

ORIGINAL ARTICLE

Shigeo Kure · Kanako Kojima · Takayuki Kudo
Kiyoshi Kanno · Yoko Aoki · Yoichi Suzuki
Toshikatsu Shinka · Yoshiyuki Sakata · Kuniaki Narisawa
Yoichi Matsubara

Chromosomal localization, structure, single-nucleotide polymorphisms, and expression of the human H-protein gene of the glycine cleavage system (*GCSH*), a candidate gene for nonketotic hyperglycinemia

Received: February 28, 2001 / Accepted: April 2, 2001

Abstract Nonketotic hyperglycinemia (NKH) is an inborn error of metabolism caused by deficiency in the glycine cleavage system (GCS); this system consists of four individual constituents, P-, T-, H-, and L-proteins. Several mutations have been identified in P- and T-protein genes, but not in the H-protein gene (*GCSH*), despite the presence of case reports of H-protein deficiency. To facilitate the mutational and functional analyses of *GCSH*, we isolated and characterized a human p1-derived artificial chromosome (PAC) clone encoding *GCSH*. *GCSH* spanned 13.5 kb and consisted of five exons. Using the PAC clone as a probe, we mapped *GCSH* to chromosome 16q24 by fluorescence in situ hybridization. The transcription initiation site was determined by the oligonucleotide-cap method, and potential binding sites for several transcriptional factors were found in the 5' upstream region. Direct sequencing analysis revealed five single-nucleotide polymorphisms. The expression profiles of P-, T-, and H-protein mRNAs were studied by dot-blot analysis, using total RNA from various human tissues. *GCSH* was expressed in all 29 tissues examined, while T-protein mRNA was detected in 27 of the 29 tissues. In contrast, the P-protein gene was expressed in a limited number of tissues, such as liver, kidney, brain, pituitary gland, and thyroid gland, suggesting distinct transcriptional regulation of each GCS constituent.

Key words Human PAC clone · Transcription initiation site · Direct sequencing analysis · Fluorescence in situ hybridization · RNA blot analysis

Introduction

Nonketotic hyperglycinemia (NKH) is an inborn error of metabolism characterized by the accumulation of glycine in body fluids (Tada et al. 1969). In patients with NKH, severe neurological symptoms, such as convulsive seizures, coma, and respiratory distress, develop within a few days after birth (Almsh et al. 1995). The fundamental defect lies in the glycine cleavage system (GCS; EC 2.1.2.10). The GCS, a multi-enzyme complex that exists in mitochondria (Motokawa and Kikuchi 1972; Kikuchi 1973), consists of four constituents; P-protein [EC 1.4.4.2], T-protein [EC, 2.1.2.10], H-protein, and L-protein [EC2.1.8.1.4]. P-protein and T-protein are also called glycine decarboxylase (GLDC) and aminomethyltransferase (AMT), respectively. H-protein is referred to as a hydrogen carrier protein that transfers the aminomethyl moiety, which is generated by the decarboxylation of glycine, to T-protein for further degradation. L-protein is dihydrolipoamide dehydrogenase, a housekeeping enzyme that serves as a component of other complex enzyme systems, such as the pyruvate dehydrogenase complex.

Enzymatic analysis has identified three metabolic lesions in NKH, deficiencies of P-, T-, and H-proteins (Almsh et al. 1995). The first mutation identified in NKH was in the P-protein gene (Kure et al. 1991b). Subsequently, some patients were found to have mutations in the T-protein gene (Nanao et al. 1994a; Kure et al. 1998; Toone et al. 2001). However, mutations in the H-protein gene have not been identified, despite reports of a patient who was enzymatically deficient in H-protein (Trauner et al. 1981; Hiraga et al. 1981). Although the genomic organization of the human H-protein gene (*GCSH*) has been partially described (Koyata and Hiraga 1991b), complete and accurate information is still missing, especially regarding the sequence information for each exon/intron boundary that is required for single-exon amplification. Furthermore, the chromosomal localization and the transcription initiation site of the *GCSH* gene remain unknown.

S. Kure (✉) · K. Kojima · T. Kudo · K. Kanno · Y. Aoki · Y. Suzuki · T. Shinka · Y. Sakata · K. Narisawa · Y. Matsubara
Department of Medical Genetics, Tohoku University School of Medicine, 1-1 Seiryomachi, Aoba-ku, Sendai 980-8574, Japan
Tel. +81-22-717-8138; Fax +81-22-717-8142
e-mail: skure@mail.cc.tohoku.ac.jp

Here we report the structure and chromosomal localization of the H-protein gene. We isolated a p1-derived artificial chromosome (PAC) clone encoding *GCSH* and determined the chromosomal localization by fluorescence in situ hybridization (FISH) analysis. We then established a system of single-exon amplification, using flanking intronic primers, and identified several single-nucleotide polymorphisms (SNP). The transcription initiation site and the expression profile of three GCS components were also studied. Our results should facilitate the studies of NKH caused by H-protein deficiency.

