboxylase or T-protein. In this respect, it seems to be convincing that in those patients the elevated concentrations of glycine in extracellular fluids could be a result of a molecular lesion involved in the glycine cleavage system itself.

There has never been any decisive study as to whether this molecular lesion is lethal for patients. Tada and Hayasaka (1987) have pointed out in a review that higher extracellular glycine concentrations might both affect neurophysiological symptoms of the patient with nonketotic hyperglycinemia and, to a certain extent, modify development of the glycinergic synapse in spinal cord and brain stem. However, in consideration of the unique presentation of this disease, it is worth assuming that in the patient with nonketotic hyperglycinemia there might be an additional crucial genomic lesion other than the H-protein gene.

As would be expected from this idea, the H-protein cDNA probe fortunately detected significantly aberrant structures in the genomes of the patients having a molecular lesion involved in glycine decarboxylase. It is most interesting that patients LS and YH showed apparently identical deletion of the 5.0-kb *Sac*I fragment. This was the only rearrangement detected by the H-protein cDNA probe in their genomes. The remaining patients appeared to carry the genomic rearrangement identified by the undetectable 5.2-kb *Eco*RI and 5.5-kb *Sac*I fragments. Immunochemical analysis suggested that the glycine decarboxylase gene of these seven patients has been impaired. Therefore, the summarized results suggest the presence of rearrangements in multiple genomic loci of all the patients tested, although there is no clear implication in the pathogenesis of this disease at present. Characterization of both the genuine H-protein gene and observed aberrant fragments is required, including elucidation of their chromosome locations.

We have already cloned human glycine decarboxylase cDNA, and it is noteworthy that no homologous sequence, which may cause cross-hybridization in Southern analysis, is included in the H-protein cDNA and glycine decarboxylase cDNA. We have found by Southern analysis using glycine decarboxylase cDNA that rearranged glycine decarboxylase genes exist in some of the seven patients examined in the present study (unpublished data). Characterization of the aberrant glycine decarboxylase genes is also in progress in our laboratory. The precise studies on the glycine decarboxylase gene and H-protein gene will lead us to the comprehensive understanding of nonketotic hyperglycinemia.

Acknowledgments

This work was supported in part by Grants-in-Aid for Scientific Research 60580146, 61219007, 62570101, and 62109001 from Ministry of Education, Science and Culture, Japan. We thank Dr. Eiji Tsukamoto for continuous support of this work and for encouragement.

References

Baumgartner R, Ando T, Nyhan WL (1969) Nonketotic hyperglycinemia. J Pediatr 75:1022–1030

Bock SC, Harris JF, Balazs I, Trent JM (1985) Assignment of the human antithrombin III structural gene to chromosome 1q23–25. Cytogenet Cell Genet 39:67–69

Brun A, Borjeson M, Hultberg B, Sjoblad S, Akesson H, Litwin E (1979) Neonatal non-ketotic hyperglycinemia: a clinical, biochemical and neuropathological study including electronmicroscopic findings. Neuropaediatrie 10:195–205

Childs B, Nyhan WL, Borden M, Bard L, Cooke RE (1961) Idiopathic hyperglycinemia and hyperglycinuria: a new disorder of amino acid metabolism. I. Pediatrics 27:522–538

DiLella AG, Woo SLC (1987) Cloning large segments of genomic DNA using cosmid vectors. Methods Enzymol 152:199–212

Fujiwara K, Okamura-Ikeda K, Motokawa Y (1986) Chicken liver H-protein, a component of the glycine cleavage system: amino acid sequence and identification of the N^{ϵ}-lipoyllysine residue. J Biol Chem 261:8836–8841

Fyrberg EA, Kindle KL, Davidson N, Sodja A (1980) The actin genes of Drosophila: a dispersed multigene family. Cell 19:365–378

Hayasaka K, Tada K, Fueki N, Nakamura Y, Nyhan WL, Schmidt MS, Packman S, et al (1987) Nonketotic hyperglycinemia: analyses of glycine cleavage system in typical and atypical cases. J Pediatr 110:873–877

Hayasaka K, Tada K, Kikuchi G, Winters S, Nyhan WL (1983) Nonketotic hyperglycinemia: two patients with primary defects of P-protein and T-protein, respectively, in the glycine cleavage system. Pediatr Res 17:967–970