Materials and methods

Isolation and characterization of PAC clone encoding *GCSH*

A human PAC genomic library was screened by polymerase chain reaction (PCR), using hH1 and hH2 primers, which amplified the entire region of intron 4 (Table 1). The hH1 and hH2 primers were designed based on the GeneBank sequence data (accession numbers, S50220 and S50222). One PAC clone (PACHH1) was isolated and subjected to structural analysis. Exon-intron boundary sequences were determined by a cycle sequencing method, using 1 µg of PAC genomic DNA and 1 µl of thermostable topoisomerase (ThermoFidélase I; Fidelity Systems, Gaithersburg, MD, USA) according to the manufacturer's protocol (<http://www.fidelitysystems.com/TF1.html>). The lengths of introns 1–4 were estimated by long-distance PCR amplification, with the primer sets, HHE1F/HHE2R, HHE2F/HHE3R, HHE3F/HHE4R, and HHE4F/HHE5R, respectively (Table 1). One microgram of the purified genomic PACHH-1 DNA and an LA PCR kit (Takara, Tokyo, Japan) were used for the PCR.

Chromosomal mapping

DNA purified from the PACHH-1 clone was labeled with digoxigenin-labeled dUTP by nick translation. The labeled probe was combined with sheared human DNA and hybridized to normal metaphase chromosomes derived from phytohemagglutinine-stimulated peripheral blood lymphocytes, in a solution containing 50% formamide, 10% dextran sulfate, and 2× standard sodium citrate (SSC). Specific hybridization signals were detected by incubating the hybridized slides in fluorescein-labeled antidigoxigenin antibodies, followed by counterstaining with DAPI (4, 6-diamidino-2-phenylindole). A total of 80 metaphase cells were analyzed for mapping. To confirm the chromosomal assignment, a DNA fragment that was previously mapped to chromosome 16p13 was hybridized together with PACHH-1.

Analysis of the transcription initiation site

The H-protein cDNA fragment containing the transcription start site was obtained by an oligonucleotide-cap method, using a Cap Site PCR kit (Nippon Gene, Toyama, Japan) (Maruyama and Sugano 1994). An H-protein cDNA fragment containing the cap site was amplified by nested PCR. The cDNA fragment containing the cap site was designed to have the 25-mer oligonucleotide linker, 5'-CAAGGTACC CCACAGCGTATGATGC-3'. H-protein cap site cDNA was then amplified with the forward primers that were complementary to the linker and H-protein-specific reverse primers. The first PCR was performed with hH6 and hH4 primers, while the second PCR used hH3 and hH5 primers (Table 1). The products of the second PCR were separated by 4% agarose gel electrophoresis. Gel bands were excised and purified with a QIAEX II Gel Purification kit (Qiagen, Hilden, Germany). Nucleotide sequences of the PCR products were determined with a dye-terminator sequencing kit

Table 1. Oligonucleotide primers

Primer	Location	Orientation	Sequence (5' to 3')
hH1	Exon4/intro4	Forward	CAAATCTTGTTATGAAGATGGTAAGC
hH2	Intron4/exon5	Reverse	TCATCTTGATCAGCCAACCTGCAACC
hH3	Cap linker	Forward	CAAGGTACGCCACAGCGTATG
hH4	Exon 1	Reverse	CACTCGCAGCGCCATGTTTCGCAG
hH5	Exon 1	Reverse	AGGCGGGGCGGGGAGGGGCAG
hH6	Cap linker	Forward	GTACGCCACAGCGTATGATGC
HHE1F	Exon 1	Forward	<u>TGTA</u> AAAACGACGGCCAGTGGCCGAGGCGTAGCGCCGCGAC
HHE1R	Intron 1	Reverse	<u>CAGG</u> AAACAGCTATGACC <u>AAATA</u> AAGAAGGGGCAGGGTCCGC
HHE2F	Intron 1	Forward	<u>TGTA</u> AAAACGACGGCCAGTGAAGCAATGGTAATGAAAAGTGCTTC
HHE2R	Intron 2	Reverse	<u>CAGG</u> AAACAGCTATGACCTCACGCAGCCTAAACT
HHE3F	Intron 2	Forward	<u>TGTA</u> AAAACGACGGCCAGT <u>TTTA</u> AGGAAAACCTCAAGAATGTAGCC
HHE3R	Intron 3	Reverse	<u>CAGG</u> AAACAGCTATGACCTTCAATGTAAACAAAATTCATGGCATGG
HHE4F	Intron 3	Forward	<u>TGTA</u> AAAACGACGGCCAGTGTATGACAGGAATCTACTTTTTTCG
HHE4R	Intron 4	Reverse	<u>CAGG</u> AAACAGCTATGACCGAAGTCACAATCAGCTAAACTTTGCT
HHE5F	Intron 4	Forward	<u>TGTA</u> AAAACGACGGCCAGTAAGAATACAAGATGTGAGGCGCTCAG
HHE5R	Exon 5	Reverse	<u>CAGG</u> AAACAGCTATGACCTATTCTAAGTCTTCTATCCACCAC

Underlining of nucleotides indicates M13 primer sequences (single underline) and reverse primer sequences (double underlines) used for direct sequencing analysis

and an ABI 310 DNA sequencer (PE Biosystems, Forester, MA, USA).

Screening for SNPs

Genomic DNA samples were purified from 20 Caucasian and 20 Japanese control subjects for the screening of SNPs. DNA fragments containing exons 1–5 were amplified by PCR with the primer sets HHE1F/HHE1R, HHE2F/HHE2R, HHE3F/HHE3R, HHE4F/HHE4R, and HHE5F/HHE5R, respectively (Table 1). Each amplicon was purified as described above and used as a sequencing template. Sequencing was performed with the ABI 310 DNA sequencer and M13 and reverse dye-primer sequencing kits, according to the manufacturer's protocol (PE Biosystems).

RNA dot blot analysis

A nylon membrane on which the total RNA of 29 various human tissues was blotted (Clontech, Palo Alto, CA, USA) was probed with ³²P-labeled cDNA fragments (~500bp) encoding human H-, P-, and T-protein cDNA, as described previously (Kure et al. 1991a). Hybridization and washing were performed according to the manufacturer's protocol. The radioactivity of each dot spot was measured with Molecular Imager FX (Bio-Rad, Hercules, CA, USA). Background radioactivity was estimated by measuring the radioactivity of a 100-ng *Escherichia coli* DNA spot and subtracting this value from the measured radioactivity of each spot. To measure the relative levels of the P-, T-, and H-protein mRNAs in each organ, we arbitrarily designated the radioactivity of the liver RNA spot as 1.0, and estimated the relative abundance of each mRNA by calculating the radioactivity of each RNA spot as a percentage of that of the liver RNA spot.

Results

Structure of GCSH

One positive PAC clone, PACHH-1, was isolated by PCR screening with the *GCSH*-specific primers. The sequences of the exon/intron boundaries and the size of each intron are shown in Fig. 1. All exon/intron boundaries adhered to the canonical gt/ag rule. The H-protein gene spans 13.5kb and consists of five exons. Exon 1 corresponds to the 5'-untranslated region and the initiator AUG codon. A lipoic acid binding the Lys¹⁰⁷ residue is located in exon 4 (Fujiwara et al. 1986). The translation termination codon is encoded in exon 5, which is followed by a 3'-untranslated sequence of 542bp. The transcription start site was then determined by the oligo nucleotide-cap method. A 143-bp fragment was generated by the nested PCR amplification, and contained a sequence from nucleotides 215 to 340 in Fig. 1. The transcription, therefore, appeared to be initiated at an adenine

nucleotide 215, 169bp upstream from the translation initiation ATG triplet. Neither a typical TATA box sequence nor its homologue was found in the 5'-flanking region, but two CCAAT boxes were present, at nucleotides 30–34 and 182–186 (Fig. 1). A potential binding site for Sp1 was present at nucleotides 7–12. The flanking region also contained DNA motifs for AP2 at nucleotides 2–28, 39–46, and 147–154.

Chromosomal mapping of GCSH

Subchromosomal position was determined by FISH analysis with the fluorescence-labeled PACHH-1 DNA. The initial experiment resulted in specific labeling of the long arm of chromosome 16, on the basis of size, morphology, and banding pattern (arrow "a" in Fig. 2). The specific labeling was observed in 76 of 80 metaphase cells. To confirm the location, we hybridized the PACHH-1 probe together with a human PAC genomic clone that had previously been mapped to chromosome 16p13 (arrow "b" in Fig. 2). Observation of the specifically labeled chromosome 16 demonstrated that the clone PACHH-1 was located at the terminus of the long arm of chromosome 16, an area that corresponds to band 16q24.

SNPs in GCSH

Twenty Caucasian and 20 Japanese control subjects were screened for SNPs by direct sequencing analysis of single-exon amplicons. We identified five SNPs, one in exon 1 and four in introns 1, 2, and 4 (Table 2). The exonic polymorphism, a T-to-C substitution in codon 21 (TTA to TCC), caused an amino acid change from leucine (L) to serine (S), which was located in the mitochondrial leader peptide of the H-protein. The S21 allele frequencies in the Caucasian and Japanese control subjects were 22.5% and 17.5%, respectively. The other four SNPs were located in introns and caused no amino-acid substitutions. The intronic SNPs were observed in both Caucasian and Japanese controls. The IVS2+54t>c polymorphism was found only in Japanese, while the other three polymorphisms, IVS1–20a>c, IVS2–84g>a, and IVS4–62t>a, were identified only in Caucasians (Table 2).