Hiraga K, Kikuchi G (1980*a*) The mitochondrial glycine cleavage system: purification and properties of glycine decarboxylase from chicken liver mitochondria. J Biol Chem 255:11664–11670

——— (1980*b*) The mitochondrial glycine cleavage system:

functional association of glycine decarboxylase and aminomethyl carrier protein. J Biol Chem 255:11671–11676

Hiraga K, Kochi H, Hayasaka K, Kikuchi G, Nyhan WL (1981) Defective glycine cleavage system in nonketotic hyperglycinemia: occurrence of a less active glycine decarboxylase and an abnormal aminomethyl carrier protein. J Clin Invest 68:525–534

Hiraga K, Kure S, Yamamoto M, Ishiguro Y, Suzuki T (1988) Cloning of cDNA encoding human H-protein, a constituent of the glycine cleavage system. Biochem Biophys Res Commun 151:758–762

Hollis GF, Hieter PA, McBride OW, Swan D, Leder P (1982) Processed genes: a dispersed human immunoglobulin gene bearing evidence of RNA-type processing. Nature 296: 321–325

Kao FT, Morse HG, Law ML, Lidsky A, Chandra T, Woo SLC (1984) Genetic mapping of the structural gene for antithrombin III to human chromosome 1. Hum Genet 67:34–36

Kikuchi G (1973) The glycine cleavage system: composition, reaction mechanism, and physiological significance. Mol Cell Biochem 1:169–187

Kume A, Kure S, Tada K, Hiraga K (1988) The impaired expression of glycine decarboxylase in patients with hyperglycinemias. Biochem Biophys Res Commun 154: 292–297

Lehrach H, Diamond D, Wozney JM, Boedtker H (1977) RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry 16:4743–4751

Mizusawa S, Nisimura S, Seela F (1986) Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucleic Acids Res 14:1319–1324

Motokawa Y, Hiraga K, Kochi H, Kikuchi G (1970) Evidence for the presence of a protein-bound intermediate in the cleavage and the synthesis of glycine. Biochem Biophys Res Commun 38:771–778

Nyhan WL (1989) Nonketotic hyperglycinemia. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) The Metabolic Basis of Inherited Disease, 6th ed. McGraw-Hill Information Services, New York, pp 743–753

Nyhan WL, Borden M, Childs B (1961) Idiopathic hyperglycinemia: a new disorder of amino acid metabolism. II. The concentrations of other amino acids in the plasma and their modification by the administration of leucine. Pediatrics 27:539–550

Piechaczyk M, Blanchard JM, Sabouty R-E, Dani C, Marty L, Jeanteur P (1984) Unusual abundance of vertebrate 3-phosphate dehydrogenase pseudogenes. Nature 312: 469–471

Prochownik EV, Markham AF, Orkin SH (1983) Isolation of a cDNA clone for human antithrombin III. J Biol Chem 258:8389–8394

Rosenberg LE, Lilljeqvist AC, Hsia YE (1968) Methylmalonic aciduria: an inborn error leading to metabolic acidosis, long-chain ketonuria and intermittent hyperglycinemia. N Engl J Med 278:1319–1322

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Shuman RM, Leech RW, Scott CR (1978) The neuropathology of the nonketotic and ketotic hyperglycinemias: three cases. Neurology 28:139–146

Tada K, Hayasaka K (1987) Non-ketotic hyperglycinemia: clinical and biochemical aspects. Eur J Pediatr 146:221–227

Trauner DA, Page T, Greco C, Sweetman L, Kulovich S, Nyhan WL (1981) Progressive neurodegenerative disorder in a patient with nonketotic hyperglycinemia. J Pediatr 98:272–275

Weiner AM, Deininger PL, Efstratiadis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem 55:631–661

Yoshida T, Kikuchi G, Tada K, Narisawa K, Arakawa T (1969) Physiological significance of glycine cleavage system in human liver as revealed by the study of a case of hyperglycinemia. Biochem Biophys Res Commun 35: 577–583

Yoshida T, Kikuchi G (1970) Major pathways of glycine and serine catabolism in rat liver. Arch Biochem Biophys 139:380–392