Expression of P-, T-, and H-protein mRNAs in various tissues

The relative amounts of H-, P-, and T-protein mRNAs in each tissue were calculated, and findings are summarized in Fig. 3. H-protein mRNA was detected in all 29 tissues examined. T-protein mRNA was expressed in all tissues, except for stomach and bone marrow. P-protein mRNA was detected in a limited number of tissues: strong gene expression was detected in liver, placenta, and kidney; moderate expression in brain, small intestine, thyroid gland, and pituitary gland; and weak expression in colon, bladder, and lung. No P-protein mRNA signals were detected in other tissues.

Fig. 1. Structure of *GCSH*. Nucleotides in exons are shown in **boldface**, and protein-coding regions are shown in **uppercase**. The transcription initiation site is marked with a *dot above the nucleotide*. Several putative motifs for transcriptional factors are underlined and denoted below the line. Four single-nucleotide polymorphisms are indicated by underlined bold face letters with the substituted nucleotide shown below. The translation termination codon is shown as *TRM*. Double underlined sequences indicate two polyadenylation signals. *SPI*, Sp1-binding site; *AP2*, AP2-binding site; *CCAAT*, CAAT box

```

1 taggcccgcgcc 13
SP1
14 cctcctcgggcagggccaatcatagccagggccctccgcgctagtctacgcctctcgggtgctgctcttccggggcgtcaggttcttcc 103
AP2 CCAAT AP2
104 ccattgcccagccctcttctcgcgtcgcgtaaccgaggtcgcgcagggcctggagtcgccaggtctccgctgctggccaatcagacc 193
AP2 CCAAT
194 tgcctcgccccctcgcctgggagttctcccccctccaagctctgcccagtcaggctctgtcccgccccggcgaccgctccgcgctcc 283
exon 1
284 atccagccggctccctccggccggaactgcccctcccccggcctccggcgcggtggccgagggctagcgcggcagccccgcac 373
374 ccoctggaacATGGCGTGGGAGTGGTGGCGAGCGTGGGGCCCTGCTCTGCACCCCTGGCGCGGTCCCGTTACCCGCCGCCCTGCC 463
Met1 C
464 GCCGAGGCCCTGGCAGCTGGGGTGGGCGCCCTCCGTACGCTGGCAGTGGACCCGCTCTGCTCTCGGgtaagcggcgggcagctggg 553
554 cggctgctccctcccgctgggctgctgtccgggcccagcgggacacgaggagcggcgggcagcgtcctgagccaccggggcgga 643
644 cgcgagcgggaccctgccctcttatttctgttgggatccctggctctcggcgcttctcccctt-----
----- (4.8 kb) -----
1 gtttaaatgctcatctctatacatgatcttcatcttctcctgagccatgcaactatlaaaggtagtgcacaggtcaaacagaaattc 90
91 tgggaaagtcttctttaaagcaatggtatgaaagtgctcttaagaagtaagctactgcttctgaaagtgttggtctctctgttttag 180
c
181 TCGTAAATTCACAGAGAAACACGAATGGGTAACAACAGAAAATGGCATTGGAACAGTGGGAATCAGCAATTTGCACAGgtattggatt 270
exon 2
271 atattgaaatatttgcctcagtgctcttgcctgatttcttctctacctttttaaagtggttaggctgctgaggtggtcag 360
c
361 cctgtaatcccaacttgggaggtgagggcagggattgcttgagccaggagttcaagaccgctggtcaacaaa-----
----- (2.5 kb) -----
1 actgagtagtttttacatgtttaaaggatttaagaaaactcaagaatgtagccctccagatataaaatataactaataatcttcatt 90
91 ataggggattcaggtttggcttattctgttactggaggttagtataactcttagagattttggttttgtcttatttttag 180
c
181 GAAGCGTTGGGAGATGTTGTTATTGTAGTCTCCCTGAAGTTGGGACAAAATGAACAACAAGgtgagtgctcttaggactcttagaat 270
exon 3
271 atccatgccaatgattttgtttacattgaatagtaatttggttgtaccctggattaattagagtgcttttttttttggagacagctct 360
361 gctttgtcaccagcgtggagtgagtgaccatctcggccactgtaacctctgcctcctgg-----
----- (2.1 kb) -----
1 ttgcccactgcactccagcctgggcaacagagcggagactccatctcaaaaaaaaaaaaaaagaaatttgaattttaaatcatagct 90
91 gtatgacaggaatctacttttctgtatttatgtttcaagtaactgacttttaacattctgatttttcaatttttttccacttag 180
181 ATGAGTTTGGTCTTTGAAAGTGTGAAACCTGCTAGTGAACCTTATCTCCTTTATCAGGAGAAGTAACGAAATTAATGAAGCTCTTG 270
exon 4 Lys107
271 CAGAAAATCCAGGACTTGTAACAACAACTCTGTTATGAAGATGgtaagctgtgctagaaattctcaaggaattactactacagtaata 360
361 ttttatcagaatttaaagcaagtttagctgattgtgactcatcttttctactcttggtaactaaatagagatgtctgctttaaag 450
451 aaaaacagaggttagttgaataactaaagttatgtttaagat-----
----- (1.1 kb) -----
1 catttggtatatacaagtatttaattcttaaggctaaaattaaagttctatgaacaaataatgtatggtttttgtaagaatacaaga 90
91 ttgagcgtcagaaatttatttaccgttttggatgaattgactttttaaagtttaagtggaatataaggtgtcttttgggtgcag 180
a
181 GTTGGCTGATCAAGATGACACTGACTAACCTTCGAACCTAGATGAACCTTATGAGTGAAGAAGCATATGAGAAATACATAAAATCTATTG 270
exon 5
271 AGGAGTGAaaatggaactcctaaataaacctagtagtaaaataogcaagccagcagagttgtcttaaattagtggtgtagaagacttag 360
TRM
361 aatagaaaacttttagtattaccgatggggaaaaaaaactactgtaaacctgctaatgaagaaaatgoccttaactttctaatgatt 450
451 atagataaataatgatcgtcttttcaataatcctatgatttttagactaggtctaggttcagaattcatgaaattatcactcatt 540
541 aaaaactagttataaaaattacataattcaagataaacattgttattcttaagccttataataattgtaacttgcatgtatccatcct 630
631 ggatttgggatgaaatacttaagatctttccattggaataactggaagtgaagaggttttggctgtgtacagtgcatgaggaac 720
721 accactatcttaatttgggatcactgcatttggctgtctatttttatacagtgaaagcaagcgtttgcagcaaaaataaaataact 810
811 tcttcgttaa:catgtttgggtttgatgtaataatttcattagtaactctgctagattttggaagtgtaactttaaactacggaaaag 900

```

Discussion

We isolated a PAC clone (PACHH-1) encoding the human H-protein gene (*GCSH*) and mapped *GCSH* on chromosome 16q24 by FISH analysis. Previous genetic analysis of the GCS revealed that the human P-protein gene (*GLDC*) consisted of 25 exons (Takayanagi et al. 2000) and was mapped to chromosome 9p22-24 (Isobe et al. 1994). The T-

protein gene (*AMT*) was shown to consist of 9 exons and was located on 3p21.1-2 (Nanao et al. 1994b). The L-protein gene (*GCSL*) was shown to have 12 exons (Feigenbaum and Robinson 1993) and resided on chromosome 7q31-32 (Scherer et al. 1991). We recently reported a processed pseudogene of *GLDC*, designated ψ *GLDC*, which showed 97.5% homology with the *GLDC* cDNA (Takayanagi et al. 2000). Neither the T-protein gene nor the L-protein gene has any known pseudogenes. When GeneBank was

screened for homologous sequences with human H-protein cDNA, three highly homologous processed type pseudogenes were found. Their chromosomal location, accession numbers, and sequence identities are: chromosome 1q22-23.3, HS702J19, 99.6%; chromosome 5p, AC011401, 96.7%; chromosome 7p, AC007551, 90.0%; and chromo-

some 19, AC008976, 98.9%. If a unit evolutionary period (UEP-time need for a 1% divergence) of 2–3.5 million years is accepted for a silent site, then these processed pseudogenes probably arose 0.8–35 million years ago (Gonzales et al. 1993).

We determined the structural organization of *GCSH* and found that its transcription start site was 168bp upstream from the translation initiation triplet, using the oligo nucleotide-cap method. Koyata and Hiraga (1991b) previously assigned the transcription initiation site to 168bp upstream of the initiation triplet, by primer extension analysis, essentially agreeing with our result. Their S1 nuclease protection

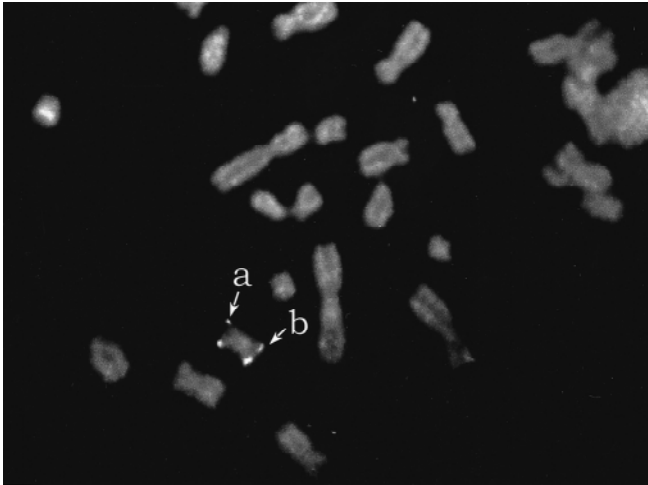


Fig. 2. Chromosomal mapping of *GCSH*. Specific hybridization pattern of the genomic clone PACHH-1 on 16q24 (a), and cohybridization with a p1-derived artificial chromosome (PAC) clone, known to map to chromosome 16p13 (b)

Table 2. Single-nucleotide polymorphisms (SNPs) in human H-protein gene

SNP	Location (position in Fig. 1)	Allele	Caucasian allele	Japanese allele
L21S	Exon 1 (Nucleotide 445)	L	31 (77.5%)	33 (82.5%)
		S	9 (22.5%)	7 (17.5%)
IVS1–20a>c	Intron 1 (Nucleotide 161)	a	36 (90%)	40 (100%)
		c	4 (10%)	0
IVS2+54t>c	Intron 2 (Nucleotide 314)	t	40 (100%)	38 (95%)
		c	0	2 (5%)
IVS2–84g>a	Intron 2 (Nucleotide 97)	g	37 (92.5%)	40 (100%)
		a	3 (7.5%)	0
IVS4–62t>a	Intron 4 (Nucleotide 119)	t	30 (75%)	40 (100%)
		a	10 (25%)	0

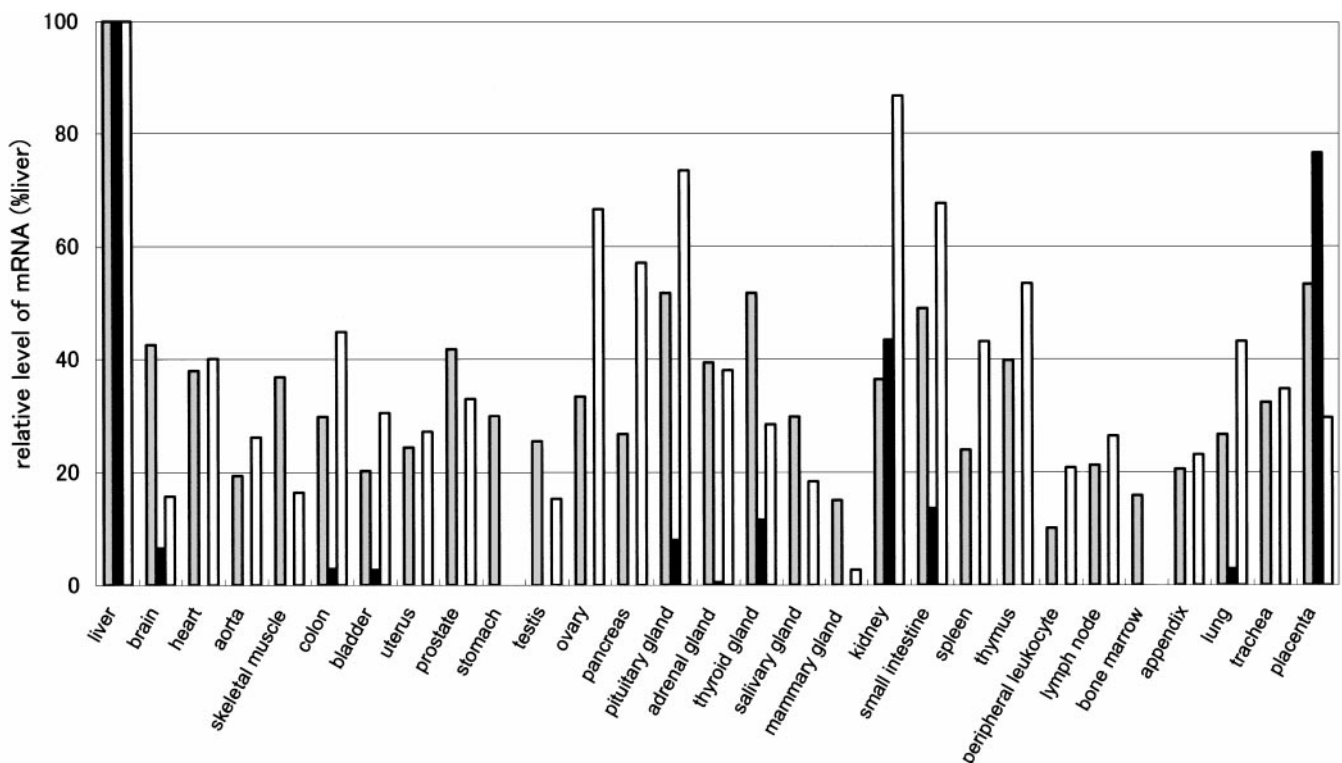


Fig. 3. Expression of P-, T-, and H-protein mRNAs in various human tissues. Dot blots of total RNAs purified from 29 various human tissues were probed with P-, T-, and H-protein cDNAs. The radioactivity of

each dot was measured, and mRNA levels relative, to those of liver, were calculated. Gray, black, and open bars represent relative levels of H-, P-, and T-protein mRNAs, respectively

analysis, however, yielded no specific band. They concluded that there was an additional noncoding exon located 5' upstream of exon 1. Because the cap site fragment in our experiment did not contain any foreign sequence and its length agreed with the results of their primer extension analysis, we cannot explain the discrepancy. The mitochondrial leader peptide of the H-protein precursor consists of 48 amino acids, and the L21S polymorphism was located in the middle of this sequence (Koyata and Hiraga 1991a). Interestingly, all of the processed pseudogenes of *GCSH* described above have the S21 allele, suggesting that the L21 allele arose later than the S21 allele. The L21S polymorphism may affect the efficiency of mitochondrial import of H-protein, causing individual variation in GCS activity and in glycine levels in serum and cerebrospinal fluid. Further functional analysis is required to elucidate the physiological significance of this polymorphism.

RNA dot-blot analysis with P-, T-, and H-protein cDNA probes revealed that the genes encoding the P-, T-, and H-proteins were not coordinately expressed. The H-protein gene was expressed in all tissues examined, suggesting that it was a housekeeping gene. The result was in line with our observation that *GCSH* had no typical TATA box, because absence of the TATA box is commonly observed in housekeeping genes. In contrast, the P-protein gene was expressed in a limited number of tissues. Enzymatic activity of the GCS was identified only in liver, brain, and kidney, in which tissues P-protein mRNA was abundantly expressed (Yoshida and Kikuchi 1973). A similar relationship has also been observed in other species, such as chickens (Kure et al. 1991a) and bovine (Okamura-Ikeda et al. 1991). Therefore, the limited expression of the P-protein gene appears to determine distribution of the functional GCS. In this context, it is interesting to note that moderate expression of the P-protein gene, together with H-protein and T-protein gene expression, was observed in the pituitary and thyroid glands. To date, GCS activity in these endocrine organs remains uninvestigated. It would be intriguing to examine whether the GCS plays an important role in the regulation of endocrine systems in the pituitary and thyroid glands.

Enzymatic analysis suggested that most patients with neonatal onset NKH had deficient P-protein or T-protein activity (Tada and Hayasaka 1987). A common *GLDC* mutation, S564I, was identified in Finland, where the incidence of neonatal onset NKH is unusually high (Kure et al. 1992). Nanao et al. (1994a) reported three *AMT* mutations in two patients, with neonatal and infantile onset NKH. A missense *AMT* mutation was also identified in a large Israeli-Arab family with neonatal onset NKH (Kure et al. 1998). Toone et al. (2000 and 2001) reported several *AMT* mutations, including a recurrent mutation, in Caucasian patients. H-protein deficiency seems to be relatively rare, compared with P- and T-protein deficiencies. Trauner et al. (1981) reported a patient with atypical NKH who was asymptomatic in the neonatal period and did not show the typical convulsive seizures, but who later showed mental retardation. This patient was proven to have deficient H-protein activity, although P- and T-protein activities were normal; almost all of the H-protein was in the apo-enzyme

form (Hiraga et al. 1981). Our results showing the structure, polymorphism, and expression of *GCSH* should facilitate the molecular analysis of such patients with variant forms of NKH that are caused by H-protein deficiency.

Acknowledgments This work was supported by grants from the Ministry of Education, Science, Sports and Culture of Japan and the Ministry of Health and Public Welfare of Japan.

References

- Almsh A, Johnston MV, Valle D (1995) Nonketotic hyperglycinemia. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) *The metabolic and molecular bases of inherited disease*, 7th edn, vol 1. McGraw-Hill, New York, pp 1337–1348
- Feigenbaum AS, Robinson BH (1993) The structure of the human dihydrolipoamide dehydrogenase gene (*DLD*) and its upstream elements. *Genomics* 17:376–381
- 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 epsilon-lipoyllysine residue. *J Biol Chem* 261:8836–8841
- Gonzalez IL, Tugendreich S, Hieter P, Sylvester JE (1993) Fixation times of retroposons in the ribosomal DNA spacer of human and other primates. *Genomics* 18:29–36
- 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
- Isobe M, Koyata H, Sakakibara T, Momoi-Isobe K, Hiraga K (1994) Assignment of the true and processed genes for human glycine decarboxylase to 9p23-24 and 4q12. *Biochem Biophys Res Commun* 203:1483–1487
- Kikuchi G (1973) The glycine cleavage system: composition, reaction mechanism, and physiological significance. *Mol Cell Biochem* 1:169–187
- Koyata H, Hiraga K (1991a) The glycine cleavage system: structure of a cDNA encoding human H-protein, and partial characterization of its gene in patients with hyperglycinemias. *Am J Hum Genet* 48:351–361
- Koyata H, Hiraga K (1991b) Partial structure of the human H-protein gene. *Biochem Biophys Res Commun* 178:1072–1077
- Kure S, Koyata H, Kume A, Ishiguro Y, Hiraga K (1991a) The glycine cleavage system. The coupled expression of the glycine decarboxylase gene and the H-protein gene in the chicken. *J Biol Chem* 266:3330–3334
- Kure S, Narisawa K, Tada K (1991b) Structural and expression analyses of normal and mutant mRNA encoding glycine decarboxylase: three-base deletion in mRNA causes nonketotic hyperglycinemia. *Biochem Biophys Res Commun* 174:1176–1182
- Kure S, Takayanagi M, Narisawa K, Tada K, Leisti J (1992) Identification of a common mutation in Finnish patients with nonketotic hyperglycinemia. *J Clin Invest* 90:160–164
- Kure S, Mandel H, Rolland MO, Sakata Y, Shinka T, Drugan A, Boneh A, Tada K, Matsubara Y, Narisawa K (1998) A missense mutation (His42Arg) in the T-protein gene from a large Israeli-Arab kindred with nonketotic hyperglycinemia. *Hum Genet* 102:430–434
- Maruyama K, Sugano S (1994) Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* 138:171–174
- Motokawa Y, Kikuchi G (1972) Isolation and partial characterization of the components of the reversible glycine cleavage system of rat liver mitochondria. *J Biochem* 72:1281–1284
- Nanao K, Okamura-Ikeda K, Motokawa Y, Danks DM, Baumgartner ER, Takada G, Hayasaka K (1994a) Identification of the mutations in the T-protein gene causing typical and atypical nonketotic hyperglycinemia. *Hum Genet* 93:655–658
- Nanao K, Takada G, Takahashi E, Seki N, Komatsu Y, Okamura-Ikeda K, Motokawa Y, Hayasaka K (1994b) Structure and chromosomal localization of the aminomethyltransferase gene (*AMT*). *Genomics* 19:27–30

- Okamura-Ikeda K, Fujiwara K, Yamamoto M, Hiraga K, Motokawa Y (1991) Isolation and sequence determination of cDNA encoding T-protein of the glycine cleavage system. *J Biol Chem* 266:4917-4921
- Scherer SW, Otulakowski G, Robinson BH, Tsui LC (1991) Localization of the human dihydrolipoamide dehydrogenase gene (*DLD*) to 7q31-q32. *Cytogenet Cell Genet* 56:176-177
- Tada K, Hayasaka K (1987) Non-ketotic hyperglycinaemia: clinical and biochemical aspects. *Eur J Pediatr* 146:221-227
- Tada K, Narisawa K, Yoshida T, Yokoyama K, Nakagawa H, Tanno K, Mochizuki K, Arakawa T, Yoshida T, Kikuchi G (1969) Hyperglycinemia: a defect in glycine cleavage reaction. *Tohoku J Exp Med* 98:289-296
- Takayanagi M, Kure S, Sakata Y, Kurihara Y, Ohya Y, Kajita M, Tada K, Matsubara Y, Narisawa K (2000) Human glycine decarboxylase gene (*GLDC*) and its highly conserved processed pseudogene (*psiGLDC*): their structure and expression, and the identification of a large deletion in a family with nonketotic hyperglycinemia. *Hum Genet* 106:298-305
- Toone JR, Applegarth DA, Coulter-Mackie MB, James ER (2000) Biochemical and molecular investigations of patients with nonketotic hyperglycinemia. *Mol Genet Metab* 70:116-121
- Toone JR, Applegarth DA, Coulter-Mackie MB, James ER (2001) Identification of the first reported splice site mutation (IVS7-1G>A) in the aminomethyltransferase (T-protein) gene (*AMT*) of the glycine cleavage complex in three unrelated families with nonketotic hyperglycinemia. *Hum Mutat* 17:76
- 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
- Yoshida T, Kikuchi G (1973) Major pathways of serine and glycine catabolism in various organs of the rat and cock. *J Biochem* 73:1013-1022